

Data management



Gruppo di lavoro

Marc Godat

Dipl. Esperto in finanza e controlling
Lostorf

Marek Gossner

Dipl. Esperto in finanza e controlling
Wallisellen

Peter Herger

Specialista in finanza e contabilità con attestato professionale federale
Bäretswil

Daniel Keller

MAS Accounting & Finance
Rafz

Dr. Abbas Tutcuoglu

PhD Space Engineering, California Institute of Technology
Wallisellen

Markus Zäch

Dipl. Ingegnere meccanico HTL
Eggersriet

Manuel Zmak

Dipl. Esperto in finanza e controlling
Bülach

Inhaltsverzeichnis

Prefazione	5
1 Introduzione	6
1.1 Definizione	6
2 Obiettivi e concetti base di data management	9
2.1 Introduzione ai database.....	9
2.2 Termini fondamentali	9
3 Sulla forma dei dati	13
3.1 Formato dei dati	13
3.2 Struttura dei dati	13
3.3 Granularità dei dati	15
3.4 Pulizia dei dati scaricati da sistemi a monte	15
4 MS Access – una panoramica	16
4.1 Tabelle	16
4.2 Relazioni	16
4.3 Query.....	17
4.4 Maschere, report, pagine e moduli	17
5 Costruzione di un data warehouse con MS Access	18
6 Sistemi ERP (Enterprise Resource Planning)	20
6.1 Cos'è un sistema ERP?.....	20
6.2 Integrazione dei dati attraverso i sistemi ERP	21
6.3 Evoluzione storica dei sistemi ERP	22
6.4 Scelta di un sistema ERP	23
6.5 Fornitori di software ERP e di soluzioni ERP	24
6.6 Produttori di ERP	24
6.7 Costi di un sistema ERP	25
6.8 Struttura del sistema ERP/ERD	25
6.9 Limiti e prospettive dei moderni sistemi ERP	29
7 Business Intelligence (BI)	30
7.1 Definizione di Business Intelligence	30
7.2 Perché la Business Intelligence è importante?	31
7.3 Perché utilizzare la business intelligence?	31
7.4 Software di business intelligence	32
7.5 Online Analytical Processing (OLAP)/Business Analytics	32
7.6 Report e query	33
7.7 Dashboard digitali o visivi	33
7.8 Integrazione	33

8	Big Data	34
8.1	Dove si usano i big data?	34
8.2	Valutazione dei big data	34
8.3	Utilizzo dei big data	35
8.4	Dove memorizzare i big data?	35
8.5 Soluzioni softw	
	data	35
8.6	Critica ai big data	36
9	SQL	37
9.1	Introduzione	37
9.2	Comando SELECT	37
9.3	Comando DISTINCT	38
9.4	Comando WHERE	38
9.5	Operatori SQL AND / OR	39
9.6	Operatore SQL IN	40
9.7	SQL-Operator BETWEEN	41
9.8	Comando SQL LIKE	42
9.9	Operatore SQL ORDER BY	43
9.10	Comando SQL GROUP BY	43
9.11	Comando SQL HAVING	44
10	Robotic Process Automation (RPA)	46
10.1	Introduzione	46
10.2	Struttura	46
10.3	Robotic Process Automation per un database di clienti	47
10.4	10.4	48
 Robotic Process Automation nel test di softw	
	are	48
10.5	10.5	48
 Introduzione di un Robotic Process Automation –	
	cosa si deve considerare?	48
11	Intelligenza artificiale e business analytics	50
11.1	Cos'è l'intelligenza artificiale?	50
11.2	Le diverse forme di IA	52
11.3	Classificazione e regressione	53
11.4	L'emergere di modelli di IA	54
11.5	La vulnerabilità di un modello di IA	61
11.6	Una strada rapida verso il primo modello	62
12	Trasformazione digitale e nuovi modelli di riferimento	64
12.1	Digitalizzazione	64
12.2	Obiettivi di una trasformazione digitale.....	65
12.3	Sfide della trasformazione digitale	67
12.4	Roadmap di una Digital Finance Transformation	67
13	Cambiamento della conduzione finanziaria e del controllo dell'organizzazione	70
13.1	Nuovi modelli di ruoli delle funzioni finanziarie nel mondo digitale	70
13.2	Change management	73

Prefazione

Gentili signore e signori

Gli esami professionali federali altamente specializzati di "Specialista in finanza e contabilità" ed "Esperto/a in finanza e controlling" si tengono da numerosi anni. Questi esami sono tra quelli proposti a livello federale dal maggior numero di anni. La base della contabilità oggi è ancora la contabilità a partita doppia, che venne descritta per la prima volta dal matematico italiano Luca Pacioli nel 1494.

La digitalizzazione della contabilità è cominciata negli anni '70, quando la contabilità ha iniziato ad essere registrata e presentata in forma elettronica, con relativa facilità, come una sequenza logica. Oggi, è ancora comune che l'intera tecnologia dell'informazione dipenda o sia subordinata al CFO e che sia proprio il CFO ad avere la responsabilità dei relativi processi digitali.

I dati sono il nuovo petrolio e i processi che generano i dati sono le nuove raffinerie. Pertanto, è ancora più importante definire chiaramente la sovranità dei dati e regolare centralmente la responsabilità dei dati e dei processi. Per questo motivo, gli esami professionali superiori in finanza e controlling sono stati completati con la nuova materia Data Management. In questa materia si sviluppano le basi per il corretto trattamento dei dati e la comprensione dei processi in modo che i diplomati siano pronti per il futuro.

Purtroppo, c'è ancora una carenza di letteratura specializzata, per cui questo opuscolo è stato redatto come un aiuto per la/il futura/o "Esperto/Esperta in finanza e controlling".

Un ringraziamento va agli autori e alla società per gli esami superiori in contabilità e controlling, senza il cui impegno non sarebbe stato possibile realizzare questo script. La collaborazione nella commissione di esperti in data management, con un grande sostegno da parte della società per gli esami superiori di contabilità e controlling, è stata molto proficua e sono orgoglioso che in poco tempo siamo stati in grado di realizzare una brochure così ampia e informativa.

Auguro ogni successo ai futuri diplomati dei nostri esami professionali. Sono convinto che con il sapere in "Data management", la contabilità in Svizzera resterà moderna anche in futuro e potrà continuare a svilupparsi in questi tempi caratterizzati da rapidi e continui cambiamenti.

Zurigo, novembre 2021

Cordialmente

Peter Herger

Membro del comitato veb.ch, Responsabile della digitalizzazione finanza e controlling.

1 Introduzione

Il termine "data management" può essere interpretato in molti modi. I principi presentati in questo script coincidono con quelli della commissione di esperti in data management, che è composta da professionisti nell'ambito della contabilità e del controlling. Le seguenti spiegazioni non pretendono di coprire completamente i numerosi temi relativi al data management, ma sono intese a tracciare un percorso attraverso i vari argomenti del data management utilizzando come esempio un database di clienti.

1.1 Definizione

Folgende Begriffe werden in diesem Standard mit der angegebenen Bedeutung in alphabetischer Reihenfolge verwendet. Die in den jeweiligen Absätzen erklärten Begriffe werden hier nicht redundant definiert.

Termini	Descrizione
APICS	Association for Operations Management (APICS) è un'organizzazione fondata negli Stati Uniti nel 1957 molto influente nella definizione, nella formazione e nella certificazione dell'organizzazione aziendale.
APS	I sistemi di Advanced Planning and Scheduling (sistemi APS) vengono utilizzati dall'inizio degli anni 2000 per ampliare le funzioni di pianificazione secondo il metodo MRP II e servono per ottimizzare i termini di consegna. Attualmente non esiste una definizione esatta e generalmente accettata e, pertanto, ogni produttore di ERP sviluppa i propri modelli.
Data Warehouse (DWH)	Un database costruito per le analisi. Spesso il DWH è alimentato da diverse fonti di dati.
DTA/SEPA	Per il traffico di pagamenti nazionali esiste un formato di scambio di supporti dati (formato DTA o DTAUS). Quest'unico strumento permette il trattamento elettronico degli ordini di pagamento (bonifici e addebiti diretti) nel traffico dei pagamenti nazionali svizzeri. Il DTA viene gradualmente sostituito dallo standard SEPA (Single Euro Payments Area [SEPA]).
EDI	EDI (Electronic Data Interchange) è la comunicazione interaziendale dei documenti di business in formato standard. EDI è un formato elettronico standard che sostituisce documenti cartacei come gli ordini di acquisto o le fatture.
ERD	La rappresentazione grafica dei tipi di relazione (entità) è denominata Entity Relationship Diagram (ERD) o diagramma delle relazioni tra entità. La rappresentazione è una panoramica grafica delle relazioni interne in un database o in un sistema di dati attraverso una struttura complessa, spesso simile a una rete.
ERP	ERP sta per Enterprise Resource Planning e definisce il contenuto di una soluzione software per la pianificazione delle risorse di un'azienda o di un'organizzazione. I sistemi ERP integrano un gran numero di processi aziendali che vengono elaborati e memorizzati in un database centrale.
ESR	In der Schweiz werden als Einzahlungsschein mit Referenznummer (ESR) zwei verschiedene Formulare verwendet, um einen Geldbetrag in Schweizer Franken zu überweisen. Der ESR wird ab 2020 durch die QR-Rechnung abgelöst.

Termine	Descrizione
ICT	Nel campo dell'elettronica e della tecnologia informatica, ICT è l'abbreviazione di Information and Communications Technology, ovvero di tecnologia dell'informazione e della comunicazione. Nota bene: ICT ha vari altri significati in altre aree professionali.
ISO 20022	ISO 20022 è lo standard internazionale basato su messaggi XML (Extensible Markup Language) per lo scambio elettronico di dati nel settore finanziario a livello globale. La piazza finanziaria svizzera ha introdotto questo standard, semplificando così le operazioni di pagamento e raggiungendo un maggior grado di automazione dei processi di pagamento. In questo contesto, il tradizionale file DTA è stato sostituito dal messaggio pain.001 secondo lo standard ISO 20022 a partire dal 30 giugno 2018.
MRP II	Manufacturing Resources Planning (MRP II) completa la pianificazione del programma di produzione con ulteriori livelli di pianificazione (piano delle vendite), che facilitano la pianificazione mobile e il budgeting. Il modello MRP, che non considerava i vincoli di capacità, è stato ottimizzato introducendo i vincoli di capacità secondo la tecnica della programmazione a ritroso.
MRP und MRP I	Material Requirements Planning (MRP) descrive la pianificazione dei fabbisogni di materiale delle distinte base a più livelli e il calcolo dei fabbisogni lordi e netti. MRP I, sviluppato alla fine degli anni '70, prende in considerazione anche le date delle lavorazioni oltre che le distinte base dei materiali. Il modello ad esso associato, che non considera i vincoli di capacità, non ha mai preso piede nella pratica ed è per questo che è stato sviluppato MRP II.
PCP	PCP è un sistema computerizzato che si occupa della pianificazione e controllo della produzione. Il sistema PCP è una componente centrale di un sistema ERP poiché può essere utilizzato per pianificare le risorse finanziarie e le risorse umane dell'azienda.
Programma di produzione	Il programma di produzione contiene l'insieme di tutti i prodotti di un'azienda, per tipo e quantità, e allo stesso tempo stabilisce l'ubicazione in cui fabbricare i prodotti in un determinato intervallo di tempo o in un lasso di tempo predefinito.
PUNS	Il software certificato «Procedura unitaria di notifica dei salari» (PUNS) regola e organizza la trasmissione dei dati salariali tra le aziende, le assicurazioni sociali e le autorità. I dati salariali sono trasmessi all'AVS, alle assicurazioni, agli uffici fiscali e all'Ufficio federale di statistica.
PVR	In Svizzera vengono usati due diversi formulari come polizza di pagamento con numero di riferimento (PVR) per trasferire una somma di denaro in franchi svizzeri. A partire dal 2020 la PVR sarà gradualmente sostituita dalla fattura QR.
Risorsa	Una risorsa (fonte) descrive il mezzo, la condizione e le caratteristiche o le proprietà necessarie per perseguire gli obiettivi, per far fronte ai requisiti, per eseguire azioni specifiche o per far funzionare un processo in modo mirato all'obiettivo. In economia aziendale, in economia politica e nelle organizzazioni, per risorse si intendono le attrezzature, i mezzi finanziari, le materie prime, l'energia, le persone e il tempo (di lavoro).

Termine	Descrizione
RFID	RFID (Radio Frequency Identification), identificazione in radiofrequenza, è una tecnologia per sistemi trasmettitore-ricevitore per l'identificazione e la localizzazione automatica e senza contatto di oggetti ed esseri viventi mediante onde radio.
SCOR	Supply Chain Operations Reference Model (SCOR) descrive tutti i processi aziendali interni ed esterni all'azienda. Dal 2014 SCOR è stata integrata in APICS. La documentazione presentata prende in considerazione le basi secondo il modello SCOR.
Simultaneous engineering	Metodo per accorciare e minimizzare i cicli di innovazione.
Software deployment	Il software, nuovo o modificato, passa attraverso un processo di deployment, cioè il software viene installato, configurato e distribuito nell'ambiente di produzione.
SSEE	SSEE è l'opposto di VUCA, cioè stable (stabile), secure (sicuro), easy (facile), explicit (esplicito) e descrive i presupposti per una gestione aziendale più semplice.
State of the Art	Corrisponde alla tecnologia attuale, all'implementazione effettuata secondo le best practice.
Supply Chain	La supply chain (catena del valore, catena logistica, catena di fornitura) di un prodotto comprende tutte le attività generatrici di valore di tutte le fasi della produzione e della vendita, a partire dalle materie prime fino ai prodotti finiti.
Supply Chain	VUCA è un acronimo per i termini inglesi volatility (volatilità), uncertainty (incertezza), complexity (complessità) e ambiguity (ambiguità). Descrive le difficili condizioni generali della gestione aziendale in tempi di globalizzazione.

2 Obiettivi e concetti base di data management

2.1 Introduzione ai database

Oggigiorno Excel è onnipresente nella contabilità e nel controlling. In questi ambiti si trovano a volte dei veri e propri «mostri Excel» che elaborano un'enorme quantità di dati. Molto spesso Excel viene usato per attività per le quali non è stato sviluppato. Bisogna sempre essere consapevoli che Excel non è un database e non dovrebbe essere usato come tale (anche se computer sempre più potenti lo rendono possibile). Qual è la raccomandazione di Microsoft su quando, o fino a dove, dovrebbe essere utilizzato Excel e da quando, invece, dovrebbe essere utilizzato Access?

"Use **Access** when you:

- Require a **relational** database (multiple tables) to store your data.
- May need to **add more tables** in the future to an originally flat or non-relational data set
- Have a **very large amount of data** (thousands of entries)
- Have data that is mostly of the **long text string type** (not numbers or defined as numbers).
- Rely on **multiple external databases** to derive and analyze the data you need
- Want to run **complex queries**
- Have **many people working** in the database and want robust options to expose that data for updating"

"Use **Excel** when you:

- Require a flat or **non-relational** view of your data (you do not need a relational database with multiple tables)
- Want to run primarily **calculations** and **statistical comparisons** on your data
- Know your **dataset is manageable in size** (no more than 15,000 rows)."

Riassumendo:

MS Excel è più adatto per compiti con complessità limitata, bidimensionali, con quantità di dati piuttosto ridotte, con alte richieste di flessibilità e bassi requisiti di integrità.

MS Access è più adatto per l'elaborazione di grandi quantità di dati, specialmente se questi sono multidimensionali e di natura relazionale (vedere i termini fondamentali indicati di seguito).

2.2 Termini fondamentali

Di seguito vengono spiegati alcuni termini fondamentali del «mondo dei database». Sembrerà un po' teorico, ma alcuni concetti diventeranno più chiari molto rapidamente, passo dopo passo, man mano che si procederà con la lettura di questo script.

Cubetools

Il termine Cubetools indica una collezione di applicazioni software (software suite o suite di software) composta da un database Access, da uno o più cubi OLAP e da uno o più Frontend. La suite di software lavora con dati che vengono trasferiti da sistemi a monte. L'architettura può essere rappresentata graficamente come nella Figura 1.

OLAP-Cube (cubo OLAP)

Un OLAP cube (denominato anche Data cube, Cube operator o cubo OLAP) è un termine usato nella teoria dei data warehouse per la rappresentazione logica dei dati. I dati sono disposti come elementi di un cubo (in inglese cube) **multidimensionale**. Le dimensioni del cubo descrivono i dati e permettono l'accesso in modo semplice. I dati possono essere selezionati attraverso uno o più assi del cubo. Il termine OLAP deriva dall'analisi dei dati e

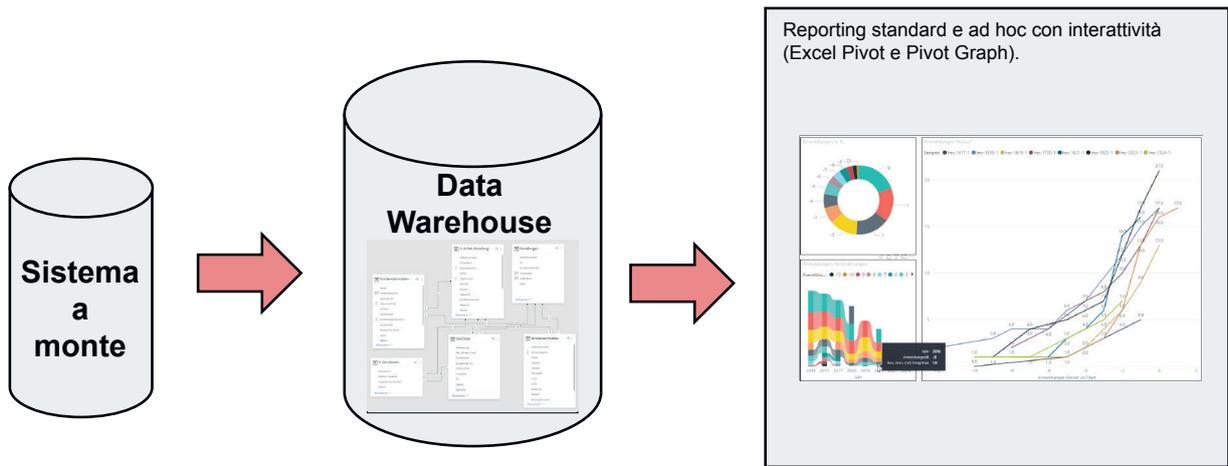


Figura 1: Architettura dei Cubetool

significa Online Analytical Processing. Questo tipo di rappresentazione è vantaggioso per l'analisi dei dati, poiché diverse caratteristiche (dimensioni) dei dati sono accessibili nello stesso modo. Da qui il suo utilizzo in applicazioni OLAP che analizzano o elaborano visivamente i dati in un data warehouse. (Fonte: Wikipedia)
in einem Data Warehouse analysieren oder visuell aufbereiten. (Quelle: Wikipedia)

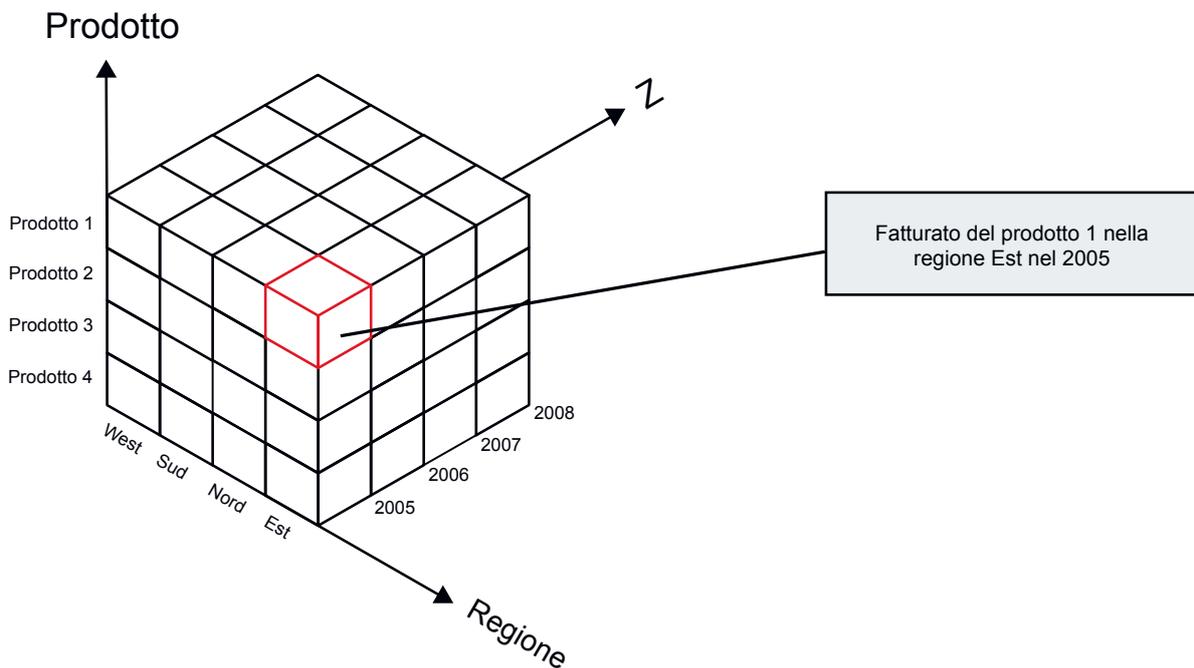


Figura 2: Esempio di cubo OLAP

Data Warehouse (DW)

Un data warehouse (in inglese letteralmente magazzino dati) è una raccolta centrale di dati (di solito un grande database) il cui contenuto è composto da dati provenienti da diverse fonti. Ad esempio, un data warehouse per il controlling può consistere in una parte di statistiche di vendita, una parte di contabilità finanziaria, una parte di gestione delle scorte, una parte di forecast, che possono provenire da un sistema ERP, e una parte di offerte, che possono provenire da un sistema CRM. I dati vengono caricati dalle diverse fonti di provenienza nel data warehouse, ove restano memorizzati a lungo termine, principalmente per l'analisi dei dati e per il supporto alle

decisioni aziendali. Il termine ha origine dalla gestione delle informazioni in economia aziendale. Un data warehouse consente di integrare le informazioni.

Attualmente non esiste una definizione omogenea del termine «Data warehouse». C'è un ampio consenso su quanto segue:

Un data warehouse permette di ottenere una visione multidimensionale di archivi di dati eterogenei e distribuiti in cui i dati rilevanti per una visione globale e provenienti da varie fonti vengono aggregati in un archivio di dati unico e consistente. Quindi, il contenuto di un data warehouse viene creato copiando ed elaborando dati provenienti da molteplici fonti. **Nella maggior parte dei casi, un data warehouse è la base per l'aggregazione di cifre chiave aziendali e per l'analisi all'interno di matrici multidimensionali** (cubo OLAP), il cosiddetto Online Analytical Processing (OLAP). Un data warehouse è spesso il punto di partenza per il data mining (letteralmente dall'inglese estrazione di dati). Di regola, le applicazioni lavorano con estratti specifici dal data warehouse, i cosiddetti data mart, che si collocano a valle di un data warehouse e ne rappresentano un sottoinsieme logico o fisico. (Fonte: Wikipedia)

Normalizzazione

La normalizzazione consiste nel dividere le colonne delle tabelle all'interno del database in nuove colonne (ad esempio dividere la colonna dell'indirizzo in tre colonne con il codice postale, la città e la via) o nel collegare queste colonne con altre tabelle (ad esempio un cliente viene collegato alla tabella clienti tramite un codice univoco).

Lo **scopo della normalizzazione** è quello di ridurre le ridondanze (ridondanza = la stessa informazione esiste più volte) e di prevenire le anomalie (ad esempio contenuti contraddittori dei dati) al fine di semplificare la manutenzione del database e garantire la consistenza dei dati.

Ciò diventa chiaro con un semplice **esempio**: un database contiene i clienti e i loro indirizzi, nonché gli ordini che sono assegnati ai clienti. Ipotizziamo che il cliente cambi il proprio indirizzo: se ogni ordine contenesse direttamente l'indirizzo del cliente, si dovrebbero cercare tutti i record degli ordini di questo cliente e modificare tutti i campi relativi all'indirizzo. In un database normalizzato, c'è solo una voce da modificare, in questo caso il record contenuto nella tabella dei clienti. Inoltre, la normalizzazione porta a un minor utilizzo di memoria in quanto il record di dati di una tabella (ad esempio "Ordini") fa riferimento a un record di dati di un'altra tabella (ad esempio "Clienti") invece di ricopiare/duplicare i campi in esso contenuti.

Poiché ci potrebbero essere numerosi ordini dello stesso cliente, in un database non normalizzato i dati del cliente verrebbero copiati ogni volta per ogni nuovo ordine (ridondanza). Se ci fosse più di un indirizzo per un cliente nel database, questo potrebbe portare a un output contraddittorio o errato; i dati non sarebbero quindi consistenti. (Fonte: Wikipedia)

Integrità

L'integrità dei dati descrive il requisito per la protezione e la qualità dei dati digitali. Per assicurare l'integrità dei dati, devono essere garantite la consistenza, la completezza, l'accuratezza e la validità dei dati durante l'intero periodo di conservazione. Tutti i cambiamenti di dati devono essere documentati in modo tracciabile, affinché i dati non possano essere modificati inavvertitamente o senza autorizzazione. L'integrità dei dati ha quindi l'obiettivo principale di proteggere i dati da violazioni interne ed esterne.

Dati di base (dati statici)

- sono caratterizzati da un'elevata staticità (invariati nel tempo) e di solito non hanno alcun riferimento temporale,
- sono spesso usati da diverse applicazioni o reparti aziendali, ad esempio da altri record (acquisti, produzione, logistica, contabilità, vendite, preparazione del lavoro),
- sono spesso i criteri usati nelle valutazioni analitiche (ad esempio prodotto, filiale, cliente) e ciò li rende candidati per la definizione delle dimensioni in un Online Analytical Processing (OLAP), e
- sono per lo più conservati nel lungo termine.

Dati di movimento

- sono caratterizzati da un'elevata dinamicità (variabili nel tempo) e di solito hanno un riferimento temporale (data di validità),
- non vengono utilizzati spesso,
- spesso forniscono i fatti nelle valutazioni analitiche e ciò li rende candidati per il contenuto delle celle nell'OLAP,
- sono di solito necessari solo per un periodo di tempo limitato e quindi
- vengono conservati a breve o medio termine.

3 Sulla forma dei dati

Il punto di partenza è: esistono uno (o più) sistemi che contengono dati con i quali l'utente desidera lavorare. La varietà di tali sistemi (qui denominati Sistemi a monte) e le strutture dei loro dati è molto ampia; per-tanto, le informazioni generali devono essere sufficienti per i punti più importanti.

