

# Sistemi Informativi Aziendali

Corso tenuto dalla Professoressa Barbara Oliboni

Università di Verona

*Alessio Gjergji*

# Indice

<b>1</b>	<b>Introduzione ai sistemi informativi aziendali</b>	<b>4</b>
1.1	La prospettiva informativa della teoria dell'organizzazione . . . . .	4
1.1.1	Sistema informativo . . . . .	4
1.1.2	Sistema informatico . . . . .	5
1.1.3	Sistema organizzativo . . . . .	5
1.2	Sistema informativo aziendale . . . . .	7
1.2.1	La piramide dell'informazione . . . . .	8
1.2.2	Informazione come risorsa organizzativa . . . . .	8
1.2.3	Quantità di informazione . . . . .	9
1.2.4	Flessibilità nell'organizzazione . . . . .	9
1.2.5	Organizzazione come sistema aperto . . . . .	9
1.2.6	Modello gerarchico ( <i>piramide di Anthony</i> ) . . . . .	10
1.2.7	Cooperazione . . . . .	10
1.2.8	Sistemi informativi verticali o orizzontali . . . . .	10
<b>2</b>	<b>Classificazione dei Sistemi Informativi Aziendali</b>	<b>11</b>
2.1	SIA lungo la piramide aziendale . . . . .	11
2.1.1	Classificazione lungo la piramide aziendale . . . . .	12
2.2	SIA rispetto alle aree funzionali . . . . .	13
2.3	Portafoglio applicativo . . . . .	13
2.3.1	Portafoglio istituzionale . . . . .	14
2.3.2	Portafoglio operativo . . . . .	14
2.3.3	Portafoglio direzionale . . . . .	14
2.4	Aree funzionali . . . . .	15
2.4.1	L'informazione del portafoglio operativo nelle imprese manifatturiere . . . . .	15
<b>3</b>	<b>Sistemi ERP e CRM</b>	<b>16</b>
3.1	Introduzione ai Sistemi ERP . . . . .	16
3.1.1	Caratteristiche Principali . . . . .	17
3.1.2	Sistema gestionale classico . . . . .	17
3.1.3	Sistema ERP . . . . .	17
3.1.4	Moduli Principali . . . . .	17
3.1.5	Architettura di un Sistema ERP . . . . .	18

3.2	Sistemi CRM . . . . .	18
3.2.1	Sistemi CRM e catene di servizio . . . . .	18
3.2.2	CRM e aziende di utility . . . . .	19
3.3	Sistemi di business intelligence . . . . .	19
3.3.1	Il portafoglio operativo delle società di servizi . . . . .	20
<b>4</b>	<b>Ingegneria dei processi gestionali</b>	<b>21</b>
4.1	Processi . . . . .	21
4.2	Business Process . . . . .	21
4.3	Catena del valore di Porter . . . . .	22
4.3.1	Strategia Buy-Side . . . . .	23
4.3.2	Strategia Inside . . . . .	23
4.3.3	Strategia Sell-Side . . . . .	23
4.4	Classificazione dei processi . . . . .	24
4.5	Scomposizione dei processi . . . . .	24
4.6	Griglia metodologica . . . . .	24
4.6.1	Variabili organizzative . . . . .	25
4.7	Fasi della metodologia di analisi . . . . .	26
<b>5</b>	<b>BPMN</b>	<b>28</b>
5.1	Business process . . . . .	28
5.2	Il ciclo di vita di un processo . . . . .	29
5.3	Concetti di base . . . . .	29
5.3.1	Elementi di base di BPMN . . . . .	29
5.3.2	Collegare attività . . . . .	30
5.4	Comportamento delle attività . . . . .	30
5.5	Gateway . . . . .	30
5.6	Rappresentare l'organizzazione di un'azienda . . . . .	31
5.6.1	Pool . . . . .	31
5.6.2	Lane . . . . .	32
5.6.3	Flusso di messaggi . . . . .	32
5.7	Artefatti . . . . .	32
5.8	Costrutti avanzati . . . . .	33
5.8.1	Specializzare tipi di task . . . . .	33
5.8.2	Attività . . . . .	34
5.8.3	Attività multi-istanza . . . . .	34
5.8.4	Eventi . . . . .	34
<b>6</b>	<b>Data Warehousing</b>	<b>36</b>
6.1	Piattaforma per la Business Intelligence . . . . .	36
6.2	Introduzioni al Data Warehousing . . . . .	37
6.2.1	OLTP e OLAP . . . . .	37
6.2.2	Il Data Warehouse . . . . .	38
6.2.3	Le interrogazioni . . . . .	38
6.3	Architetture . . . . .	38

---

6.3.1	Architettura a un livello . . . . .	39
6.3.2	Architettura a due livelli . . . . .	39
6.3.3	Architettura a tre livelli . . . . .	40
6.3.4	ETL - Extraction, Transformation and Loading . . . . .	40
6.4	Modello multidimensionale . . . . .	42
6.5	Tecniche di analisi dei dati . . . . .	42
6.6	Progettazione logica del data warehouse . . . . .	43
6.7	Il ciclo di vita del data warehouse . . . . .	43
6.7.1	Il ciclo di sviluppo . . . . .	44
6.7.2	Progettazione del data mart . . . . .	44
6.7.3	Analisi e riconciliazione delle sorgenti eterogenee . . . . .	45
6.7.4	Relazioni tra concetti comuni . . . . .	46
6.7.5	Fasi dell'integrazione . . . . .	46
6.7.6	Analisi dei Requisiti . . . . .	47
6.8	Progettazione Concettuale . . . . .	48
6.8.1	DFM - Dimensional Fact Model . . . . .	48
6.8.2	Costrutti di Base . . . . .	49
6.8.3	Relazioni tra gli Elementi . . . . .	50
6.8.4	Naming Conventions . . . . .	50
6.8.5	Eventi e Aggregazioni . . . . .	50
6.8.6	Additività delle Misure . . . . .	51
6.8.7	Passi di Progettazione di uno Schema Concettuale . . . . .	51
6.8.8	Problema della Sparsità . . . . .	51
6.9	Progettazione Logica . . . . .	52
6.9.1	MOLAP . . . . .	52
6.9.2	ROLAP: Schema a Stella . . . . .	52
6.9.3	ROLAP: Schema a Fiocco di Neve ( <b>Snowflake Schema</b> ) . . . . .	52
6.9.4	Le Viste . . . . .	53

# Capitolo 1

## Introduzione ai sistemi informativi aziendali

### 1.1 La prospettiva informativa della teoria dell'organizzazione

Ci occupiamo di capire perché è importante l'informazione in ambito organizzativo, e per capirlo dobbiamo comprendere quali problematiche possono sorgere all'interno di un'organizzazione.

Quando le persone lavorano insieme, si crea la necessità di organizzarle e coordinarle. Questo è il compito della **teoria dell'organizzazione**, che studia come le persone si organizzano per raggiungere obiettivi comuni.

#### 1.1.1 Sistema informativo

##### Sistema informativo

Il **sistema informativo** è la componente (*sottosistema*) di un'organizzazione che gestisce le informazioni di interesse.

Un'azienda è un insieme di persone e tecnologie che operano in sinergia per raggiungere obiettivi comuni. All'interno di un'azienda, l'**informazione** rappresenta una risorsa fondamentale, in quanto da essa è possibile estrapolare conoscenza utile per prendere decisioni strategiche.

Ad esempio, se in un determinato periodo dell'anno si verifica un calo delle vendite, grazie all'utilizzo di **data warehouse** è possibile analizzare i dati storici per individuare le cause del calo e attuare le opportune azioni correttive.

Le informazioni di interesse sono quelle prodotte e utilizzate durante l'esecuzione dei processi aziendali. I processi rappresentano il modo in cui vengono svolte le attività finalizzate al raggiungimento di specifici obiettivi. I dati e i processi rappresentano le due facce della stessa medaglia.

Il sistema informatico è un sottoinsieme del sistema informativo, che a sua volta è un sottoinsieme del sistema organizzativo e che a sua volta è un sottoinsieme del sistema azienda.

### 1.1.2 Sistema informatico

Quando è necessario organizzare la risorsa **informazione**, si crea un sistema informativo. L'automatizzazione del sistema informativo porta alla creazione di un sistema informatico.

### 1.1.3 Sistema organizzativo

#### Sistema Organizzativo

Il sistema organizzativo è l'insieme di risorse e regole per lo svolgimento coordinato delle attività (*processi*) al fine del perseguimento degli scopi.

Le risorse di un'azienda sono di quattro tipi:

- **Risorse umane:** persone che lavorano all'interno dell'azienda.
- **Risorse finanziarie:** denaro necessario per finanziare le attività aziendali.
- **Risorse materiali:** beni tangibili che l'azienda possiede.
- **Risorse informative:** dati e informazioni utili per prendere decisioni.

### Organizzazione

In economia aziendale il termine **organizzazione** ha due significati:

- Dato l'insieme di persone che con il loro lavoro partecipano allo svolgimento dell'attività aziendale, l'organizzazione è il processo attraverso il quale tale insieme di persone viene strutturato secondo i principi di divisione del lavoro e coordinamento. Grazie all'organizzazione tale insieme acquisisce una struttura e diventa sistema.
- Il risultato del processo di divisione del lavoro e coordinamento. In questo senso l'organizzazione è sinonimo di **azienda**.

Ai fini dello studio della sua organizzazione, l'azienda può essere considerata un sistema socio-tecnico costituito da:

- **Persone:** le risorse umane che lavorano all'interno dell'azienda e che portano il **know-how** necessario per svolgere le attività.
- **Tecnologie:** mezzi tecnici e strumenti che supportano le attività svolte all'interno dell'azienda.

La conoscenza si struttura in tre livelli:

- **Sapere:** conoscenza codificata, attinente a discipline per le quali esistono comunità di studiosi e di esperti (*ciò che si impara*).
- **Saper fare:** conoscenza operativa e procedurale, abilità pratiche, esperienza professionale specifica (*applicare i concetti*).
- **Saper essere:** conoscenza relativa a valori, atteggiamenti, comportamenti, motivazioni, personalità (*collaborazione*).

Le risorse aziendali sono importanti, poiché sono le persone che sanno fare le cose. Abbiamo quindi bisogno che le risorse umane utilizzino in maniera appropriata le tecnologie per raggiungere l'obiettivo posto.

### Organizzazione aziendale

In azienda, ogni persona ricopre un ruolo specifico, con funzioni spesso legate al ruolo stesso. L'organizzazione si basa su fini, metodi e regole, fondamentali per mettere in atto ruoli e funzioni. Per raggiungere i propri obiettivi, l'azienda si avvale di processi, ovvero insiemi di attività coordinate per perseguire un fine comune.

I processi si suddividono in:

- **Operativi o produttivi:** si occupano della produzione di beni e servizi.
- **Di controllo e gestione:** si tratta della parte amministrativa.

Le risorse a disposizione dell'azienda includono:

- Risorse umane.
- Risorse materiali.
- Risorse informative.

L'azienda definisce i propri obiettivi in base alle opportunità offerte dall'ambiente esterno, tenendo conto anche dei vincoli imposti da quest'ultimo. L'azienda interagisce con l'ambiente, in particolare l'ambiente con cui interagisce è il mondo, formato quindi da tanti attori.

### Tecnologie informatiche

Le tecnologie informatiche sono gli strumenti, sistemi e tecniche per automatizzare il trattamento delle informazioni. Nelle aziende che gestiscono l'informazione come scopo, il sistema informativo è importante.

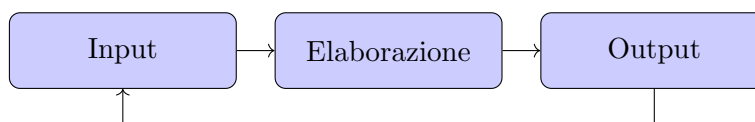
## 1.2 Sistema informativo aziendale

### Sistema informativo aziendale

È un sistema di elementi interconnessi che ci permettono di raccogliere, catalogare, ricercare, elaborare, memorizzare e distribuire i dati trasformandoli in informazioni utili per supportare le attività decisionali di controllo di un'azienda.

