

A.A. 2020/21*

SISTEMI INFORMATIVI AZIENDALI

PROF. BARBARA OLIBONI

FABS :)

NOTA

Questi appunti/sbobinatura/versione “discorsiva” delle slides sono per mia utilità personale,
quindi pur avendole revisionate potrebbero essere ancora presenti typos, commenti/aggiunte personali (che anzi, lascio
di proposito) e nel caso peggiore qualche inesattezza!

Comunque spero siano utili! 

*

Non ho seguito il corso in diretta, ma solo dalle registrazioni; in particolare, non sono riuscita a scaricare in tempo quelle
del 2021/22 quindi anche se queste sbobine sono state fatte quest’anno
le lezioni di riferimento sono quelle dell’anno 2020-21.

- Personalmente trovo che occasionalmente la prof/le slides abbiano una struttura poco chiara, quindi ho dato io un po’ di struttura al contenuto, a volte muovendo un po’ in giro gli argomenti per raggrupparli in maniera (per me) più sensata.
- I grafici sono stati realizzati con **Visual Paradigm Online** (bless VP, è draw.io ma con tutti i preset fatti. Davvero, bless VP). Sono visualizzabili e modificabili a [questo link](#).
- Crediti al prof. Domenico Benevento per i disegnetti in DFM

Appunti e puttanate sono scritte *principalmente* in questo stile.

**Questo file fa parte della mia collezione di sbobinature,
che è disponibile (e modificabile!) insieme ad altre in questa repo:
<https://github.com/fabfabretti/sbobinamento-seriale-uniVR>**

INDICE

NOTA	2
Indice.....	4
Introduzione.....	5
Prospettiva informativa della teoria dell'organizzazione	6
Classificazione dei Sistemi Informativi	10
Sistemi ERP – Enterprise resource planning	13
Sistemi CRM - Client relations management.....	15
Ingegneria e trasformazione dei processi gestionali	18
Trasformazione dei processi: casi reali	22
Modellazione dei processi: BPMN.....	25
Sistemi di supporto alle decisioni	34
Progettazione di un data warehouse	39
DFM – Dimensional fact model	42
Progettazione concettuale	45
Progettazione logica	48
Analisi in un Data warehouse	52
Qualità dei dati.....	56

INTRODUZIONE

L1 - 12/10/2020

In questo corso farà una panoramica di tutti i sistemi informativi da usare in azienda per gestire le informazioni di interesse in un'azienda, per permettere di raggiungere un certo obiettivo prefissato. Le informazioni di interesse sono usate durante l'esecuzione dei processi aziendali.

Ci occuperemo di modellare e gestire i processi aziendali.

<i>Def: Sistema informativo</i>	<i>Def: Processi aziendali</i>	<i>Def: Organizzazione</i>
<p>Un sistema informativo è la componente o sottosistema di un'organizzazione che gestisce le informazioni di interesse.</p> <p>Non per forza il sistema informativo è automatizzato: qualunque cosa che gestisce l'informazione, anche senza l'uso di una macchina, è un sistema informativo; lo siamo anche noi quando appuntiamo cose.</p>	<p>Insieme di attività da eseguire per ottenere un determinato risultato. Parliamo di processi aziendali, ma tutto può essere rappresentato come processo, anche ad esempio il processo "preparazione all'esame".</p>	<p>Insieme di persone e tecnologie, usate per gestire le risorse e raggiungere obiettivi.</p> <p>Ci occuperemo di capire come sono organizzate.</p>

Gestire informazioni nei DB classici è qualcosa che sappiamo già; a livello aziendale è interessante riuscire a partire dallo storico per analizzare le informazioni e fare previsione. Dunque, è molto importante il concetto di Data Warehouse, che sono sistemi informativi volti al supporto alle decisioni. Sono fondamentali, perché tutte le attività aziendali si basano sulle decisioni 😊

Quindi in questa materia ci si occupa di:

- Aspetti organizzativi dell'azienda
- Modellazione dei processi aziendali
- Data warehouse.
- Qualità dei dati.