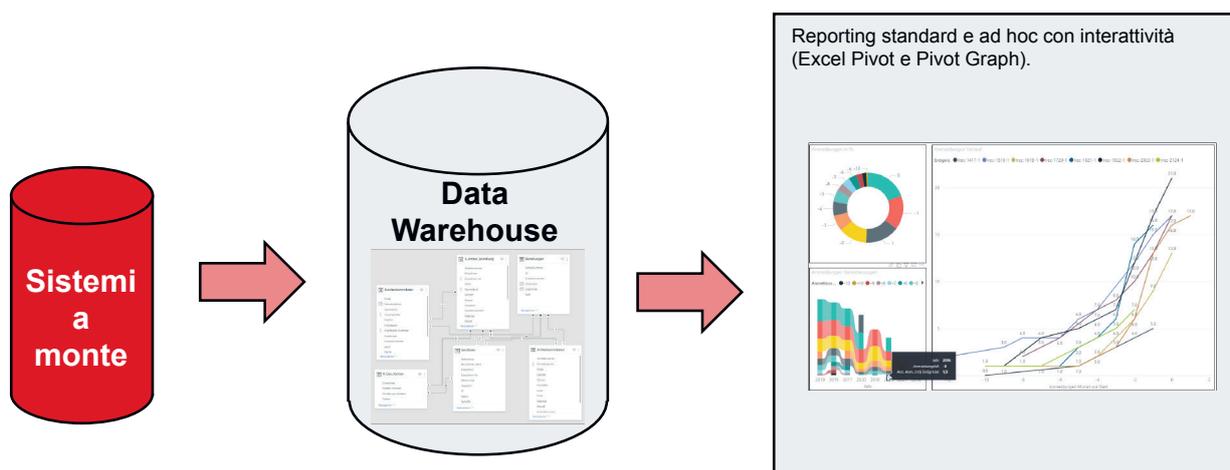


Figura 3: Sistemi a monte e data warehouse

3.1 Formato dei dati

MS Access può lavorare con i seguenti formati di dati: MS Access, dBASE III / IV / V, MS Excel, Exchange, HTML, Lotus 1-2-3, Outlook, Paradox, file di testo, XML, database ODBC. I dati possono essere scaricati in uno o più di questi formati da quasi tutti i sistemi di dati esistenti (come SAP). A volte è necessario un pas-saggio intermedio; ad esempio, se si utilizza SAP può essere opportuno esportare dapprima i dati SAP come file di testo, quindi ripulirli in Excel e, infine, importare il file Excel in MS Access.

3.2 Struttura dei dati

Di principio, i dati dei sistemi a monte devono essere disponibili in forma normalizzata. Per un data ware-house, è necessario disporre di record di dati con un campo per ogni caratteristica rilevante. Tutte le carat-teristiche con cui si desidera lavorare devono essere contenute nei record di dati. Ad esempio, un conto eco-nomico e un bilancio strutturati in un sistema contabile possono avere la seguente forma (dettaglio):

Conto economico febbraio 2020 Year-to-date				
Mandante (azienda)	Gruppo di genere di costo	Numero e nome del genere di costo	Importo anno corrente	Importo anno precedente
BU21	Personalkosten			
		450400 Salari	47'123.–	45'231.–
		450300 Casa pensione	13'652.–	13'211.–
		450450 Abbonamenti metà prezzo	0.–	450.–
	Totale costi del personale		60'775.–	58'892.–

Il modello riportato **non è adatto** come base per il trasferimento di dati in un data warehouse per i seguenti motivi:

- Il **mandante (azienda)** (BU21) è solitamente importante per le analisi, ma qui non è parte integrante di ogni record di dati.
- Il **nome del genere di costo** non è necessario, poiché ha una relazione 1:1 con il numero del genere di costo e può quindi essere richiamato dalla tabella dei dati di base dei generi di costo (**normalizzazione**).
- Una suddivisione dei valori in **anno corrente e anno precedente** non è utile allo scopo, perché l'es-senza del contenuto del campo è la stessa, ovvero il valore delle transazioni su cui si basa il conto eco-nomico. Sarebbe più appropriato associare il campo a una **data di registrazione**.
- In una valutazione, si potrebbe anche voler avere la possibilità di eseguire un **drill down** fino al livello della singola registrazione. Questo significa che, invece di valori aggregati, si deve poter disporre delle singole **transazioni**.
- Se si desiderasse anche sapere quali costi del personale sono stati sostenuti a gennaio e quali a febbraio, sia nell'anno corrente che nell'anno di confronto, la **data di registrazione** consentirebbe di raggiungere questo scopo.
- Non c'è bisogno di un **totale (somma)** dei costi del personale, in quanto questo totale può essere cal-colato come somma di tutti i record di dati (o tutti i record di dati con una o più caratteristiche comuni) di un data warehouse.

La **soluzione** potrebbe consistere in un record con i seguenti campi:

Nome del campo	Esempio del nome del campo	Commento
Mandante	BU 21	Il codice del mandante può essere collegato al nome dell'azienda attraverso una tabella di dati di base contenuta nel data warehouse; questo rende le valutazioni più leggibili.
Genere di costo	450400	Il numero del genere di costo è sufficiente perché il nome del genere di costo può essere ripreso dalla corrispondente tabella di dati di base.
Data di registrazione	14.02.2012	Con la data di registrazione come caratteristica, si può logicamente derivare in quale mese, trimestre ed anno una transazione ha influito sul risultato economico.
Testo della registrazione	Accantona-mento pro rata del bonus	Se si ha la possibilità di effettuare drill-down fino al livello della singola transazione, il testo della registrazione fornisce informazioni aggiuntive utili per l'analisi.
Importo	3'000.–	Valore della singola transazione. Questo può essere aggregato nel database attraverso tutte le dimensioni delle tabelle di dati.

Se non si dispone delle conoscenze per creare le proprie query personalizzate in un sistema a monte, lo si può commissionare al fornitore del software o al consulente software. I costi per questa attività sono di solito conte-nuti, poiché anche i sistemi a monte sono costruiti secondo i principi dello sviluppo di database, con dati nor-malizzati. Pertanto, le valutazioni riassuntive sono di solito create partendo da dati normalizzati.

3.3 Granularità dei dati

La scelta della granularità dei dati è un compromesso tra le possibilità di **valutazione** da un lato e **le prestazioni / la velocità** dall'altro. MS Access può elaborare un notevole volume di dati; se si memorizzano diverse centinaia di migliaia o un milione di record di dati, si dovrebbe però porre attenzione alle possibilità di ottimizzazione. Prendendo come esempio la contabilità finanziaria, può aver senso lavorare a livello delle singole transazioni solo per l'anno corrente; quando poi i dati, nel corso del tempo, diventano i valori dell'anno precedente si potrebbe considerare la possibilità di memorizzare solo i saldi per mese. In questo caso, le transazioni individuali possono essere **condensate / raggruppate**, cioè sommate secondo criteri ragionevoli, riducendo così il numero di record di dati.

3.4 Pulizia dei dati scaricati da sistemi a monte

Anche se le query necessarie per recuperare i record di dati normalizzati sono disponibili nei sistemi a monte, è spesso necessario un piccolo lavoro di "pulizia" da eseguirsi dopo l'estrazione dei dati. I possibili motivi che portano a dover svolgere questa attività potrebbero essere molteplici, ad esempio: la query del sistema a monte fornisce più campi di quelli necessari per il data warehouse o i formati dei contenuti dei campi devono essere adattati.

Ad esempio, **SAP** fornisce dati in cui i valori numerici superiori a 1000 sono separati, di default, da una virgola. Tuttavia, MS Access non legge correttamente i numeri formattati in questo modo e ciò causa messaggi di errore durante l'importazione nel data warehouse.

Poiché le importazioni di dati in un data warehouse devono essere fatte a cadenza regolare (mensilmente, settimanalmente o su richiesta), la pulizia manuale dei dati richiede tempo e comporta un certo rischio di **manipolazioni errate**. È consigliabile, se possibile, effettuare tale pulizia dei dati utilizzando le **macro di Excel**.

4 MS Access – una panoramica

Un database MS Access è composto da diversi tipi di oggetti, che sono brevemente spiegati di seguito.

4.1 Tabelle

Una tabella è il luogo dove i dati sono fisicamente memorizzati. La tabella è composta da campi. Il tipo di campo è determinato in base al contenuto previsto nel campo. Una riga della tabella, con due o più campi, è un record. I tipi di campo più comuni sono:

Feldart	Kommentar
Testo	Questo tipo di campo è indicato, in pratica, per tutti i campi che non appartengono agli altri tre tipi. La lunghezza massima del campo è 255 caratteri.
Numero	Questo tipo di campo è indicato per tutti i campi che devono essere utilizzati nei calcoli numerici.
Data / Ora	Questo tipo di campo è utilizzato per le date e può essere utilizzato per aggregare i valori sulla base di intervalli temporali.
Numerazione automatica	La numerazione automatica è un contatore che garantisce l'univocità di un record di dati all'interno di una tabella. Ad ogni nuovo record aggiunto nella tabella viene assegnato automaticamente un numero sequenziale univoco. Se il record di dati viene cancellato, anche il numero di sequenza scompare con esso e non sarà più utilizzato da record aggiunti successivamente. Per questo motivo, il tipo "numerazione automatica" è di solito utilizzato come chiave primaria.

La **chiave primaria** è il campo della tabella che assicura che il record sia unico. In un database di clienti, la chiave primaria potrebbe essere il numero AVS, poiché questo è un numero unico per ogni persona. Se non si dispone di un valore unico, si può anche utilizzare un contatore (numerazione automatica).

La chiave primaria non è necessaria se l'univocità di un record di dati è già garantita da un sistema a monte. In un sistema contabile, ad esempio, non è possibile memorizzare due generi di costo che hanno lo stesso numero di genere di costo ma denominazioni diverse. In questo caso, non è necessario definire una chiave primaria per i generi di costi nella tabella dei dati di base del data warehouse.

4.2 Relazioni

Le relazioni definiscono il rapporto / legame tra le singole tabelle di un database. Nell'esposizione del tema della normalizzazione, è stato citato l'esempio di un database con dati relativi ai clienti e agli ordini. È stato spiegato perché ha senso memorizzare i dati dei clienti in una tabella e i dati degli ordini in un'altra tabella.

Poiché un singolo cliente può effettuare più ordini, ma al contrario un ordine può provenire solo da un cliente, si ha una **relazione 1:n** tra la tabella dei clienti e la tabella degli ordini. Questo è il tipo di relazione più comune in un database.

Un database con relazioni tra le proprie tabelle è denominato **database relazionale**.

4.3 Query

Una query è una **rappresentazione** dei dati memorizzati in una tabella.¹ Le query permettono di selezionare, raggruppare, combinare e presentare i dati in una forma ottimale per l'analisi. Una query può essere

- direttamente su una tabella o
- basata su un'altra query.

In quest'ultimo caso, si parla di una **gerarchia di query**.

4.4 Maschere, report, pagine e moduli

Le **maschere** sono utilizzate per semplificare l'inserimento e la ricerca dei dati direttamente in MS Access. Con la **funzione report**, i dati possono essere presentati in formati di report predefiniti. Le **pagine** hanno la stessa funzione delle maschere, ma in un formato web. I **moduli** possono essere utilizzati per scrivere i propri programmi o parti di programmi.

¹ Questo vale per i tipi *query di selezione* e *query crosstab*.

5 Costruzione di un data warehouse con MS Access

La parte più complessa del lavoro di sviluppo è certamente la costruzione di un data warehouse all'interno di un progetto di sviluppo Cubetool.

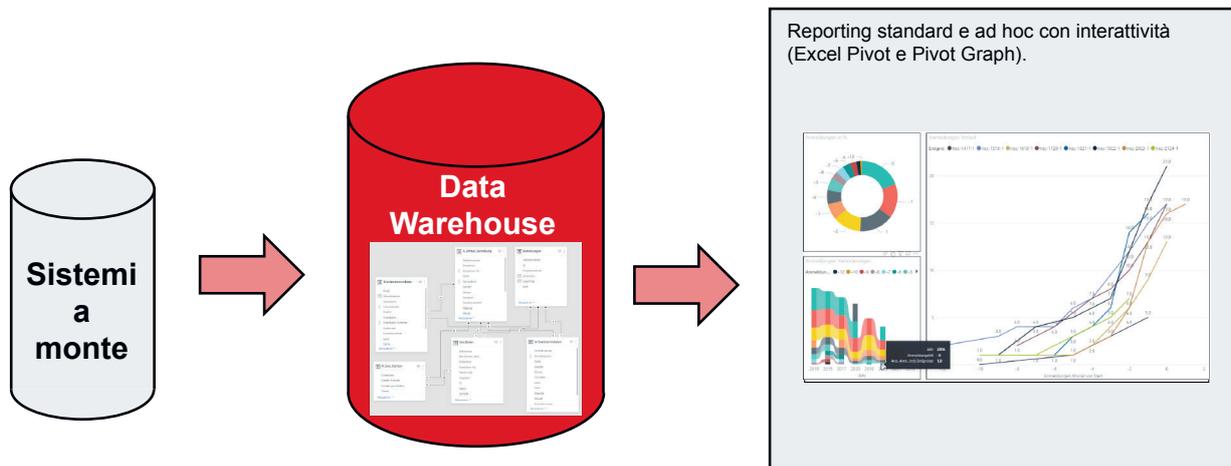


Figura 4: Cubetool – Data warehouse

Il modello mostrato in questo esempio è la versione adattata di un database che è stato utilizzato con successo per diversi anni in tre aziende (circa 400 dipendenti, circa 100 milioni di franchi di fatturato) appartenenti a un grande gruppo. Il controlling di queste tre aziende è centralizzato. Access è usato come data warehouse, in cui i dati provenienti dal sistema a monte (SAP) sono elaborati e successivamente presentati attraverso un frontend Excel.

Secondo il responsabile del controlling, questo Cubetool ha **ridotto di circa il 90%** lo sforzo richiesto per preparare i dati necessari per il controlling. Il database in questione è stato progettato come un data warehouse e contiene solo tabelle e query; nessuna maschera, report o macro. MS Excel (grafico pivot) serve come front-end.

Dopo ogni chiusura mensile, i valori effettivi del **sistema a monte SAP** vengono trasferiti al data warehouse sotto forma di un **calcolo delle voci dell'ordine del cliente** (ricavi e costi diretti) e di un **calcolo del centro di profitto** (conto economico completo) con voci individuali. Inoltre, vengono aggiornati gli aggiornamenti della pianificazione per ogni singolo piano - forecast 1, 2 e 3 (effettivo + previsione a fine dell'anno) così come target setting (3 anni) e budget (3 anni). Questi aggiornamenti della pianificazione (incluso il budget) non vengono memorizzati nel sistema SAP, ma vengono salvati direttamente nel data warehouse. Nel data warehouse vengono trasferiti solo i dati del conto economico e nessun dato del bilancio.

Di seguito è mostrata la struttura delle tabelle e delle query del data warehouse. I tre oggetti di database inferiori **TAggiornamentiPianificazioneNuovo**, **TKE5Z_con_ordiniclienti** e **T_KE24_nuovo** sono tabelle; gli altri oggetti del database sono query. Le frecce mostrano il flusso di dati nella gerarchia della query. I nomi delle tabelle e delle query sono mostrati nelle caselle.

La figura 5 mostra la struttura sistematica di un tale data warehouse. Nel settore sinistro vengono immessi i dati della pianificazione, nei settori centrale e destro i dati effettivi. Nel settore superiore, i dati effettivi sono combinati con i dati di pianificazione e arricchiti con i dati di base. Infine, la query superiore costituisce la base per un front-end Excel

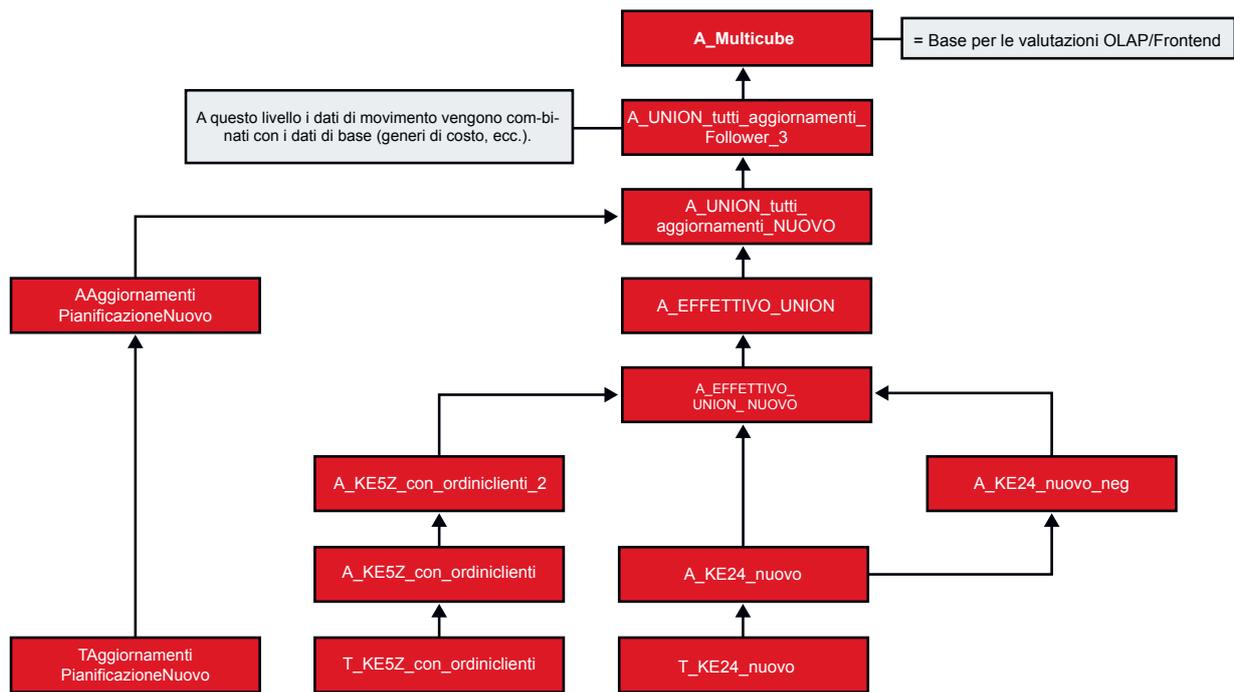


Figura 5: Esempio della struttura di un data warehouse

Il grande vantaggio di una tale architettura è certamente la facilità di manutenzione. Il completamento o l'aggiornamento dei dati avviene solo nelle tre tabelle **TAggiornamentiPianificazioneNuovo**, **T_KE5Z_con_ordiniclienti** e **T_KE24_nuovo**. Non bisogna cambiare nulla nelle altre query, a condizione che non ci siano cambiamenti strutturali. I dati caricati vengono elaborati automaticamente.

Questo esempio mostra chiaramente come una PMI possa creare un management information system (MIS) basato su un data warehouse in un modo molto economico (usando solo le risorse esistenti).

6 Sistemi ERP (Enterprise Resource Planning)

Negli ultimi decenni le aziende sono cresciute sempre di più. Sono comparse imprese molto grandi, con strutture molto confuse, enormi flussi di cassa e quindi anche grandi fabbisogni di capitale. Tali gruppi sono già molto complessi da gestire in tempi normali. Ma nel mondo di oggi, in cui l'ambiente cambia sempre più rapidamente e sia i cicli di vita dei prodotti che i trend diventano sempre più brevi, è quasi impossibile mantenere una visione d'insieme di queste imprese con gli strumenti usuali. È qui che si può creare un collegamento con i sistemi ERP. Grazie alla loro capacità di mappare l'azienda nella sua totalità e in tempo reale, i sistemi Enterprise Resource Planning sono diventati indispensabili nella prassi decisionale di oggi.

6.1 Cos'è un sistema ERP?

La maggior parte delle persone, specialmente nella finanza, ha sentito parlare di sistemi ERP. Ma cosa fa effettivamente un sistema ERP? Fondamentalmente, un sistema ERP mappa le risorse disponibili nell'azienda e fornisce al management, in modo tempestivo, le informazioni appropriate per poterle gestire. Queste risorse includono, ad esempio:

- Capitale
- Collaboratori
- Apparecchiature

Gli obiettivi principali di un sistema ERP sono il miglioramento dei processi e delle strutture organizzative, la rapida adattabilità ai cambiamenti dell'azienda e del mercato e l'ottimizzazione dei processi aziendali.

I sistemi ERP sono programmi integrati basati sulla gestione centralizzata dei dati. Questo supporta, mappa, ottimizza e documenta i flussi di informazione all'interno di un processo aziendale.

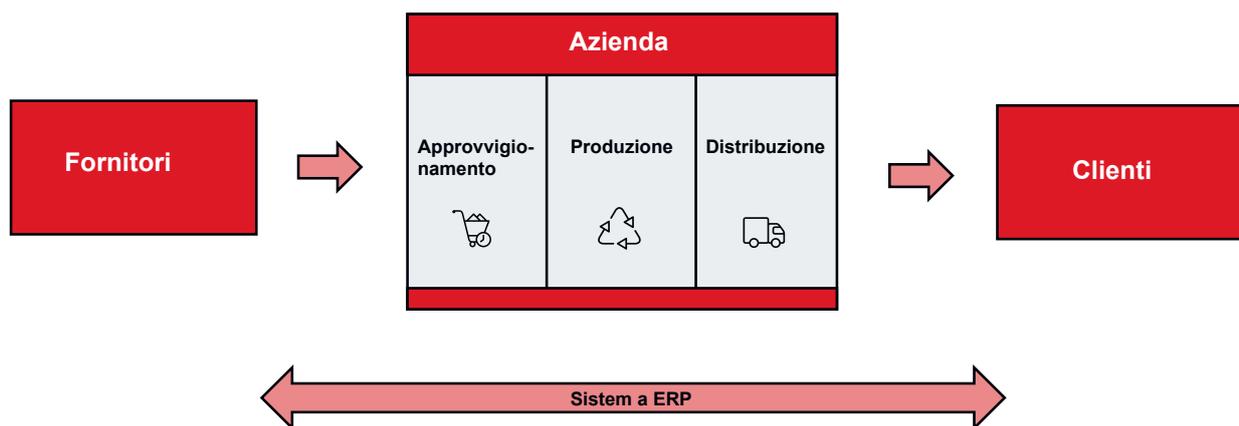


Figura 6: Processo di un sistema ERP

La grande innovazione di un sistema ERP rispetto ai sistemi tradizionali è che i processi, descritti nell'immagine, sono mappati nella loro intera catena. Questo consente al management di rilevare e analizzare l'azienda nella sua totalità.

Inoltre, consente al management di reagire molto più rapidamente ai cambiamenti del mercato e quindi di neutralizzare o ridurre i possibili impatti negativi sul risultato aziendale.

Da un punto di vista organizzativo, un sistema ERP è composto dalle divisioni aziendali di seguito raffigurate. Attraverso un design integrativo, tutte le aree aziendali condividono lo stesso database e comunicano tra loro.

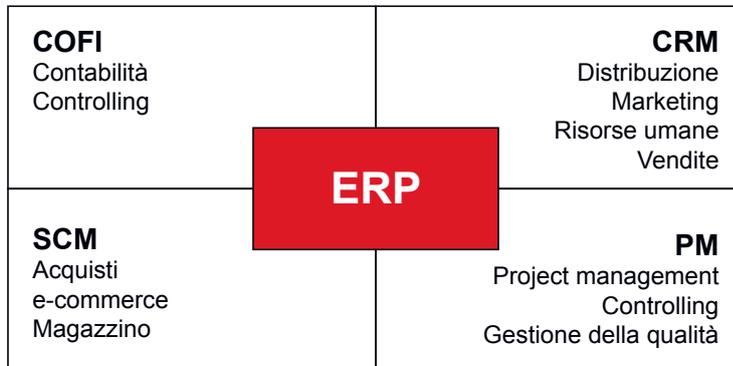


Figura 7: Organizzazione di un ERP

COFI **Contabilità finanziaria**

CRM **Customer Relationship Management.** Attraverso l'allineamento sistematico dei processi interni dell'azienda e con l'aiuto di software speciali (sistemi CRM), l'acquisizione di nuovi clienti e la cura dei clienti esistenti vengono sostenute e ottimizzate. L'obiettivo è quello di costruire sistematicamente relazioni con i clienti e mantenerle a lungo termine.

GeScore **Indica un sistema di gestione delle scorte** supportato da un software che mappa il flusso dei prodotti e delle merci di un'azienda attraverso i vari livelli commerciali. Le componenti centrali sono le vendite, gli acquisti, la logistica in entrata, la logistica in uscita e la tenuta del magazzino. I sistemi di gestione delle scorte possono creare offerte, elaborare ordini di produzione e di acquisto in modo completamente automatico mediante componenti del flusso di lavoro (workflow), nonché elaborare bollettini di consegna e fatture. Attraverso il corrispondente modulo per la gestione del magazzino, le scorte di merci e di prodotti finiti possono essere controllate in modo proattivo, gestendo i relativi inventari. I sistemi di gestione delle scorte sono spesso parte integrante del sistema ERP (a volte come modulo aggiuntivo per il quale si paga un supplemento) e possono registrare le consegne di merci eseguite, i livelli di stock attuali e le merci necessarie in futuro per poter evadere gli ordini ed elaborarli automaticamente.

PCP **Sistema di pianificazione e controllo della produzione.** In un sistema PCP, la pianificazione della produzione è supportata da un software. Questo permette alle aziende di ridurre i costi nell'area della produzione e di aumentare la prontezza di consegna grazie a questa forma di pianificazione. Con un'ottimizzazione supportata dal PCP, le aziende possono ridurre il tempo di lavorazione della loro produzione, ottenere un carico uniforme delle macchine, aumentare la prontezza di consegna, aumentare il rispetto delle date di consegna, ridurre e ottimizzare le scorte e quindi anche il vincolo di capitale.

6.2 Integrazione dei dati attraverso i sistemi ERP

L'immagazzinamento tempestivo di quantità enormi di dati, la loro rilevazione e analisi rappresenta una grande sfida per i moderni sistemi ERP in grandi imprese e gruppi d'impresa. Tutti i processi, che vengono eseguiti in background e assicurano un'elaborazione dei dati senza intoppi, sono enormemente complessi ed enormemente costosi. Interi reparti aziendali hanno il compito di garantire un'elaborazione dei dati regolare.

Benché un sistema ERP sia composto da numerosi moduli software, esso può svolgere i propri compiti in tutte le aree aziendali solo se i dati sono univoci e disponibili solo una volta, in modo non ridondante, in tutta l'azienda. Le informazioni ottenute da questi dati sono il bene più prezioso di un'azienda. Ecco perché l'integrità dei dati è importante e perché la mancanza di integrità dei dati può causare grandi danni e perdite di prestazioni.

6.3 Evoluzione storica dei sistemi ERP

Per molto tempo, le singole aree funzionali delle aziende come la distribuzione, la gestione dei materiali o la gestione delle risorse umane sono state gestite separatamente l'una dall'altra e supportate dalle cosiddette soluzioni isolate (dette anche soluzioni isola). Viste singolarmente, queste soluzioni isolate fornivano dati assolutamente corretti che tuttavia, a causa della mancanza di collegamenti, portavano spesso a prendere decisioni inconsistenti con conseguenti perdite (di tempo o denaro).