In azienda è importante prendere decisioni, poiché tutto ciò che viene fatto è frutto di decisioni. Le decisioni sono **strategiche**, poiché impattano sulle decisioni. Il fatto che sia a **supporto** permette di gestire la risorsa **informazione** per prendere le decisioni, e in ambito aziendale è fondamentale.

Il sistema informativo prende in input dei dati, li elabora, produce un output che può essere utilizzato come nuovo input.



In azienda tutto quello che viene svolto in un'organizzazione produce informazione, quando si arriva al punto di averne molta, quell'informazione se non si riesce a gestire diventa un problema, non più una risorsa. Con il supporto del sistema informativo si riesce a gestire l'informazione, e a trasformarla in risorsa. A seconda di ciò che si deve fare bisogna avere il sistema informativo adeguato.

### Data Warehouse

Il **Data Warehouse** è un sistema informativo che permette di aggregare l'informazione, prendendo come input l'output della base dati classica.

Il sistema informativo aziendale permette quindi di partire dai dati e trasformarli in informazioni, e quindi in conoscenza.



### 1.2.1 La piramide dell'informazione



La piramide dell'informazione rappresenta un modello che descrive il percorso attraverso cui i dati grezzi vengono trasformati in valore.

Questo processo avviene attraverso diverse fasi:

- **Dati:** Alla base della piramide troviamo i dati, raccolti tramite basi di dati. In questa fase, i dati sono grezzi e privi di significato (*Basi di dati*).
- **Informazioni:** I dati elaborati e strutturati diventano informazioni utili e comprensibili (*Basi di dati*).
- **Conoscenza:** Interpretando le informazioni, possiamo acquisire una comprensione profonda e significativa, utile per decisioni strategiche (*Data Warehouse*).
- **Saggezza:** All'apice, la saggezza consente di formulare previsioni e anticipare scenari futuri grazie a strumenti avanzati (*Cruscotti aziendali che permettono di fare previsioni*).

La forma a piramide evidenzia come, man mano che si sale, la quantità diminuisce ma il valore aumenta, richiedendo maggiore capacità elaborativa, grazie alla conoscenza aggregata. L'informazione proviene dai livelli sottostanti, perciò non si corre il rischio di perdere informazioni.

### 1.2.2 Informazione come risorsa organizzativa

L'informazione è la risorsa principale nell'attività di gestione e controllo. L'informazione viene prodotta anche dal processo produttivo ma anche dalle attività organizzative, di controllo e amministrative. Le aziende non manifatturiere l'informazione è la risorsa principale. L'informazione è importante perché permette di prendere decisioni, dato che le decisioni sono il

risultato di un processo di valutazione delle informazioni disponibili, anche dall'esterno. L'informazione non è facilmente divisibile o appropriabile, può essere soggetta ad obsolescenza, ma non diventa mai inutile.

L'informazione ha delle caratteristiche distintive che la rendono diversa dalle altre risorse, dato che non si distrugge dall'uso, ma si auto rigenera (*tipico del sistema informativo*).

La capacità auto generativa dell'informazione permette di instaurare circoli virtuosi di generazione di conoscenza e di arricchimento dell'informazione disponibile.

I circoli conoscitivi virtuosi si traducono in un incremento di prestazioni dei progetti gestionali.

### 1.2.3 Quantità di informazione

Sulla base della quantità di informazione, possiamo distinguere due possibili situazioni:

- **Overload informativo:** Si verifica quando l'informazione disponibile è eccessiva rispetto alle esigenze dell'organizzazione. Eccedendo la capacità di elaborazione, l'informazione in eccesso diventa inutile e dannosa.
- **Underload informativo:** Si verifica quando l'informazione disponibile è insufficiente rispetto alle esigenze dell'organizzazione. In questo caso, l'informazione inadeguata può portare a decisioni errate, semplificando quindi decisioni complesse.

Lo scenario ideale è un trade-off tra le due situazioni, in cui l'informazione disponibile è adeguata alle esigenze dell'organizzazione, permettendo di prendere decisioni corrette e tempestive.

### 1.2.4 Flessibilità nell'organizzazione

Organizzazione significa essere in grado di adattarsi ai cambiamenti dell'ambiente esterno, sono quindi sistemi aperti. Devono adattarsi anche a nuovi requisiti che esso impone (*ad esempio normative europee*). La flessibilità organizzativa deve essere garantita dalle tecnologie informatiche a supporto delle attività gestionali.

Le informazioni che arrivano dall'esterno devono essere date in input al sistema informativo, che le trasforma in informazioni utili per prendere decisioni.

### 1.2.5 Organizzazione come sistema aperto

Gli stakeholder sono coloro che hanno interesse nell'azienda, e sono coloro che possono influenzare le decisioni dell'azienda. Gli stakeholder si interfacciano con l'organizzazione e l'organizzazione ha il proprio sistema informativo per gestire le informazioni. L'eccezione che arriva dall'ambiente non per forza è qualcosa di negativo, ma può essere anche qualcosa di positivo.

L'organizzazione come sistema aperto e tutto ciò che influisce o arriva dall'esterno si chiama **incertezza ambientale**. L'incertezza ambientale determina i requisiti di capacità elaborativa delle organizzazioni e l'adeguatezza della struttura del sistema informativo.

L'**incertezza** genera delle **eccezioni**, che richiedono delle **decisioni**.

### 1.2.6 Modello gerarchico (*piramide di Anthony*)

Se rappresento l'organizzazione con un modello gerarchico posso pensare di avere livello operativo, direzionale e strategico. A seconda del tipo di input che arriva dall'esterno dovrò prendere decisioni diverse a livello strategico, livello direzionale e livello operativo.

Questi livelli possono essere mappati in parte sulla piramide dell'informazione. A livello operativo ho le informazioni transazionali, a livello direzionale ho le informazioni di supporto alle decisioni, a livello strategico ho le informazioni di supporto alle decisioni strategiche. Vi è una corrispondenza anche con i tipi di processi aziendali.

### 1.2.7 Cooperazione

La cooperazione fra più individui determina la suddivisione del compito elaborativo e la suddivisione delle informazioni. ognuno necessita di un sotto-insieme di informazioni adatte allo scopo del sotto-compito di cui è responsabile.

La capacità elaborativa dell'organizzazione dipende dall'organizzazione delle persone a livello decisionale.

### 1.2.8 Sistemi informativi verticali o orizzontali

Tutto ciò si riflette nei sistemi informativi nel modo in cui l'informazione fluisce all'interno delle gerarchie decisionali dell'organizzazione.

- **Sistemi informativi verticali:** Sono sistemi informativi che supportano le decisioni a livello **gerarchico**, e quindi supportano le decisioni a livello operativo, direzionale e strategico.
- **Sistemi informativi orizzontali:** Sono sistemi informativi che supportano le decisioni a livello di **processo**, e quindi supportano le decisioni a livello operativo, direzionale e strategico.

Al giorno d'oggi possiamo pensare a decisioni che hanno necessità di risalire e poi scendere o di rimanere a un certo livello. I sistemi informativi devono essere progettato per supportare il flusso di informazione in entrambe le direzioni.

Nell'azienda non abbiamo solo **incertezza ambientale**, ma anche **incertezza comportamentale**, che riguarda il comportamento delle persone all'interno dell'organizzazione. Se possiamo essere confidenti sul comportamento di un'applicazione, ma rispetto al comportamento delle persone è più difficile.

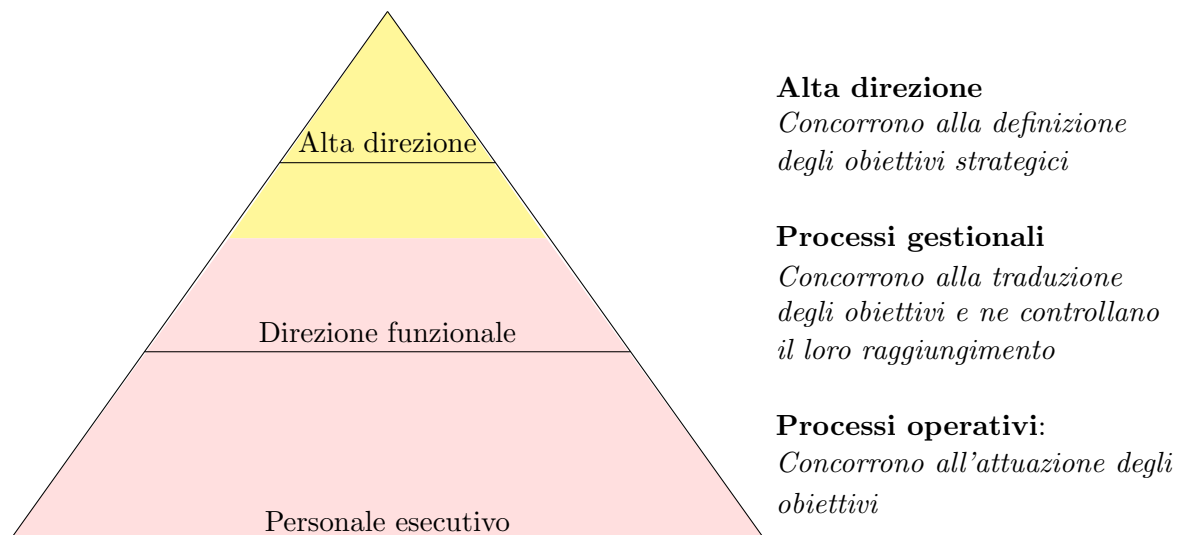
## Capitolo 2

# Classificazione dei Sistemi Informativi Aziendali

I sistemi informativi aziendali possono essere classificati rispetto a due dimensioni:

1. Sistemi informativi aziendali disposti lungo la piramide aziendale (*piramide di Anthony*): la loro definizione e le funzioni attribuite dipendono dal livello al quale i Sistemi Informativi sono collocati.
2. Sistemi informativi aziendali disposti lungo le aree gestionali dell'impresa: la loro definizione e le loro funzioni attribuite dipendono dall'area gestionale su cui i Sistemi Informativi sono collocati.

### 2.1 SIA lungo la piramide aziendale



Il fatto che i tre livelli facciano cose diverse, implica che gestiscano informazioni diverse, implementando quindi processi diversi. Su un certo livello della piramide possiamo ragionare rispetto a:

- Ruolo del personale.
- Decisioni che intraprende.
- Processi che esegue.
- Informazioni necessarie.

Le decisioni strategiche prese nella parte direzionale vengono tradotte al livello sottostante in veri e propri processi gestionali, che vengono poi tradotti in processi operativi.

### Esempio

- **Alta direzione:** decisione sulla produzione di un pandoro senza glutine.
- **Direzione funzionale:** decisione sulla quantità di pandori senza glutine da produrre, entrata sul mercato, ecc.
- **Personale esecutivo:** Attuazione della produzione del pandoro senza glutine.

Oltre alle classificazioni relative a sistemi informativi verticali e orizzontali, possiamo classificare i sistemi informativi aziendali in:

- **Sistemi informativi operazionali:** si occupano di gestire le informazioni quotidiane dell'azienda, si trovano nella metà inferiore della piramide di Anthony. Non riguarda solo la parte esecutiva, ma anche una parte di gestione.
- **Sistemi informativi informazionali:** si occupano di gestire le informazioni, aiutando ad eseguire le attività di business intelligence, si trovano metà superiore della piramide di Anthony.