+ seminari aziendali: organizza seminari in cui le aziende ci raccontano la ricaduta pratica di quel famo 😊

PROSPETTIVA INFORMATIVA DELLA TEORIA DELL'ORGANIZZAZIONE

Vogliamo capire come gestire, produrre, etc. l'informazione. Ci racconta "quattro cose" sull'organizzazione dell'azienda, ma con la prospettiva che siamo interessati all'informazione.

TL;DR: L'informazione è una risorsa aziendale importantissima.

Definizioni varie

Def: Organizzazione

In lingua italiana, organizzazione può essere intesa sia in senso di azienda che in senso di suddivisione e coordinamento di un'attività. Ci interessano entrambe le accezioni, anche perché sono un po' inscindibili: in un'azienda ho bisogno di coordinamento.

Organizzazione è anche la suddivisione e coordinamento delle persone; quindi un'organizzazione stabilisce come organizzo le risorse umane. Nel momento in cui riesco a organizzare le risorse riesco a costruire una struttura.

Def: Sistema informativo aziendale

È un sistema che mi permette di gestire, elaborare e produrre informazioni.

Questo vuol dire per esempio che se ho notato che in un certo periodo il fatturato si abbassa, riesco ad estrapolare un'informazione e prendere decisioni/mettere in atto politiche.



Def: Azienda

Sistema sociotecnico costituita dove coordino **persone** e **tecniche**.

- L'azienda ha un obiettivo da raggiungere attraverso le sue risorse e attraverso un processo aziendale. Questo insieme di attività deve essere fatto in maniera coordinata con le risorse umane
- Le attività delle persone sono quelle attività modellate dal processo aziendale.
- In alcuni casi, per migliorare le prestazioni delle risorse che ho a disposizione posso sfruttare le risorse informatiche.

Dall'interazione fra risorse umane e tecnologie deriva il **comportamento aziendale**, che produce i risultati. Alla base di tutto questo ho i dati: nessun processo aziendale può essere portato avanti senza informazioni.

Teoria dell'organizzazione

Per la definizione economica, **studio scientifico e progettazione scientifica dei compiti per il miglioramento delle prestazioni**.

Ci interessa perché le attività aziendali sono modellate da processi aziendali, e dunque vogliamo capire quale sia il processo da modellare. Il compito di raccogliere i requisiti e modellare un processo efficiente spetta a noi informatici.

Organizzazione aziendale

- Prevede ruoli e funzioni diversificate
- Caratterizzata da fini, metodi, regole.
- Opera attraverso processi
 - Parte produttiva e operativa
 - Parte amministrativa di controllo e gestione
- Con risorse a disposizione:
 - Umane
 - Materiali
 - Informative: in un'azienda che produce servizi sostituiscono anche la parte di materiali. È quello che interessa a noi.

Def: Tecnologie informatiche

Insieme di **strumenti, sistemi e tecniche** per **automatizzare il trattamento delle informazioni**. Tutti i sistemi di cui parleremo fanno parte delle IT.

Le tecnologie informatiche supportano il ciclo di vita dell'informazione come risorsa aziendale; l'informazione è utile per coordinare le attività e prendere decisioni.

Le tecnologie informatiche sono lo **strumento** dei sistemi informativi.

Def: Sistema informativo aziendale

Insieme di elementi interconnessi che **raccolgono, catalogano, ricercano, elaborano, ricercano, memorizzano, distribuiscono dati**, trasformandoli in **informazioni utili per supportare le attività decisionali** di controllo di un'azienda.

La differenza fra SIA e DB è che in DB si parte dal presupposto che si voglia organizzare l'informazione d'interesse, facendo tutte le fasi (requisiti, modellazione, ..) e poi la **uso nelle attività quotidiane**, come fare le fatture.