I primi sistemi simili agli ERP (a quel tempo avevano un nome diverso) sono comparsi nell'industria negli anni '60. Questi sistemi erano utilizzati principalmente per ottimizzare i processi di produzione. I cosiddetti sistemi MRP (Material Resource Planning), che sono considerati i primi precursori dei moderni sistemi ERP, venivano utilizzati per ottenere una pianificazione standardizzata dei fabbisogni di materiale.

L'introduzione di questi sistemi ha portato grandi guadagni di efficienza e presto ci si è chiesti se questi sistemi non potessero essere estesi ad altre aree dell'azienda. Questo obiettivo è stato raggiunto negli anni '80, quando i sistemi MRP sono stati ampliati per includere la pianificazione della produzione; questi sistemi sono denominati sistemi

MRP II. Con il tempo, sono state aggiunte sempre più funzioni, come le risorse umane, il CRM e la logistica, e presto si è iniziato a parlare, per la prima volta, di sistemi ERP.

I sistemi ERP si sono sviluppati rapidamente, sono in grado di elaborare sempre più dati e coprire sempre più aree. I moderni sistemi ERP sono spesso indicati come sistemi ERP II, la cosiddetta seconda generazione. Oltre ai processi interni, coprono anche aree trasversali lungo la catena del valore e si basano su un'architettura web. Funzioni come Customer Relationship Management (CRM) e Supply Chain Management (SCM) sono integrate nel sistema ERP. Questo permette di gestire i processi, indipendentemente dal fatto che abbiano luogo internamente o esternamente all'azienda. Oggi, per esempio, un moderno sistema ERP consente facilmente di riprendere automaticamente gli ordini avviati dall'esterno ed elaborarli all'interno dell'azienda.

Inoltre, le soluzioni innovative dei produttori di ERP hanno creato opportunità per rendere i sistemi ERP accessibili anche alle PMI. Proprio in questo contesto si deve menzionare il nuovo ERP cloud. Con la soluzione ERP dal cloud (Software-as-a-Service, SaaS), i fornitori offrono soluzioni mature e spesso a più livelli per le PMI di media dimensione. Oltre a una soluzione ERP matura e standardizzata, l'azienda gode anche, ad esempio, di soluzioni all'avanguardia (state of the art) nell'ambito della sicurezza dei dati, che spesso superano di molte volte il livello di sicurezza tenuto nell'azienda stessa.

L'ERP dal cloud è ancora un argomento relativamente nuovo e deve fare i conti con alcuni pregiudizi, soprattutto per quanto riguarda la disponibilità e la sicurezza dei dati. Tuttavia, per le piccole imprese che non possono permettersi né il software né l'hardware, e tanto meno un proprio reparto IT, può essere una soluzione efficiente e relativamente poco costosa per stare al passo con le grandi aziende in termini di tecnologia.

Oltre ai trend SaaS e cloud ERP, il mobile computing domina il tema dell'ERP. Il fenomeno BYOD (Bring Your Own Device) risponde al bisogno di molti collaboratori e aziende di poter lavorare con i loro smart gadget usati privatamente – che sia uno smartphone, un tablet o un notebook – e di poter accedere ai dati aziendali da qualsiasi luogo. Con l'ERP mobile basato sul cloud, i sistemi ERP diventano ancora più flessibili e possono essere ancora

più personalizzati alle esigenze delle imprese e dei loro collaboratori.

Lo sviluppo più recente dei sistemi ERP è l'integrazione di elementi dei social media. Questa integrazione trasforma il sistema ERP in uno strumento che soddisfa anche la disponibilità e la capacità dei collaboratori ad interagire come un processo e risorsa aziendale. Le generazioni più giovani in particolare sono nate in un mondo permeato dalla tecnologia e usano le piattaforme social con grande naturalezza. Spesso rappresentano il modo ideale in cui le informazioni vengono generate, fornite, scambiate e inoltrate.

Sempre più fornitori ERP stanno seguendo questa tendenza, integrando elementi dei social media nelle loro soluzioni software. Le informazioni possono quindi essere messe a disposizione centralmente e distribuite ovunque, in qualsiasi momento, dove e quando sono necessarie. Le informazioni relative agli individui, provenienti dai social media, possono essere integrate e raccolte nel CRM, ad esempio come base per creare nuove campagne pubblicitarie.

La tendenza emersa più di recente è la crescente difficoltà dei produttori di ERP nel soddisfare tutte le richieste delle imprese riguardo a CRM, GeScorte e PCP, a causa della crescente complessità. Questa circostanza ha aperto la porta a produttori di software specializzati che sono stati in grado di assicurarsi un posizionamento di nicchia, specializzandosi molto abilmente in un singolo modulo e potendo così offrire soluzioni all'avanguardia. Un esempio di spicco è certamente la società di software Salesforce, che si è specializzata in soluzioni CRM completamente integrabili e nel cloud computing, offrendo sul mercato sistemi di grande successo.

6.4 Scelta di un sistema ERP

Considerando le circa 250 diverse soluzioni standard ERP e le circa 1000 soluzioni di software aziendale, presenti nella sola Svizzera e specifiche a determinati settori, la selezione della soluzione ERP appropriata pone spesso importanti difficoltà a un'azienda.

Al momento di prendere una decisione sull'introduzione di un nuovo sistema ERP, è consigliabile utilizzare un processo di valutazione standardizzato; se il know-how necessario non è disponibile all'interno dell'azienda, si può ottenerlo esternamente. Il successo o il fallimento in questa fase è solitamente determinato dal porsi le domande giuste, insieme al tempo sufficiente per trattarle con accuratezza. Oltre al management, devono essere coinvolti in questo processo anche altri responsabili e portatori di know-how.

Quali criteri devono essere considerati da ciascuna singola azienda dipende, tra l'altro, dalla decisione di scegliere una soluzione on-premise, cioè un'installazione fissa sul server aziendale, o una soluzione software-as-a-service nel cloud. Non tutte le aziende vogliono esternalizzare i propri dati; inoltre, vincolarsi a un fornitore è spesso visto con scetticismo, soprattutto per quanto riguarda la scelta, l'operatività e la manutenzione di una soluzione ERP. Molte imprese temono di doversi assumere costi eccessivi per l'acquisto come pure per i costi di licenza e di aggiornamento.

Qual è la durata d'utilizzo di un software ERP?

I cicli di cambiamento degli ERP indicano per quanto tempo le aziende utilizzano i loro sistemi ERP prima di passare a nuove soluzioni. Questo lasso di tempo non si riferisce agli aggiornamenti delle release, ma al passaggio a un sistema completamente nuovo. La tendenza è che i cicli di cambiamento degli ERP diventino sempre più brevi.

Una delle ragioni principali di questa tendenza, cioè la durata d'utilizzo sempre più breve del precedente sistema ERP, è da ricercare in un cambiamento di paradigma. Questo è dovuto al fatto che sempre più aziende stanno optando per soluzioni ERP cloud. Contrariamente alle soluzioni on-premise (installazione fissa sui propri

server), che prima della messa in funzione richiede elevati costi d'installazione, il software ERP cloud è immediatamente pronto per l'uso.

Altri fattori che portano ad avere cicli di cambiamento sempre più brevi sono il trasferimento dei dati sempre più semplificato, i periodi di introduzione più brevi e la tendenza alla riduzione dei costi.

6.5 Fornitori di software ERP e di soluzioni ERP

Poiché oggi i sistemi ERP non sono più progettati solo per le grandi imprese, il numero di produttori e fornitori è diventato molto alto. Le offerte differiscono considerevolmente in termini di tipo, prestazioni e portata.

Fondamentalmente, ci sono due tipi di soluzioni ERP:

Da un lato, ci sono le **soluzioni ERP cross-industry** (indipendenti dal settore) che coprono la maggior parte delle aree funzionali e dei processi operativi all'interno di un'azienda; si tratta di soluzioni molto meno complesse e meno costose dei sistemi specifici al settore. Sono utilizzate soprattutto dalle PMI.

D'altro lato, ci sono le **soluzioni software ERP specifiche per un settore**. Queste versioni specializzate sono sempre più utilizzate, per esempio, da imprese commerciali, istituti finanziari, compagnie di assicurazione e, in particolar modo, nell'industria. In questo caso stiamo parlando di soluzioni molto complesse, che sono in grado di mappare i processi in modo molto preciso e anche di gestire un'intera catena (ad esempio la catena di approvvigionamento dal fornitore al cliente). Nel caso delle soluzioni ERP specifiche per un settore, c'è anche la tendenza a preferire soluzioni ERP cross-company, che collegano più imprese. Queste soluzioni non sono più utilizzate solo all'interno dell'azienda, come accadeva in passato, ma sono collegate in rete con tutte le imprese della catena di approvvigionamento. Questo significa che non solo i fornitori diretti, ma anche i loro fornitori sono integrati nel sistema. In questo modo, i colli di bottiglia o i cambiamenti nelle condizioni di produzione e consegna possono essere rilevati con ampio anticipo e, quindi, possono essere considerati tempestivamente nella pianificazione.

6.6 Produttori di ERP

Il mercato ERP per le grandi imprese e gruppi è chiaramente dominato solo da pochi produttori di ERP:

- SAP
- Oracle
- Microsoft

Il mercato delle medie imprese, invece, è distribuito su un numero molto più grande di produttori.

Questi includono, tra gli altri:

- Abacus
- Sage
- Proffix
- IFS

6.7 Costi di un sistema ERP

Quando un'impresa valuta l'introduzione di un sistema ERP, di solito non trascorre molto tempo prima che venga posta in discussione la trattanda «costi». Per poter stimare i costi dell'introduzione di un ERP, si devono prendere in considerazione diversi fattori. Questi sono, ad esempio:

- Dimensione dell'impresa
- Requisiti posti al software ERP
- Modello operativo

Oltre ai costi esterni per il sistema ERP, si devono considerare anche ulteriori costi per

- Prestazioni di consulenza
- Formazione
- Manutenzione del sistema ERP

Questi costi possono essere considerevoli. Per evitare un sovraccarico finanziario in relazione alla selezione e all'implementazione del sistema ERP, si dovrebbero pianificare sia i costi a breve termine che i costi a lungo termine. Sfortunatamente, non ci sono valori di riferimento generalmente validi per i costi. I valori derivanti dall'esperienza forniscono delle stime solo approssimative, che possono essere considerate nei propri calcoli. Per le grandi imprese, i costi di un tale progetto ERP possono ammontare a più di mezzo milione di franchi (per i grandi gruppi tale ammontare può raggiungere rapidamente le otto cifre); le imprese di media dimensione devono prevedere un budget di diverse centinaia di migliaia di franchi, mentre le piccole imprese devono attendersi costi da cinque cifre.

L'alternativa a questi elevati costi di investimento, specialmente per le piccole e medie imprese, sono i modelli di noleggio/abbonamento del software (Software-as-a-Service, SAAS), che sono spesso offerti dai fornitori di software ERP basati sul cloud. Con questa forma di servizio, vengono meno gli elevati costi d'introduzione e l'abbonamento mensile che ne consegue è relativamente più gestibile per le piccole imprese. Allo stesso tempo, l'azienda può crescere con il software e adeguare il software alla propria crescita.

6.8 Struttura del sistema ERP/ERD

I moderni sistemi ERP sono tutti costruiti secondo lo stesso concetto. Le relazioni sono utilizzate per collegare tabelle, dati e informazioni. La visualizzazione della rete di tabelle e dati risultanti dai collegamenti relazionali è denominata **diagramma ERD** (Entity Relationship Diagram, diagramma entità relazione) e rappresenta tutte le informazioni memorizzate nel sistema ERP.

Un sistema ERP può rapidamente crescere e comprendere molte tabelle, come mostra la figura 8. È molto importante strutturare il database in modo ottimale, poiché strutture efficienti consentono di evitare la ridondanza dei dati e contribuiscono a realizzare un'elevata performance nell'accesso ai dati.

Struttura di elaborazione nei sistemi ERP

La visualizzazione e l'elaborazione delle informazioni nei sistemi ERP avviene su diversi livelli. Per l'utente esperto e gli esperti che lavorano a livello concettuale è importante conoscere le funzioni e le differenze tra questi livelli, poiché ciò è cruciale nella ricerca dei problemi e nell'ulteriore sviluppo del software. Oggi i collaboratori non hanno più bisogno di una comprensione approfondita dei processi. L'alto livello di integrazione dei processi implica che i collaboratori non devono più occuparsi del dare e dell'avere, dei conti o di tutte le registrazioni dell'imposta sul valore aggiunto. Questo viene realizzato attraverso gruppi di registrazioni predisposte in modo differenziato e presentate automaticamente in background (a condizione che i dati di base e i parametri siano stati impostati correttamente).

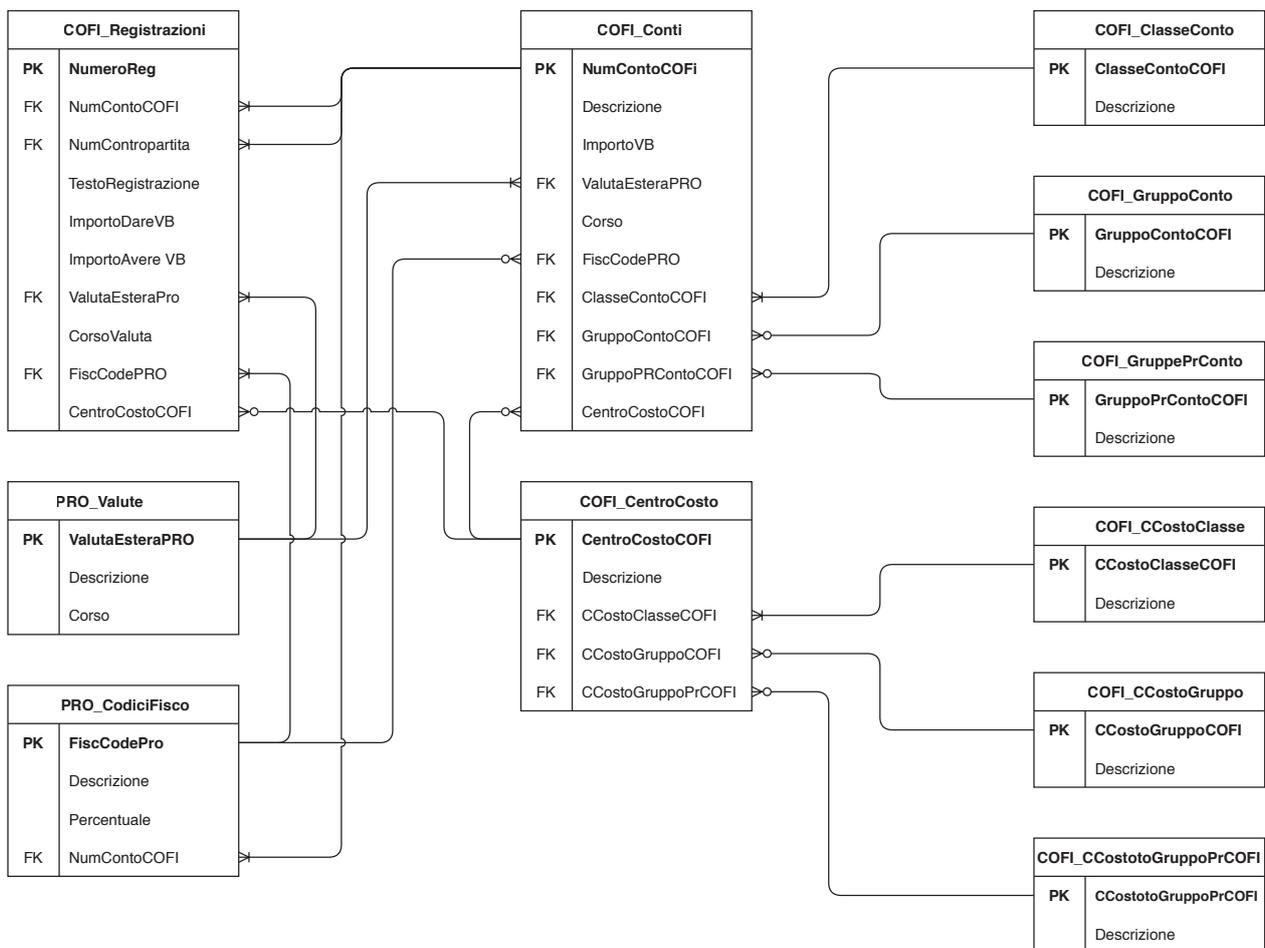


Figura 8: Diagramma ERD (tradotto da Proffix Px5 Finanzbuchhaltung, 2022)

Per la comprensione da parte di un utente esperto, è determinante conoscere i livelli procedurali relativi al software e, in caso di adattamenti, saper distinguere tra approcci di implementazione buoni e cattivi.

La maggior parte della letteratura si riferisce all'inserimento dei dati utilizzando i termini interfaccia utente (user interface) o GUI. Questi dati vengono normalmente inseriti nel sistema mediante maschere o formulari (1° e 2° livello della rappresentazione di seguito presentata). Nei processi di registrazione, questi dati vengono modificati (calcolati, copiati, mutati) e scritti in nuove tabelle. Nel caso della contabilità dei salari, nella maggior parte dei casi questi dati non sono più modificabili. Pertanto, è importante essere in grado di controllare prima del processo di registrazione definitiva (al livello 3) ciò che alla fine sarà registrato e, allo stesso tempo, garantire che i dati inseriti e registrati come costi, scadenze, materiali e dimensioni (centri di costo, prodotti, ecc.) siano verificabili e referenziabili.

Tutti i software funzionano in modo tale che le informazioni vengono scambiate tra i singoli livelli. Normalmente, l'informazione viene trasferita dall'alto verso il basso e, tra i singoli livelli, i dati non possono più essere modificati. Tutti i sistemi ERP hanno in comune il fatto che dopo la registrazione (dal 3° al 4° livello) le informazioni inserite vengono (per lo più) cancellate nel 3° livello e vengono messe a disposizione (registrate) nel 4° livello senza ulteriore possibilità di modifica. Non è possibile, o lo è solo con grande difficoltà, cancellare le informazioni nel 4° livello, poiché i dati possono essere scritti in più tabelle (per esempio dalla contabilità dei salari alle registrazioni riepilogative della contabilità finanziaria e simultaneamente alla contabilità dei costi).

A dipendenza del sistema, i dati del 4° livello possono essere ritrasferiti al 3° livello utilizzando la funzione di storno. A causa delle registrazioni riepilogative, viene riscritta o stornata una serie di record. I record modificati

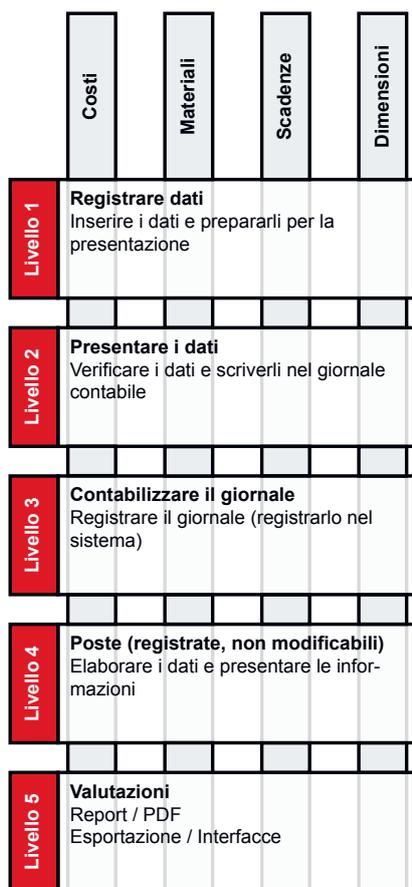


Figura 9: Livelli di elaborazione software (Zäch, 2021)

non vengono fisicamente cancellati, ma contrassegnati come inattivi o stornati (ossia con una registrazione di compensazione con segno invertito).

È quindi fondamentale essere in grado di controllare tutte le informazioni già al 3° livello, prima di inviarle al 4° livello. Per questo motivo, vengono messi a disposizione i cosiddetti rapporti provvisori o protocolli in modo da controllare le registrazioni e, ad esempio, verificare i conti COFI.

Quando si sviluppano nuove funzioni è importante considerare che esse intervengano nel software al giusto livello e che sia possibile risalire ai processi di modifica delle registrazioni. Per questo motivo, i produttori di ERP proteggono sempre più spesso parti del software rendendolo non modificabile, in modo che gli errori di sviluppo non abbiano impatti sostanziali. Questo riguarda soprattutto le registrazioni nel giornale contabile COFI e le registrazioni del materiale nella contabilità di magazzino.

Per questo motivo, si dovrebbe sempre ricorrere alle conoscenze e all'esperienza degli esperti per modificare i processi contabili.

Esempio di processo di registrazione dalla contabilità dei debitori alla COFI

Nella figura 10 sono mostrate le relazioni tra le principali tabelle e le elaborazioni nella contabilità dei debitori.

- Le informazioni generali vengono memorizzate nei dati di base (ad esempio informazioni su indirizzo, valuta, gruppi di registrazioni, gruppi di registrazione dei debitori, gruppi di registrazioni IVA, dimensioni come listino prezzi / centri di costo e unità di costo, ecc.).

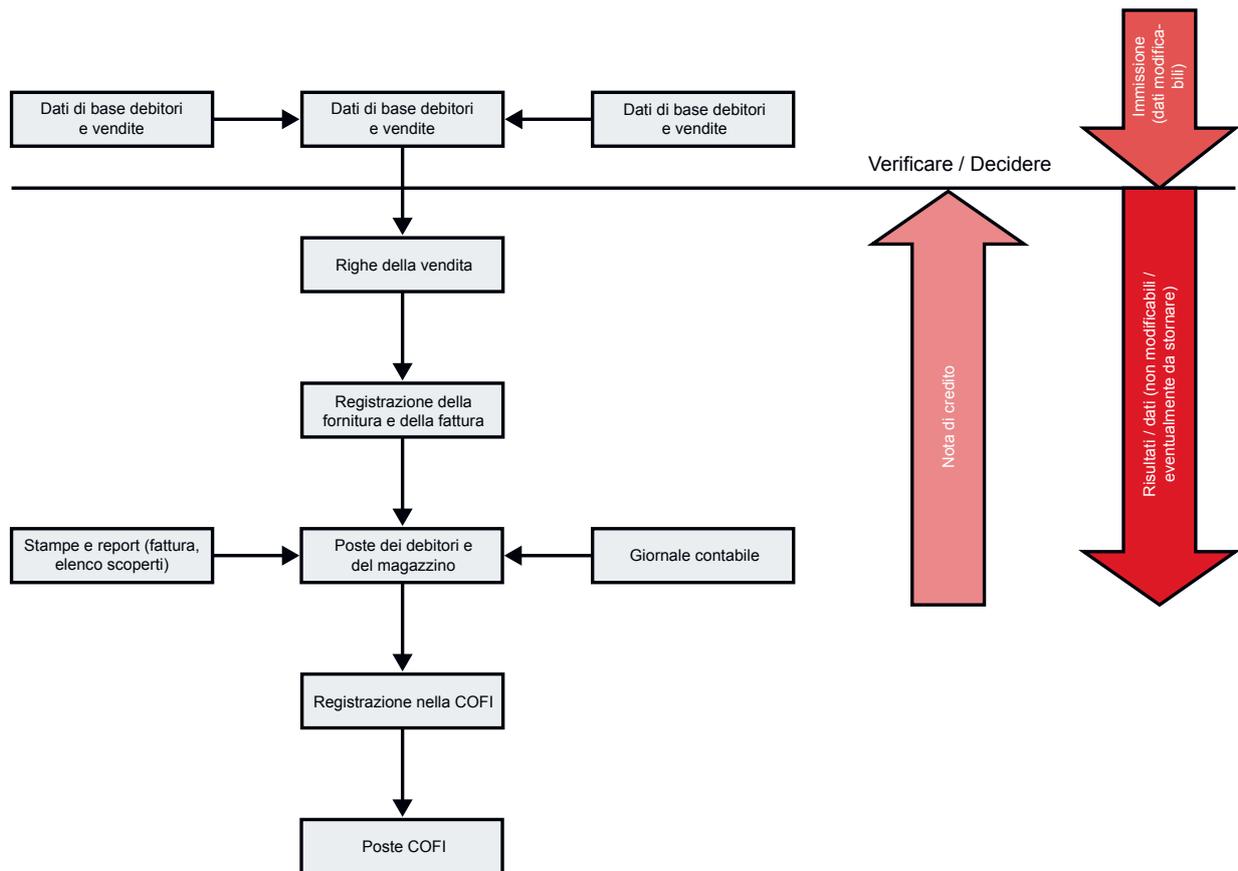


Figura 10: Struttura della contabilità dei debitori (Zäch, 2021)

- I gruppi di registrazioni accedono a loro volta alle matrici di registrazioni, che a loro volta fanno riferimento ai conti COFI.
- Nella tabella di Intestazione dell'ordine di vendita, vengono trasferite le informazioni attuali del cliente al momento dell'ordine corrente (di solito come una copia dei dati di base).
- Nella tabella Righe dell'ordine di vendita vengono assegnate al cliente le registrazioni degli articoli (di solito come copia dei dati di base).
- La funzione Registra (consegna e/o fatturazione) esegue la registrazione e trasferisce l'ordine alle consegne effettuate e alle fatture registrate. Vengono, inoltre, calcolate e contabilizzate le registrazioni COFI (ad esempio scarico del magazzino, IVA dovuta e ribassi). Una consegna la cui fattura è già stata contabilizzata non può più essere modificata, ma deve essere compensata mediante l'emissione di una nota di credito.
- La maggior parte dei report nella contabilità dei debitori si basa sulle forniture e fatture contabilizzate.
- Ogni fornitura e fattura viene documentata nel giornale COFI e negli altri conti ausiliari interessati (magazzino, IVA, debitori).
- A dipendenza del sistema ERP, è possibile definire se i dati provenienti dal modulo delle vendite vengono automaticamente registrati nelle voci (poste) dei debitori e del magazzino e come poste COFI nella contabilità finanziaria. In caso di registrazione automatica si parla di sistema integrato.

6.9 Limiti e prospettive dei moderni sistemi ERP

I limiti dei sistemi ERP di oggi risiedono nel fatto che considerano principalmente l'interno, cioè l'azienda, e quindi rappresentano simbolicamente un cosmo interno all'azienda. Allo stesso tempo, questa prospettiva mostra una fotografia istantanea dell'azienda, con lo svantaggio che determinate situazioni non possono essere riprodotte retrospettivamente (ad esempio i calcoli dei bisogni in quantità della logistica, le carenze di quantità, i fabbisogni ad una determinata data o la situazione prima del processo di registrazione). La sola prospettiva interna e l'immagine istantanea dell'azienda non sono più sufficienti in un mondo globalizzato.

Nel migliore dei casi, la prospettiva esterna viene assicurata unidirezionalmente attraverso interfacce ai processi rilevanti per l'azienda come le operazioni di pagamento secondo ISO 20022 (fattura QR, SEPA), EDI / ELM / operazioni doganali. Poiché nella prospettiva esterna i bisogni delle aziende sono molto diversificati, assume grande importanza lo sviluppo di interfacce. I produttori di software si sforzano sempre più di mappare la prospettiva esterna attraverso interfacce nel sistema ERP e di progettare le interfacce in modo riproducibile e trasparente.