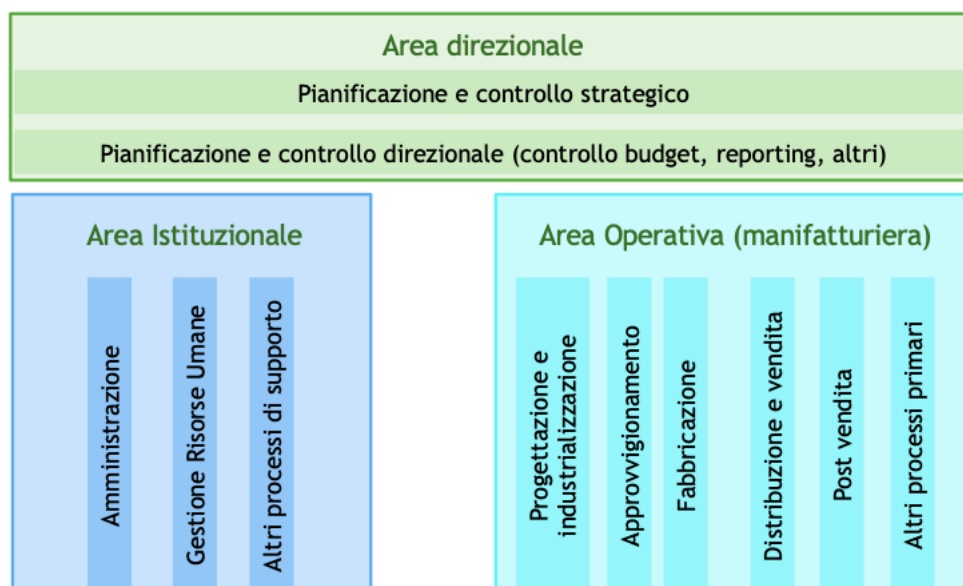
#### 2.1.1 Classificazione lungo la piramide aziendale

Partendo dalla base abbiamo:

- **Transaction Processing System (TPS):** sono le basi di dati.
- **Management Information System (MIS):** sono i sistemi informativi gestionali (*costruzione di report periodici*).
- **Decision Support System (DSS):** sono i sistemi per il supporto alle decisioni. Utilizzano i dati prodotti dall'attività quotidiana e li vanno a storicizzare e fare delle analisi.
- **Executive Information System (EIS):** sono i sistemi informativi direzionali (*cru-scotti aziendali che utilizzano anche dati su fonti esterne*).

Il punto in cui i sistemi informativi aziendali si trovano ci dice anche chi lo usa e che competenze sono necessarie per gestirli.

## 2.2 SIA rispetto alle aree funzionali



Mentre la piramide è più generale, il diagramma delle aree funzionali varia rispetto alla tipologia di azienda, dato che le funzionalità possono variare.

La suddivisione viene fatta su:

- Area operativa
- Area istituzionale
- Area direzionale

## 2.3 Portafoglio applicativo

Il **portafoglio applicativo** di un'azienda sono l'insieme di applicazioni da usare per svolgere le attività aziendali. Esso può essere suddiviso in:

- **Portafoglio direzionale:** applicazioni per la direzione per la pianificazione strategica e di pianificazione e di controllo delle risorse aziendali.
- **Portafoglio istituzionale:** applicazioni per la parte istituzionale, alla gestione delle risorse umane e contabilità.
- **Portafoglio operativo:** applicazioni per la parte operativa per i processi principali dell'azienda.

Il portafoglio operativo è **verticale**, tipico di ogni singolo settore, mentre i portafogli direzionale e istituzionale sono **orizzontali** indipendenti alle specifiche caratteristiche del settore.

### 2.3.1 Portafoglio istituzionale

Il portafoglio istituzionale è stato il primo tipo di attività che è e stata informatizzata, dato che ha grandi volumi di dati e necessità di controllo. D'altra parte è un processo ripetitivo e quindi facilmente automatizzabile (*es. contabilità, buste paga*).

### 2.3.2 Portafoglio operativo

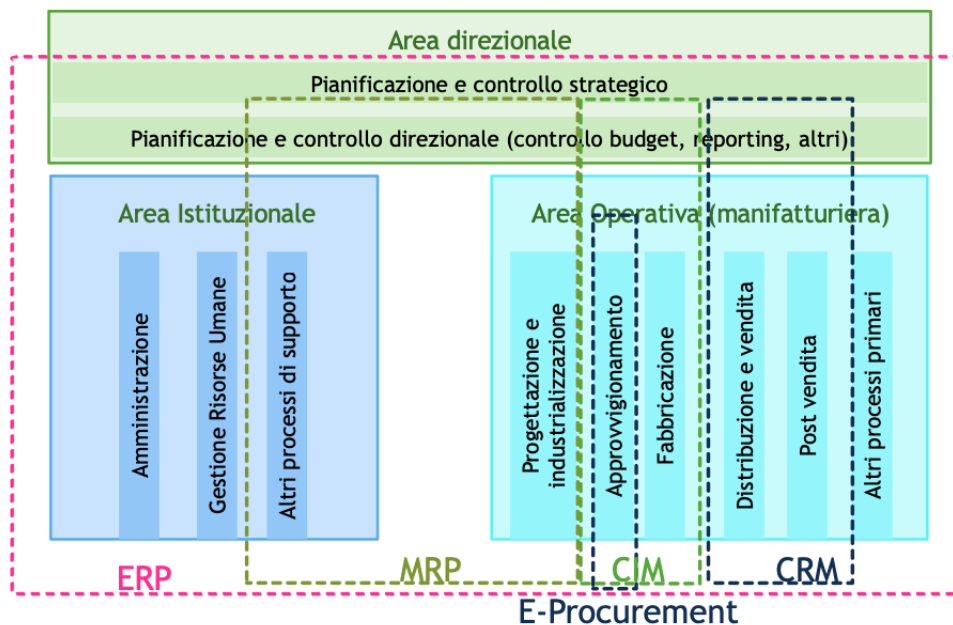
Il portafoglio operativo varia a seconda del settore. Comprende le applicazioni informatiche a supporto dei processi primari della **catena del valore di Porter**, ovvero l'azienda vista come un insieme di attività che creano valore. Ad altissimo livello, per un'azienda manifatturiera, potrò pensare che avrò come attività:

- L'acquisizione delle materie prime.
- La loro trasformazione.
- Marketing e vendita.
- Distribuzione.
- Assistenza post-vendita.

### 2.3.3 Portafoglio direzionale

Il portafoglio direzionale è composto da applicazioni che supportano la direzione dell'azienda. Queste applicazioni sono utilizzate per la pianificazione strategica e di controllo delle risorse aziendali.

## 2.4 Aree funzionali



### 2.4.1 L'informazione del portafoglio operativo nelle imprese manifatturiere

- **Progetti custom (1960-1990):** soluzioni ad hoc spesso in COBOL, sistemi centralizzati.
- **Manufacturing Resource Planning (MRP) (1970-1995):** progetti integrati, database, spesso centralizzati.
- **Computer Integrated Manufacturing (CIM) (dal 1980):** integrazione di sistemi, database distribuiti, sistemi di controllo di processo.
- **Enterprise Resource Planning (ERP) (dal 1990):** utilizzo del modello client-server, database distribuiti, integrazione di processi aziendali.
- **Customer Relationship Management (CRM) (dal 1995):** integrazione di processi di marketing, vendita e assistenza.
- **E-Procurement (dal 1995):** integrazione di processi di acquisto.



## Capitolo 3

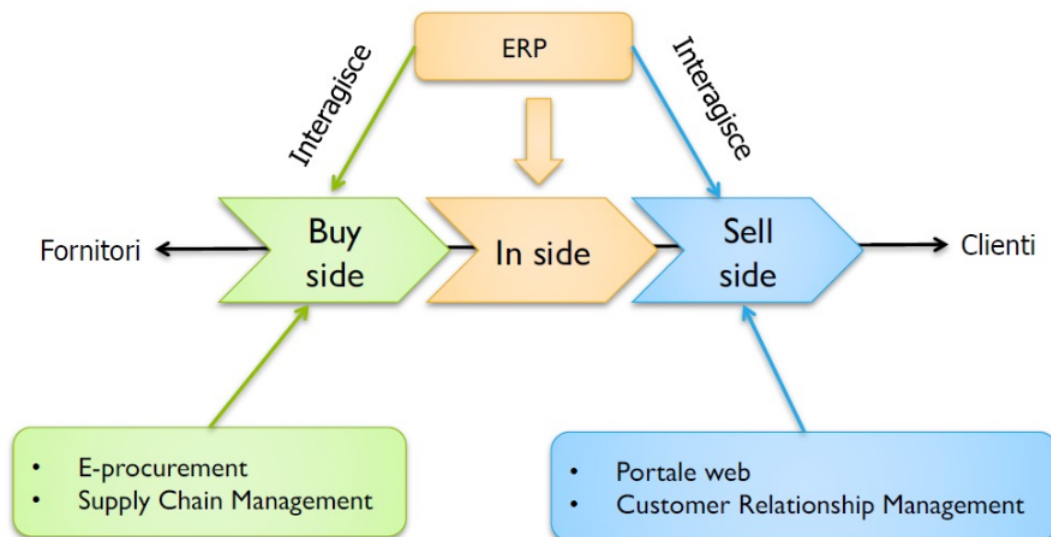
# Sistemi ERP e CRM

### 3.1 Introduzione ai Sistemi ERP

I sistemi ERP (*Enterprise Resource Planning*) rappresentano una suite di moduli applicativi che supportano l'intera gamma dei processi aziendali.

Oggi possiamo pensare ai sistemi ERP come sistemi a pacchetti che possono coprire più aree funzionali dell'azienda.

Il sistema ERP è un sistema informativo, operativo integrato che sfrutta un'unica base di dati e sfrutta **un'unica base di dati** e permette di avere procedure interagenti e cooperanti tra i vari moduli.



Dove è necessario gestire l'informazione in maniera integrata, il sistema ERP è la soluzione ideale.

### 3.1.1 Caratteristiche Principali

- **Sistema informativo integrato:** gli ERP combinano dati e processi aziendali in un'unica piattaforma condivisa.
- **Base di dati unica:** garantisce la consistenza e l'integrità delle informazioni aziendali, permettendo una gestione dei processi aziendali integrata.
- **Sistemi modulati:** i moduli applicativi possono essere integrati in base alle esigenze aziendali. Sono quindi altamente configurabili fino ad un certo punto, il core del sistema è sempre lo stesso, è quindi configurabile a seconda dei pacchetti che si acquistano.
- **Periodo di avviamento iniziale:** l'introduzione di un sistema ERP richiede un periodo di implementazione e formazione del personale, ridefinendo i processi aziendali.

### 3.1.2 Sistema gestionale classico

Isole informatiche autonome e specializzate (*Legacy Systems*). le varie aree funzionali interagiscono, ma le basi di dati sono molteplici e non integrate tra loro.

Il vantaggio di tale soluzione è che è possibile gestire in maniera automatizzata i processi aziendali. Siccome le costruzioni sono ad hoc, i sistemi non sono pensati per essere modulari e le basi di dati non sono integrate tra loro, si ha una serie di problemi:

- Duplicazione di dati.
- Ridondanza di informazioni.
- Incoerenza dei dati.
- Difficoltà di integrazione.

### 3.1.3 Sistema ERP

Nel momento in cui riesco ad avere un sistema in cui riesco ad integrare tutte le informazioni in un'unica base di dati, riesco a gestire in maniera integrata i processi aziendali.

### 3.1.4 Moduli Principali

Gli ERP sono costituiti da moduli indipendenti che coprono aree aziendali critiche:

- **Amministrazione e Finanza:** gestione della contabilità generale, controllo di gestione, bilanci e reportistica.
- **Logistica e Supply Chain:** gestione dei materiali, magazzini, inventario, ordini e distribuzione.
- **Vendite e Marketing:** gestione delle relazioni con i clienti, offerte, contratti e analisi di mercato.
- **Produzione:** pianificazione e controllo dei processi produttivi, gestione delle risorse e ottimizzazione della capacità produttiva.

- **Risorse Umane:** gestione del personale, formazione, pianificazione dei turni e calcolo delle retribuzioni.

### 3.1.5 Architettura di un Sistema ERP

Il sistema ERP è composto da:

- **Presentation Server:** interfaccia utente per l'accesso ai moduli applicativi.
- **Application Server:** logica di business e processi aziendali.
- **Database Server:** base di dati centralizzata e condivisa.

## 3.2 Sistemi CRM

I sistemi CRM (*Customer Relationship Management*) supportano la gestione delle relazioni con i clienti, concentrandosi sulla fidelizzazione e sulla personalizzazione dei servizi offerti. L'obiettivo è raccogliere e analizzare i dati dei clienti per migliorare la customer experience e la soddisfazione del cliente.

I sistemi CRM sfruttano i datawarehouse per l'analisi dei dati e gli strumenti di Business Intelligence per l'elaborazione delle informazioni.