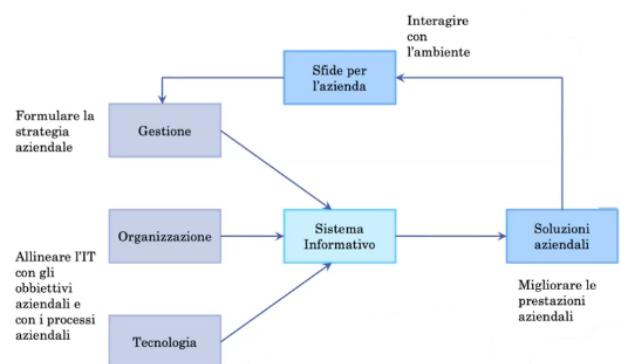
Quando voglio invece fare un passetto oltre, ovvero considerare anche informazioni volte alle decisioni, mi serve il sistema informativo aziendale, che usa le informazioni raccolte con il DB per **prendere decisioni** (come ad esempio cambiare una linea di produzione o migliorare le prestazioni ecologiche).

È progettato per svolgere tre macroprocessi fondamentali:

- **Input:** acquisizione dei dati
- **Elaborazione:** trasformazione dei dati
- **Output:** restituzione dei dati.

Il SIA deve **automatizzare la gestione e l'elaborazione dei dati**, rendendoli disponibili al management per trasformarsi in informazioni utili per prendere decisioni efficaci, tempestive e nel rispetto dell'efficienza globale dell'azienda.

L'azienda usa il SIA per affrontare le sfide e gestire i cambiamenti.



Informazione come risorsa organizzativa

L'informazione è risorsa fondamentale nella presa di organizzazione. È presente anche nella parte bassa (operativa) della piramide, ma è una risorsa molto meno principale. Diventa fondamentale a piani più alti, per controllare e coordinare.

- È la **risorsa principale**, scambiata selezionata ed elaborata nelle attività di controllo e coordinamento.
- È **prodotta da qualunque attività**, anche operativa.
- Si distingue dalle altre risorse per essere **immateriale**; include anche la conoscenza delle risorse umane.
- **Non** è facilmente **divisibile** (=passabile) o **appropriabile** (=rubabile)
- **Soggetta all'obsolescenza**
- **Non è distrutta dall'uso.**
- È **self-generating**: l'informazione **non si esaurisce, si autogenera**. Continuo ad elaborare info per estrarne altra; posso instaurare circoli virtuosi di generazione di conoscenza.
Di solito questi circoli si traducono in un *incremento delle prestazioni* dei processi decisionali; banalmente, anche solo risparmiare tempo aumenta il margine di guadagno.

Quantità di informazione

È rilevante, perché deve essere appropriata alla capacità elaborativa dell'organizzazione per poter essere sfruttata.

Overload informativo:

Aumento incontrollato dell'informazione disponibile.

- Disponibilità di informazioni che eccedono le capacità di elaborazione individuale
- Rallentamento e peggioramento delle decisioni

Underload informativo:

Disponibilità di informazioni al di sotto delle capacità di elaborazione.

- Semplificazione delle decisioni
- Decisioni in tempi brevi.

Flessibilità nell'organizzazione: organizzazione come sistema aperto

L'organizzazione è un **sistema aperto**, ovvero deve interfacciarsi con l'ambiente. L'organizzazione riesce a raggiungere i suoi obiettivi nonostante l'essere un sistema aperto quando è **sufficientemente flessibile** da adattarsi al sistema che le sta intorno.

L'organizzazione è resa flessibile dal suo SIA, perché è il supporto che permette di prendere decisioni veloci e corrette usando tutto ciò che ho a disposizione. L'informazione mi permette di avere la flessibilità necessaria per riuscire a procedere.

→ !!! fa la domanda “cosa vuol dire che l'organizzazione è un sistema aperto” vuole sentirsi dire “è qualcosa che sta immerso in un ambiente e vi si interfaccia; in particolare, vi si trovano fornitori, ambiente, clienti, autorità, azionisti, concorrenti, stakeholders e l'azienda deve prendere decisioni per stare al passo”.

Def: Incertezza ambientale

L'insieme dei cambiamenti imprevisti dati dall'esterno è chiamato **incertezza ambientale**.

Non sempre posso davvero decidere cosa fare (es. le leggi sono imposte!), ma posso cercare di adattarmi. Gestire queste eccezioni significa cambiare il proprio processo aziendale per raggiungere ugualmente l'obiettivo nonostante l'incertezza ambientale.