Diverse prospettive stanno al centro delle interfacce:

1. Interfacce di database aperte e mappatura intelligente delle informazioni da parte dei produttori di software. Collegamenti già predisposti con i sistemi di concorrenti o di terze parti.
2. Integrazione di informazioni esterne nel sistema ERP (ad esempio dati demografici, sociali o congiunturali).
3. Creazione di visualizzazioni configurabili e personalizzabili in dashboard / cockpit per consentire ai collaboratori di disporre di una visione personalizzata dell'importanza e della priorità dei compiti in base all'area di responsabilità e alla funzione.
4. Automatizzazione bidirezionale dello scambio di dati con i partner commerciali per aumentare l'efficienza, evitare dati ridondanti, migliorare e controllare costantemente la qualità dei dati.
5. Esportazione di informazioni per scopi interni ed esterni (ad esempio valutazioni interne o notifiche a partner come autorità e assicurazioni)

L'ulteriore sviluppo dei sistemi ERP non è mai concluso perché anche le esigenze delle imprese sono in continua evoluzione. Parallelamente allo sguardo verso l'esterno, alla globalizzazione e alle crescenti incertezze del mondo VUCA, cresce la necessità di confrontare le informazioni, riconoscere i trend e, nel migliore dei casi, essere in grado di avviare i provvedimenti operativi giusti, al momento giusto e attraverso previsioni significative.

Mentre la gestione dei dati e delle informazioni viene semplificata per l'utente anche attraverso la standardizzazione delle interfacce e dei tool, le strutture dei dati delle organizzazioni diventano più complesse e la quantità di dati da gestire cresce in modo esponenziale. Il monitoraggio e la gestione dei dati, così come la loro qualità, avranno in futuro maggiore priorità e porranno nuove sfide alle organizzazioni.

In queste organizzazioni, i compiti del controller cambieranno sostanzialmente attraverso la valutazione globale e l'analisi dei dati, l'ulteriore sviluppo mediante nuovi approcci e prospettive di controlling. Il controller diventerà l'analista dei dati e delle informazioni operative e strategiche, colui che guida lo sviluppo aziendale operativo e strategico.

7 Business Intelligence (BI)

7.1 Definizione di Business Intelligence

La Business Intelligence è un processo guidato dalla tecnologia per analizzare i dati e presentare informazioni pratiche che aiutano i quadri dirigenti, i manager e gli altri utilizzatori finali a prendere decisioni aziendali fondate su solide basi.

La Business Intelligence come processo include tutte le forme di raccolta sistematica, valutazione, analisi e presentazione delle informazioni. La domanda sempre in primo piano è: come possono le aziende gestire grandi quantità di dati e usare questi dati per un miglior processo decisionale?

In questo contesto, i dati devono sempre essere intesi in senso stretto come informazioni che vengono utilizzate per il processo decisionale.

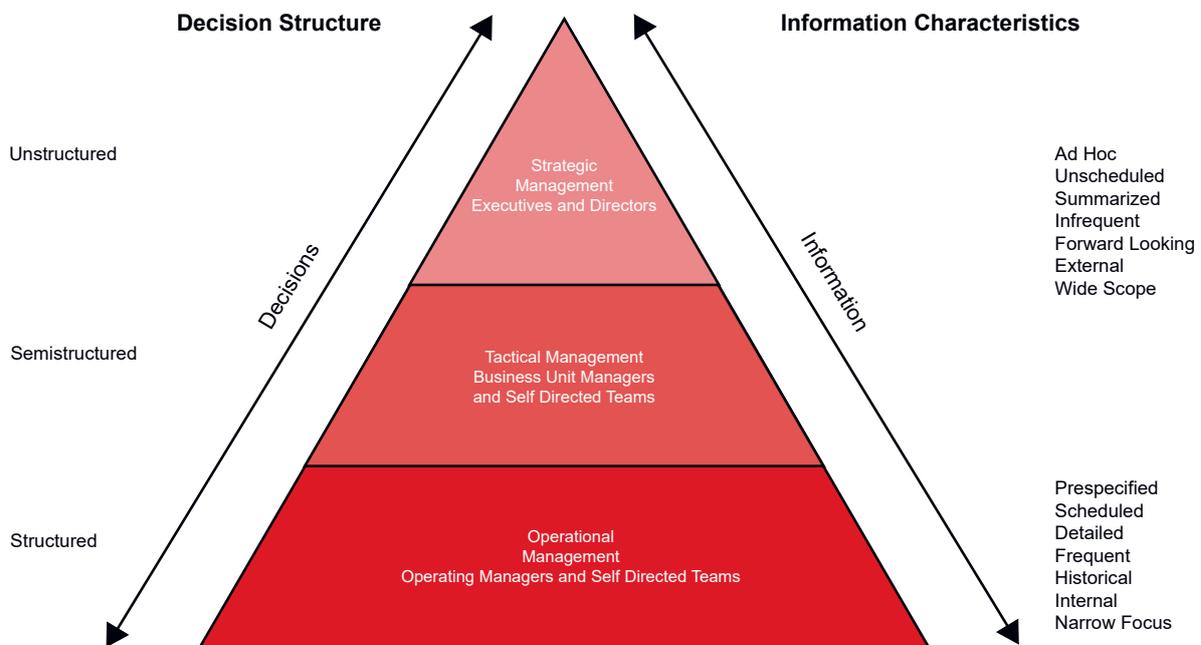


Figura 11: Livello decisionale e informazione (O'Brien, 2010)

Dalla figura sopra riportata, si può osservare che la BI è di fondamentale importanza per la gestione strategica. Questo va visto nel contesto dell'elaborazione delle informazioni e delle decisioni che ne derivano. Per semplificazione, si può considerare il seguente **esempio**: il management operativo riceve l'informazione molto strutturata, che indica che un cliente è in ritardo con i suoi pagamenti. La decisione che ne deriva è relativamente semplice: il cliente deve ricevere un richiamo di pagamento. Per prendere questa decisione non è necessario ricorrere alla BI, poiché l'informazione è disponibile in modo molto strutturato.

La situazione è diversa per quanto riguarda le decisioni strategiche, in cui l'informazione è di solito poco strutturata e risulta pertanto molto più complesso prendere una decisione. Ecco un altro esempio:

Le aziende sono oggi sommerse da dati. Dispongono di dati sui profili dei clienti, ERP, programmi di fidelizzazione, analisi di mercato, ecc. Vengono immagazzinati più dati di quanto l'essere umano sia in grado di analizzare. Inoltre, la capacità di stoccaggio dei dati continua ad aumentare e i costi della capacità di stoccaggio

continua a diminuire. Quale impatto ha questo sviluppo per la BI? Conclusione: non appena le informazioni diventano più destrutturate, il processo decisionale diventa molto più complesso.

Diamo un breve sguardo allo sviluppo storico. Il termine «BI» è stato utilizzato per la prima volta nell'articolo di Hans Peter Luhn «A Business Intelligence System» pubblicato nel 1958 ed è stato ripreso e coniato dal Gartner Group nei primi anni '90.

La business intelligence si è sviluppata inizialmente in modo relativamente indipendente dal data warehousing, dall'Enterprise Content Management e dal Knowledge Management. I sistemi di business intelligence erano intesi come sistemi informativi analitici, specialmente nel mondo di lingua tedesca. Nel frattempo, il data warehousing e i sistemi di business intelligence hanno sperimentato un'enorme crescita, una crescente importanza per il management delle informazioni, nonché diversi cambiamenti di paradigma ed ampliamenti.

7.2 Perché la Business Intelligence è importante?

In origine, gli strumenti di BI erano utilizzati principalmente da analisti di dati e da altri esperti IT che eseguivano analisi e creavano rapporti per gli utenti aziendali attraverso i risultati delle query. Le aree classiche di utilizzo sono principalmente gli ambiti aziendali della finanza e del controlling. Sempre più spesso, però, i manager e i collaboratori utilizzano le piattaforme di BI in modo autonomo, in parte grazie allo sviluppo della BI self-service, dei tool di rilevamento dati e dei dashboard per il rilevamento dei dati.

Per far fronte alla quantità sempre crescente di dati, negli ultimi anni si sono affermati sistemi di business intelligence supportati dall'IT.

Soprattutto in relazione con i big data, la componente analitica diventa sempre più importante (tableau.com, 2021). Inoltre, c'è il cambio di paradigma: da un sistema automatizzato per la diffusione delle informazioni ai diversi reparti delle organizzazioni industriali, scientifiche e governative alla misurazione delle prestazioni (performance) aziendali.

7.3 Perché utilizzare la business intelligence?

I benefici potenziali degli strumenti di business intelligence includono l'accelerazione e il miglioramento dei processi decisionali, l'ottimizzazione dei processi interni all'azienda, l'aumento dell'efficienza operativa, la realizzazione di nuove vendite e l'ottenimento di un vantaggio competitivo sui concorrenti.

Se consideriamo la figura 12 che presenta l'architettura di BI, osserviamo che la BI accede a più fonti (elencate a sinistra), le riunisce in un data warehouse, quindi elabora i dati in diversi cubi OLAP suddivisi tematicamente e, infine, mette a disposizione i dati attraverso diversi frontend. Il grande vantaggio rispetto a un ERP puro è che l'accesso non è limitato solo ai dati statici. Con la BI si ha l'opportunità di percorrere strade innovative e anche di raccogliere dati da programmi di fidelizzazione, social network, ecc., elaborarli in modo strutturato e utilizzarli per prendere decisioni previsionali basate sul riconoscimento di modelli piuttosto che decidere sempre in modo reattivo. Questo mette l'azienda nella posizione di sapere cosa vorrà il cliente in futuro.

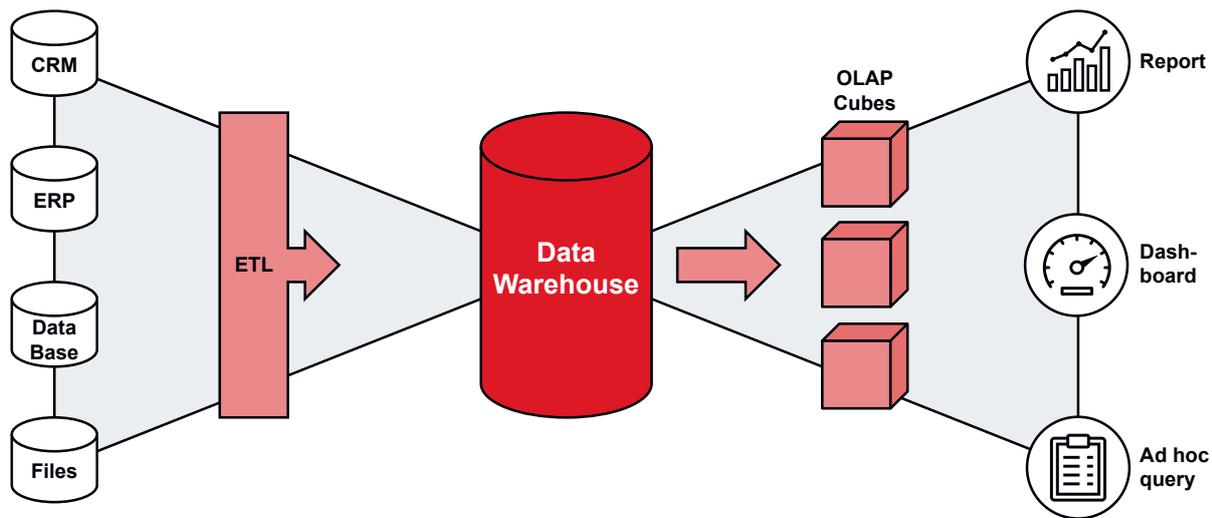


Figura 12: BI back / Frontend

7.4 Software di business intelligence

Le soluzioni di business intelligence hanno una miriade di funzioni, ma in generale, possono essere utilizzate per eseguire tutti o alcuni dei seguenti compiti:

- Confronto e regolazione di diversi risultati tramite test A/B o test multivariati
- Previsione dei risultati attraverso la modellazione predittiva
- Approfondimenti su modelli o relazioni
- Presentazione di scenari di causa-effetto

Quando si confrontano le soluzioni di business intelligence, assicuratevi che includano tutte o almeno alcune delle funzioni di seguito elencate.

7.5 Online Analytical Processing (OLAP) / Business Analytics

OLAP comprende l'intera gamma di preparazione, elaborazione e messa a disposizione di dati per il reporting. Il nucleo delle soluzioni BI è formato da tool OLAP con i quali è possibile analizzare dati multidimensionali in tempo reale o contemporaneamente ad altri procedimenti di analisi. In generale, più possibilità si hanno per isolare e visualizzare i dati da un dataset (il cosiddetto cubo OLAP), più flessibile è l'interpretazione dei dati.

La business analytics è un processo di raccolta ed elaborazione di dati per la gestione strategica dell'azienda. È un processo digitale in cui vengono raccolti grandi volumi di dati, che vengono arricchiti, elaborati e valutati con metodi statistici e metodi iterativi.

La business analytics raccoglie e analizza i dati come la business intelligence, riconosce le correlazioni tra i dati e i modelli nei big data mediante l'utilizzo di algoritmi e, soprattutto, esegue analisi predittive per determinare la probabilità di risultati futuri. Come per la business intelligence, questi rapporti sono visualizzati in dashboard personalizzati dall'utente

7.6 Report e query

Il reporting e le query sono un altro punto fondamentale del software di BI che genera e ordina dati complessi. Nel frattempo, le moderne soluzioni di BI dispongono di query intuitive che permettono agli utenti di mappare dati complessi anche senza dover scrivere istruzioni SQL o disporre di conoscenze nella programmazione. OLAP prepara automaticamente i record per ogni nuova query. Questo permette alle aziende di utilizzare i tool di BI in tutta l'impresa e oltre il tradizionale regno del reparto IT.

7.7 Dashboard digitali o visivi

La visualizzazione dei dati è la fase finale del reporting ed è, in effetti, uno degli elementi più identificabili delle soluzioni di BI. I manager e gli utenti non tecnici spesso vedono la BI esclusivamente da questo punto di vista. Le visualizzazioni avanzate includono rapporti geo spaziali, mappe di calore (heat map), ecc.

7.8 Integrazione

Le soluzioni di BI si basano su più fonti di dati per fornire intuizioni significative e, pertanto, l'integrazione è di fondamentale importanza. Infatti, molti venditori offrono tool di BI collegati con altre applicazioni aziendali come CRM, automazione della forza vendite (salesforce) e software per il carrello della spesa.

8 Big Data

Quando si parla di big data (grande massa di dati o megadati), si intendono volumi di dati così grandi da non poter essere elaborati con normali software e hardware. Come nascono questi volumi di dati così grandi? I tempi attuali sono caratterizzati da una frenetica raccolta di dati. Al giorno d'oggi, tutto quello che facciamo viene memorizzato da qualche parte. Ad esempio, facendo acquisti online o cercando qualcosa su internet, inserendo un percorso in un software di navigazione o facendo una transazione finanziaria, una telefonata o andando in palestra; non importa cosa, tutto viene memorizzato. Questo porta a dover gestire enormi volumi di dati che devono anche essere elaborati. Si stima che la stessa quantità di dati che l'umanità ha raccolto dalla sua nascita fino al 2002 sia stata creata in dieci minuti nel 2014 e, in seguito, sia raddoppiata ogni due anni.

Tuttavia, con big data non si intende soltanto la quantità di dati in sé, ma soprattutto la loro analisi per riconoscere o ricavare modelli da tali dati. Un chiaro esempio di ciò è lo shopping online, che si sperimenta quasi quotidianamente. Una persona decide di comprare l'articolo X. Non appena lo aggiunge al suo carrello, vede subito la proposta di diversi prodotti che hanno comprato gli altri clienti che hanno acquistato lo stesso articolo che ha messo nel carrello (e quindi potrebbero piacere anche a questa persona?). Tali informazioni vengono calcolate in tempo reale e indicano il livello di probabilità che anche questo articolo possa piacere alla persona che sta eseguendo l'acquisto.

8.1 Dove si usano i big data?

- **Ricerca medica:** analizzando grandi volumi di dati, i medici possono trovare le migliori soluzioni terapeutiche e i migliori piani di cura per i loro pazienti.
- **Industria:** grazie all'analisi dei dati dei propri macchinari, le imprese possono aumentare l'efficienza della loro produzione e quindi lavorare in modo più sostenibile.
- **Economia:** i big data permettono alle aziende di conoscere meglio i loro clienti e di adattare meglio le loro offerte.
- **Energia:** per adeguare il consumo di energia alle esigenze individuali, è necessario conoscere tali esigenze. I dati di consumo raccolti assicurano un approvvigionamento energetico sostenibile a lungo termine.
- **Marketing:** i big data sono spesso utilizzati nel settore del marketing per scopi di targeting. L'obiettivo è di solito quello di potenziare il customer relationship e migliorarne l'implementazione in varie misure di marketing.
- **Lotta contro il crimine:** anche il governo e i servizi di sicurezza statali ricorrono ai big data, ad esempio nella lotta contro il terrorismo.
- **Politica:** i big data sono utilizzati nelle campagne elettorali per attirare nuovi potenziali elettori, per prevedere i risultati e, se necessario, influenzarli.

8.2 Valutazione dei big data

In questo contesto possiamo fare il collegamento con il capitolo precedente. La business intelligence è il processo attraverso il quale i big data possono essere raccolti, analizzati e utilizzati. I vantaggi descritti precedentemente emergono attraverso la valutazione sistematica dei big data nelle soluzioni di BI, creando quindi uno strumento importante nella guerra per le quote di mercato. Presentando e prevedendo opportunità e sfide, la business intelligence permette alle organizzazioni di utilizzare in modo ottimale i loro big data per il loro successo.

8.3 Utilizzo dei big data

Come procede un'impresa se vuole utilizzare i big data? La priorità assoluta è certamente una buona pianificazione. Ma quali elementi bisogna prendere in considerazione nella pianificazione dei big data? Gli esperti parlano spesso delle 5 V:

- **Volume:** pianificare la gestione del volume di dati previsti. Come si vogliono memorizzare questi dati e, soprattutto, dove?
- **Varietà:** identificare il luogo / la posizione di tutte le fonti di dati e valutare i migliori tool per preservare o raccogliere i dati.
- **Velocità:** procurarsi la tecnologia appropriata per analizzare i dati e creare con essi analisi significative.
- **Veridicità:** assicurarsi assolutamente che i propri dati siano puliti e corretti.
- **Valore:** progettare un ambiente big data che permetta di creare analisi significative e di distinguere ciò che è importante da ciò che non lo è. Non tutti i dati raccolti sono importanti per l'azienda.

8.4 Dove memorizzare i big data?

Fondamentalmente ci sono due possibilità:

Data Warehouse

In un data warehouse si possono immagazzinare grandi quantità di informazioni e renderle disponibili per ulteriori valutazioni. Il data warehouse riunisce diversi dati in formati e strutture uniformi, basandosi sul tipo di analisi che deve essere svolta. Questo significa che un data warehouse contiene esclusivamente dati aggregati come le cifre chiave o i dati transazionali. Solo i dati che sono già stati elaborati e che servono a uno scopo prestabilito possono essere memorizzati. È difficile modificare le informazioni memorizzate in un data warehouse.

Data Lake

Il data lake è un repository di dati in grado di archiviare set di dati di grandi dimensioni (big data), provenienti da molte fonti, in un formato grezzo e granulare. Può memorizzare dati strutturati, semi-strutturati o non strutturati. Questo significa che i dati possono essere memorizzati in un formato più flessibile per un successivo utilizzo. Quando si memorizzano i dati, il data lake li associa a identificatori e tag di metadati per velocizzarne il successivo recupero. I data scientist possono utilizzare i data lake per richiamare, preparare e analizzare i dati più rapidamente e con maggiore precisione. Gli esperti di analisi possono richiamare i dati da questo pool di dati in base alle necessità, per diversi casi d'uso (use cases), come la sentiment analysis (analisi del sentiment) o il rilevamento di frodi.

8.5 Soluzioni software per i big data

Tra i tool più noti ci sono sicuramente:

- Hadoop: soluzione software open source per big data (Apache Software Projects)
- MapReduce (Apache Software Projects)
- Spark: soluzione software open source per big data (Apache Software Projects)

8.6 Critica ai big data

A dispetto di tutti i vantaggi dei big data, si pongono anche alcune questioni critiche:

- A chi appartengono questi dati?
- Vogliamo che solo poche aziende controllino questi dati? Un tipico esempio è google. Google dispone di un'enorme quantità di dati e viene criticata per la sua posizione monopolistica e il conseguente rischio di un utilizzo improprio dei dati.
- Vogliamo che il comportamento umano sia prevedibile?
- Cosa succede se questi dati cadono nelle mani sbagliate?

In questo contesto, i big data possono essere visti sia come una maledizione che come una benedizione. Ciò dipenderà principalmente da come questi dati saranno utilizzati in futuro. Tuttavia, nonostante tutte le critiche, bisogna anche dire che i big data possono essere molto utili. Ad esempio, grazie ai big data è stato possibile fare importanti progressi nella ricerca contro il cancro, che non sarebbero stati possibile senza avere a disposizione questa moltitudine di dati. Allo stesso modo, l'approvvigionamento energetico e i trasporti vengono costantemente ottimizzati e sviluppati attraverso la valutazione di grandi quantità di dati, a beneficio di tutta la popolazione. In ogni caso, è certo che il nostro futuro sarà plasmato dai big data.

9 SQL

9.1 Introduzione

SQL è l'abbreviazione di Structured Query Language (linguaggio per interrogazioni strutturate) ed è un linguaggio dei database che consente di creare strutture di database relazionali, nonché di eseguire elaborazione e query su dati esistenti. SQL è composto da istruzioni in inglese ed è prevalentemente basato sulla lingua inglese.

Con SQL è possibile creare query che riprendono i dati da database relazionali e li rendono disponibili all'utente tramite un software utente, rispettivamente un front-end.

SQL è oggi lo standard comune per la realizzazione di query nei database relazionali. Oltre a SQL, esiste anche il linguaggio di database MDX, utilizzato nei database multidimensionali.

9.2 Comando SELECT

Il comando SQL SELECT viene di regola utilizzato come base per le query SQL che devono essere eseguite su un database. Mediante questo comando è possibile selezionare i dati da una o più tabelle collegate attraverso una JOIN (collegamento, relazione).

SQL SELECT, esempio

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Per mostrare l'esempio dell'istruzione SQL SELECT viene utilizzata la tabella Datibaseclienti, sopra riportata. Se si vogliono estrarre da questa tabella solo i campi Nome e Luogo, l'istruzione SQL da eseguire è la seguente:

```
SELECT NOME, LUOGO FROM DATIBASECLIENTI
```

Il risultato della query SQL SELECT è il seguente:

Vorname	Ort
Monika	Niederglatt
Thomas	Zürich
Felix	Niederglatt
Regula	Zürich
Reto	Niederglatt
Stefano	Flaach
Barbara	Diesldorf

9.3 Comando DISTINCT

Il comando SQL DISTINCT viene indicato direttamente dopo il comando SQL SELECT di una query. Il comando DISTINCT viene utilizzato per eliminare i valori duplicati dal risultato della query.

SQL DISTINCT, esempio

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Con la query SQL SELECT DISTINCT possiamo verificare se i dati contenuti nelle colonne presentano delle duplicazioni e, qualora ce ne fossero, possiamo eliminare tali ridondanze. Nel nostro esempio della tabella Datibaseclienti, il campo ID_CI non dovrebbe presentare alcuna ridondanza poiché il numero di cliente (ID_CI) deve essere univoco per ciascun cliente (di solito un ID è sempre una chiave primaria e non può contenere valori duplicati).

Il comando per creare l'elenco dei dati di base dei clienti senza ridondanze è il seguente:

```
SELECT DISTINCT ID_CL FROM DATIBASECLIENTI
```

Il risultato di questa query SQL è il seguente:

KD_ID
4721
5397
6423
8174
9371
7348

Naturalmente, il problema della duplicazione del numero di cliente (ID_CI) non è ancora risolto. Da un lato, è necessario ripulire i dati anagrafici e dall'altro è indispensabile costruire i dati anagrafici dei clienti in modo tale che non sia possibile duplicare il numero di cliente (ID_CI).

9.4 Comando WHERE

Con il comando SQL WHERE è possibile selezionare record in base a criteri di ricerca mirati. Il comando SQL WHERE funziona in realtà come un filtro in cui è possibile visualizzare solo i record che rispettano determinati criteri. Con il comando WHERE è possibile specificare i seguenti operatori:

- Uguale (=) o diverso (<>)
- Maggiore di (>) o minore di (<)
- Maggiore o uguale (>=), minore o uguale (<=)

SQL WHERE, esempio

Torniamo all'esempio della nostra tabella dei dati di base dei clienti (Datibaseclienti):

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

Desideriamo visualizzare solo i record (tutti i campi) che hanno numero di avviamento postale NAP = 8172. A tale scopo si utilizza il seguente comando SQL:

```
SELECT NOME, COGNOME, GENERE, ID_CL, VIA, NUM, NAP, LUOGO, CLIENTE_DAL, TRANSAZIONI FROM DATIBASECLIENTI WHERE NAP = 8172
```

Il risultato di questa query SQL è il seguente:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20

9.5 Operatori SQL AND / OR

Gli operatori logici SQL AND & OR sono di regola utilizzati nell'impostazione di criteri di selezione in modo da poter ottenere determinati risultati della query. Essi possono essere paragonati, a grandi linee, alla funzione Excel SE. Gli operatori AND & OR si basano sull'algebra booleana. L'operatore AND consente di combinare, rispettivamente legare, due o più condizioni SQL. L'operatore OR permette di differenziare due o più condizioni SQL. L'operatore OR prevale sempre sull'operatore AND; questo significa che, in caso di conflitto tra operatore AND e operatore OR, l'operatore OR è sempre preponderante e prevale sull'operatore AND.

SQL WHERE AND, esempio

Iniziamo con un esempio della nostra tabella dei dati di base dei clienti (Datibaseclienti).

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

Desideriamo estrarre solo i record (tutti i campi) relativi a clienti che hanno NAP = 8172 e sono di genere maschile (genere = m). A tale scopo si utilizza il seguente comando SQL:

```
SELECT NOME, COGNOME, GENERE, ID_CL, VIA, NUM, NAP, LUOGO, CLIENTE_DAL, TRANSAZIONI FROM
DATIBASECLIENTI WHERE NAP = 8172 AND GENERE = "m"
```

Il risultato di questa query SQL è il seguente:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20

SQL WHERE OR, esempio

Iniziamo con un esempio della nostra tabella dei dati di base dei clienti (Datibaseclienti).