### Processo CRM

Il processo CRM è un processo integrato e strutturato di gestione della relazione con la clientela, che mira a costruire relazioni personalizzate di lungo termine, migliorare la soddisfazione del cliente e il valore generato per l'azienda, e integrare tutti i punti di contatto tra azienda e cliente.

Il cliente è elemento centrale della strategia commerciale.

### 3.2.1 Sistemi CRM e catene di servizio

Le strategie CRM sono infattibili senza sistemi CRM e in generale senza un sistema integrato:

- I sistemi CRM formano il front end delle catene di servizio.
- I sistemi ERP formano il back end delle catene di servizio, supportano l'esecuzione del servizio.
- I sistemi SCM supportano la gestione della catena di fornitura.

L'era CRM nasce negli anni '90, quando nasce la rete internet.

Il CRM è un sistema necessario per aziende che necessitano continuità e frequenza di relazioni con i clienti, considerando anche il numero di clienti che si hanno. Per sviluppare questo sistema di relazioni, è necessario disporre di diversi canali di comunicazione con il cliente, per poterlo raggiungere in maniera efficace a seconda delle sue esigenze.

Nel momento in cui è necessario proporre un'offerta devo avere i dati **aggregati** dei clienti, per poter fare delle analisi e delle previsioni, ricavando quindi informazione utile.

### 3.2.2 CRM e aziende di utility

Le aziende di utility rappresentano un esempio significativo dell'importanza dei sistemi CRM, spinti da diversi fattori:

- Clientela molto numerosa.
- Rapporto continuativo nel tempo con i clienti.
- Necessità di sviluppare canali virtuali, come:
  - Web.
  - Call center.
- Crescente concorrenza nel settore.
- Esigenza di sviluppare e mantenere un'offerta personalizzata.

I principali vantaggi offerti dall'implementazione di un sistema CRM nelle aziende di utility includono:

- **Abbattimento dei costi di relazione:** I sistemi automatizzati e l'utilizzo di canali digitali riducono significativamente i costi operativi.
- **Incremento dei profitti:** La personalizzazione dell'offerta e l'efficienza nella gestione del cliente hanno un impatto diretto sui risultati economici.
- **Efficacia del CRM:** Dipende dal grado di personalizzazione offerto ai clienti e dai costi associati alla sua implementazione e gestione.

## 3.3 Sistemi di business intelligence

Nel momento in cui devo gestire tutte le informazioni, proponendo offerte personalizzate, gestire i canali, per riuscire a gestire questa grandissima mole di dati, è necessario avere dei sistemi di business intelligence che mi permettano di fare analisi.

### Business Intelligence

I sistemi di business intelligence si basano sui sistemi di supporto alle decisioni, ovvero ai datawarehouse.

Si tratta di applicazioni che estraggono informazioni dai sistemi di supporto alla produzione per elaborarle e fornire supporto alle decisioni. Si occupano di passare da dati a informazioni sintetiche (*i dati provengono da fonti di dati diverse ho bisogno di **integrare** queste informazioni*). Rispondono a richieste complesse: le operazioni decisionali sono non standard e fatte da personale non informatico, non hanno competenze informatiche. Hanno un focus maggiore

sulle interrogazioni rispetto agli aggiornamenti. L'architettura dei sistemi di BI comprende una serie di data warehouse e una serie di datamart.

### 3.3.1 Il portafoglio operativo delle società di servizi

I settori che offrono principalmente servizi, come il settore bancario, assicurativo, delle telecomunicazioni, delle utility, ecc., sono caratterizzati da una alta intensità informativa, quindi la catena del valore di Porter è differente rispetto a quella delle aziende manifatturiere.

- **Processo produttivo:** predisposizione delle condizioni produttive, mi metto in condizione di poter offrire il servizio, anche con la stipulazione di contratti con altri operatori e Back office: attività svolte in seguito a un ordine ma non in presenza del cliente come la valutazione del rischio nella polizza assicurativa.
- **Processo distributivo:** attività di front office in cui erogo il servizio di fronte al cliente come l'attività di sportello della banca e procacciamento della clientela in cui porto alla stipulazione dei contratti di servizio o all'ordine da parte dei clienti.

È importante la gestione della conoscenza (**knowledge management**)

- Elaborazione dell'informazione acquisita durante la produzione e l'erogazione del servizio.
- Trasformazione in conoscenza accessibile alle attività operative e direzionali.

Ma perché è importante? Nelle società di servizi rappresenta il vero capitale dell'azienda. La gestione della conoscenza ha l'obiettivo di trasformare l'informazione non strutturata in informazione comprensibile a qualunque esperto.

## Capitolo 4

# Ingegneria dei processi gestionali

### 4.1 Processi

#### Processo

Un **processo** è un insieme di attività **coordinate** con un **input** e un **output** per raggiungere un obiettivo. Il **cliente** di un processo è chi usufruisce del processo.

I processi rappresentano il modo di operare di un'azienda.

Talvolta è necessario re-ingegnerizzare un processo per renderlo più efficiente, efficace e flessibile può essere necessario per l'avvento di nuove tecnologie, cambio di personale, cambio di mercato o per la necessità di ridurre i costi. A differenza delle base di dati, nella progettazione dei processi, non è così immediato.

I processi possono essere classificati in:

- **Materiali:** flusso di materiali e attività (*es. produzione di pandori*).
- **Informativi:** flusso di informazioni (*es. assicurazione*).
- **Business process:** è un insieme coordinato di attività, finalizzato alla realizzazione di un ben definito risultato di interesse per l'organizzazione (*ingloba le due cose*).

### 4.2 Business Process

In ogni caso, che sia la produzione di un bene materiale, che di un servizio abbiamo comunque un flusso informativo, perché il processo utilizza, gestisce e produce informazione oltre a tutto il resto; perciò non distinguiamo processi strettamente materiali da processi strettamente informativi.

Il generale un processo aziendale è formato da attività, realizzate come processi materiali o processi informativi, che sono collegate fra loro nel tempo e nello spazio e svolte dalle risorse aziendali.

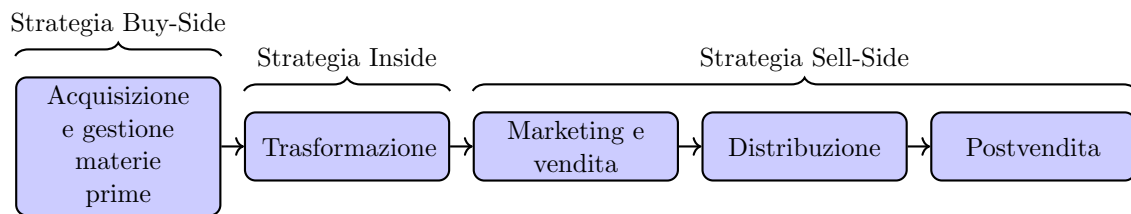
Come detto precedentemente, i clienti sono coloro chi usufruisce dell'output del processo, anche interno all'azienda.

$$\mathcal{BP} = (A, I, O, C)$$

- $A$ : insieme delle attività.
- $I$ : insieme degli input.
- $O$ : insieme degli output.
- $C$ : insieme dei clienti.

### 4.3 Catena del valore di Porter

La catena del valore di Porter è l'azienda vista secondo una prospettiva di processo, ovvero una successione di attività finalizzate a produrre valore per il cliente. Il valore è misurato dal prezzo che il cliente è disposto a pagare per il prodotto o servizio ricevuto.



Rispetto alla reingegnerizzazione dei processi, ho diverse possibili strategie di ottimizzazione che sono volte ad ottimizzare le sottoparti del macro processo.

- **Strategia Buy-Side:** ottimizzazione della catena di approvvigionamento, i sistemi di e-procurement sono un esempio.
- **Strategia Inside:** ottimizzazione della produzione, un esempio è l'ERP o le infrastrutture Internet. I sistemi ERP perché sono altamente configurabili e si mappano su tutti i processi aziendali.
- **Strategia Sell-Side:** ottimizzazione della distribuzione e vendita del prodotto, un esempio è il CRM.

Dato che la catena del valore di Porter ci permette di raccontare molte aree funzionali dell'azienda manifatturiera, possiamo pensare che ad ognuna di queste macro-attività corrispondano uno o più processi.

C'è un forte legame tra i processi aziendali e i sistemi informativi. La gestione dei processi aziendali è strettamente influenzata dall'uso dei sistemi informativi, e viceversa. I sistemi

informativi permettono di accelerare, ottimizzare e gestire in maniera più efficiente i processi, migliorandone la velocità e l'efficacia. Allo stesso tempo, l'esecuzione dei processi aziendali può influenzare i sistemi informativi, richiedendo l'integrazione di nuove funzionalità, l'accesso a dati aggiornati o l'elaborazione di informazioni aggiuntive.

In questo contesto, la catena del valore di Porter rappresenta un quadro di riferimento per comprendere come le attività aziendali siano organizzate. Ogni attività nella catena del valore copre una o più aree funzionali dell'azienda, evidenziando come i diversi processi aziendali interagiscano e contribuiscano alla creazione di valore.

#### 4.3.1 Strategia Buy-Side

È una strategia mirata all'interazione con il mondo dei fornitori. Ha lo scopo di trasformare i processi che hanno a che fare con l'approvvigionamento delle materie prime e con l'interazione con i fornitori, definizione di prezzi e condizioni, ordinazione, ricezione dei beni e servizi ordinati.

##### Benefici

- Riduzione dei costi di approvvigionamento.
- Miglioramento della qualità dei prodotti.
- Riduzione dei tempi di consegna.
- Maggiore flessibilità e reattività.

La trasformazione di questi processi si appoggia a sistemi di *e-procurement*, sistemi che supportano i mercati elettronici, infrastrutture internet.

#### 4.3.2 Strategia Inside

È una strategia mirata alla trasformazione dei processi interni all'azienda.

##### Benefici

- Riduzione di funzionamento
- Riduzione della durata dei processi
- Miglioramento della qualità

La trasformazione di questi processi si appoggia a sistemi ERP e infrastrutture internet.

#### 4.3.3 Strategia Sell-Side

Orientata ai processi di marketing, vendita, distribuzione del prodotto e servizio di post-vendita e assistenza al cliente (*customer care*). Insomma, tutti quei processi mirati all'interazione con il cliente.



**Benefici**

- Maggior valore del prodotto percepito dal cliente.
- Abbattimento dei costi di transizione.

La trasformazione di questi processi si appoggia a sistemi CRM e sistemi di business intelligence.

## 4.4 Classificazione dei processi

Oltre alla classificazione dei processi in materiali e informativi, possiamo classificarli in base a diversi criteri:

- **Processi intersettoriali:** processi generici che descrivono le pratiche di molteplici settori (*orizzontali*). Come ad esempio la gestione delle risorse umane, acquisto di materie prime.
- **Processi settoriali:** processi specifici di un settore (*verticali*). Come ad esempio un processo sanitario.
- **Processi aziendali:** processi di una specifica azienda o di una sua parte. Sono una specializzazione dei processi settoriali.
- **Processi normativi:** processi di riferimento, in particolare quelli considerati come processi ottimali per un determinato aspetto. Sono importanti per capire se il mio processo è migliorabile rispetto alla best practice.

## 4.5 Scomposizione dei processi

Il processo può essere scomposto a seconda del livello di dettaglio che si vuole raggiungere.

- **Macro-processo:** primo livello di segmentazione di un'azienda (*es. catena del valore di Porter*).
- **Processo:** illustrano il modo di operare di un'azienda
- **Fase:** Illustrano il modo in cui un processo è implementato.
- **Attività:** livello minimo di analisi normalmente adottato nello studio dei processi.
- **Operazione:** passi elementari per eseguire una attività.

Ciascun livello di processo può essere descritto da diagrammi di flusso, diagrammi gerarchici, tabelle di proprietà, ecc.

## 4.6 Griglia metodologica

Lo strumento che viene usato per la reingegnerizzazione dei processi è la **griglia metodologica**.

### Griglia metodologica

La **griglia metodologica** fornisce una visuale a griglia che permette di capire come migliorare il processo. Per ogni variabile ho una fase di analisi.