Modello gerarchico di Anthony

In ognuno di questi livelli abbiamo un SIA diverso.

- **Livello operativo:**

Le DB tradizionali stanno nel livello più basso, perché sono quelle che mi permettono di gestire operazioni quotidiane e usuali fatte a livello operativo. esempio: finisco la carta, lo vedo dal DB, lo faccio direttamente al livello operativo.

- **Livello direzionale**

Se invece voglio recuperare dei report mensili, mi sto spostando di un livello.

Prendo decisioni relative al miglioramento di una linea produttiva. Non serve scomodare il CEO 😊

- **Livello strategico:**

Infine, a partire da un'aggregazione dei report mensili posso pensare di prendere decisioni a livello strategico. Voglio fare una scelta strategica (es. nuova linea produttiva senza glutine), che avrà ricadute anche sul direzionale e operativo..



La piramide mi permette di capire le aree, i processi e le informazioni necessarie. Le decisioni sono diverse in base al livello, hanno ricadute sui livelli sotto e hanno un SIA specifico.

L'incertezza ambientale causa **riplanificazione**: a seconda del tipo di incertezza posso avere delle decisioni da prendere a un certo livello della piramide. Ad esempio:

- Cambiamento di tendenza di mercato → ripianificazione strategica degli obiettivi
- Malfunzionamento HW sulla linea di produzione → problema operativo

Maggiore è l'incertezza ambientale, più è necessario sia intensa l'attività informativa per potermici adeguare. Ho bisogno, dunque, di avere una capacità elaborativa maggiore, che è una conseguenza dell'avere un SI adeguato.

Capacità elaborativa e collaborazione

Def. Capacità elaborativa

La **misura di quanto un'azienda è in grado di elaborare informazioni** a essa imposte dai propri obiettivi e dal contesto in cui opera. La capacità elaborativa dell'organizzazione si basa sui SIA.

La capacità elaborativa di un'organizzazione **non è la mera somma** delle capacità elaborative di quelli che partecipano: ancor di più se ho un'organizzazione gerarchica sbilanciata su un unico decisore, la capacità elaborativa è bottlenecked dal decisore! Voglio che la capacità elaborativa sia **più della somma**, perché ho le capacità elaborative individuali e un potenziamento dato dal coordinamento e dalla collaborazione. **L'organizzazione vuole superare i limiti individuali.**

Modificare l'organizzazione aziendale implica riuscire a far collaborare le risorse.

Cooperazione

Arrivare alla cooperazione significa **suddividere i compiti in sottocompiti**.

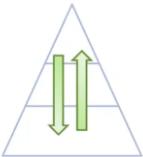
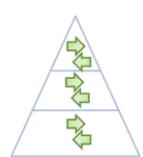
- Nel momento in cui faccio cooperare più individui riesco a **suddividere il compito elaborativo** e anche le informazioni.
 - Ognuno necessita di un sottoinsieme di informazioni, adatte al sottocompito di cui è responsabile.
- Nel momento in cui c'è della cooperazione si **generano anche delle interdipendenze informative**;
 - Chi fa la documentazione si interfaccia con chi scrive il codice

Per raggiungere una capacità elaborativa ottimale, dunque, tutto sta nel SIA. All'interno dell'organizzazione ce ne sono diversi; esistono anche sistemi che permettono di gestire contemporaneamente più necessità. La progettazione dei SIA dunque è molto importante ed è strettamente correlata all'organizzazione dell'azienda; più i compiti saranno divisi, più dovrò occuparmi di condividere l'informazione.

Di conseguenza, deve esserci un forte legame fra Sistema Informativo, Struttura Organizzativa e livello di incertezza.

Tipi di sistema informativo

Rispetto alle cose raccontate, storicamente ci sono a disposizione sistemi informativi verticali e orizzontali.