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

Desideriamo estrarre solo i record (tutti i campi) che hanno NAP = 8172 o NAP = 8000. A tale scopo si utilizza il seguente comando SQL:

```
SELECT NOME, COGNOME, GENERE, ID_CL, VIA, NUM, NAP, LUOGO, CLIENTE_DAL, TRANSAZIONI FROM
DATIBASECLIENTI WHERE NAP = 8000 OR NAP = 8172
```

Il risultato di questa query SQL è il seguente:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20

9.6 Operatore SQL IN

L'operatore SQL IN può essere utilizzato nelle query SQL per combinare diversi criteri di ricerca specificati nelle condizioni SQL Where. Quindi consente di sostituire diversi operatori OR, semplificando così notevolmente la struttura dei criteri di selezione.

Operatore SQL IN, esempio

Continuiamo con il nostro esempio dei dati di base dei clienti (condizione WHERE) e mostriamo come la problematica di molti operatori OR può essere risolta con l'operatore IN

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

Desideriamo estrarre solo i record (tutti i campi) che hanno NAP = 8172 o NAP = 8000. A tale scopo si utilizza il seguente comando SQL:

```
SELECT NOME, COGNOME, GENERE, ID_CL, VIA, NUM, NAP, LUOGO, CLIENTE_DAL, TRANSAZIONI FROM
DATIBASECLIENTI WHERE NAP IN (8000, 8172)
```

Il risultato di questa query SQL è il seguente:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20

9.7 SQL-Operator BETWEEN

Se in una query SQL si desidera selezionare un intervallo di dati (data range), si può utilizzare l'operatore BETWEEN. Con questo operatore è possibile specificare, in modo semplice, un intervallo di selezione.

Operatore SQL BETWEEN, esempio

Continuiamo con il nostro esempio dei dati di base dei clienti. Per disporre di un campo in formato Data, abbiamo inserito anche il campo Cliente_dal.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

Desideriamo estrarre solo i record (tutti i campi) relativi ai clienti che sono stati registrati a partire dal 2011 e fino al 2015. A tale scopo si utilizza il seguente comando SQL:

```
SELECT NOME, COGNOME, GENERE, ID_CL, VIA, NUM, NAP, LUOGO, CLIENTE_DAL, TRANSAZIONI FROM
DATIBASECLIENTI WHERE CLIENTE_DAL BETWEEN 2011 AND 2015
```

Il risultato di questa query SQL è il seguente:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15

9.8 Comando SQL LIKE

Finora abbiamo introdotto solo comandi e operatori che si basano su un valore fisso o che consentono di specificare un criterio fisso (ad esempio il genere, il codice postale, il cognome del cliente, ecc.). Ma come si procede per selezionare i dati in base a un frammento di contenuto, ad esempio se si desidera selezionare tutti i clienti il cui nome inizia con la lettera R? In questo caso possiamo utilizzare il comando SQL LIKE.

Con il comando LIKE possiamo cercare dati come negli esempi seguenti (questi esempi si riferiscono a ricerche in base al cognome dei clienti):

- "B_L": Tutti i cognomi di tre lettere che iniziano con la lettera B e finiscono con la lettera L
- "BE*": Tutti i nomi che iniziano con BE (anche quelli che hanno più di due caratteri)
- "*LE": Tutti i nomi che finiscono con LE (anche quelli che hanno più di due caratteri)
- "**SP*": Tutti i nomi che contengono i caratteri SP (anche quelli che hanno più di due caratteri).

Comando SQL LIKE, esempio

Torniamo al nostro esempio dei dati di base dei clienti. Desideriamo estrarre dai nostri dati di base tutti i clienti il cui cognome finisce con CH.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

A tale scopo si utilizza il seguente comando SQL:

```
SELECT NOME, COGNOME, GENERE, ID_CL, VIA, NUM, NAP, LUOGO, CLIENTE_DAL, TRANSAZIONI FROM
DATIBASECLIENTI WHERE COGNOME LIKE "*ch"
```

Il risultato di questa query SQL è il seguente:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

9.9 Operatore SQL ORDER BY

Sinora ci siamo concentrati solo sull'estrazione di dati che rispettano determinate condizioni (funzionamento analogo alla funzione Excel SE). Naturalmente, SQL consente anche di eseguire l'ordinamento dei dati. Come in Excel, è possibile ordinare i dati in ordine crescente o decrescente. Il corrispondente comando SQL è:

- ASC: in ordine ascendente (crescente)
- DESC: in ordine discendente (decrescente)

SQL ORDER BY, esempio

Desideriamo ordinare i dati di base dei nostri clienti in ordine decrescente di identificativo cliente (ID_CL):

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12

A tale scopo si utilizza il seguente comando SQL:

```
SELECT NOME, COGNOME, GENERE, ID_CL, VIA, NUM, NAP, LUOGO, CLIENTE_DAL, TRANSAZIONI FROM DATIBASECLIENTI ORDER BY ID_CL DESC
```

Il risultato di questa query SQL è il seguente:

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesldorf	2016	12
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2

9.10 Comando SQL GROUP BY

Se desideriamo raggruppare i nostri dati in base a determinati parametri, possiamo ricorrere al comando SQL GROUP BY. Spesso questo comando viene utilizzato insieme alle funzioni di aggregazione SQL. Esempi di funzioni di aggregazione sono: AVG, COUNT, MAX, MIN e SUM.

SQL GROUP BY, esempio

Desideriamo sapere, partendo dai nostri dati di base di tutti i clienti, quanti nuovi clienti abbiamo acquisito ogni anno a partire dal 2014.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Lorenz	Fischer	m	5401	Fischerweg	7	8000	Zürich	2014	14
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Sabine	Meier	w	6355	Meierstrasse	1	8812	Horgen	2015	4
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

A tale scopo si utilizza il seguente comando SQL:

```
SELECT CLIENTE_DAL, COUNT (CLIENTE_DAL) AS NUMERO_CLIENTI FROM DATIBASECLIENTI
WHERE CLIENTE_DAL >= 2014 GROUP BY CLIENTE_DAL
```

Il risultato di questa query SQL è il seguente:

Kunde seit	Anzahl_Kunden
2014	2
2015	2
2016	1

9.11 Comando SQL HAVING

Il comando SQL HAVING corrisponde al comando SQL WHERE in un SQL GROUP BY. Vogliamo realizzare un'ulteriore aggregazione del comando SQL GROUP BY eseguito in precedenza, utilizzando un ulteriore attributo? In questo caso dobbiamo utilizzare il comando SQL HAVING. Questo comando può restringere ulteriormente la selezione e produrre risultati sulla base di funzioni di aggregazione (AVG, COUNT, MAX, MIN, SUM).

SQL HAVING, esempio

Desideriamo sapere, dai nostri dati di base di tutti i clienti, quanti nuovi clienti che hanno eseguito almeno 15 transazioni abbiamo acquisito ogni anno.

Vorname	Name	Geschlecht	KD_ID	Strasse	Nr.	PLZ	Ort	Kunde seit	Transaktionen
Monika	Weissnicht	w	4721	Weissnichtweg	18	8172	Niederglatt	2010	2
Thomas	Muster	m	5397	Musterweg	3	8000	Zürich	2011	8
Felix	Beispiel	m	6423	Beispielweg	5	8172	Niederglatt	2012	23
Regula	Possible	w	5397	Possiblestrasse	23	8007	Zürich	2013	33
Lorenz	Fischer	m	5401	Fischerweg	7	8000	Zürich	2014	14
Reto	Unmöglich	m	8174	Unmöglichstrasse	8	8172	Niederglatt	2014	20
Sabine	Meier	w	6355	Meierstrasse	1	8812	Horgen	2015	4
Stefano	Possible	m	9371	Possiblestrasse	13	8416	Flaach	2015	15
Barbara	Möglich	w	7348	Möglichstrasse	2	8153	Diesdorf	2016	12

A tale scopo si utilizza il seguente comando SQL:

```
SELECT CLIENTE_DAL, COUNT (CLIENTE_DAL) AS NUMERO_CLIENTI FROM DATIBASECLIENTI  
WHERE CLIENTE_DAL >= 2014 GROUP BY CLIENTE_DAL, HAVING TRANSAZIONI >=15
```

Il risultato di questa query SQL è il seguente:

Kunde seit	Anzahl_Kunden
2014	1
2015	1

10 Robotic Process Automation (RPA)

10.1 Introduzione

Ci sono attività ricorrenti nella vostra giornata lavorativa che trovate noiose e impegnative allo stesso tempo? Non è raro che i processi di lavoro monotoni richiedano un alto livello di concentrazione per poter produrre risultati abbastanza precisi. Alle persone non piace lavorare su uno stesso lavoro per un'infinità di volte e quindi, nel tempo, tendono a perdere la concentrazione.

Nel controlling si investe parecchio tempo in attività manuali per eseguire la riconciliazione tra le contabilità ausiliarie e la contabilità finanziaria, tra la chiusura di gruppo e le chiusure individuali, tra i conti dei flussi, ecc. In presenza di elevati volumi di registrazioni, le attività di riconciliazione e di controllo vengono spesso limitate a controlli a campione, poiché non è praticamente possibile riconciliare tutte le registrazioni o tutti i conti.

Il controlling è sfidato dalle tendenze future come, ad esempio, i big data. I nuovi concetti nel reporting sono molto richiesti e da essi nascono molti altri progetti di sviluppo. Allo stesso tempo, il lavoro di routine dovrebbe rimanere invariato. Qui, il controlling ha bisogno di soccorso per essere in grado di investire le sue risorse in ulteriori sviluppi.

La robotic process automation automatizza le attività manuali con l'aiuto di software e, in alcuni casi, di hardware. I vantaggi sono evidenti: il controlling è sgravato dal lavoro di routine e il «robot» ha un enorme potenziale di scalabilità. A differenza degli umani, i robot sono in grado di eseguire la stessa attività ripetitiva x volte, il che aumenta significativamente la quantità e la qualità. Con la RPA, non è più necessario limitarsi a controlli a campione. Inoltre, di regola, si raggiunge un grado di standardizzazione significativamente più alto nei processi di lavoro, il che è molto vantaggioso per l'azienda, ad esempio, nel caso di cambiamenti di personale.

10.2 Struttura

Per strutturare un RPA, è necessario scomporre un processo nelle sue sottofasi più piccole (microfasi). Si devono quindi definire le condizioni di ogni microfase in termini di:

- Come iniziare la microfase?
- Quando iniziare la microfase?
- La microfase dipende da una microfase precedente?
- La microfase può essere eseguita in parallelo con un'altra microfase?
- Cosa succede se la microfase si interrompe a causa di altri problemi tecnici?
- Come deve essere protocollata la microfase?
- La microfase richiede un'interazione manuale?

La maggior parte degli strumenti RPA rappresenta la definizione del processo mediante un diagramma di flusso, in modo che le condizioni possano essere riconosciute a colpo d'occhio. Al più tardi nella definizione di una microfase diventa evidente che l'automazione dei processi è raccomandata solo per processi con un'elevata standardizzazione.

10.3 Robotic Process Automation per un database di clienti

I possibili utilizzi di RPA con un database di clienti sono molti e variegati e possono quindi essere personalizzati ai bisogni dell'azienda. Per fornire alcune idee, di seguito vengono descritti alcuni possibili casi d'utilizzo in relazione con un database di clienti. I casi d'utilizzo sono ordinati in base alla complessità del database dei clienti, partendo da un semplice elenco di clienti fino a un database "state of the art" e a una soluzione di BI.

Controllo d'integrità

In un semplice elenco di clienti, è essenziale che l'elenco sia redatto secondo il principio dell'integrità dei dati. Integrità dei dati significa che i dati dei clienti sono validi e non contraddittori. Con l'RPA, ad esempio, è possibile verificare in modo completamente automatico se un cliente è memorizzato più di una volta nell'elenco dei clienti e se tutti i campi obbligatori (indirizzo, numero di cliente, ecc.) sono stati completati. Gli importi delle fatture aperte possono essere confrontati con la situazione dei crediti per forniture e prestazioni. Le vie e gli altri elementi dell'indirizzo possono essere controllati con un elenco di indirizzi esterno. Le eventuali discrepanze riscontrate possono essere segnalate a un utente per la correzione dei dati o, in una certa misura, possono essere corrette direttamente dal sistema stesso.

Trasferimento di dati

Può accadere che i dati dei clienti debbano essere inseriti manualmente nel database (ad esempio, acquisite un nuovo cliente via e-mail). L'RPA può supportare questo processo convertendo automaticamente tutte le e-mail dei clienti in dati strutturati, scrivendoli nel database. È anche ipotizzabile che l'inserimento dei dati delle richieste telefoniche dei clienti possa essere automatizzato con l'RPA sulla base del riconoscimento vocale e di altri dati (ad esempio, il numero di telefono chiamante).

Consolidamento

Quando due imprese eseguono una fusione, i dati dei clienti devono essere uniti e consolidati. Si tratta di un'attività complessa che può richiedere molto tempo e ostacolare l'operatività quotidiana. Con l'RPA è possibile automatizzare logiche di migrazione complesse, riducendo così di molto la durata del processo e aumentando in modo significativo la qualità.

Azioni verso i clienti

Di solito è necessario contattare regolarmente i clienti, ad esempio mediante newsletter, e-mail e simili. Con la RPA, l'elaborazione di massa di e-mail, lettere, post online e altre comunicazioni può essere eseguita in modo completamente automatico e ricorrente.

Automatizzazioni di soluzioni BI

Se il database dei clienti è collegato, per le valutazioni, a fonti di dati esterne e sono state create soluzioni di BI estese, spesso si utilizza un data warehouse per garantire una valutazione performante. Un data warehouse consente inoltre di ampliare strutturalmente un database di clienti operativi per eseguire valutazioni e per consentire una maggiore ricostruzione storica dei dati. Ad esempio, un ordine di un cliente deve essere collegato all'indirizzo del cliente che era valido al momento dell'ordine.

Nei database operativi, l'attenzione è rivolta alle attività quotidiane e il database non deve essere gravato da numerose relazioni. L'utilizzo di un data warehouse richiede un gran numero di lavori tecnici per l'estrazione, la trasformazione e il caricamento dei dati nel data warehouse (ETL). Inoltre, spesso è necessario informare gli utilizzatori dei dati sulle nuove valutazioni dopo la preparazione dei dati o fornire loro immediatamente i file corrispondenti. È anche possibile impostare una dashboard per il management, che viene aggiornata periodicamente con le cifre chiave attuali. Nell'intera catena di processi di arricchimento (data enrichment), valutazione e pubblicazione dei dati, l'RPA è in grado di orchestrare in modo completamente automatico le singole fasi di lavoro. È inoltre possibile definire valori di tolleranza e, se un valore di tolleranza viene violato, il robot lo segnala automaticamente al controlling.

10.4 10.4 Robotic Process Automation nel test di software

Quanto più personalizzati e complessi sono i database e le valutazioni dei dati, tanto maggiore è l'impegno nell'eseguire i test per le nuove release e gli ulteriori sviluppi. Poiché il database dei clienti è una tra le aree più critiche per un'impresa, le relative modifiche devono essere sottoposte a un approfondito controllo di qualità. Con la RPA, gli input dell'utente possono essere eseguiti x volte e valutati sistematicamente per individuare eventuali errori. Se tali test vengono integrati nel processo di distribuzione delle modifiche, è possibile garantire, in modo completamente automatico, che ogni modifica del software passi attraverso una serie di test prima di essere resa operativa.

10.5 10.5 Introduzione di un Robotic Process Automation – cosa si deve considerare?

L'implementazione di un RPA richiede un approccio orientato al progetto. È essenziale strutturare il progetto in fasi, in cui si devono considerare anche i seguenti punti:

1. Scegliere un prodotto standard

Sul mercato esistono già molti prodotti RPA standard; quindi, è meglio evitare, se possibile, di sviluppare qualcosa da soli.

2. Esempi di fornitori di RPA

- UiPath (Gartner, 2020)
- Automation Anywhere (Gartner, 2020)
- Atomic Software

3. Pay back period molto basso

Prima di investire in un RPA, si deve eseguire un calcolo d'investimento. La RPA può richiedere rapidamente investimenti elevati, che dovrebbero essere economicamente sostenibili. Il software ha una durata di vita breve: un periodo di ammortamento superiore a due o tre anni è critico. Se si vogliono includere anche motivazioni non monetarie (ad esempio, garantire il know-how), si raccomanda un'ulteriore analisi dei benefici.

4. Un approccio a tappe

Non si deve cercare di automatizzare tutti i processi o intere catene di processi in una sola volta. È meglio suddividere un processo in sotto-processi, che poi dovranno essere gradualmente automatizzati, in modo da acquisire esperienza rapidamente e trarne beneficio per adattare la portata della soluzione RPA desiderata.

5. Fasi manuali in modo consapevole

La RPA non deve far perdere il controllo sui processi. È consigliabile inserire deliberatamente delle fasi manuali per essere in grado di soddisfare i requisiti del sistema di controllo interno (SCI) e rispettare il principio dei quattro occhi (ad esempio, l'approvazione di una nota di credito al cliente).

6. Autorizzazione incluso tenuta di protocolli

Il software RPA necessita spesso di ampie autorizzazioni in diversi sistemi. Pertanto, è necessario creare un concetto di autorizzazione rigoroso per identificare chiaramente la persona che sta dietro al software RPA. Tutte le fasi del lavoro devono essere registrate in protocolli, in modo da poter risalire in seguito alle transazioni.

7. Error Handling (Gestione degli errori)

I processi RPA vengono spesso eseguiti al di fuori dell'orario di lavoro per non influenzare negativamente l'attività operativa (ad esempio, a causa di un aumento del carico di sistema). Occorre quindi prestare particolare attenzione alla capacità di reazione in caso di errore. Cosa bisogna fare in caso di interruzione? In caso di errore, è possibile inviare messaggi SMS e simili per ogni fase del lavoro. I tempi di reazione al di fuori dell'orario di lavoro devono essere concordati separatamente.

8. **Essere consapevoli** che la RPA non tollera alcuna deviazione dalla norma. Se il software non trova la situazione iniziale prevista, la sottofase si interrompe. In questo caso, si devono definire misure quali notifiche via e-mail all'amministratore, ecc.

9. Tempo d'esecuzione

Prestare attenzione ai tempi di consegna con l'RPA. Più lungo è il tempo d'esecuzione di un processo automatizzato, più alto è il rischio di problemi tecnici. In primo luogo, la frequenza dei time-out e delle situazioni indesiderate aumenta con l'aumentare del tempo d'esecuzione. In secondo luogo, quanto più lunghi sono i tempi d'esecuzione, tanto maggiore è l'impatto negativo sulle attività operative quotidiane. È consigliabile fissare fin dall'inizio dei valori target per il tempo d'esecuzione e misurare regolarmente i tempi d'esecuzione effettivi paragonandoli con i valori target.

10. RPA non è solo un tool

Le soluzioni RPA sono spesso composte da diversi tool. Nella maggior parte dei casi, si utilizza un software in grado di comunicare con un'ampia gamma di applicazioni e di riunire le singole fasi delle applicazioni in una procedura. Inoltre, nelle sottofasi vengono utilizzati tool specifici, come ad esempio tool specializzati in confronti e in abbinamenti. Solo l'insieme di questi tool consente di automatizzare la riconciliazione tra le contabilità ausiliarie e la contabilità generale.

11. Meno "doing", ma più "monitoring"

La RPA alleggerisce il controlling nei lavori di routine, ma aumenta lo sforzo di monitoraggio. Si deve cercare di contenere il più possibile i costi di monitoraggio. Ad esempio, il controlling può essere dotato di un cockpit di monitoraggio che mostra lo stato di tutti i processi RPA a colpo d'occhio.

12. Conoscere i limiti

È molto difficile automatizzare completamente l'interpretazione e il commento dei dati: questo necessita un controlling e un know-how adeguati all'azienda.

11 Intelligenza artificiale e business analytics

11.1 Cos'è l'intelligenza artificiale?

Quando inseriamo «conto» in un motore di ricerca, ci viene subito suggerito un completamento della ricerca del tipo «aprire conto». Quando carichiamo una foto delle nostre vacanze in un tool foto sul cloud, viene suggerito il termine «spiaggia» come parola chiave. Quando richiediamo una carta di credito e inseriamo i nostri dati, ci viene automaticamente calcolato un punteggio di solvibilità. Quando sottoscriviamo un abbonamento di qualsiasi tipo, iniziamo a ricevere raccomandazioni personalizzate per noi da parte del fornitore dell'abbonamento.

Tutti questi processi hanno una cosa in comune: i modelli che eseguono queste proposte e i calcoli sottostanti sono stati dotati di dati per il riconoscimento di pattern. Il motore di ricerca ha appreso dallo storico delle ricerche che molti utenti scrivono «aprire» prima di scrivere «conto». Il tool foto sul cloud ha appreso l'aspetto di una spiaggia dopo aver esaminato un gran numero di fotografie con l'etichetta «spiaggia» e ora è in grado di valutare se una spiaggia è visibile o meno in una fotografia. Un modello per la valutazione del punteggio di solvibilità ha imparato, sulla base dei punteggi precedentemente calcolati per altre persone, che una persona con un reddito elevato e stabile riceve un punteggio di solvibilità più alto rispetto a quello di una persona che ha appena terminato gli studi. Gli abbonati vengono suddivisi in gruppi o segmenti in base ai loro dati personali e alla loro attività, in quanto i provider dispongono dei dati di migliaia di abbonati e possono quindi riconoscere modelli e pattern utilizzando questi dati.

In base a dati e informazioni storiche, è possibile imitare modelli di comportamento, leggi della natura o altre

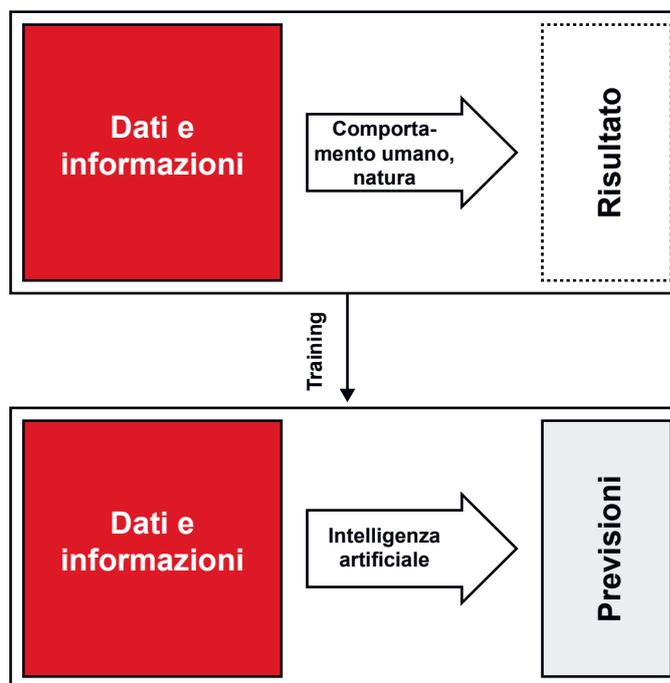


Figura 13: Natura di un'intelligenza artificiale

regole.

L'intelligenza artificiale, abbreviata in IA (in inglese AI), descrive il processo di riconoscimento di modelli e la loro applicazione a nuovi dati per poter eseguire una previsione o una valutazione grazie all'aiuto dei modelli imparati. Il riconoscimento dei modelli e degli schemi sottostanti ai modelli viene anche definito apprendimento o

training, ossia addestramento di un modello di intelligenza artificiale. Poiché di solito il training non viene eseguito manualmente, ma da un computer, si parla anche di machine learning o apprendimento automatico.

Nel linguaggio quotidiano, termini come advanced analytics, data mining e predictive analytics sono usati come sinonimi, anche se questi tre termini rappresentano, in realtà, delle aree dell'IA.

Per comprendere ancor meglio questo riconoscimento di pattern (modelli e schemi), generiamo un esempio artificiale. Ipotizziamo che la nostra azienda conceda prestiti, ovviamente solo a persone che siano in grado di rimborsarli. Per ogni richiedente il prestito, disponiamo solo di tre informazioni: se ha già un conto presso di noi, se gli piace il colore arancione e quante passeggiate fa in un anno. Abbiamo a disposizione il seguente set di dati storici:

Già cliente?	Ti piace l'arancione?	Passeggiate all'anno?	Prestito rimborsato?
Sì	No	3	Sì
No	No	9	No
No	No	10	Sì
Sì	Sì	5	Sì
No	Sì	9	No
Sì	No	15	Sì
Sì	Sì	10	Sì
No	Sì	11	No
No	No	8	No
No	No	11	Sì

Se desideriamo riconoscere dei modelli in base alle tre informazioni disponibili e al risultato dell'evento prestito rimborsato, possiamo riuscire a trovarli rapidamente. Se il richiedente è già cliente, il prestito sarà sempre rimborsato. Se il richiedente non è ancora un cliente e gli piace il colore arancione, il prestito non sarà mai rimborsato. Se il richiedente non è ancora un cliente e non ama l'arancione, il prestito sarà rimborsato nel caso in cui il cliente fa almeno 10 passeggiate ogni anno. L'esempio presentato (certamente un po' irrealistico) mostra chiaramente cosa significa per noi riconoscere i modelli. Questo è esattamente ciò che fa l'intelligenza artificiale. L'IA cerca di riconoscere le regole in base alle quali può elaborare le informazioni in suo possesso in modo da generare una previsione. Pertanto, se ci viene chiesto un prestito da una persona che non è nostro cliente, non ama l'arancione e fa 12 passeggiate ogni anno e desideriamo sapere se questa persona potrà rimborsarlo, l'IA, dopo aver riconosciuto gli schemi in una sessione di training, è in grado di tranquillizzarci e prevedere che probabilmente il prestito sarà rimborsato.

Naturalmente, l'IA è in grado di riconoscere anche situazioni più complesse, che noi potremmo identificare solo con un grande sforzo o, addirittura, non potremmo riconoscere affatto. Proprio in queste situazioni, l'IA ci semplifica la vita. Per poter comprendere meglio le diverse sfaccettature dell'IA per le prossime sfide come accountant nel controlling o in altri ambiti, e quindi essere in grado di valutare meglio dove ha senso usare l'IA e cosa deve essere considerato, vogliamo affrontare esplicitamente questo argomento nei prossimi para-grafi.

In questo capitolo analizzeremo i diversi aspetti dell'IA. Nel paragrafo «Le diverse forme di IA», introdurremo una struttura nel vasto mondo dell'IA e alcuni termini importanti. Successivamente, nel paragrafo «L'emergere di modelli di IA», discuteremo come vengono creati i modelli di IA, cosa rende buono un modello di IA e come possiamo valutarlo e validarlo. I rischi dei modelli di IA sono illustrati nel paragrafo «Le vulnerabilità di un sistema di IA». Infine, «Una strada rapida verso il primo modello» metterà in evidenza quali strumenti possiamo utilizzare per ottenere un primo modello (forse non ancora perfetto) nel più breve tempo possibile.

Termini importanti da questo paragrafo: apprendere, training, previsione, IA, AI, apprendimento automatico, machine learning, advanced analytics, data mining, predictive analytics.

11.2 Le diverse forme di IA

Gli esempi precedenti illustrano l'idea centrale dell'IA: riconoscere i modelli e applicare questi modelli a nuovi casi. Tuttavia, i singoli casi differiscono in molti aspetti e danno solo una minima idea della portata dell'IA.