La metodologia comprende:

- Descrizione delle variabili di analisi (*variabili organizzative*).
- Descrizione delle fasi di analisi.
- Identificazione degli strumenti a supporto delle attività di analisi come i diagrammi di flusso, organigrammi, ecc.

La trasformazione di processi per avere successo deve integrare **innovazione tecnologica** e **innovazione organizzativa**.

Le variabili organizzative, insieme con la tecnologia informatica, sono le leve attraverso cui sono trasformati i processi.

Le variabili organizzative possono essere modificate per ottimizzare il processo.

#### 4.6.1 Variabili organizzative

Le variabili che possiamo considerare per l'ottimizzazione del processo sono:

- **Flusso delle attività:** sequenza delle attività attraverso cui è svolto il processo, determina la durata del processo. Posso ottimizzare se più attività sono fatte in parallelo, è importante risparmiare tempo in quanto il tempo è denaro. Il flusso delle attività determina:
  1. Durata del processo: a seconda del numero di attività previste.
  2. Livello di servizio: a seconda del grado di flessibilità, è importante essere flessibili. Per flessibilità si intende la capacità di prevedere che succedono delle cose.
  3. Qualità del prodotto: quando ottimizzo il processo voglio mantenere alta la qualità del prodotto.

La modellazione dei flussi può essere fatta attraverso:

1. Schemi di sequenza: indicano solo la sequenza che formano un processo. Più o meno ricchi in base al livello di dettaglio che voglio dare.
2. Schemi più ricchi: indicano anche gli attori coinvolti, gli eventi che determinano le attività, le informazioni utilizzate. Richiedono una raccolta di informazione che può essere lunga e costosa.

Il flusso può essere fisico, informatico o di controllo. Gli oggetti possono essere di natura fisica o informativa e deve essere indicato il profilo temporale (*informazioni permanenti o temporanee*).

- **Organizzazione del processo:** divisione operativa del lavoro nell'ambito del processo, come i processi sono mappati sui ruoli, in che modo l'organizzazione dell'azienda impatta sul processo.

L'organizzazione può impattare sul modo di operare dell'azienda. Esempio: se ho solo una persona che sa fare una cosa, in ogni processo in cui c'è quella attività devo usare quella risorsa.

Per descrivere le organizzazioni si possono usare gli **organigrammi**. Negli organigrammi viene illustrata **la gerarchia delle responsabilità e delle autorità di una organizzazione a vari livelli di dettaglio**. Esempio: un autista risponde a chi si occupa di trasporti e questi ultimi rispondono a chi si occupa di logistica. L'organigramma rappresenta **come viaggia l'informazione**. Inoltre, nelle **tabelle delle proprietà**, le proprietà indicano le caratteristiche fondamentali delle strutture. LRC offre una visione tabellare della responsabilità organizzativa che integra quella dell'organigramma. Tabella che per ciascun processo specifica il ruolo svolto da ogni struttura nel processo.

- **Competenza delle risorse umane che operano nel processo:** l'adeguatezza delle competenze è condizione fondamentale per la trasformazione del processo a seguito di una innovazione tecnica. Ad esempio, se viene introdotto un data warehouse, ci deve essere qualcuno con le competenze necessarie per utilizzarlo. La variabile risorse umane è essenziale per il successo del processo correttamente ingegnerizzato. L'innovazione tecnologica causa la necessità di avere a disposizione figure professionali specializzate. Queste figure possono essere create internamente riconvertendo il personale o acquisite sul mercato.
- **Sistema di misura e controllo delle prestazioni:** utilizzato per governare il processo e valutare gli attori aziendali. Questo sistema comprende il sistema di pianificazione e controllo, il sistema di promozione e di incentivazione e il sistema dei valori.

## 4.7 Fasi della metodologia di analisi

La metodologia di analisi prevede un approccio sia bottom-up che top-down per il miglioramento dei processi aziendali esistenti. Questi approcci si distinguono per le modalità operative:

### Approccio bottom-up:

- **Rilevazione della situazione esistente:**
  1. Identificazione dei macroprocessi.
  2. Dettaglio dei processi.
  3. Incrocio processi / unità organizzative.
  4. Valutazione del processo.
- **Confronto con altre imprese e diagnosi dei problemi:**

1. Confronto quantitativo e parametrizzazione: devo decidere a priori su cosa concentrarmi dato che non posso ottimizzare tutto.
2. Confronto qualitativo.

- **Ridefinizione dei processi.**

**Approccio top-down:** Disegno del processo con applicazione di criteri di ottimizzazione noti, limitata analisi della situazione esistente a quanto necessario per poter attuare il progetto.

### Griglia fasi-variabili

Variabili	Rilevazione	Confronto e diagnosi	Riprogettazione
Flussi di attività	Macroprocessi, processi, attività.	Benchmarking e analisi di integrazione.	Personalizzazione best practice e simulazione.
Struttura aziendale	Struttura organizzativa, stile di direzione.	Ruolo delle funzioni nell'integrazione.	Cambiamenti organizzativi e operativi.
Risorse umane	Inventario di competenze.	Diagnosi rispetto ai migliori concorrenti.	Programmi di formazione e assunzioni.
Misurazione delle prestazioni	Identificazione prestazioni strategiche.	Diagnosi e confronto con il mercato.	Nuovi obiettivi e indici di prestazione.

Tabella 4.1: Griglia fasi-variabili

L'informatica consente una maggiore integrazione inter-funzionale e inter-organizzativa per due ragioni fondamentali: aumenta la disponibilità di informazioni e supporta l'esecuzione dei compiti individuali di natura decisionale (*possibilità di condividere informazioni prima non disponibili*).

# Capitolo 5

## BPMN

### 5.1 Business process

Il vantaggio di BPMN è che permette di comprendere i processi aziendali in maniera semplice e intuitiva, anche per chi non è esperto di informatica.

#### BPMN

BPMN è una notazione per la modellazione grafica dei processi di business che definisce i simboli, il loro significato e le regole per le loro combinazioni, permettendo la descrizione standardizzata e comprensibile dei processi aziendali.

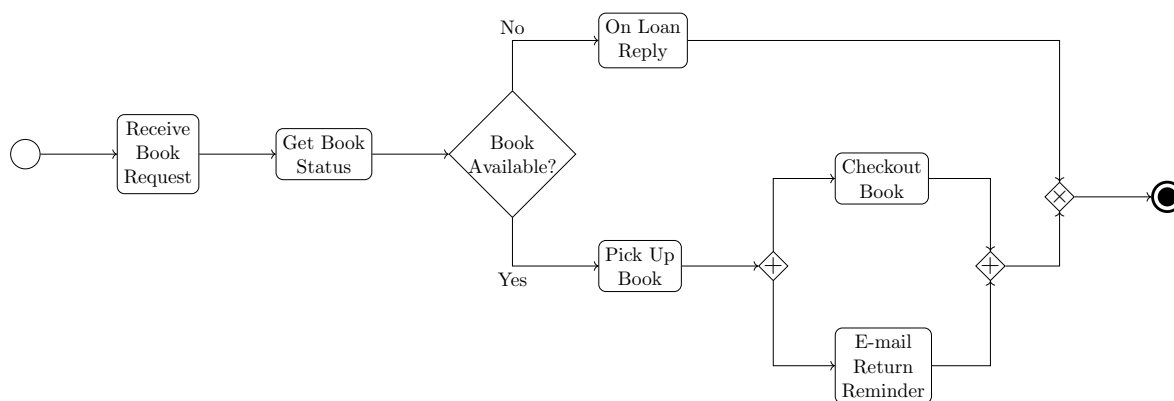


Figura 5.1.1: Diagramma BPMN del processo di prestito di un libro

Guardando tale rappresentazione risulta facile capire come funziona il processo di prestito di un libro. Si parte dallo start event, che rappresenta l'inizio del processo, e si passa attraverso le varie attività, rappresentate da task, fino ad arrivare ai vari gateway, che rappresentano delle decisioni. In questo caso, se il libro non è disponibile, si invia una risposta di prestito, altrimenti si procede con il ritiro del libro. Infine, si arriva ai parallel gateway, che rappresen-

tano attività che verranno eseguite in parallelo. Solo dopo aver completato queste attività, si può procedere con la terminazione del processo.

## 5.2 Il ciclo di vita di un processo

Il processo segue un ciclo di vita che comprende le seguenti fasi:

- Identificazione del processo
- Scoperta del processo (*as-is*)
- Analisi del processo
- Reingegnerizzazione del processo (*to-be*)
- Implementazione del processo
- Monitoraggio del processo

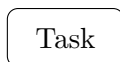
## 5.3 Concetti di base

### 5.3.1 Elementi di base di BPMN

- **Eventi:** rappresentano azioni che avvengono in maniera istantanea all'interno di un processo.



- **Task:** rappresentano unità di lavoro atomiche che hanno una durata.



- **flusso:** rappresenta una dipendenza temporale tra elementi di un processo.

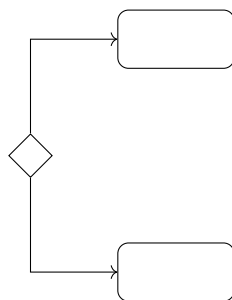


- **Gateway:** rappresentano decisioni o punti di convergenza all'interno di un processo.

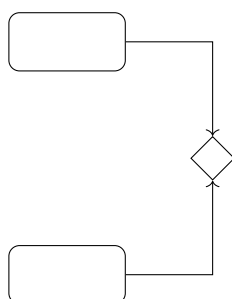


I gateway possono essere di diversi tipi:

- **Gateway di split:** rappresentano un punto di divergenza all'interno di un processo.



- **Gateway di merge:** presente in caso di gateway parallelo, ma anche per quando ho gateway di split per la rappresentazione di processi ben strutturati.



Certe attività possono essere rappresentate sia come task che come eventi, cambia solo il significato che voglio dare a quella particolare attività. Se voglio rappresentare un'attività che ha una durata, la rappresento come task, altrimenti come evento.

### 5.3.2 Collegare attività

Il flusso di sequenza definisce l'ordine in cui le attività vengono eseguite.

Gli eventi di start hanno un solo arco in uscita, mentre gli eventi di fine hanno un solo arco in entrata. Le attività possono avere più archi in entrata e in uscita, ma per una modellazione per strutturata si cerca di evitare.

## 5.4 Comportamento delle attività

Per raccontare come si comporta un processo esiste un oggetto teorico chiamato **token**. Esso non fa parte dei costrutti BPMN, ma è un'idea che permette di pensare a come si comporta un processo.

## 5.5 Gateway

I gateway sono elementi che permettono di prendere decisioni all'interno di un processo. Esistono diversi tipi di gateway:

- **Gateway esclusivo:** rappresenta una decisione che ha un solo percorso possibile.



- **Gateway parallelo:** provvede a un meccanismo di sincronizzazione di più flussi paralleli. Se un token arriva al gateway di merge, esso aspetta che anche l'altro token arrivi prima di proseguire.



- **Gateway inclusivo:** rappresenta una decisione che può avere più percorsi possibili. Se un token arriva al gateway di merge, esso aspetta che anche gli altri token arrivino prima di proseguire.



- **Gateway basato su eventi:** rappresenta un punto di decisione basato su eventi.



In caso di gateway, è possibile specificare anche un cammino di default, che verrà seguito se nessuna delle condizioni specificate è soddisfatta, attraverso un trattino sull'arco di uscita dal gateway.

## 5.6 Rappresentare l'organizzazione di un'azienda

A livello di BPMN possiamo rappresentare:

- **Ruoli** (*skill, competenze, qualifiche*): rappresentati come *lane*, che identificano le responsabilità di una specifica unità organizzativa.
- **Gruppi** (*dipartimenti, team, uffici, unità organizzative*): rappresentati come *pool*, che rappresentano un'unità organizzativa in grado di contenere più ruoli.

### 5.6.1 Pool

#### Pool

Una **pool** rappresenta un'unità organizzativa indipendente, che interagisce con altre pool attraverso dei messaggi.



### 5.6.2 Lane

#### Lane

Una **lane** rappresenta un ruolo o una risorsa all'interno di una pool.