<p>Progettazione di Sistemi Informativi Verticali.</p> 	<p><i>Sistemi informativi verticali</i></p> <p>Permettono scambi informativi lungo la catena delle relazioni gerarchiche, e portano ad un accentramento delle decisioni.</p> <ul style="list-style-type: none">• Sono i primi ad essere supportati, perché rispecchiavano l'organizzazione gerarchica stretta• Al crescere dell'incertezza e del numero di eccezioni, i livelli gerarchici superiori vengono sovraccaricati dai compiti decisionali.• Il limite, dunque, è che la capacità elaborativa dell'organizzazione coincide con la capacità elaborativa dell'unità gerarchicamente più elevata, che potrebbe anche essere un singolo individuo.
<p>Progettazione di Sistemi Informativi Orizzontali.</p> 	<p><i>Sistemi informativi orizzontali</i></p> <p>Permette all'informazione di viaggiare in orizzontale, fra unità organizzazione dello stesso livello (anche se dipendenti gerarchicamente da unità distinte).</p> <ul style="list-style-type: none">• Nascono poiché con l'ingrandirsi dell'azienda l'organizzazione gerarchica diventa un problema, e parte il meccanismo dei team e delle deleghe.

Incidenza comportamentale

È un altro fattore problematico, oltre all'incertezza ambientale; fa riferimento a tutti i problemi dell'ingegneria sociale.

Implica un flusso informativo, perché per limitare l'incertezza legata al fattore umano posso organizzare l'azienda con dei meccanismi di controllo – ovvero vado a definire dei meccanismi di controllo fra risorse umane. Questo, ancora, implica un altro flusso di controllo e di informazioni; per esempio, report settimanali di ogni dipendente su cosa ha fatto.

Assume che:

1. Gli individui non perseguono necessariamente solo gli obiettivi **organizzativi**, e hanno interesse personale che predomina.
2. La gerarchia in sé non è sufficiente a eliminare l'**opportunismo** e servono altre forme di coordinamento.

È un costo:

- **Costo di controllo**: verifica del comportamento (capo che legge i report)
- **Costo di garanzia**: produzione di documentazione (dipendente fa report settimanali)

Ovviamente questo tempo poteva essere speso in maniere più remunerative, **but alas**. Posso minimizzare questi costi sfruttando il SIA per minimizzare il costo di coordinamento, ovvero il tempo da sfruttare in azienda per questo processo.

A livello di tecnologie informatiche, a noi interessa il ruolo delle tecnologie nel supporto alle decisioni. Dal punto di vista informativo, il comportamento opportunistico pone il problema di:

- | | |
|---|---|
| <ul style="list-style-type: none">• Bilanciare costi di controllo e garanzie. | <ul style="list-style-type: none">• Trovare il trade off fra centralizzazione e delega. |
|---|---|

CLASSIFICAZIONE DEI SISTEMI INFORMATIVI

Esistono più classificazioni a seconda del punto di vista.

Classificazione dei SIA disposti lungo la piramide aziendale

Definizione e funzioni attribuite dipendono dal livello al quale i SI sono collocati.

Classificazione dei livelli: piramide di Anthony

Processi operativi	Produrre	Attività operative e processi operativi, aka che concorrono all'attuazione degli obiettivi. Negli esempi della volta scorsa possiamo pensare alla produzione di beni e servizi. A questo livello, i processi hanno lo scopo di raggiungere gli obiettivi definiti nei livelli superiori. Il personale che segue i processi è il personale esecutivo.
Processi gestionali	Tradurre da strategico a operativo	Processi che permettono di tradurre gli obiettivi strategici dell'azienda, e verificare il raggiungimento degli obiettivi. È il livello intermedio, che mappa le decisioni strategiche sulle attività da svolgere. Per esempio, la decisione strategica "nuova linea senza glutine" a livello funzionale si traduce nelle scelte che permettono di portare a casa questo obiettivo
Processi direzionali	Prendere decisioni	Si occupa di prendere decisioni strategici e prende decisioni strategiche