Apprendimento supervisionato e apprendimento non supervisionato

Un elemento fondamentale nell'IA riguarda come avviene l'apprendimento. I dati utilizzati per l'apprendimento vengono chiamati **dati di training**. I dati di training contengono i dati che consentono di identificare il modello. Nell'esempio del punteggio di solvibilità, i dati che consentono di identificare il modello sono molteplici: reddito, patrimonio, durata del domicilio nel comune attuale, numero di impieghi negli ultimi cinque anni e molto altro ancora. I dati che sono utilizzati come input per il riconoscimento del modello vengono denominati **feature** (caratteristiche).

L'esempio del punteggio di solvibilità non contiene però sole feature. I dati di training comprendono anche la solvibilità, spesso misurata in CHF. Tuttavia, disponiamo di questo «risultato» solo per i dati di training: se un nuovo cliente richiede un conto, la sua solvibilità non è ancora conosciuta. Questo «risultato» viene anche denominato **label**. L'IA, che viene addestrata su dati di training che hanno etichette (label) e caratteristiche (feature), è denominata anche **apprendimento supervisionato** (supervised learning). Il termine «supervisionato» si riferisce semplicemente al fatto che è già disponibile il risultato, ossia la label. Anche il riconoscimento di oggetti su immagini rientra nella categoria dell'apprendimento supervisionato, poiché il modello impara sulla base di immagini etichettate con parole chiave (label) come spiaggia, palma, montagna, ecc.

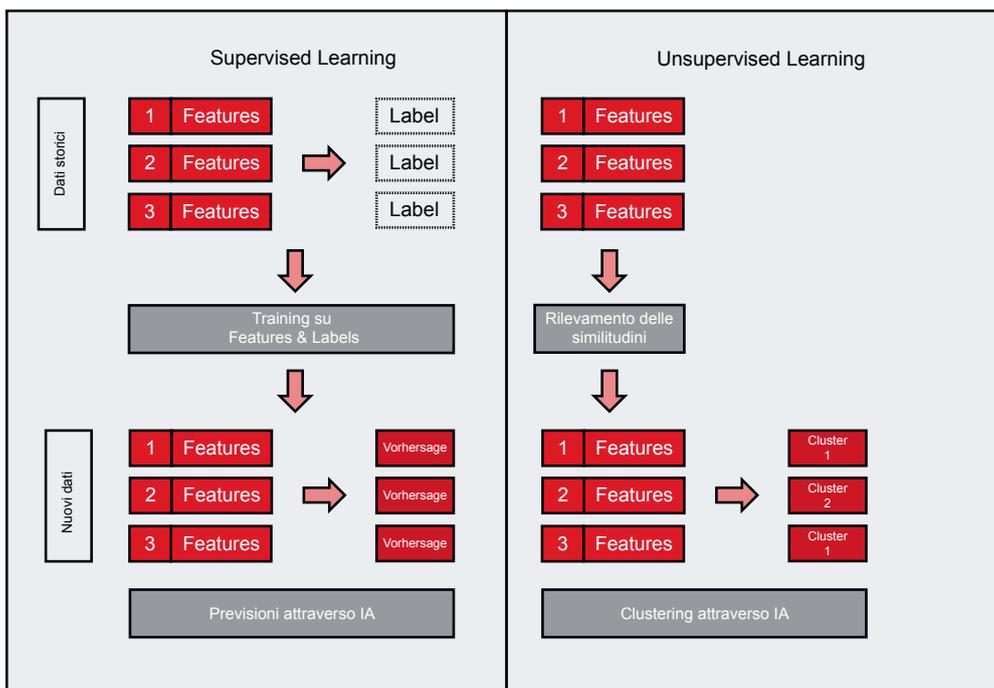


Figura 14: Differenza Supervised Learning / Unsupervised Learning

L'immagine mostra la differenza fondamentale tra apprendimento supervisionato (a sinistra) e apprendimento non supervisionato (a destra). Il primo dispone a priori di una label (ad esempio, è solvibile o non è solvibile) per

ogni record (ad esempio, ogni potenziale interessato a un conto); attraverso l'IA si vanno ad imitare le regole che hanno determinato la mappatura dalle feature alla label, in modo da poter eseguire una previsione sulla solvibilità. L'apprendimento non supervisionato non dispone di alcuna label nei dati su cui viene addestrato e, quindi, non esiste un risultato conosciuto a priori. In questo caso l'IA cerca di ricavare dai dati, a posteriori, una label sotto forma di appartenenza a un cluster.

In alcuni esempi, non esistono tali label a priori. Supponiamo di essere un fornitore di telefonia mobile e di voler capire meglio i nostri clienti. Naturalmente, non ha senso esaminare e comprendere ogni singolo cliente nel dettaglio: richiederebbe troppo tempo. Vogliamo invece dividere i nostri clienti in gruppi; ogni gruppo dovrebbe essere più o meno omogeneo. In altre parole: cerchiamo dei cluster in cui i clienti abbiano molte affinità / somiglianze tra loro. Questo tipo di IA che riconosce schemi / gruppi / cluster / segmenti è denominato anche **apprendimento non supervisionato**, poiché non esiste una label. L'apprendimento non supervisionato può essere applicato anche all'esempio di nuovi interessati a ottenere un prestito. I dati di training potrebbero includere le feature sopra menzionate e il modello di IA potrebbe riconoscere i relativi rag-gruppamenti, come ad esempio *i domiciliati in uno stesso luogo da numerosi anni, con un reddito stabile e un patrimonio modesto oppure i lavoratori indipendenti con reddito instabile e patrimonio elevato o anche i pensionati con reddito basso ma stabile e domiciliati nello stesso luogo da molti anni.*

L'apprendimento non supervisionato o *unsupervised learning*, tuttavia, non comprende solo il rilevamento dei cluster. Anche il rilevamento dei valori fuori norma, i cosiddetti *outlier* o *anomaly detection* e le *association rules* fanno parte dell'*unsupervised learning* poiché i dati di training sottostanti non sono stati dotati a priori di una label.

Termini importanti da questo paragrafo: dati di training, feature, label, apprendimento supervisionato, supervised learning, apprendimento non supervisionato, unsupervised learning, outlier detection, anomaly detection, association rules.

11.3 Classificazione e regressione

Un'ulteriore distinzione tra le diverse forme di IA può essere eseguita anche con riferimento alla label nel caso dell'apprendimento supervisionato. Riprendiamo l'esempio della solvibilità. Come descritto in precedenza, le feature possono includere reddito, patrimonio o altri dati finanziari personali. La label può essere descritta, ad esempio, attraverso la solvibilità in franchi. Consideriamo quindi una label che descrive l'intervallo continuo tra un valore minimo (ad esempio, 0 franchi) e un valore massimo (ad esempio, 1 milione di franchi). In uno scenario alternativo, tuttavia, potrebbe essere sufficiente rispondere alla domanda sulla solvibilità di un cliente con un «sì» o un «no». I dati di training includeranno gli stessi feature descritti in precedenza, ma la label includerà solo uno dei due valori «sì» o «no». Mentre il primo esempio di IA, che prevede una solvibilità compresa tra zero e un milione di franchi, è denominato **regressione**, nel secondo caso in cui si deve prevedere quale sia la classe corretta tra due possibili classi (in questo caso «sì» o «no») si utilizza il termine **classificazione** o **classification**. La classificazione multilabel descrive la classificazione con più di due classi, ad esempio un'intelligenza artificiale per il riconoscimento delle immagini che valuta se un'immagine mostra una palma, una spiaggia, una montagna e/o una grotta.

Termini importanti da questo paragrafo: regressione, classificazione, classification.

Machine learning tabellare, riconoscimento di immagini e riconoscimento di testo

L'IA può anche essere classificata in base all'uso a cui è destinata. I tre ambiti principali sono il **riconoscimento di testo (o riconoscimento vocale)** o **Natural Language Processing (NLP)**, il **riconoscimento di immagini** o **Visual Recognition** e il **Machine Learning tabellare** o **Tabular Machine Learning**. NLP e Visual Recognition simulano le capacità cognitive umane, che corrisponde a una «concezione classica» dell'IA.

Il **machine learning tabellare** comprende tutti i casi in cui le caratteristiche (feature) possono essere tipicamente rappresentate in forma tabellare. Il caso della solvibilità è un esempio classico: un record corrisponde a una riga le cui colonne rappresentano le caratteristiche (feature), ad esempio il reddito, il patrimonio, ecc.

Il **riconoscimento di immagini** descrive l'apprendimento basato su una serie di immagini con label. La label di un'immagine può contenere l'oggetto visibile nell'immagine. Alternativamente, la label può includere la posizione di un oggetto nell'immagine per consentire il riconoscimento dell'oggetto. Nel primo caso, il modello imparato viene utilizzato per completare con una stima le immagini che non includono ancora una label. Un'app che riproduce il nome di un fiore non appena viene fotografato si basa proprio su questo concetto. L'app è stata precedentemente addestrata su un ampio set di immagini di fiori con la rispettiva label (ad esempio margherita, denti di leone, ecc.).

Il **riconoscimento di testo** descrive il riconoscimento di modelli da testi. Questo può assumere varie forme. Un esempio classico è la traduzione di testi. I più noti fornitori online, come DeepL, si basano su un principio piuttosto semplice: i dati di training sono costituiti da testi di siti web scritti in più lingue. Su questa base, il sistema di riconoscimento del testo può imparare a tradurre parole e frasi. Un altro classico campo di applicazione è la cosiddetta sentiment analysis. In questo caso, viene esaminata una frase per comprenderne l'orientamento positivo o negativo, che può essere utile, ad esempio, per il team di assistenza clienti di un'azienda per poter fornire risposte automatiche migliori ai messaggi degli utenti.

Termini importanti da questo paragrafo: riconoscimento di testo, riconoscimento vocale, natural language processing (NLP), riconoscimento di immagini, visual recognition, machine learning tabellare, tabular machine learning.

L'intelligenza artificiale come base per l'automazione/supporto

Vogliamo ancora discutere un'ultima distinzione. L'intelligenza artificiale è spesso considerata la base per l'automazione. Ciò è parzialmente corretto, ad esempio quando un editor di testi apporta automaticamente correzioni sulla base di dati storici o quando un sistema di gestione dei clienti invia automaticamente una mail quando la probabilità di perdere un cliente (calcolata tramite l'intelligenza artificiale) è alta. Tuttavia, l'intelligenza artificiale non deve necessariamente sostituire i processi nella loro totalità; in molti casi ha più senso utilizzare il risultato del modello di IA come supporto per il processo decisionale. Anche in questo caso, l'esempio della solvibilità ci aiuta. Un modello di intelligenza artificiale potrebbe calcolare la solvibilità in franchi per ogni potenziale cliente e quindi generare automaticamente un contratto. In alternativa, il risultato potrebbe anche servire al collaboratore della banca come base per le sue considerazioni e per le trattative contrattuali con il cliente; il modo in cui utilizzare queste informazioni aggiuntive spetta al collaboratore.

11.4 L'emergere di modelli di IA

Finora abbiamo visto diversi esempi in cui un modello riconosce schemi di riferimento sulla base di dati storici e successivamente applica questi stessi modelli a nuovi dati per eseguire previsioni e stime. Tuttavia, all'interno di questo processo, è necessario considerare diverse aree:

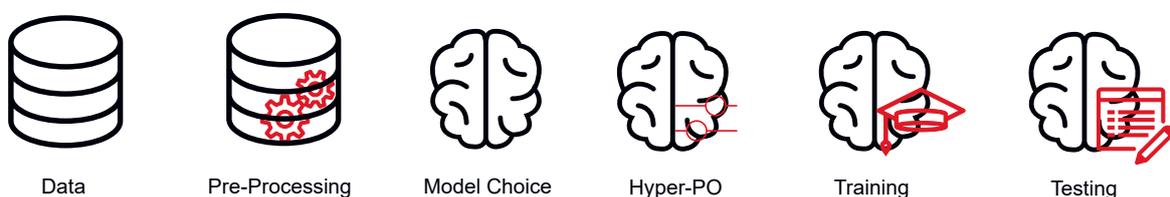


Figura 15: Fasi fino al testing

1. **Data Pre-Processing:** come possiamo adattare i nostri dati in modo che il modello possa riconoscere al meglio le feature?
2. **Training e scelta del modello:** quali dati forniamo al training? Quale modello si adatta meglio al problema che stiamo affrontando? Posso configurare il modello e, se sì, qual è la configurazione migliore?
3. **Testing:** quali dati utilizzo per il testing? Come misuro la bontà del modello o, in altre parole, quando ho ottenuto un buon modello?

In questo paragrafo vogliamo affrontare queste tre aree che dobbiamo percorrere per arrivare al nostro modello finale.

Data Pre-Processing

I dati rappresentano la base con cui alimentiamo un modello, in modo tale che esso possa riconoscere gli schemi. Si applica il principio molto semplice del *garbage in, garbage out*. Fino a quando i dati non riconoscono gli schemi o non sono in una forma che consenta al modello di imparare, il modello non riconoscerà questi schemi.

Spetta quindi a noi preparare i dati in modo che il modello possa estrarre gli schemi. Per farlo, possiamo scegliere tra diversi strumenti. In questa sede esamineremo i metodi più importanti:

- **Filtrare i record** o *come ottengo i record migliori?* Il filtraggio dei record consente di conservare solo i record più importanti e di qualità più elevata. Questo può aiutare a identificare più facilmente gli schemi partendo da questi dati di elevata qualità. In particolare, si vogliono eliminare i cosiddetti valori anomali o outlier. Questa fase viene anche chiamata *rilevamento delle anomalie*, un'operazione per la quale esistono numerose librerie già pronte e accessibili liberamente. È importante osservare che il filtraggio dei dati avviene nel quadro del training (addestramento). Nel funzionamento, questo filtraggio può essere adottato e dotato di determinate regole oppure può essere omissivo. In quest'ultimo caso, tuttavia, dobbiamo analizzare come il modello reagisce a queste anomalie in quanto, dopo tutto, non le ha conosciute durante il training.
- **Arricchimento dei record** o *come procediamo con i dati mancanti?* A volte mancano alcuni dati o campi. Consideriamo, ad esempio, l'individuazione di *transazioni fraudolente*, cioè transazioni di natura illegale, che si verificano molto meno frequentemente rispetto alle transazioni convenzionali. Per un modello, la presenza di molti più casi di una classe (*transazioni fraudolente*) che di un'altra può comportare varie difficoltà. In questo caso, esistono diverse opzioni. Due di queste opzioni prevedono la riduzione dei dati di test della classe più grande o la replica dei dati di test della classe più piccola. Questo processo viene denominato *resampling* (ricampionamento). Naturalmente esistono numerosi altri metodi, come la generazione di record sintetici utilizzando tecniche come SMOTE (*Synthetic Minority Oversampling Technique*), che vanno oltre lo scopo di questo script.
- **Feature Engineering** o *come possiamo tradurre e ampliare le nostre colonne in modo che anche il modello le capisca?* Spesso nei nostri record ci sono colonne (o feature) che contengono dati in un formato che il modello non riesce a interpretare facilmente. Supponiamo che quando un cliente si registra per aprire un conto, debba inserire tutti i suoi tipi di investimento preferiti in un campo di testo libero; ad esempio, una persona inserisce «azioni, immobili, criptovalute», un'altra persona inserisce «azioni e immobili» e un'altra ancora «azioni e criptovalute». Per il modello, queste voci non sono altro che classi diverse. Per aiutare il modello a capire che si tratta di alcune classi («azioni», «immobili», «criptovalute») che si ripetono più volte, dovremmo creare nuove feature come «Nel tipo di investimento preferito ci sono le azioni», «Nel tipo di investimento preferito ci sono gli immobili», «Nel tipo di investimento preferito ci sono le criptovalute». A noi queste nuove feature non danno alcuna nuova informazione. Tuttavia, queste nuove feature permettono al modello di identificare più facilmente gli schemi legati al tipo di investimento preferito.

- **Imputing** o *cosa facciamo con i campi vuoti?* L'imputing è il processo di riempimento dei dati mancanti. Ad esempio, supponiamo che il campo età nel formulario di registrazione dell'account non sia un campo obbligatorio. Molti clienti lo compilano, altri no. Tuttavia, non si possono fornire al modello campi vuoti o, in caso contrario, bisogna comunicare al modello come devono essere interpretati. Potremmo, ad esempio, provare a sostituire il contenuto del campo con un valore costante, come la media dei dati di training. Tuttavia, questo non è sempre raccomandato e, in ogni caso, il valore della costante deve essere scelto con attenzione. In alternativa, si può stimare l'età in base agli altri dati indicati nel formulario di registrazione. Ne risulta un'interessante costellazione, in quanto utilizziamo l'IA per arricchire i dati che vengono utilizzati per addestrare un'altra IA.
- **Normalization** o *perché il modello valuta anche la somiglianza?* Questo procedimento si riferisce alla normalizzazione dei dati, in particolare alla compressione e all'ampliamento delle voci delle diverse colonne in modo che, ad esempio, la media e la deviazione standard delle colonne corrispondano a valori continui. Ma perché questo è importante? Ri-prendiamo ancora una volta l'esempio della solvibilità di un potenziale cliente. Una feature di questo cliente potrebbe essere la sua età. Un valore numerico che varia tra 18 e 100. Un altro valore può essere il patrimonio del cliente. Questo valore può essere pari a 0, per molti sarà compreso tra 10'000 e 1'000'000 franchi, ma ci saranno sicuramente clienti con patrimoni superiori a 1'000'000 franchi. Un'altra colonna può contenere la media dei voti dell'ultimo anno scolastico del cliente. Questo valore è solitamente compreso tra 4 e 6. I diversi modelli di IA possono gestire questi intervalli in modo molto diverso. Per alcuni modelli significa semplicemente un periodo di training più lungo per riconoscere gli schemi; altri modelli riescono a malapena a riconoscere gli schemi o lo fanno ma con una qualità insufficiente. Pertanto, di norma è consigliabile scalare e spostare le rispettive colonne in modo che il valore medio sia pari a 0 e la deviazione standard sia pari a 1. Questo permette di forzare una certa somiglianza nella distribuzione dei dati. Il modello di solito lo accetta e riesce ad operare con tempi di addestramento più brevi e risultati migliori.

Termini importanti: pre-processing, anomalie, outlier, anomaly detection, arricchimento dei dati, oversampling, feature engineering, imputing, normalizzazione.

Training e scelta del modello

Non appena i dati sono pronti, il modello può essere addestrato (rispettivamente può essere eseguito il training del modello). Prima di ciò, tuttavia, è necessario scegliere il modello appropriato. Esistono diversi modelli per i diversi tipi di intelligenza artificiale. Di seguito è riportata una breve selezione dei modelli che si dovrebbero conoscere:

- **Classificazione:** decision tree, random forest, Xgboost, logistic regression, GLM, neural networks.
- **Regressione:** regressione lineare, regressione ridge, regressione lasso, regressione polinomiale, regressione lineare bayesiana, regressione della rete elastica.
- **Clustering:** KMeans, Nearest Neighbor.

La comprensione dei singoli modelli va oltre lo scopo di questo script. In questa sede ci focalizziamo su un unico modello per capire cosa significhi esattamente intelligenza artificiale e riconoscimento di pattern: l'albero decisionale (decision tree).

Un decision tree (albero decisionale) può essere inteso come una sequenza di domande poste ai dati, a ciascuna delle quali si risponde con sì o no. A dipendenza della sequenza di risposte date, si ottiene una classe. La Figura 16 illustra questo procedimento con l'esempio della solvibilità. Supponiamo di essere uno studente con un reddito annuo di 30'000 franchi e di non avere debiti. Il primo nodo chiede se il reddito è superiore a 80'000 franchi. Poiché, nel nostro caso, la risposta è no, si passa al nodo successivo, che chiede se la persona è uno studente. Rispondiamo di sì, e per questo motivo si pone un'altra domanda, ossia se abbiamo dei debiti. Poiché non abbiamo debiti, in base alla valutazione di questo albero decisionale risultiamo solvibili (S).

In base alle feature reddito, patrimonio, è studente e ha debiti, l'albero decisionale decide se una persona è solvibile (S) o non è solvibile (N).

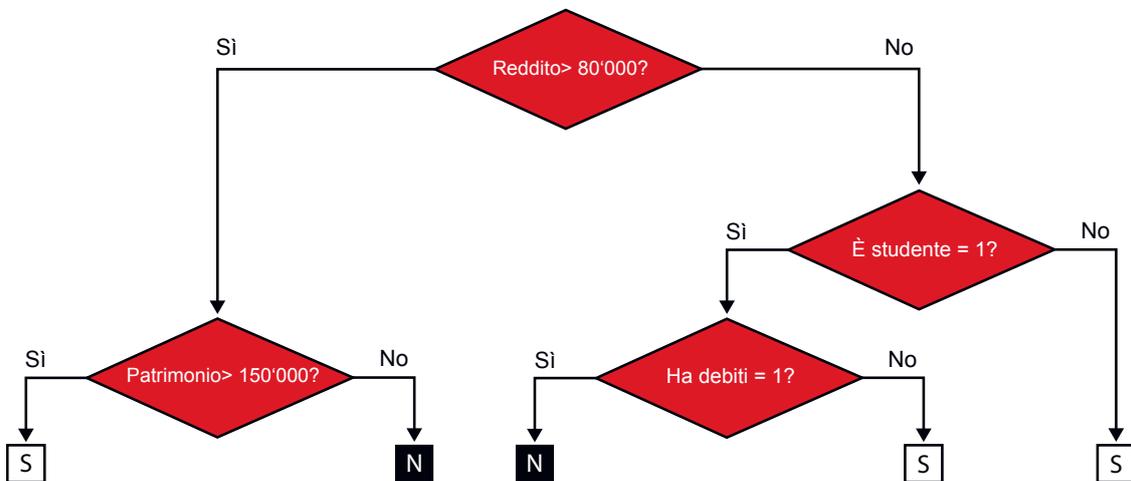


Figura 16: Decision tree

Il principio è semplice; ma dove entra in gioco l'intelligenza artificiale? Per rispondere a questa domanda, poniamocene un'altra: perché viene posta prima la domanda sul *reddito* e poi quella sui *debiti*? Perché viene posta propria la soglia di 80'000 franchi e non, ad esempio, di 50'000 franchi? Perché dopo il reddito viene richiesto il *patrimonio* o se è *studente*? È qui che entra in gioco l'intelligenza artificiale: il contenuto delle domande, le rispettive soglie e l'ordine in cui vengono poste le domande sono stati determinati in fase di training sulla base dei dati storici, in modo tale che il modello esegua la previsione giusta per la maggior parte dei dati di training. In altre parole, le domande e le rispettive soglie sono state sviluppate in modo da riflettere nel miglior modo possibile gli schemi di dati sottostanti.

Il modello è quindi inizialmente solo la struttura che attraverso il training, cioè il riconoscimento degli schemi, viene portata in una forma in cui può svolgere la propria funzione, cioè la previsione o la classificazione di nuovi dati.

Ma un modello non è solo un modello: prima ancora del training, dobbiamo configurarlo. Ma cosa c'è da configurare nel nostro esempio? Per rispondere a questa domanda, ci poniamo un'altra domanda: se il modello viene davvero ottimizzato rispetto ai dati di training durante l'addestramento, perché scendiamo al massimo di tre livelli (layers)? Non potremmo estendere l'albero decisionale a cinque livelli? Oppure a 50 livelli? O addirittura a tanti livelli quanti sono i record, in modo da ottenere la risposta giusta per ogni singolo record utilizzato nel training? Di sicuro il modello sarebbe decisamente migliore, vero? Possiamo trovare una risposta analizzando i dati di training. L'adattamento illimitato (in inglese *fitting*) del modello ai dati di training (e non ai dati generali al di fuori dei dati di training) porta al cosiddetto *overfitting*, per cui il modello fornisce una stima perfetta dei dati di training. Per i nuovi dati, tuttavia, è spesso difficile per il modello effettuare una stima corretta poiché non ha registrato la natura del problema, ma si è allenato ostinatamente alla perfezione con riferimento ai dati di training. Per evitare che ciò accada e per incoraggiare il modello a considerare solo le informazioni più importanti dai dati di training, limitiamo il numero massimo di layer.

Questo *numero massimo di layer* è un parametro che possiamo impostare o configurare. Se lo impostiamo troppo basso, l'algoritmo potrebbe non essere in grado di riprodurre l'intera logica del problema. Se lo impostiamo troppo alto, c'è il rischio di *overfitting*. Possiamo definire molti altri parametri, per ognuno dei quali dobbiamo trovare il valore ottimale.

La ricerca della configurazione ottimale, l'ottimizzazione dei parametri, è chiamata anche Hyperparameter Optimization. In questo contesto si analizzano diverse combinazioni di parametri e si cerca di trovare i parametri ottimali per ottenere il modello migliore.

Termini importanti: decision tree, random forest, Xgboost, logistic regression, GLM, Neural Networks, Linear Regression, Ridge Regression, Lasso Regression, Polynomial Regression, Bayesian Linear Regression, Elastic Net Regression, KMeans, Nearest Neighbor, Overfitting, Hyperparameter Optimization.

Testing

Una parte dell'ultima affermazione del paragrafo precedente non è ancora stata trattata esaurientemente: cosa significa esattamente migliore, cosa significa bene? Quando un modello è migliore dell'altro? Per rispondere a queste domande, dobbiamo innanzitutto introdurre due nuovi termini:

Dati di test: per valutare il modello, dobbiamo applicarlo ai dati (per i quali conosciamo il risultato reale) e confrontare il risultato calcolato dal modello con il risultato reale. Si deve tuttavia considerare che è pericoloso eseguire la valutazione del modello utilizzando gli stessi dati del training, poiché abbiamo già addestrato il modello proprio su questi dati (si potrebbe paragonare a un esame di guida eseguito sullo stesso percorso e con lo stesso traffico delle ore di training o scuola guida; in questo modo, non si verifica la capacità del guidatore di gestire situazioni nuove). È invece una buona idea prendere nuovi dati per valutare il modello. Questi dati sono denominati dati di test (test data).

Metrica: la metrica è la misura con cui valutiamo un modello. Per semplificazione, qui ci limitiamo alle metriche per la valutazione di classificatori binari (cioè classificatori con due classi come possibile risultato, ad esempio classe sì o classe no). Va detto che, naturalmente, esistono anche altre metriche per la regressione e il clustering che, a dipendenza dal problema, possono rappresentare una misura di confronto adeguata.

Una possibile metrica può essere definita nel modo seguente: *considerando tutte le valutazioni, quale percentuale di valutazioni è stata eseguita correttamente?* Possiamo esprimere questa metrica in un altro modo, utilizzando ancora l'esempio della solvibilità:

		Realtà	
		Positiva	Negativa
Previsione attraverso IA	Positivo	True Positive	False Positive
	Negativo	False Negative	True Negative

Abbildung 17: Confusion Matrix

La variabile che vogliamo prevedere è quindi la solvibilità effettiva. Denominiamo i casi:

- True positive: sono i casi che il modello classifica come solvibili (positivi) e che sono effettivamente solvibili.
- True negative: sono i casi che il modello classifica come non solvibili (negativi) e che sono effettivamente non solvibili.
- False positive: sono i casi che il modello classifica come solvibili (positivi), ma che non sono solvibili.
- False negative: sono i casi che il modello classifica come non solvibili (negativi) e che invece sono solvibili.