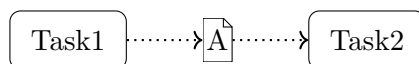
### 5.6.3 Flusso di messaggi

#### Esterno alla pool

I messaggi rappresentano le comunicazioni tra le varie pool. Essi sono rappresentati come frecce tratteggiate tra le varie pool. Non è possibile rappresentare un flusso di messaggi tra lane diverse all'interno della stessa pool, dato che l'informazione fluisce all'interno della stessa unità organizzativa, l'informazione è **implicita**.

#### Interno alla pool

È possibile esplicitare il flusso di messaggi tra lane diverse all'interno della stessa pool, con degli artefatti, ovvero i *data object*. In questo caso l'arco è puntinato.



## 5.7 Artefatti

Gli artefatti sono quell'insieme di costrutti che permettono di aggiungere informazione al processo. Essi possono essere di diversi tipi:

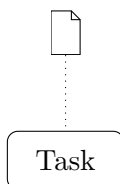
- **Data object**: rappresenta un'informazione che viene utilizzata all'interno del processo.



- **Data store**: rappresenta un'informazione che viene utilizzata all'interno del processo, ma che viene memorizzata.



- **Associazione**: permette di collegare un artefatto a un elemento del processo. Che può essere diretta o indiretta.



## 5.8 Costrutti avanzati

### 5.8.1 Specializzare tipi di task

È possibile specializzare tipi di task in 7 categorie principali:


- **Nessuno:** rappresenta un task generico che non ha una funzione specifica. Può essere utilizzato come base per altri tipi di task.

Task

- **Utente:** rappresenta un task che richiede l'interazione diretta di un utente umano. Questo tipo di task è utilizzato per attività manuali o che richiedono decisioni specifiche da parte di una persona.

Utente


- **Ricezione:** rappresenta un task che si occupa di ricevere informazioni, dati o messaggi da un'altra entità (ad esempio un sistema o un utente). È utile per gestire input.

Ricezione

- **Invio:** rappresenta un task che si occupa di inviare informazioni, dati o messaggi a un'altra entità. È utilizzato per gestire output.

Invio

- **Servizio:** rappresenta un task automatizzato che si avvale di un servizio o di un'applicazione per svolgere operazioni specifiche. È utile in processi che richiedono integrazioni con sistemi esterni.

Servizio

- **Script:** rappresenta un task che esegue uno script o un programma. Questo tipo di task è utile per automatizzare processi personalizzati o eseguire operazioni predefinite.

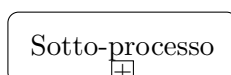
Script

- **Manuale:** rappresenta un task che richiede un'azione manuale da parte di un utente, ma che non richiede necessariamente una decisione. È utilizzato per rappresentare attività operative o ripetitive.



### 5.8.2 Attività

Un'attività può essere atomica o composta. Un'attività atomica è un'attività che non può essere suddivisa in attività più piccole, mentre un'attività composta è un'attività che può essere suddivisa in attività più piccole. In BPMN le attività composte sono rappresentate come sotto-processi.



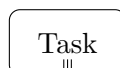
I sotto-processi possono essere di due tipi:

- **Embedded:** sono processi immersi nel loro processo padre. Tale processo non può essere riutilizzabile e tutto quello che accade viene usato dal padre, in questo caso il sottoprocesso inizia solo con l'evento generale.
- **Independent:** sono processi che possono essere riutilizzati in altri processi. In questo caso il sotto-processo può iniziare anche con eventi indipendenti dal processo padre.

### 5.8.3 Attività multi-istanza

Le attività multi-istanza sono attività che vengono eseguite più volte in parallelo o in sequenza. Esistono due tipi di attività multi-istanza:

- **Parallela:** rappresenta un'attività che viene eseguita più volte in parallelo. In questo caso, tutte le istanze vengono eseguite contemporaneamente.



- **Sequenziale:** rappresenta un'attività che viene eseguita più volte in sequenza. In questo caso, le istanze vengono eseguite una dopo l'altra.

### 5.8.4 Eventi

Gli eventi sono rappresentati come cerchi e rappresentano azioni che avvengono in maniera istantanea all'interno di un processo. Esistono diversi tipi di eventi:

- **Inizio:** rappresenta l'inizio di un processo. Gli eventi di inizio possono essere attivati da eventi o messaggi.



- **Fine:** rappresenta la fine di un processo. Tutti gli eventi di fine lanciano risultati.



- **Intermedi (di catch o throw):** rappresentano eventi che avvengono durante il processo. Gli eventi intermedi possono essere di due tipi:
  - **Catch:** rappresentano eventi che vengono catturati durante il processo.
  - **Throw:** rappresentano eventi che vengono lanciati durante il processo.



Gli eventi possono essere vari (*per semplicità non li rappresentiamo per inizio, fine e intermedi*):

- **Normale:** rappresenta un evento che avviene in maniera normale.
- **Messaggio:** rappresenta un evento che avviene in seguito a un messaggio.
- **Timer:** rappresenta un evento che avviene in seguito a un timer.
- **Condizionale:** rappresenta un evento che avviene in seguito a una condizione.
- **Segnale:** rappresenta un evento che avviene in seguito a un segnale.
- ...

## Capitolo 6

# Data Warehousing

Fino a qualche anno fa, la funzione svolta dalle basi di dati in ambito aziendale era solo quella di memorizzare dati operazionali, ovvero dati generati da operazioni svolte all'interno dei processi gestionali. Quindi, l'informatica veniva vista semplicemente come un supporto per rendere più rapide ed economiche le operazioni di gestione dell'informazione, senza creare effettivamente ricchezza.

### Evoluzione dei Sistemi Informativi

Dai primi anni '70, il ruolo dei sistemi informatici è cambiato. Essi sono stati utilizzati come sistemi di supporto alle decisioni. Oltre a migliorare l'efficienza dei processi, assumono quindi un ruolo centrale nell'organizzazione aziendale.

- **Tecnologia di supporto:** Supporta la gestione del sistema informativo aziendale.
- **Disciplina organizzativa:** Influenza i processi, i servizi e la struttura aziendale.

Un sistema informativo può essere migliorato per:

- **Trasformazione economica**
- **Trasformazione d'impresa**
- **Globalizzazione**

### 6.1 Piattaforma per la Business Intelligence

In relazione al portafoglio applicativo dell'organizzazione, sappiamo che il portafoglio direzionale è l'insieme degli applicativi direzionali, utilizzati dai manager per analizzare lo stato dell'azienda, prendere decisioni rapide, prendere le decisioni migliori.

### Piattaforma per la business intelligence

Si parla di piattaforma per la **business intelligence**, ossia la disciplina che consente a chi deve decidere in azienda di capire, attraverso soluzioni software, i fattori chiave del business e conseguentemente di prendere le migliori decisioni in quel momento.

Si parla di **piattaforma** poiché per consentire ai manager analisi potenti e flessibili è necessario definire un'apposita infrastruttura hardware e software. Questa piattaforma è composta da:

- Hardware dedicato
- Infrastrutture di rete
- DBMS
- Software di back-end
- Software di front-end

Il ruolo chiave di una piattaforma di business intelligence è la trasformazione dei dati aziendali in informazioni fruibili a diversi livelli di dettaglio.

Per ogni azienda è fondamentale poter disporre in maniera rapida e completa delle informazioni necessarie al processo decisionale.

## 6.2 Introduzioni al Data Warehousing

### Data Warehousing

Il **data warehousing** è una tecnica che consente di integrare e organizzare i dati provenienti da sorgenti di varia natura e di renderli disponibili per analisi e valutazioni finalizzate alla pianificazione e al processo decisionale.

#### 6.2.1 OLTP e OLAP

Mescolare interrogazioni *analitiche* e *transazionali* su un unico database è inefficiente. Per questo motivo, si separa l'elaborazione di tipo OLTP (Online Transaction Processing) da quella di tipo OLAP (Online Analytical Processing), costruendo un nuovo raccoglitore di informazioni che integri i dati provenienti da sorgenti di varia natura, li organizzi in una forma appropriata e li renda disponibili per scopi di analisi e valutazione finalizzate alla pianificazione e al processo decisionale.

Una collezione di metodi, tecnologie e strumenti di ausilio al knowledge worker (*dirigente, amministratore, gestore, analista*) per condurre analisi dei dati finalizzate all'attuazione di processi decisionali e al miglioramento del patrimonio informativo.

### Caratteristiche del processo di datawarehousing

- **Accessibilità** a utenti con conoscenze limitate di informatica e strutture dati.
- **Integrazione dei dati** sulla base di un modello standard dell'impresa.
- **Flessibilità di interrogazione** per trarre il massimo vantaggio dal patrimonio informativo esistente;
- **Sintesi** per permettere analisi mirate ed efficaci;
- **Rappresentazione multidimensionale** per offrire all'utente una visione intuitiva ed efficacemente manipolabile delle informazioni;
- **Correttezza e completezza** dei dati integrati.

#### 6.2.2 Il Data Warehouse

##### Processo Data Warehouse

Un **Data Warehouse** è una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche:

- è orientata a i soggetti di interesse: i soggetti sono coloro che fanno analisi.
- è integrata e consistente: raccoglie informazioni da fonti di dati eterogenee.
- è rappresentativa dell'evoluzione temporale: mantiene lo storico, dato che non elimino informazione.
- non volatile: il data warehouse non richiede tecniche sofisticate di gestione delle transazioni, ma richiede solo query-throughput elevato e resilienza.

#### 6.2.3 Le interrogazioni

Le interrogazioni possono essere di due tipi:

- **Interrogazioni di tipo OLTP:** le interrogazioni eseguono transazioni che leggono e scrivono un ridotto numero di record da diverse tabelle legate da semplici relazioni.
- **Interrogazioni di tipo OLAP:** Le interrogazioni effettuano un'analisi dinamica e multidimensionale che richiede scansione di un'enorme quantità di record per calcolare un insieme di dati numerici di sintesi che quantifichino le prestazioni dell'azienda.

### 6.3 Architetture

I requisiti architetturali sono:

- **Separazione:** L'elaborazione analitica e quella transazionale devono essere mantenute il più possibile separate.
- **Scalabilità:** L'architettura hardware e software deve poter essere facilmente ridimensionata per far fronte alla crescita dei volumi di dati e del numero di utenti.

- **Estendibilità:** Deve essere possibile integrare nuove applicazioni e tecnologie senza la necessità di una riprogettazione completa del sistema.
- **Sicurezza:** Il controllo sugli accessi è essenziale, data la natura strategica dei dati memorizzati.
- **Amministrabilità:** La complessità delle attività di amministrazione non deve risultare eccessiva.

### 6.3.1 Architettura a un livello

I **dati operazionali** sono quelli nelle basi di dati tradizionali che devono essere travasate nel data warehouse. Il livello warehouse è implementato per mezzo di **middleware** che simula la presenza del datawarehouse. Nel livello di analisi abbiamo strumenti di reportistica e strumenti OLAP.

#### Svantaggi

Non c'è una vera separazione tra OLAP e OLTP, poichè la presenza del data warehouse è solo simulata.

### 6.3.2 Architettura a due livelli

A livello di sorgente dati, i dati operazionali e i dati esterni vengono trasformati e caricati nel data warehouse. La trasformazione avviene mediante strumenti di ETL (Extract, Transform, Load). A livello del warehouse i dati vengono organizzati in **data mart**.

I **data mart** sono sottoinsiemi o aggregazioni dei dati presenti nel data warehouse primario, che contengono l'insieme delle informazioni rilevanti per una particolare area del business, divisione aziendale, categoria di soggetti.

A livello di analisi abbiamo strumenti di reportistica e strumenti OLAP, strumenti di data mining e per l'analisi what-if.

I data mart si classificano in due categorie:

- **Data mart dipendenti:** sono data mart che dipendono dal data warehouse primario.
- **Data mart indipendenti:** sono data mart che non dipendono dal data warehouse primario, sono alimentati direttamente dalle sorgenti, detti indipendenti.