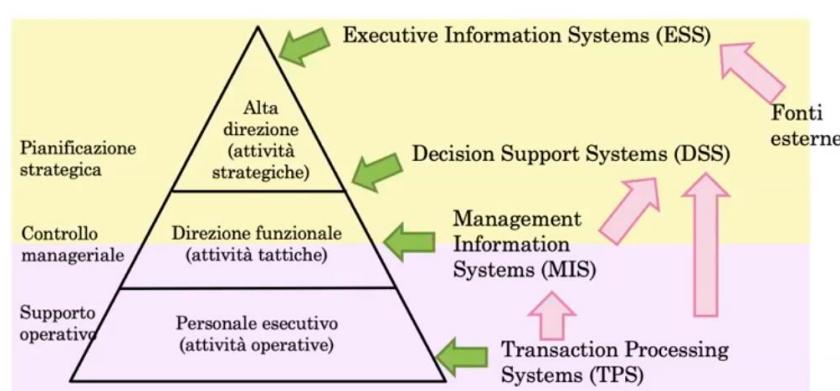
Tutti questi, comunque, restano processi aziendali; quindi sono processi che puntano al raggiungimento di un obiettivo.



Per quanto riguarda il livello dei processi gestionali, c'è una sovrapposizione fra i due tipi; è un po' a cavallo, in base al processo specifico che stiamo mettendo in atto; c'è una parte che prende decisioni più vicine al management e un'altra parte più vicina all'esecutivo.

SIA lungo la piramide

- TPS Transaction Processing Systems:** sistemi DB tradizionale
- MIS Management Information Systems:** produzione di report periodici. Questo tipo di sistema usa le informazioni che stanno nel DB, aka informazioni riassuntive di routine che potrebbero tranquillamente essere fatte in SQL.
- DSS Decision Support Systems:** si collocano fra il livello medio e l'alto della piramide. Questo sistema informativo prende i dati dai due tipi di SI sottostanti, e prende informazioni che non per forza sono di routine; ad esempio per accorgersi che un mese c'è stato meno fatturato del solito. È orientato ad un utente che difficilmente avrà competenze informatiche; sarà un manager!!



Non scrivono SQL ma hanno una UI grafica che permette di concatenare analisi.

Usa i dati interni in maniera interattiva, e permette a chi deve prendere decisioni – anche non di routine – con cognizione di causa.

Sono importanti, perché la quantità di informazioni nel DB è enorme (non solo attuali, ma anche tutto lo storico!)

Permettono anche di simulare ipotesi.

- **ESS Executive Information Systems o Cruscotti Aziendali:** usano solo fonti esterne, ovvero pescano dal sistema circostante e riescono a dare una visione di cosa sta succedendo nel mercato. Anche questi permettono di prendere decisioni strategiche, in base ad esempio a esigenze di mercato o competitors che fanno cose. Servono alle decisioni **strategiche**.

Classificazione dei SIA lungo le aree gestionali dell'impresa

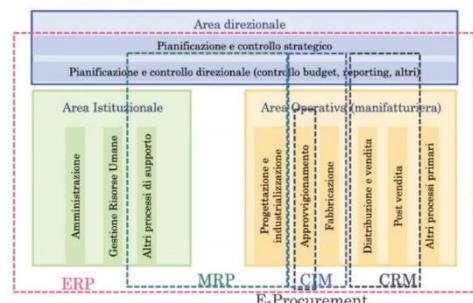
Classificazione delle aree funzionali

Qui non abbiamo una piramide, ma c'è corrispondenza:

- **Area direzionale** = metà **alta** della piramide
- **Area istituzionale e operativa** = metà **inferiore** della piramide.

Qui, però, la metà inferiore è divisa in

- Area **operativa** – ovvero quella in cui le funzionalità sono volte alla **produzione**, e che corrisponde al livello basso
- Area **istituzionale** – che gestisce la **parte burocratica** e corrisponde al livello basso e al livello medio.



Portafoglio applicativo

Def Portafoglio applicativo

È un insieme di applicazioni utili in azienda, dunque sarebbero i nostri SI. Per il nostro punto di vista, sono i sistemi già visti sopra.

Rispetto alle tre aree, abbiamo **un portafoglio per ciascuna delle tre aree**.