Nel caso ottimale, un modello riconosce correttamente tutti i veri positivi (true positive) e anche tutti i veri negativi (true negative). Il numero di false positive (riconosciuti come positivi, anche se negativi) e di false negative (riconosciuti come negativi, anche se positivi) deve sempre essere mantenuto il più basso possibile.

Pertanto, la metrica precedentemente indicata può anche essere definita come (true positive + true negative) / (true positive + true negative + false positive + false negative). Ci riferiamo a questa metrica anche con il termine *accuratezza* (inglese *Accuracy*). Un'elevata accuratezza è in generale da considerarsi buona; ad esempio, un'**accuratezza** del 95 % significa che il 95 % di tutti i dati è stato riconosciuto correttamente.

Tuttavia, il 95% non garantisce necessariamente un modello *buono* in ogni caso. Supponiamo, ad esempio, che una banca non conceda un prestito solo al 5% di tutti i richiedenti. Al contrario, concede un prestito al 95% dei clienti. Se dovessimo costruire il più semplice di tutti i modelli, ovvero un modello che esegue la valutazione della solvibilità per ogni cliente, avremmo anche un'accuratezza del 95%; tuttavia, il modello sarebbe tutt'altro che buono: se l'obiettivo dell'IA fosse quello di identificare i clienti non solvibili, il risultato sarebbe completamente catastrofico.

Un'alternativa che non influisce sul cosiddetto Imbalance – quindi sulla diversa ponderazione delle due classi – è formata dalla coppia precisione (inglese *Precision*) e *recall*. Cosa sono queste due metriche?

- **Precisione:** la precisione indica quale quota di tutti i casi contrassegnati come positivi (true positive) è effettivamente positiva, quindi $[\text{true positive}] / [\text{true positive} + \text{false positive}]$.
- **Recall:** il recall misura la quota di casi da accettare che sono stati effettivamente accettati, quindi $[\text{true positive}] / [\text{true positive} + \text{false negative}]$.

Per entrambe le metriche, è ovviamente auspicabile ottenere il valore più elevato possibile. La cosiddetta media armonica di precisione e recall $[2 * \text{precisione} * \text{recall}] / [\text{precisione} + \text{recall}]$ che è denominata anche **F1 Score** ed è una delle metriche più diffuse per i problemi con classi sbilanciate. Anche in questo caso, tuttavia, occorre prestare attenzione ad alcuni elementi, come illustriamo nell'esempio seguente.

Esempio: una banca concede un prestito a 9'500 clienti solvibili su 10'000 potenziali clienti. Consideriamo la problematica dal punto di vista in cui siamo particolarmente interessati, ossia alla corretta previsione della solvibilità. Quindi consideriamo **solvibile** come la label **positiva** e **non solvibile** come la label **negativa**. Supponiamo che il modello classifichi **9'600 clienti** come solvibili (label positiva). Di questi, **9'400** erano effettivamente solvibili (**true positive**) e **200** sono stati erroneamente classificati come tali (**false positive**). Tra le **400** persone classificate come non solvibili dal modello, **100** erano solvibili (**false negative**) e **300** non erano solvibili (**true negative**).

- Anche la **precisione** sarebbe elevata: $9'400 / (9'400 + 100) = 98.9 \%$.
- Il **recall** sarebbe leggermente più basso: $9'400 / (9'400+200) = 97.9 \%$.
- **F1 Score** si calcola come media armonica di precisione e recall ed è circa del **98.4 %**. Si tratta di un valore piuttosto elevato. Questo è particolarmente evidente se lo confrontiamo con un modello che assegna la label solvibile a ogni persona; in questo caso il punteggio F1 sarebbe del **97,4%**. Anche se decidessimo a caso la solvibilità di ogni persona con una probabilità del 95%, non saremmo in una posizione miglio-re con un F1 score del **95,0%**.

Ma come si presenta tutto questo se il nostro obiettivo è quello di scartare il maggior numero possibile di clienti non solvibili? In questo caso, consideriamo i clienti **non solvibili** come la label positiva (cioè l'etichetta di nostro interesse).

- L'**accuratezza** è: $(9'400 + 300) / 10'000 = 97.0 \%$, quindi va ancora bene.
- La **precisione** è $300 / (300 + 200) = 60.0 \%$, significativamente inferiore alla precisione calcolata prima.
- Anche il **recall** è solo $300 / (300 + 100) = 75.0 \%$.
- **F1 Score** è solo del **66.7 %**.

Questo risultato è migliore rispetto a quello di indovinare a caso la solvibilità o meno di un cliente. Resta tuttavia da valutare se il risultato può essere ritenuto sufficientemente buono e, alla fine, questo dipende dagli obiettivi della banca.

L'esempio precedente ha mostrato che il punteggio F1 score è una buona metrica per la valutazione di un modello. Tuttavia, le metriche non rappresentano solo una possibilità per valutare un modello. Piuttosto, consentono di valutare e confrontare più modelli utilizzando proprio la stessa metrica. In particolare, quando si tratta di configurare un modello, la metrica aiuta a trovare la configurazione migliore, perché anche in questo caso vale quanto segue: tra due modelli dello stesso tipo, con configurazioni diverse, di solito è preferibile il modello con la metrica migliore: la sua configurazione è migliore di quella dell'altro modello.

Riprendiamo ancora l'esempio precedente, in cui abbiamo considerato un albero decisionale (decision tree). Come facciamo a sapere se dobbiamo selezionare 5, 50, 500 o 5000 layer come profondità massima dell'albero decisionale? Questo è difficile da valutare a priori: dobbiamo addestrare il modello per ogni singola configurazione (5, 50, 500 o 5000 layer) e confrontare i diversi modelli utilizzando una metrica come il punteggio F1 score. Questo tipo di ottimizzazione della configurazione, già introdotto nello scorso capitolo, è denominato Hyperparameter Optimization.

Dopo aver evidenziato gli elementi tecnici della creazione di un'IA, vorremmo eseguire un'introduzione al *Cross Industry Standard Process for Data Mining* (CRISP-DM). Questo standard, spesso utilizzato (consapevolmente o inconsapevolmente), comprende le seguenti fasi:

1. **Business Understanding:** comprensione professionale del problema
2. **Data Understanding:** comprensione dei dati
3. **Data Preparation:** modifica e arricchimento dei dati
4. **Modeling:** creazione del modello, incluso il training
5. **Evaluation:** valutazione del modello in base a KPI professionali e tecnici
6. **Deployment:** entrata in produzione del modello, ad esempio locale o sul cloud

Richiamare regolarmente questo standard è utile per due motivi: in primo luogo, questo approccio sottolinea che la creazione di un'IA senza una comprensione tecnica ha poco senso. Anche se un data scientist si sente a proprio agio nel suo settore: in ogni progetto che prevede un valore aggiunto attraverso l'IA, il data scientist dovrebbe acquisire almeno una comprensione di base del background professionale. L'ingegnerizzazione delle

feature è una parte importante nel pre-processing; quanto migliore è la comprensione professionale, tanto più facile è generare nuove feature che facilitano il riconoscimento degli schemi per il modello.

D'altra parte, questo processo sottolinea anche che un progetto con l'obiettivo di un'IA non si esaurisce con la hyperparameter optimization, rispettivamente l'addestramento del modello ottimale. La messa in funzione del modello, ossia la sua introduzione nei processi esistenti, la sostituzione di processi e le interfacce grafiche che visualizzano i risultati devono essere attentamente pensati e non devono essere sottovalutati.

Termini importanti: dati di test, metrica, accuratezza, precisione, recall, F1 Score, true positive, true negative, false positive, false negative.

11.5 La vulnerabilità di un modello di IA

Nella realizzazione di un'IA bisogna essere sempre consapevoli, in ogni momento, dei limiti di un modello di IA. La validità di un modello di questo tipo dipende dai dati che sono stati utilizzati per il suo addestramento. Se i dati con cui viene addestrata un'IA non hanno la forma giusta o sono comunque di bassa qualità, non avrà alcuna importanza per quanto tempo si addestrerà il modello: non si otterrà mai un buon risultato. Ma quali sono le vulnerabilità di un'IA?

Overfitting: abbiamo già introdotto la nozione di overfitting, secondo la quale durante l'addestramento il modello viene orientato in modo eccessivo ai dati di training. In questo caso, le performance sui dati di training sono generalmente molto elevate, ma allo stesso tempo sono molto basse sui dati di test. L'overfitting può essere aggirato scegliendo i parametri in modo tale che il modello debba eseguire una generalizzazione sufficiente del problema. L'hyperparameter optimization ci aiuta a trovare questa configurazione dei parametri. Valutando le prestazioni del modello non sui dati di addestramento, bensì sui dati di test, l'hyperparameter optimization riesce a trovare i parametri appropriati che consentono una generalizzazione sufficiente su tutti i dati.

Tuttavia, anche questo presenta dei rischi. Per evitare che il modello si addestri su dati di test costanti, in questo processo viene solitamente utilizzata la cosiddetta cross validation. L'idea della cross validation è quella di dividere i dati a disposizione in un determinato numero di gruppi (indichiamo con n il numero di gruppi) di uguale dimensione. Ora, per ogni valutazione del modello, l'addestramento viene eseguito n volte, sempre su $n-1$ gruppi, in modo che i gruppi restanti fungano da dati di test. La bontà di un modello, misurata ad esempio sotto forma di F1 score, non viene più valutata su un singolo training, ma sugli n risultati degli n processi di training. Un passo importante per evitare l'overfitting.

Bias: il bias si riferisce alla condizione in cui l'IA tratta determinati attributi in modo «unfair». In questo contesto, unfair deve essere inteso nel senso che ai record che soddisfano un determinato attributo viene assegnata quasi sempre una classe dal modello, anche se in molti casi questa assegnazione è sbagliata.

Prendiamo ad esempio la categorizzazione della solvibilità. Nelle domande poste al cliente, egli dichiara di vivere in un luogo il cui codice postale ha un'alta densità di studenti al primo anno le cui richieste sono state respinte. Il modello è stato addestrato, tra l'altro, su questi dati. Nonostante il cliente abbia un lavoro stabile e abbia già risparmiato un piccolo patrimonio, gli viene negato il credito: solo a causa del codice postale e dell'errore di valutazione del sistema, basato solo su questo attributo.

Il bias si presenta in molte forme diverse. Spesso il bias è il risultato di dati di addestramento non bilanciati. Poiché qualsiasi sistema ha l'obiettivo di raggiungere la massima fairness, la composizione dei dati di training deve essere scelta con attenzione. Non solo la composizione dei dati di training può aiutare. Si deve anche considerare che alcune caratteristiche non dovrebbero essere incluse se non assolutamente necessarie; tra queste l'età, il genere e la provenienza.

Mancanza di trasparenza: l'intelligenza artificiale riesce ad apprendere e ad eseguire previsioni sulla base di dati storici. Tuttavia, in molti casi la sola previsione non è sufficiente. Le decisioni non possono sempre essere prese sulla base di una singola valutazione, ma devono essere giustificate. Un'idea del perché gli algoritmi decidano a favore (o contro) una specifica previsione è fornita dalla cosiddetta Explainable AI (AI = Artificial intelligence).



Figura 18: Illustrazione di un Explainable AI per un caso

Il lato sinistro mostra le feature che, in questo caso, sono contro la solvibilità, mentre il lato destro mostra le feature che sono a favore della solvibilità. La lunghezza delle barre è legata alla rilevanza della feature pro/contro. In questo esempio, gli elementi "contro" sembrano essere predominanti, almeno per i primi 5 pro e contro.

Concept drift: indica che le label, che il modello sta cercando di prevedere dai dati di input con l'apprendimento guidato, non sono sempre costanti ma cambiano nel tempo. Utilizziamo ancora una volta l'esempio della solvibilità. Assumiamo che il nostro modello sia stato addestrato su dati che hanno avuto origine in un'epoca in cui le banche utilizzavano regole molto conservative per valutare la solvibilità di un cliente. Il modello apprende quindi esattamente questo approccio conservativo. Quindi, se la banca decide di cambiare approccio e di non tenere più conto, ad esempio, dell'età del richiedente, questo cambiamento di approccio non sarà considerato dal modello fino a quando non forniremo nuovi dati di addestramento al modello (e, in condizioni ottimali, i dati obsoleti devono essere rimossi dal modello).

Termini importanti: overfitting, bias, explainable AI, concept drift.

11.6 Una strada rapida verso il primo modello

Questa breve digressione sull'intelligenza artificiale offre solo una panoramica delle diverse aree e delle possibilità che ci vengono offerte. Lo sviluppo degli ultimi anni ha messo alla prova due scenari. Da un lato, l'IA ha avuto e ha sempre maggior impatto sulla nostra vita quotidiana. D'altro canto, però, anche il clamore suscitato dall'intelligenza artificiale è stato relativizzato.

Quindi, se nella vita lavorativa di tutti i giorni riconosciamo processi che mostrano determinati schemi che potrebbero essere appropriati per l'intelligenza artificiale, dovremmo assolutamente considerare l'idea di valutare l'impiego dell'IA. Questa verifica può avvenire in tempi relativamente brevi - con l'aiuto dei cosiddetti approcci AutoML, che sono offerti sia on premise (h2o, TPOT, autosklearn) che nel cloud (Google Cloud AutoML, Azure

Auto ML, Amazon SageMaker Autopilot), in cui l'utente non deve fare altro che definire il problema. Hyperparameter Optimization e altri adattamenti vengono gestiti dagli algoritmi di AutoML.

Allo stesso tempo, però, è importante considerare esattamente il valore aggiunto che l'IA può apportare. Un altro punto centrale, inoltre, è il modo in cui la nuova IA può essere integrata nei processi, se i processi possono essere adattati o addirittura completamente sostituiti o rinnovati tenendo conto della nuova IA.

Quindi, se ci troviamo in un progetto di questo tipo in cui abbiamo individuato dei vantaggi provenienti dal riconoscimento automatico degli schemi e quindi dall'utilizzo dell'IA, dobbiamo tenere sempre in considerazione standard come il CRISP-DM o processi simili. Senza i futuri data scientist, il modello faticherà ad apprendere e a riconoscere questi schemi nella produzione. Tuttavia, senza esperti professionali e responsabili tecnici, di solito si fa fatica a valutare il modello, a ottimizzarlo, ad orientarlo al problema obiettivo e a inserirlo adeguatamente nei processi esistenti o addirittura a sostituirli o ridefinirli.

L'interdisciplinarietà è una componente centrale di un progetto di IA di successo, che deve essere curata e vissuta dall'inizio alla fine del progetto. Infine, un progetto di IA non è unidimensionale. Si basa sul feedback costante dell'azienda e sulla traduzione dei requisiti professionali in nuove feature tecniche da parte dei data scientist.

Termini importanti: AutoML, CRISP-DM.

12 Trasformazione digitale e nuovi modelli di riferimento

La digitalizzazione è onnipresente e già profondamente radicata nelle imprese come «megatrend». Da anni ormai non è quasi più possibile sfuggirne, né in ambito privato né in ambito aziendale, e da numerosi punti di vista la maggioranza di tutte le generazioni non lo vuole (più); la sua diffusione e i vantaggi che ne derivano sono troppo grandi. Naturalmente, in alcuni ambiti della digitalizzazione ci sono sviluppi discutibili o addirittura assurdi, oltre alle condizioni talvolta poco chiare che riguardano la protezione e la sicurezza dei dati. A parte questo, la generale accettazione della digitalizzazione mostra come essa sia decisamente giustificata in tutte le generazioni, come è peraltro dimostrato dalla sua rapida diffusione negli ultimi anni o addirittura negli ultimi decenni. Ma cosa significa effettivamente «digitalizzazione»? Come interagiscono «digitalizzazione» e «trasformazione digitale»? E quali nuovi modelli di riferimento acquisteranno importanza, in futuro, nel mondo finanziario digitale? Questo capitolo esplorerà queste e altre questioni con la necessaria profondità, spiegandole con esempi pratici e grafici.

12.1 Digitalizzazione

Sin dalle sue origini, con digitalizzazione si intende la trasformazione di informazioni analogiche in informazioni digitali, nonché la loro successiva elaborazione e memorizzazione su un supporto (media) digitale. Per digitalizzazione si intendono quindi le tecnologie che ci circondano come privati e aziende in diversi ambiti e in varia misura, influenzando in modo fondamentale il comportamento, i processi, la collaborazione, le condizioni competitive e molto altro ancora.

La digitalizzazione crea quindi le condizioni necessarie per la conseguente trasformazione digitale attraverso la conversione tecnologica dei prodotti esistenti e la creazione di nuovi prodotti.



Figura 19: Trasformazione digitale come parte della digitalizzazione

12.2 Obiettivi di una trasformazione digitale

Come suggerisce il nome, la trasformazione digitale è un processo di cambiamento innescato dalla digitalizzazione. I driver di questo cambiamento possono essere molteplici, per cui la velocità della trasformazione aumenta con l'aumentare del numero di driver e il contemporaneo aumento della pressione per l'attuazione. I principali motori di una trasformazione digitale possono essere, tra gli altri, i seguenti:

Driver esterni	Driver interni
Economia	Capacità dei collaboratori e loro comportamenti di lavoro
Concorrenza	Strategia dell'azienda
Clienti	Ricerca e sviluppo
Ecologia	ecc.
ecc.	

Per sviluppare una miglior comprensione della trasformazione digitale, si può ricorrere all'aiuto del **modello dei tre livelli di conoscenza digitale** (Martin, 2008). Il modello divide il percorso verso l'effettivo processo di trasformazione in tre fasi principali e le descrive in modo tale che si debba dapprima costruire la competenza digitale (level 1: Digital Competence – skills, concepts, approaches, attitudes, ecc.). In seguito, si trova il secondo livello (level 2: Digital Usage – professional / discipline application), che prevede l'uso innovativo e creativo di applicazioni digitali significative e ben congegnate, messe a disposizione come strumenti di lavoro in linea con la strategia generale. Solo al terzo livello avviene l'effettiva trasformazione digitale (level 3: Digital transformation).

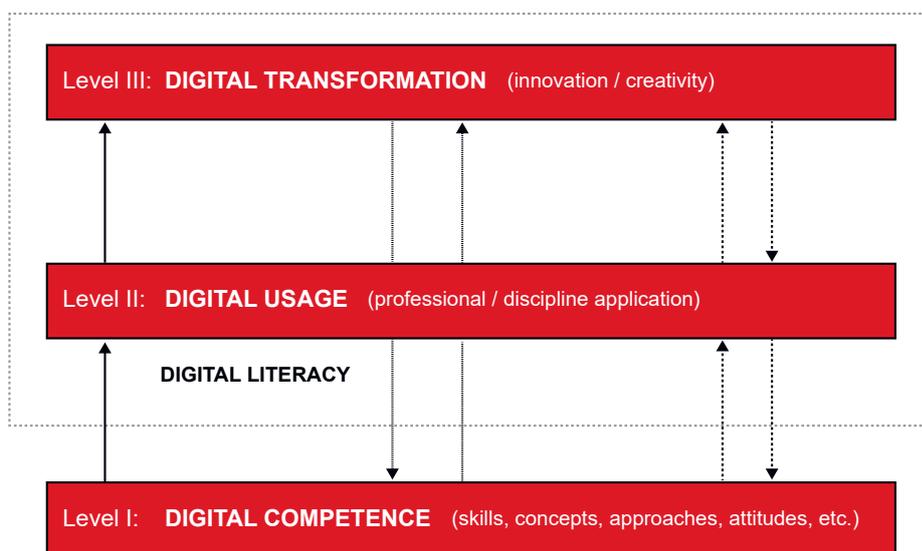


Figura 20: Levels of Digital Literacy (Martin, 2008)

Con il raggiungimento del terzo livello, la trasformazione digitale vera e propria (Ruoss, 2015) consente di raggiungere gli obiettivi elencati nella figura 21 nelle aree fondamentali «aumento della fidelizzazione dei clienti», «aumento dell'efficienza interna» e «aumento delle vendite attraverso nuovi prodotti».

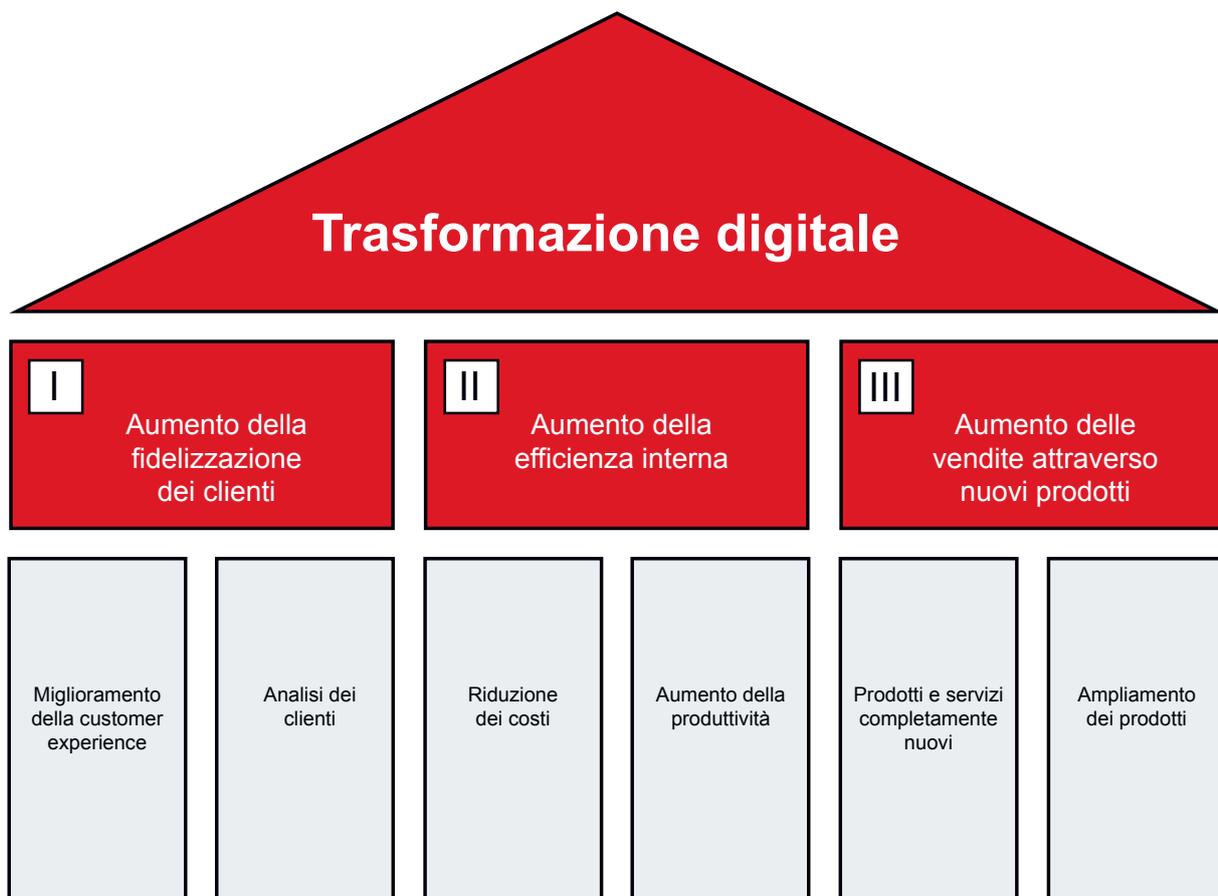


Figura 21: Obiettivi della trasformazione digitale (Ruoss, 2015)

Aumento della fidelizzazione dei clienti

Il comportamento collaborativo delle persone e il modo in cui si raccolgono le informazioni che le riguardano nella sfera privata e/o nella vita professionale sono già cambiati in modo considerevole grazie alle possibilità create dalla digitalizzazione e continueranno a farlo in futuro, a una velocità che si prevede in continuo aumento. Ad esempio, i clienti vengono contattati attraverso diversi canali (social media) e piattaforme digitali e questo anche a causa delle diverse aspettative dei clienti; questo significa che la complessità per le aziende è in costante aumento. L'azienda deve allineare alle esigenze dei clienti tutti i processi relativi alla customer experience. Altrettanto complesso è il processo che le aziende devono svolgere per acquisire le informazioni che consentono loro di comprendere le esigenze attuali e future dei clienti e per adattare di conseguenza i propri prodotti e servizi.

Aumento dell'efficienza interna all'azienda

L'aumento dell'efficienza all'interno dell'azienda viene realizzato attraverso una riduzione dei costi, a sua volta ottenuta grazie alla digitalizzazione e al conseguente aumento della produttività. Benché richieda un investimento iniziale e successivamente comporti costi correnti, la digitalizzazione consente alle aziende di ottenere dei benefici elevatissimi rispetto al capitale investito; questi benefici sono realizzabili grazie alla riduzione dei costi conseguente l'automazione dei processi. Lo stesso vale anche per l'aumento della produttività. Grazie alla digitalizzazione l'azienda riesce a ridurre i costi e, contemporaneamente, a diventare più produttiva; questo si traduce in un elevato rendimento sotto forma di maggiore efficienza che, in ultima analisi, impatta anche sulla liquidità e sul risultato economico dell'azienda.

Aumento delle vendite attraverso nuovi prodotti

La digitalizzazione è un driver essenziale dell'innovazione. A dipendenza del tipo e delle caratteristiche dell'innovazione, la digitalizzazione consente di creare prodotti e servizi completamente nuovi oppure di migliorare i prodotti esistenti. È così possibile aprire nuovi mercati grazie alla digitalizzazione.

A questo punto è importante sottolineare che la trasformazione digitale può essere attuata con successo solo se tutte le aree principali di un'azienda interagiscono in modo significativo. Di conseguenza, la trasformazione digitale fallisce se i prerequisiti dei livelli I e II non sono soddisfatti e/o le aree principali della trasformazione digitale non sono coordinate tra loro. Inoltre, la trasformazione digitale porta automaticamente un'azienda a doversi occupare intensamente del proprio modello di business e dei relativi processi operativi.

12.3 Sfide della trasformazione digitale

Le sfide della trasformazione digitale (Ruoss, 2015) provengono da tre aree: «Avvio», «Implementazione» e «Coordinamento»:

Avvio

Mancanza di slancio, vincoli normativi, preoccupazioni per la sicurezza e la protezione dei dati, reputazione, business case poco chiaro.

Implementazione

Mancanza di know-how, mancanza di cultura del cambiamento, difficoltà IT.

Coordinamento

Pensiero a compartimenti stagni, scarso coordinamento generale, «management attention» insufficiente, conduzione top-down.