#### Vantaggi

- A livello del warehouse è continuamente disponibile informazione di buona qualità, anche quando l'accesso alle sorgenti è temporaneamente precluso per motivi tecnici o organizzativi.
- L'interrogazione analitica sul Data Warehouse non interferisce con la gestione delle transazioni operazionali, la cui affidabilità è essenziale per il funzionamento aziendale.
- L'organizzazione logica del data warehouse è basata sul modello multidimensionale, mentre le sorgenti dati tipicamente adottano modelli relazionali o semi-strutturati.
- Esiste una discordanza temporale e di granularità: i sistemi OLTP gestiscono dati correnti e di dettaglio massimo, mentre i sistemi OLAP operano su dati storici e sintetici.
- A livello del warehouse è possibile utilizzare tecniche specifiche per ottimizzare le prestazioni delle applicazioni di analisi e reportistica.

### 6.3.3 Architettura a tre livelli

Rispetto all'architettura a due livelli, abbiamo i dati riconciliati che mettono insieme i dati dalle sorgenti.

I **dati riconciliati** sono dati operazionali ottenuti attraverso **integrazione** e **ripulitura** dei dati sorgente. Sono quindi dati integrati, consistenti e di qualità.

#### Vantaggi

- Creazione di un modello di dati comune e di riferimento per l'intera azienda.
- Introduzione di una separazione tra le problematiche legate all'estrazione e integrazione dei dati dalle sorgenti e quelle inerenti all'alimentazione del Data Warehouse.

#### Svantaggi

- Introduzione di ridondanza rispetto ai dati operazionali sorgente.

### 6.3.4 ETL - Extraction, Transformation and Loading

I processi ETL comprendono tre fasi principali: Estrazione, Trasformazione e Caricamento dei dati, fondamentali per la gestione di un Data Warehouse.

## Estrazione

### Estrazione

I dati vengono estratti dalle sorgenti.

- **Estrazione statica:** eseguita solo alla prima popolazione del DW, rappresenta una fotografia completa dei dati operazionali.
- **Estrazione incrementale:** cattura solo i cambiamenti dalle ultime estrazioni, basandosi sui log del DBMS e time-stamp.
- La scelta dei dati estratti dipende principalmente dalla loro qualità.

## Pulitura

### Pulitura

La pulitura dei dati migliora la loro qualità eliminando errori e incoerenze.

- Rimozione di duplicati e dati mancanti.
- Correzione di inconsistenze tra valori logicamente connessi.
- Uniformazione di convenzioni differenti (*es. abbreviazioni negli indirizzi*).
- Identificazione e correzione di valori fuori range o errati.

## Trasformazione

### Trasformazione

I dati vengono convertiti dal formato operativo a quello del DW.

- **Integrazione di fonti eterogenee:** risolve differenze di formato e convenzioni.
- **Conversione e normalizzazione:** *es. unità di misura unificate.*
- **Matching:** allinea campi equivalenti tra sorgenti diverse.
- **Selezione:** riduce il numero di campi e record.
- **Aggregazione:** nei DW i dati vengono sintetizzati e denormalizzati per ottimizzare analisi e reporting.

## Caricamento

### Caricamento

Processo di inserimento dei dati trasformati nel DW.

- **Refresh:** sostituisce interamente i dati nel DW (*usato solo nella prima popolazione, per evitare la perdita della storicizzazione*).

- **Update:** aggiorna il DW aggiungendo solo le modifiche avvenute nei dati sorgente (*metodo usato per aggiornamenti periodici*).

## 6.4 Modello multidimensionale

Il modello multidimensionale è il fondamento per la rappresentazione e l'interrogazione dei dati nei data warehouse.

I concetti principali a cui tener conto quando si modellano i dati all'interno del data warehouse sono:

- **Il fatto di interesse:** il concetto che si vuole analizzare.
- **Dimensioni:** sono i diversi punti di vista utilizzati per l'analisi.
- **Misura:** è il valore numerico che si vuole analizzare.
- **Gerarchie di attributi:** ogni dimensione può essere radice di una gerarchia di attributi. Essa permette di aggregare l'informazione. (*es. anno, trimestre, mese, giorno*). Aggregare è importante per avere una visione sintetica dei dati, fornendo una vista analitica.

Fissate le coordinate sulle dimensioni di analisi ho delle misure quantitative.

Nella modellazione di un data warehouse dobbiamo avere in mente:

- Qual è il fatto d'analizzare.
- Quali sono i punti di vista rispetto ai quali vogliamo analizzare il fatto (*dimensioni*)
- Quali sono le misure quantitative che vogliamo analizzare.
- Quali sono le gerarchie di attributi che vogliamo utilizzare per aggregare le informazioni.

Gli operatori OLAP permettono di gestire le informazioni del cubo per ricavarne informazioni utili.

## 6.5 Tecniche di analisi dei dati

Una volta che i dati sono stati ripuliti, integrati e trasformati, occorre capire come trarne il massimo vantaggio informativo. Esistono tre approcci differenti, supportati da altrettante categorie di strumenti, all'interrogazione di un data warehouse da parte degli utenti finali:

- **reportistica:** non richiede conoscenze informatiche.
- **OLAP:** richiede all'utente di ragionare in modo multidimensionale, permettendo di manipolare il cubo per informazione di interesse.
- **data mining:** richiede all'utente di avere conoscenze statistiche e matematiche, creando regole associative tra i dati con supporti e confidenze.

## 6.6 Progettazione logica del data warehouse

Dal punto di vista della progettazione concettuale, i data warehouse possono essere classificati in due categorie:

- **ROLAP: *Relational OLAP***, utilizza un database relazionale per memorizzare i dati del data warehouse. La sua diffusione è legata alla popolarità dei database relazionali. Tuttavia, presenta problemi di prestazione dovuti alle costose operazioni di join tra tabelle di grandi dimensioni, spesso mitigati attraverso la denormalizzazione dei dati.
- **MOLAP: *Multidimensional OLAP***, utilizza un database multidimensionale per memorizzare i dati. Qui non esiste un modello di dati standard, ma si basa su un modello logico ad hoc che permette di rappresentare direttamente dati e operazioni multidimensionali. I dati sono fisicamente memorizzati in vettori, con accesso posizionale, garantendo alte prestazioni e facilitando le operazioni multidimensionali senza necessità di join. Tuttavia, la mancanza di uno standard unificato rende le implementazioni MOLAP eterogenee e complesse da ottimizzare, specialmente nella gestione della scarsità dei dati (*ad esempio, celle vuote in analisi di vendita*).

Avremo quindi una base di dati relazionale che rappresenterà il fatto d'interesse, in cui abbiamo quindi la modellazione delle dimensioni con le determinate gerarchie e delle misure. Gli operatori ROLAP potranno quindi essere traslati in query SQL.

## 6.7 Il ciclo di vita del data warehouse

Molte organizzazioni mancano della necessaria esperienza e capacità per affrontare con successo le sfide implicite nei progetti di data warehousing.

Uno dei fattori che maggiormente minaccia la riuscita dei progetti è la mancata adozione di un **approccio strutturato**, che minimizza i rischi di insuccesso essendo basato su un'analisi costruttiva degli errori commessi.

Nella progettazione bisogna tener conto dell'utente finale che utilizzerà il prodotto, e la sua esperienza informatica. Inoltre, è importante considerare i requisiti di ciò che si vorrà fare con i dati e i dati a disposizione.

Ipotizzando di voler progettare di data warehouse ci sono due approcci possibili:

- **Approccio top-down**: considerando la definizione di data mart, che rappresenta un fatto di interesse, progettare tutti i data mart relativi a tutte le aree di analisi in maniera globale.

### Vantaggi

Visione globale del data warehouse, quindi i data mart saranno ben integrabili.

**Svantaggi**

Prima di mostrare i risultati all'utente finale, bisogna attendere la realizzazione di tutti i data mart, non riesco a fornire un prototipo in tempi brevi.

- **Approccio bottom-up:** si progettano i data mart uno alla volta, in modo incrementale, e si integrano successivamente. Si parte quindi dai data mart più significativi.

**Vantaggi**

Possibilità di fornire un prototipo in tempi brevi all'utente finale.

**Svantaggi**

Determina una visione parziale del dominio di interesse.

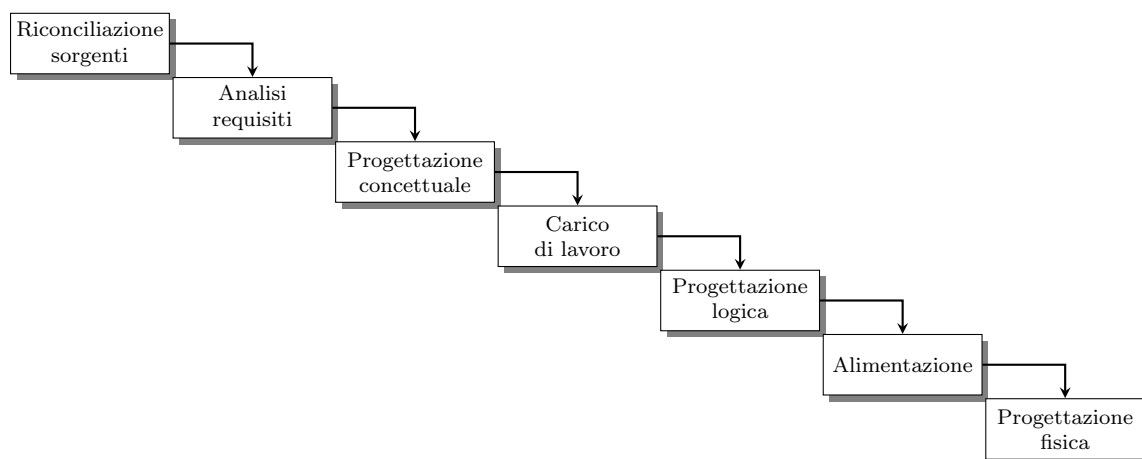
### 6.7.1 Il ciclo di sviluppo

Il ciclo di sviluppo di un data warehouse è composto da diverse fasi:

- **Definizione degli obiettivi e pianificazione**
- **Progettazione dell'infrastruttura**
- **Progettazione e sviluppo dei data mart**

### 6.7.2 Progettazione del data mart

La progettazione dei data mart segue diverse fasi:



L'interlocutore da tener in considerazione oltre all'utente finale è l'**amministratore della base di dati**, dato che rispetto alle richieste dell'utente finale, saprà dove sono i dati e come sono strutturati.

La parte di analisi dei requisiti viene giocata su due fronti:

- Chi ci dice cosa vuole che il sistema faccia.
- Chi ci dice quali sono i dati a disposizione e dove sono.

Prima di procedere con la progettazione concettuale bisogna analizzare e riconciliare i dati provenienti dalle sorgenti. I dati potrebbero provenire da sorgenti eterogenee. Il risultato sarà lo **schema riconciliato**.

La progettazione può essere fatta tenendo in considerazione solo l'utente finale, progettando quindi il data warehouse senza preoccuparsi dell'informazione disponibile o partendo dai dati disponibili e genera lo schema di fatto. Per la progettazione concettuale si utilizzerà il **dimensional fact model** che conterrà:

- **Fatto:** il fatto d'interesse.
- **Dimensioni:** i punti di vista di analisi.
- **Misura:** i valori numerici che si vogliono analizzare.
- **Gerarchie di attributi:** permettono di aggregare l'informazione.

Queste quattro componenti ci permettono di costruire lo schema di fatto del data mart. Questo schema di fatto deve poi essere mappato in uno schema logico, che nel nostro caso sarà di tipo relazionale.

Nella progettazione dell'alimentazione bisogna pensare a come travasare i dati dalle sorgenti al data warehouse. Questo processo è complesso e richiede l'utilizzo di strumenti ETL.

### 6.7.3 Analisi e riconciliazione delle sorgenti eterogenee

#### Riconciliazione e Normalizzazione

Nella prima fase, è necessario progettare il livello riconciliato e produrre lo schema riconciliato. Questo processo inizia con l'analisi delle sorgenti disponibili e la successiva integrazione degli schemi, tenendo conto delle diverse fonti di dati coinvolte.

Una volta completata l'integrazione, si può procedere alla definizione dello schema di mappatura. Per farlo, è fondamentale avere una chiara comprensione delle informazioni presenti nelle sorgenti di partenza. In questo, risultano di grande aiuto sia l'amministratore del sistema di origine sia la documentazione prodotta nella fase di progettazione dei sistemi informativi.