! Al giorno d'oggi l'idea di portafoglio applicativo lascia il tempo che trova: ad esempio con i sistemi di ERP copre quasi tutte le funzionalità di cui ho bisogno, perché i sistemi di ERP sono quelli che si sono allargati sempre più a coprire tutti gli aspetti dell'azienda. !

Def Portafoglio direzionale

Applicazioni informatiche a supporto dei cicli di **pianificazione strategica** e di pianificazione e controllo delle risorse aziendali. Ricalca quello che si è detto prima: cerca di essere di supporto alle decisioni. Sono i sistemi informazionali

Def Portafoglio istituzionale

Applicazioni informatiche per i processi di **supporto all'amministrazione**, alla **gestione delle risorse umane** e alla **contabilità**. Qui ci sono gli applicativi a supporto dei processi istituzionali:

- **Adempimenti di legge:** contabilità, retribuzione, previdenza
- **Amministrazione di infrastrutture:** immbii
- **Amministrazione di fattori di produzione:** personale, scorte, denaro, impianti

È molto orizzontale, ovvero **indipendente dal settore industriale**.

È il primo ad essere **informatizzato** e reso disponibile per le aziende, grazie alle sue caratteristiche: grandi quantità di dati, ma da elaborare in maniera procedurale, periodica e ripetitiva. Di conseguenza, la stessa attività viene fatta in meno tempo e con meno errori. Dagli anni 90 questi sistemi diventano parte integrante degli ERP.

Def Portafoglio operativo

Applicazioni informatiche per i processi primari dell'azienda.

È di tipo **verticale**: è specifico per ciascun settore. A seconda del settore abbiamo un'attrattiva informatica variabile.

Informatizzazione del portafoglio applicativo

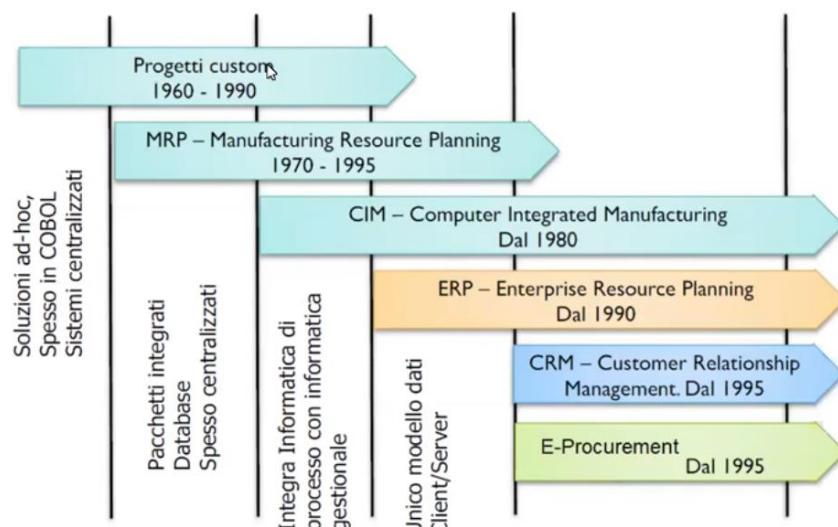
- All'inizio c'erano solo soluzioni ad hoc, spesso in cobol.
- Dagli anni '70, con gli RDBMS, iniziano dei pacchetti integrati per la pianificazione delle risorse, soprattutto in area operativa
- Si allarga, con l'informatica gestionale, dal 1980: i CIM. Sono pacchetti orientati all'automazione delle linee, e le tecnologie si basano su reti di fabbrica e elaborazione in tempo reale.
- Dagli anni 90 iniziano anche gli ERP; che coprono tutte le funzionalità con un'organizzazione componibile. A differenza delle soluzioni precedenti, tutti i pacchetti fanno riferimento a un unico modello di dati, e quindi ho una sola DB per tutta l'azienda.

È una soluzione integrata, partita come istituzionale e poi allargandosi a coprire tutti i portafogli.

Qui finalmente si passa al client server, mentre prima era tutto centralizzato.