12.4 Roadmap di una Digital Finance Transformation

Oltre a una cultura aziendale favorevole, un importante prerequisito per una Digital Finance Transformation strutturata è una strategia ben ponderata e definita chiaramente, che prenda anche una posizione chiara in merito alla visione digitale dell'azienda. Una volta stabiliti questi prerequisiti generali, l'implementazione operativa si ottiene al meglio con una roadmap (ExpertFocus, 2019 / 12) che definisce le milestone essenziali. L'intero processo è accompagnato in tutte le fasi da una comunicazione continua e trasparente con i collaboratori. Questo rende la trasformazione digitale più accessibile ai collaboratori e riduce eventuali incertezze o timori.

La sequenza delle milestone segue una certa logica. Il successo dell'implementazione della trasformazione digitale è garantito quando ogni fase, in sequenza, viene completata e implementata con successo.

Archiviazione elettronica

Prima di iniziare l'automazione vera e propria, l'archivio fisico deve essere convertito in formato digitale. Nel fare ciò, si deve prestare attenzione a rispettare i requisiti interni dell'azienda e i requisiti normativi. Un ulteriore guadagno di efficienza si ottiene concordando con i diversi produttori di documenti (fatture e altri documenti) di eseguirne lo scambio solo in forma elettronica.

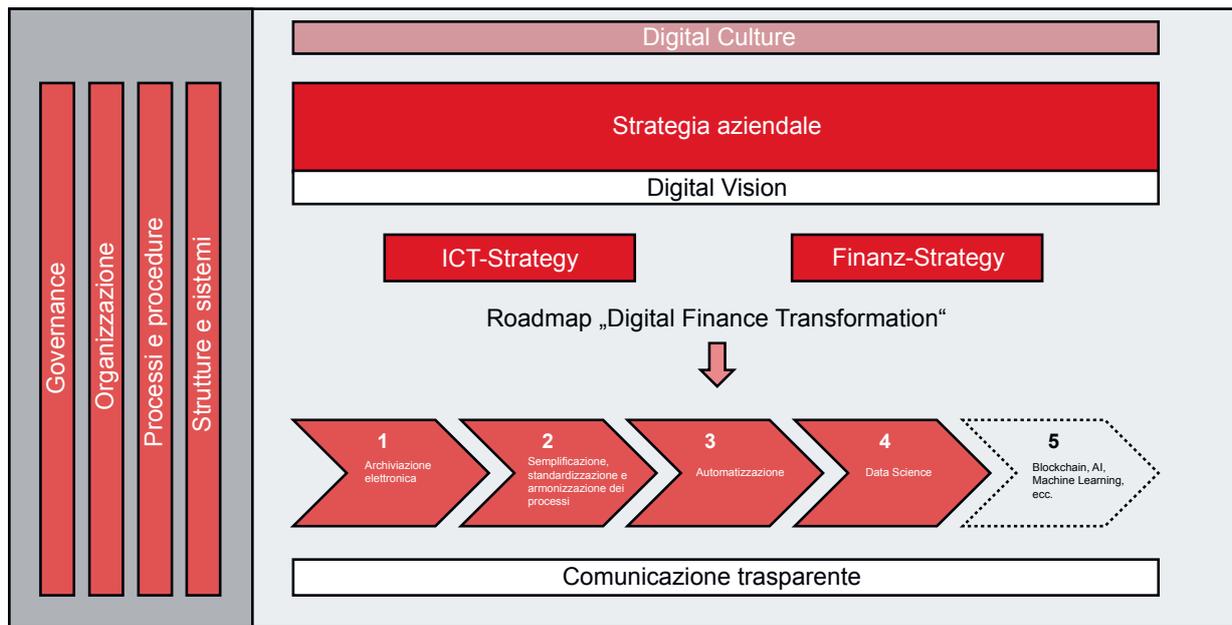


Figura 22: Digital Finance Transformation

In questo caso, un'azienda dovrebbe porsi le seguenti domande, in particolare per l'implementazione dell'archiviazione elettronica (ExpertFocus, 2019 / 12):

- Qual è l'attuale politica di archiviazione? Come deve essere adattata all'era della digitalizzazione e quali reparti interni all'azienda devono essere coinvolti?
- Quale sistema centrale viene utilizzato per l'archiviazione elettronica?
- Dove vengono memorizzati e conservati i dati?
- Come è possibile trasformare in formato digitale tutti i documenti in entrata?

Semplificazione, standardizzazione e armonizzazione dei processi

Nella seconda fase, i processi semplici vengono standardizzati all'interno di un Enterprise Resource Planning System (ERP), in modo da aumentare l'efficienza e ridurre gli errori; in seguito, devono essere trasferiti in routine programmate. Grazie alla semplificazione, alla standardizzazione e all'armonizzazione, nella fase successiva è possibile automatizzare soprattutto i processi end-to-end più semplici.

Automatizzazione

L'automazione dei processi può avvenire in diversi modi. In generale, si tratta di processi end-to-end che prima venivano elaborati manualmente e che ora vengono elaborati automaticamente senza l'intervento umano. In termini di tecnologia, l'automazione può avvenire attraverso le possibilità offerte dal sistema ERP centrale e/o attraverso l'uso della Robotic Automation (RPA) o dell'intelligenza artificiale. Si può, pertanto, ipotizzare che in futuro i sistemi ERP combineranno tutte le possibilità di automazione.

Data Science

Una volta che la contabilità aziendale è stata completamente automatizzata, entra in gioco la disciplina superiore, il **data science**. Questo comporta l'utilizzo di tecnologie di data science per ricavare conoscenze, sotto forma di schemi e modelli, sia dai dati precedenti e ora completamente disponibili in formato digitale sia dai nuovi dati che vengono costantemente aggiunti. Tutto ciò sarà ulteriormente utilizzato in azienda in modo vantaggioso. Le conoscenze acquisite possono essere utilizzate per lo sviluppo di nuovi prodotti e servizi o per l'ampliamento di quelli esistenti.

Tuttavia, il data science non si limita ad acquisire conoscenze attraverso la valutazione dei dati dei sistemi in-terni esistenti. Si spinge molto oltre e combina i dati interni con quelli disponibili all'esterno, provenienti da diversi canali. Solo in questo modo è possibile accertare i bisogni complessivi di tutti gli stakeholder rilevanti e adeguare i prodotti e i processi dell'azienda.

Blockchain, AI, Machine Learning

L'ultimo passo della roadmap si riferisce al continuo e ulteriore miglioramento dei processi e alla valutazione dei dati attraverso strumenti come il process mining e l'uso di nuove tecnologie concepibili in futuro.

Tutte le fasi della roadmap fanno parte di un processo di controlling agile e devono essere costantemente controllate per verificare la presenza di cambiamenti tecnologici o di nuovi requisiti. Da questo punto di vista, la trasformazione digitale non è un processo statico una tantum, ma deve essere attuata come un processo di controlling continuo anche dopo la sua implementazione con successo.

13 Cambiamento della conduzione finanziaria e del controllo dell'organizzazione

Il data mining (valutazione automatizzata di grandi quantità di dati per riconoscere determinati schemi), associato al data science, consente alle aziende di riconoscere gli schemi dai dati raccolti internamente ed esternamente e combinarli tra loro in modo significativo, per sviluppare scenari e infine prendere decisioni (Data Driven Decision Making). La sfida del controlling e, in ultima analisi, anche della conduzione finanziaria dell'azienda consiste nel raccogliere in qualsiasi momento i dati rilevanti dalle corrispondenti fonti interne ed esterne, valutarli e trarne le conclusioni strategicamente e operativamente rilevanti. La conduzione finanziaria si estende quindi a tutti i canali fino alla valutazione dei prodotti digitalizzati, aumentando così le esigenze post-est alla funzione di controller.

13.1 Nuovi modelli di ruoli delle funzioni finanziarie nel mondo digitale

Introduzione e comprensione dei ruoli

Il concetto di conduzione/direzione, ancora oggi molto diffuso, si riferisce al potere della posizione di manager. I manager avevano il potere attraverso la loro funzione; questo significa che potevano lodare, sanzionare o punire. Alcune generazioni fa, l'autorità dei manager non avrebbe mai potuto essere messa in dubbio, poiché la forza del loro «potere» prevaleva sempre. Nella nuova concezione di direzione, ai manager non viene più attribuito il potere, bensì vengono assegnati compiti e ruoli. Oggi i manager non sanno più tutto e non sono più infallibili, anzi, commettono anche degli errori. Ammettere i propri errori, tornare sulle decisioni sbagliate e correggerle è percepito come un punto di forza dalla grande maggioranza dei collaboratori. L'autoriflessione critica dei manager ha un effetto esemplare. I manager affermati interpretano la loro funzione come supporto ai collaboratori dei team. Ciò significa che il manager, che sia gerarchicamente legittimato o meno, è anche uno sviluppatore di team.

Gli esperti di organizzazione concordano oggi sul fatto che le organizzazioni divisionali e a matrice non sono in grado di sopravvivere dal punto di vista operativo, perché non possono resistere a mercati dinamici. Le organizzazioni gerarchiche vengono sostituite da organizzazioni reticolari modulari e decentralizzate.

La comprensione del ruolo dei collaboratori si basa su compiti, responsabilità, diritti e competenze. Il collaboratore ha il diritto (e il dovere) di muoversi liberamente all'interno del proprio ruolo, di interpretare i propri compiti, di svolgere la propria mansione e di ottimizzarla se necessario.

La **posizione** descrive il posto formale che una persona occupa all'interno di un'organizzazione, ad esempio direttore delle vendite, CEO, project manager, team leader. La **funzione** definisce lo scopo e gli obiettivi dell'attività. Nel **ruolo** sono descritti compiti, diritti, doveri e competenze. Questi sono estratti anche dalle descrizioni dei processi dell'organizzazione. Tutte le descrizioni dei ruoli descrivono l'intera organizzazione e tutti i compiti.

È opportuno sottolineare i seguenti punti:

- Le descrizioni dei ruoli non sono legate alle persone
- Una persona non può ricoprire più ruoli
- I ruoli possono essere ricoperti anche solo temporaneamente.

Le descrizioni dei ruoli non sono interpretabili liberamente, bensì sono delimitate da obblighi e competenze. Il concetto di ruolo presuppone che nessuno sia posto al di sopra degli altri, ma che nell'organizzazione tutti abbiano lo stesso valore.

Nel ruolo sono descritti i compiti. È quindi più probabile che i collaboratori competenti siano predestinati ad assumere un ruolo nuovo o diverso o addirittura si impongano su di esso.

La precedente concezione di manager è sotto pressione. Non si tratta più di «superiori», ma di «manager leader» che, nello svolgimento del loro ruolo, attivano e promuovono il potenziale dei collaboratori e rimuovono gli ostacoli presenti nei team.

Il ruolo del superiore non è più quello di ricevere risultati dai propri collaboratori, bensì di sostenere i suoi collaboratori e i suoi team come leader, facendo leva sulle loro potenzialità.

Queste considerazioni influenzano anche la gestione delle risorse umane. In passato, le aziende davano molta importanza alla compatibilità dei candidati con il team e l'organizzazione. Oggi, oltre alle considerazioni sociali e culturali, l'attenzione si concentra sempre più sulla promozione dell'innovazione e sulla guida del team per ottenere le massime prestazioni. Ciò richiede attriti e ricchezza di prospettive. L'obiettivo del reclutamento non è più quello di cercare persone «uguali», ma «diverse». Molti manager affermati hanno un problema con questo nuovo approccio, perché si considerano i decisori in caso di divergenze di opinioni. Questo è sbagliato perché i collaboratori sono sempre più spesso in grado di risolvere autonomamente i conflitti e anche di cercare soluzioni imprenditoriali.

Modelli di ruolo nell'era della digitalizzazione

I nuovi modelli di ruolo derivanti dalla digitalizzazione richiedono sempre più capacità analitiche e diagnostiche in relazione a componenti e funzioni meccaniche ed elettroniche. In questo senso, la complessità del nostro lavoro sta aumentando. I tool IT, i robot e i processori diventeranno sempre più pervasivi, cambiando le nostre vite e le nostre competenze professionali. Se un tempo questi sistemi erano in grado di svolgere compiti piuttosto semplici, oggi sono in grado di svolgere compiti ripetitivi e sempre più complessi. L'analisi delle nuove esigenze del mondo del lavoro mostra che, con l'eliminazione di molti lavori e attività semplici, i profili professionali esistenti cambieranno ed emergeranno nuove attività e nuovi profili professionali.

Mondo VUCA

Oltre ai vantaggi della globalizzazione, la nostra società ha sperimentato anche alcuni svantaggi. Si pensi al fatto che, per le generazioni passate, solo i disastri naturali e le malattie avevano un impatto globale. Più recentemente, il crollo del blocco orientale all'inizio degli anni Novanta ha dimostrato, ad esempio, che anche i cambiamenti nei sistemi politici hanno un impatto globale sulle economie nazionali e sulla società.

L'acronimo adottato per descrivere il mondo multilaterale è il termine VUCA, coniato in base alle iniziali delle parole Volatility, Uncertainty, Complexity e Ambiguity (volatilità, incertezza, complessità e ambiguità).

A livello aziendale, il mondo VUCA è caratterizzato dal fatto che le strategie e i mezzi di ieri non sono più sufficienti per affrontare e risolvere le sfide di oggi, soprattutto per quanto riguarda la pianificazione a lungo termine e l'orientamento agli obiettivi. Mentre in passato la pianificazione annuale, basata sulla gestione finanziaria e sul confronto tra cifre preventive e cifre consuntive, creava sufficiente trasparenza, al giorno d'oggi gli eventi a breve termine possono stravolgere l'intero processo di pianificazione. I manager e i dirigenti devono far fronte a una pletora di informazioni e requisiti, poiché non hanno più alcuna forza di previsione. In questa situazione, il management agile aiuta ad alleggerire le richieste poste ai singoli manager.

In futuro le aziende e le organizzazioni avranno bisogno di collaboratori che pianifichino meno, ma più a breve termine, che pensino in modo aperto, che raccolgano molteplici conoscenze specialistiche da prospettive diverse, che imparino rapidamente attraverso la sperimentazione e gli errori condivisi e che puntino all'eccezionalità operativa e strategica. Affrontare il mondo VUCA è molto impegnativo e riguarda soprattutto il livello strategico dell'azienda.



Figura 23: Mondo VUCA (Zäch, 2021)

Tuttavia, i giochi mentali associati alla gestione del mondo VUCA sono strettamente intrecciati con il mindset dell'agilità.

Condizione agile

Condizione agile significa lavorare in modo da garantire che l'azienda possa adattarsi ai cambiamenti il più rapidamente possibile. La condizione agile sostiene i propri collaboratori nel trovare insieme le soluzioni migliori per affrontare le sfide. In questo, tiene conto delle esigenze di tutti gli stakeholder. I valori centrali della condizione agile sono l'apertura alle nuove idee, la comunicazione e le gerarchie piatte.

I principi più importanti della condizione agile sono:

1. Apertura ai cambiamenti
2. Approccio olistico
3. Fiducia nel team e nel lavoro svolto
4. Decisioni prese in modo agile (rapidamente)
5. Le iterazioni possono essere utili
6. Creazione di una forte cultura del feedback
7. Affrontare gli errori in modo positivo (cultura dell'errore)
8. Promozione dei contatti e della comunicazione all'interno del team
9. Cambiamento della concezione della gerarchia (condizione gerarchica versus gerarchia piatta)
10. Utilizzo di metodi agili

Introduzione di valori agili

Oltre alla moltitudine di competenze, la trasformazione digitale riguarda anche lo sviluppo di valori agili.

Questi valori possono essere distinti in tre livelli. Si deve considerare che con l'introduzione della condizione agile è indispensabile raggiungere almeno il livello intermedio. Se ciò non fosse possibile, questi livelli dovrebbero essere sviluppati preventivamente. Si deve dapprima determinare il livello attuale dell'organizzazione e mapparli sulla base della rappresentazione qui presentata. Diventa subito visibile se ci sono diversi livelli nell'azienda. Questo renderebbe difficile il cambiamento culturale desiderato, perché non c'è una base comune.



Figura 24: Elementi dei valori agili (Zäch, 2021)

Il fondamento formale della conduzione agile è lo scopo del lavoro e l'obiettivo comune del lavoro. In molti casi, non si tiene conto del fatto che tutti i soggetti coinvolti devono costruire una base di fiducia a livello fattuale e a livello sociale. Le riserve e i dubbi devono essere dissipati in modo che si possa sviluppare una collaborazione di fiducia.

Nel costruire la fiducia è importante sostenere i collaboratori nelle seguenti competenze:

- Promuovere l'apertura, ad esempio attraverso scambi informali, progetti più semplici e condivisione delle conoscenze
- Disponibilità a impegnarsi nel conflitto, ad esempio attraverso la definizione di regole, spazi protetti / opportunità di fare un passo indietro
- Impegno personale e commitment, ad esempio attraverso obiettivi personali condivisi e comunicati, nonché distinzione tra obiettivi operativi e obiettivi comportamentali
- Impegno reciproco, ad esempio attraverso obblighi comunicati
- Orientamento agli obiettivi, ad esempio attraverso ritiri per concentrarsi sul lavoro

Le relazioni di fiducia nei team sono caratterizzate dal fatto che i membri del team si conoscono molto bene, conoscono cose personali l'uno dell'altro e trattano le conoscenze personali in modo confidenziale, anche nelle crisi. I cambiamenti descritti possono avvenire solo attraverso l'evoluzione e non possono mai essere ordinati. Dal punto di vista della conduzione, la fiducia si manifesta nella progressione della fiducia stessa, nella delega di competenze, nella responsabilità e nell'apprezzamento per il lavoro svolto.

13.2 Change management

«L'unica costante nella vita è il cambiamento.» (Eraclito, 600 a. C.)

La citazione del filosofo greco Eraclito è oggi più che mai attuale. Pone l'accento sul fatto che il cambiamento è inevitabile e che le persone devono imparare ad accettarlo.

Come già detto, la digitalizzazione sta creando un cambiamento globale a una velocità senza precedenti. Né le dimensioni dell'azienda, né il marchio, né la tradizione, né l'esperienza possono proteggere le imprese dalla digitalizzazione. Questo ambiente imprevedibile e in rapida evoluzione viene definito anche ambiente VU-CA. VUCA sta per ambiente volatile, incerto, complesso e ambiguo.

Il cambiamento stesso è un fenomeno complesso, che dovrebbe essere semplificato attraverso l'uso di metodologie come il change management. Tuttavia, a causa della complessità, della velocità e della profondità dei cambiamenti, il change management già sperimentato non è all'altezza della trasformazione digitale. Il motivo è che in molte aziende i «vecchi» approcci al cambiamento sono vissuti come un processo progettato ed eseguito in modo top-down. Inoltre, la gestione classica del change management risale a un'epoca in cui i cambiamenti tecnologici non erano così forti come oggi.

Il change management non può più essere visto come un'azione sporadica, perché oggi tutto cambia più rapidamente. Il mondo sta appena iniziando a scoprire e sperimentare le possibilità tecnologiche. Inoltre, si sta passando da un ambiente analogico a un ambiente digitale. Le nuove tecnologie offrono una moltitudine di mezzi digitali con i quali è possibile evitare in futuro i problemi tipici del change management, come ad esempio i problemi di comunicazione.

Nel quadro della trasformazione digitale, non è più sufficiente affrontare una semplice riprogettazione o meccanizzazione dei processi (Capgemini-Consulting, 2017). Un prerequisito essenziale richiesto è anche l'inclusione della cultura aziendale e la proiezione di quale sia la cultura ottimale per l'azienda nel mondo digitale. È inoltre emerso che, di regola, durante la fase di trasformazione manca la comunicazione con i collaboratori e che, in relazione a ciò, la gestione delle paure dei collaboratori è inadeguata. A tale proposito, il modello in 8 fasi (Kotter, 1996) si è dimostrato molto efficace anche nella gestione del change management agile. Di seguito sono riportate le 8 fasi:

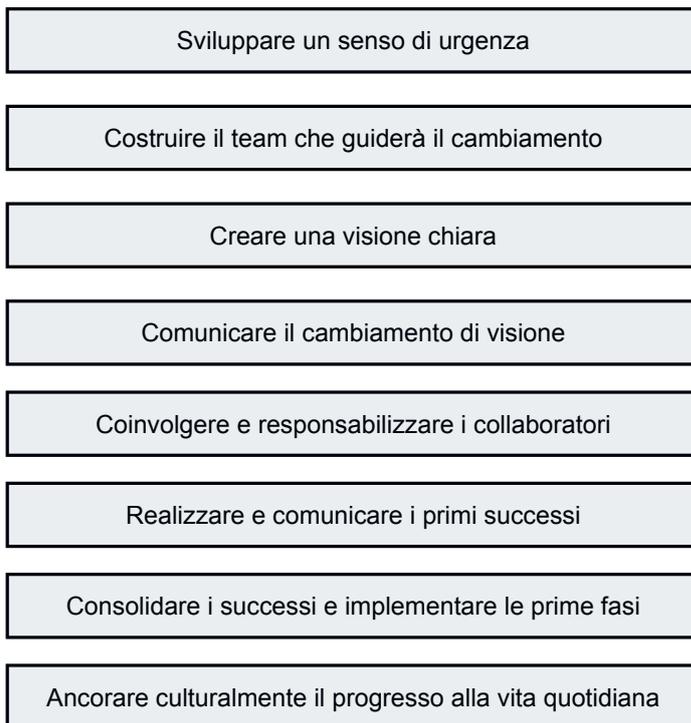


Figura 25: Modello delle 8 fasi (Kotter, 1996)

È importante osservare che le 8 fasi dovrebbero essere integrate nel processo di cambiamento sin dalla fase iniziale, per meglio conciliare i requisiti di un ambiente dinamico in costante evoluzione.

Nella nostra era digitale, le iniziative di cambiamento ampie e lente appartengono ormai al passato. Come accade per la collaborazione nei processi fondamentali di un'azienda, anche il change management deve diventare più agile e scontato.

La gestione agile del cambiamento implica che le fasi di cambiamento definite vengano attuate in tempi brevi, focalizzandosi sul risultato desiderato. Le modifiche devono essere comunicate ai collaboratori su base continuativa. Il cambiamento viene suddiviso in piccoli passi per renderlo tangibile ai collaboratori e dare loro la possibilità di avere sicurezza e fiducia. La gestione continua degli stakeholder è fondamentale. I momenti di review alla fine di ogni fase di cambiamento aiutano a evidenziare costantemente i benefici del cambiamento e offrono anche la possibilità di apportare rapidamente delle modifiche. Inoltre, è possibile identificare la fase di cambiamento logica e sostenibile per l'azienda e pianificarla di conseguenza.

14 Bibliografia

Capgemini-Consulting. (2017).

ExpertFocus. (2019/12). EXPERTsuisse.

Gartner. (27. Juli 2020). Gartner Research. Von

<https://www.gartner.com/en/documents/3988021/magic-quadrant-for-robotic-process-automation>
abgerufen

Kotter, P. (1996). 8-Stufen-Modell.

Martin, A. (2008). Digital Literacy for the Third Age: Sustaining Identity in an Uncertain World.

Microsoft Business Solutions. (2005). ERD Microsoft NAV.

O'Brien, J. A. (2010). Management Information Systems.

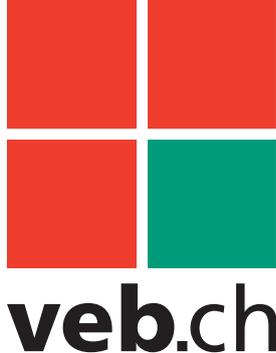
Ruoss, S. (2015). Von <https://svenruoss.ch> abgerufen

[tableau.com](https://www.tableau.com/de-de/learn/articles/was-ist-big-data). (2021). Von <https://www.tableau.com/de-de/learn/articles/was-ist-big-data> abgerufen

Zäch, M. (2021). Projektmanagement in ERP-Systemen. St. Gallen, Schweiz.

15 Sommario delle figure

Figura 1: Architettura dei Cubetools	10
Figura 2: Esempio di cubo OLAP	10
Figura 3: Sistemi a monte e data warehouse.....	13
Figura 4: Cubetool – Data warehouse	18
Figura 5: Esempio della struttura di un data warehouse	19
Figura 6: Processo di un sistema ERP	20
Figura 7: Organizzazione di un ERP	21
Figura 8: Diagramma ERD (Microsoft Business Solutions, 2005).....	26
Figura 9: Livelli di elaborazione software (Zäch, 2021).....	27
Figura 10: Struttura della contabilità dei debitori (Zäch, 2021)	28
Figura 11: Livello decisionale e informazione (O’Brien, 2010)	30
Figura 12: BI back / Frontend	32
Figura 13: Natura di un’intelligenza artificiale	50
Figura 14: Differenza Supervised Learning / Unsupervised Learning	52
Figura 15: Fasi fino al testing	54
Figura 16: Decision Tree	57
Figura 17: Confusion Matrix.....	58
Figura 18: Illustrazione di un Explainable AI per un caso	62
Figura 19: Trasformazione digitale come parte della digitalizzazione	64
Figura 20: Levels of Digital Literacy (Martin, 2008)	65
Figura 21: Obiettivi della trasformazione digitale (Ruoss, 2015)	66
Figura 22: Digital Finance Transformation	68
Figura 23: Mondo VUCA (Zäch, 2021)	72
Figura 24: Elementi dei valori agili (Zäch, 2021).....	73
Figura 25: Modello delle 8 fasi (Kotter, 1996)	74



veb.ch – Conoscenze professionali ed esperienza di prima mano

veb.ch è il vostro primo punto di contatto per tutte le do-mande in materia di tenuta dei conti, contabilità e con-trolling.

La nostra associazione conta oltre 9'000 membri in tutta la Svizze-ra. Tra questi vi sono, in particolare, gli esperti in finanza e control-ling, nonché i possessori dell'attestato professionale in finanza e contabilità. In Svizzera, sono tra gli esperti riconosciuti a livello fe-derale per tutte le questioni relative alla contabilità, a tutti i livelli aziendali. Possono aderire alla nostra associazione anche spe-cialisti in contabilità di comprovata esperienza e con altre forma-zioni. Tutte le persone collegate a noi sono invitate a diventa-re membri passivi di veb.ch.

La formazione ci sta particolarmente a cuore. Per questo motivo, come associazione, garantiamo costantemente che la formazione e gli esami siano adeguati alle esi-genze attuali. Insieme alla Società impiegati di commer-cio organizziamo, ogni anno, gli esami di Specialista in finanza e contabilità.

Alla nostra offerta di formazione continua:

 www.veb.ch / **Seminari e percorsi formativi**

veb.ch

Schweizerischer Verband der dipl. Experten in Rechnungslegung und Controlling und der Inhaber des eidg. Fachausweises in Finanz- und Rechnungswesen. Seit 1936

Talacker 34

8001 Zürich

Telefono 043 336 50 30

info@veb.ch

Leggi il nostro blog a

blog.veb.ch

Visita il nostro sito digitale a
www.veb.digital

Seguici su:



EDU  **UA**

Schweizerisches Qualitätszertifikat für Weiterbildungsinstitutionen
Certificat suisse de qualité pour les institutions de formation continue
Certificato svizzero di qualità per istituzioni di formazione continua