Infine, è necessario normalizzare i dati, con l'obiettivo di correggere gli schemi locali per rappresentare in modo più accurato il dominio applicativo. Questo processo consente di ottenere una struttura più coerente e ottimizzata, migliorando la qualità complessiva dei dati integrati.

Disaccoppiare le informazioni è utile nel contesto dell'analisi dimensionale.

## Integrazione

In schemi con sorgenti sovrapposte è utile trovare punti di congiunzione, sovrapponendo i dati per trovare le corrispondenze, integrando anche le informazioni presenti anche in una sola sorgente. Rappresentando i dati con un unico formalismo.

### 6.7.4 Relazioni tra concetti comuni

- **Identità:** Vengono usati gli stessi costrutti, il concetto è modellato dallo stesso punto di vista. Non vengono commessi errori di specifica.
- **Equivalenza:** Sono stati utilizzati costrutti diversi ma equivalenti e non ci sono errori di specifica o diversa percezione (prospettiva).
- **Comparabilità:** I costrutti utilizzati e i punti di vista dei progettisti non sono in contrasto tra loro.
- **Incompatibilità:** Gli schemi sono in contrasto a causa dell'incoerenza nelle specifiche.

### 6.7.5 Fasi dell'integrazione

- **Pre-integrazione**

#### Pre-integrazione

Viene definita la strategia di integrazione.

Ad esempio, si decide se integrare gli schemi uno alla volta oppure analizzarli tutti insieme per un'integrazione globale.

- **Comparazione degli schemi** Analisi comparativa dei diversi schemi per identificare le correlazioni e i conflitti tra i concetti espressi.
- **Allineamento degli schemi**

#### Allineamento degli schemi

Risoluzione dei conflitti evidenziati nella fase precedente.

Applicando primitive di trasformazione agli schemi sorgenti o allo schema riconciliato temporaneamente definito.

- Esempio: cambio dei nomi e dei tipi di attributi.
- Alcuni conflitti potrebbero essere irrisolvibili a causa di inconsistenze di base del sistema informativo. In questi casi, la soluzione va discussa con i clienti.

- **Fusione degli schemi**

#### Fusione degli schemi

Gli schemi allineati vengono fusi per formare un unico schema riconciliato.

I concetti comuni vengono sovrapposti e collegati ai concetti provenienti dagli schemi locali.

- Dopo questa operazione si effettuano ulteriori trasformazioni per migliorare la struttura dello schema rispetto a:
  - \* Completezza.
  - \* Minimalità (evitare ridondanze non necessarie).
  - \* Leggibilità.
- Si integrano dati provenienti da sorgenti diverse mettendoli a fattor comune.
- Possibilità di costruire le *view* per la gestione dei dati integrati.

### 6.7.6 Analisi dei Requisiti

#### Obiettivo

##### Obiettivo

La fase di analisi dei requisiti ha l'obiettivo di raccogliere le esigenze di utilizzo del data mart espresse dai suoi utenti finali.

Non solo vogliamo capire quali sono i dati a disposizione, ma anche quali mettere nel DW per soddisfare le esigenze dei clienti.

#### Fonti

##### Fonti

Le principali fonti per la raccolta dei requisiti sono:

- **Business users:** utenti del data mart che usano il DW per analisi, spesso privi di competenze informatiche.
- **Amministratori del sistema informativo:** conoscono i dati disponibili e le loro caratteristiche.

#### Interviste

##### Interviste

Due approcci principali:



- **A piramide** (approccio induttivo): si parte da domande dettagliate per poi ampliare l'argomento con domande aperte. Adatto a soggetti riluttanti.
- **A imbuto** (approccio deduttivo): si parte da domande generali per poi restringere l'argomento a temi specifici, riducendo la tensione dell'intervistato.

Le domande variano in base al ruolo dell'intervistato:

- **Dirigente**: obiettivi aziendali e problemi principali.
- **Direttore di reparto**: obiettivi, analisi di routine, analisi desiderate, livello di dettaglio necessario.
- **Amministratore del sistema informativo**: strumenti di analisi, problemi di qualità dei dati, principali fonti disponibili.

## Fatti

### Fatti

I fatti sono i concetti su cui gli utenti baseranno il processo decisionale.

L'analisi dei requisiti individua i fatti di interesse attraverso sorgenti informative e interviste. Per ogni fatto si determina:

- **Dimensioni** (dimensioni del cubo): granularità dei dati.
- **Intervallo di storicizzazione**: arco temporale coperto dagli eventi memorizzati.

Ogni data mart rappresenta un fatto che può essere analizzato con il cubo.

## 6.8 Progettazione Concettuale

La progettazione concettuale di un Data Warehouse non segue un unico formalismo universalmente riconosciuto. Sebbene il modello multidimensionale sia il riferimento principale, non esiste una metodologia standardizzata.

### 6.8.1 DFM - Dimensional Fact Model

#### DFM - Dimensional Fact Model

Il DFM è un modello concettuale grafico pensato per i data mart. Permette di:

- formulare interrogazioni in modo intuitivo;
- facilitare il dialogo tra progettista e utente finale per affinare i requisiti;
- creare una piattaforma stabile per la progettazione logica;
- generare documentazione espressiva e non ambigua.

Lo schema concettuale generato dal DFM è costituito da un insieme di **schemi di fatto** che modellano:

- **fatti**;
- **misure**;
- **dimensioni**;
- **gerarchie**.

### 6.8.2 Costrutti di Base

#### Fatto

Un **fatto** rappresenta un concetto di interesse per il processo decisionale (es. *VENDITA*). Esso corrisponde a eventi aziendali che evolvono nel tempo.

#### Misura

La **misura** è una proprietà numerica associata a un fatto, che ne descrive un aspetto quantitativo (es. *INCASSO* per una vendita).

#### Dimensione

La **dimensione** funge da coordinata per l'analisi di un fatto (es. *PRODOTTO*, *NEGOZIO*, *DATA*).

#### Attributi Dimensionali

Gli **attributi dimensionali** descrivono le dimensioni. Ad esempio, un *PRODOTTO* può essere caratterizzato da *marca*, *tipo*, *categoria*, ecc.

#### Gerarchia

La **gerarchia** è una struttura ad albero in cui:

- i nodi rappresentano attributi dimensionali;
- gli archi modellano associazioni *molti-a-uno* tra coppie di attributi.

**Esempio:** La gerarchia della dimensione *DATA*:

- Una *data* appartiene a un *mese* (un mese contiene più date).
- Una *data* appartiene a una *settimana* (anche se una settimana può estendersi su due mesi).

### 6.8.3 Relazioni tra gli Elementi

#### Relazioni nel modello ER

Nel modello ER:

- Le **dimensioni** sono rappresentate come entità.
- Le **misure** sono attributi della relazione.
- Il **fatto** corrisponde alla relazione tra le entità, esprimendo un'associazione *multi-a-molti* tra le dimensioni.

### 6.8.4 Naming Conventions

#### Naming Conventions

- Gli attributi dimensionali devono avere nomi univoci all'interno di ogni schema di fatto.
- Attributi con lo stesso significato, presenti in schemi differenti, devono mantenere lo stesso nome.

### 6.8.5 Eventi e Aggregazioni

#### Evento Primario

Un **evento primario** è un'occorrenza specifica di un fatto, identificata da una n-upla di valori per ciascuna dimensione, con un valore associato per ogni misura.

#### Evento Secondario

Un **evento secondario** è l'aggregazione di eventi primari basata su un insieme di attributi dimensionali, che riassume i valori delle misure.

Le **gerarchie** definiscono come gli eventi primari possano essere aggregati e analizzati: la dimensione alla radice stabilisce la granularità più fine, mentre gli altri attributi indicano granularità crescenti.

### 6.8.6 Additività delle Misure

#### Additività delle Misure

Le misure possono essere classificate in:

1. **Misure di flusso:** valutate su un periodo di tempo e cumulabili (es. numero di prodotti venduti al giorno, incasso mensile).
2. **Misure di livello:** valutate in un istante specifico (es. numero di prodotti in inventario, popolazione di una città).
3. **Misure unitarie:** valutate in un istante ma espresse in termini relativi.

Una misura è definita **additiva** lungo una dimensione se può essere sommata lungo la gerarchia; altrimenti è **non-additiva**. Se nessun operatore di aggregazione è applicabile, la misura è **non-aggregabile**.

### 6.8.7 Passi di Progettazione di uno Schema Concettuale

#### Passi di Progettazione

1. **Definizione dei fatti.**
2. Per ciascun fatto:
  - Costruzione dell'albero degli attributi.
  - Editing dell'albero degli attributi.
  - Definizione delle dimensioni.
  - Definizione delle misure.
  - Creazione dello schema di fatto.

**Nota:** I fatti rappresentano concetti di interesse primario per il processo decisionale e, nello schema relazionale, corrispondono a relazioni. Ogni fatto diventa la radice di un nuovo schema, e la dimensione temporale deve essere sempre inclusa.

### 6.8.8 Problema della Sparsità

Nel modello multidimensionale, ad ogni combinazione di coordinate corrisponde un possibile evento, anche se non si verifica realmente. Di solito, il numero di eventi reali è molto inferiore a quello dei possibili, causando:

- Spreco di risorse.
- Riduzione delle prestazioni del sistema.

#### Gestione della Sparsità

- **ROLAP:** memorizza solo gli eventi accaduti.
- **MOLAP:** richiede tecniche avanzate per minimizzare lo spazio necessario a rappresentare gli eventi non avvenuti.

La sparsità diminuisce all'aumentare del livello di aggregazione dei dati.

## 6.9 Progettazione Logica

La modellazione concettuale è indipendente dal modello logico scelto per l'implementazione. La struttura multidimensionale dei dati può essere rappresentata mediante:

1. **MOLAP** (Multidimensional On-Line Analytical Processing): utilizza strutture multidimensionali native, come vettori multidimensionali.
2. **ROLAP** (Relational On-Line Analytical Processing): sfrutta il modello relazionale per rappresentare i dati multidimensionali.

### 6.9.1 MOLAP

I sistemi MOLAP offrono prestazioni elevate, poiché le operazioni multidimensionali sono native e non richiedono query SQL complesse. Tuttavia, presentano il problema della sparsità (solitamente solo il 20% delle celle contiene dati). Tecniche per gestire la sparsità includono:

- **Suddivisione delle dimensioni:** il cubo n-dimensionale viene partizionato in sottocubi (*chunk*) per una gestione efficiente.
- **Compressione dei chunk:** le celle non informative sono rappresentate in forma compressa tramite un indice degli offset rilevanti.

Un ulteriore limite del MOLAP è la mancanza di standardizzazione, dovuta all'uso di strutture dati proprietarie.

### 6.9.2 ROLAP: Schema a Stella

Nel modello ROLAP la sparsità viene gestita memorizzando solo le tuple esistenti, anche se ciò comporta query più complesse per simulare operazioni multidimensionali. Il modello si basa sullo **schema a stella**, composto da:

- **Dimension Table (DT):** tabella per ogni dimensione, con chiavi surrogate e attributi descrittivi.
- **Fact Table (FT):** tabella contenente le chiavi surrogate di tutte le DT e gli attributi di misura.

### 6.9.3 ROLAP: Schema a Fiocco di Neve (Snowflake Schema)

In questo schema le DT vengono ulteriormente normalizzate e suddivise in sottotabelle, riducendo la ridondanza ma aumentando il numero di join necessari:

- **Dimension Table primarie:** tabelle le cui chiavi vengono importate direttamente nella FT.
- **Dimension Table secondarie:** tabelle che contengono dettagli aggiuntivi per ridurre la ridondanza.

#### 6.9.4 Le Viste

Le viste rappresentano porzioni della base di dati ottenute tramite query precalcolate. Esse si distinguono in:

- **Viste primarie:** rappresentano il livello minimo di aggregazione.
- **Viste secondarie:** rappresentano livelli di aggregazione superiori.

L'uso delle viste facilita l'analisi dei dati a diversi livelli di dettaglio, migliorando le prestazioni delle interrogazioni OLAP.