Usano un unico modello dei dati, quindi posso integrare tutti i processi.

- Dal 95 partono i CRM, che sono volti alla gestione del cliente. È molto importante (es. compagnie telefoniche). CRM in realtà ora fa parte di ERP. Permettono di automatizzare distribuzione, vendita e post vendita.
Sono pacchetti integrati e architetture client server che usano il web.
- E-procurement sono piazze virtuali dalle quali posso avere materia prima.



Catena del valore di Porter

Un'azienda manifatturiera può essere descritta con la catena del valore di porter, che descrive la serie di operazioni che faccio per raggiungere l'obiettivo.



SISTEMI ERP – ENTERPRISE RESOURCE PLANNING

(fyi ERP non sta per “erotic roleplay”, come qualcuno suggerisce. 😊)

L3 - 19/10/2020

[Questo video ha un ottimo recappino sulla storia](#) (first 2 sections)

Sono sistemi informativi usati in ambito aziendale andando a sostituire i sistemi precedente e permettendo finalmente all’azienda di avere un solo DB condiviso. Gli ERP avevano iniziato a sostituire l’area istituzionale, ma ormai hanno inglobato tutto. Dunque, chi oggi adotta un ERP può scegliere quali funzionalità acquistare.

Def: *ERP- Enterprise Resource Planning*

ERP indica una suite di **moduli applicativi** che supportano l’intera gamma dei processi di un’azienda.

- **Resource** = persone, materiali, impianti, capitale
- **Planning** = tempi, luoghi, quantità.

Di base è un **sistema informativo operazionale integrato**. Parliamo di operazionale perché è partito come tale; i pacchetti per la business intelligence e supporto alle decisioni gli permettono di diventare anche informazionale. Per accontentare le aziende nelle loro particolarità, anche se il sistema ERP si pone come sistema informativo universale poi ha dei singoli pacchetti acquistabili. Parte dalle aziende manifatturiere, e poi viene applicato anche a sistemi assicurativi, bancari, di servizi...

La caratteristica principale è che **si basa su un’unica base di dati**, ovvero che lo schema è unico e tutti i dati sono gestiti allo stesso modo e in maniera coerente. Essendo un sistema modulare, le sue procedure fanno parte dello stesso SIA e quindi è facilissimo farle interagire fra loro.

L’obiettivo primario è la **gestione ottimale di tutte le risorse utilizzate nei sistemi aziendali**.

Le soluzioni sul mercato sono SAP, Oracle, Peoplesoft...

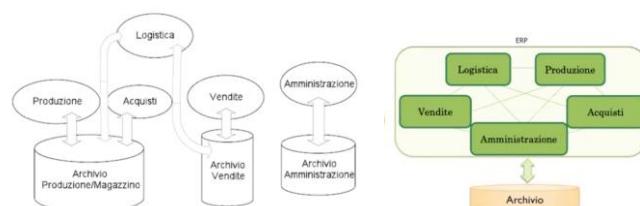
Caratteristiche

Le caratteristiche sono:

- **Sistemi modulari, con pacchetti specifici:** gestione del personale, finanziaria, produzione, ciclo di vendita...
 - Vendita e post-vendita ci ricorderebbero i CRM! Il vantaggio dell’averne un pacchetto CRM è che può sfruttare la base di dati che già ha innestandoci più funzionalità.
Questo implica che ho una base di dati minima e con garanzie sulla coerenza, non servendo la ridondanza
- Interagiscono tramite **interfacce**, e questo semplifica l’estensione
- **Altamente configurabili**
- Richiedono un **periodo di avviamento**, per esempio per l’istruzione alle risorse umane. L’acquisto di un sistema ERP induce anche la redifinizione e reingegnerizzazione dei processi aziendali. Questo è costoso, ma porta l’azienda ad un miglioramento delle prestazioni.
- Sono **preconfigurati**, quindi il margine di personalizzazione è relativo.

Confronto con legacy systems / isole informative

Ancora, in alcune aziende con informatizzazione parziale, ci troviamo ad avere sistemi di questo tipo.



Ovviamente **fa cacher**; abbiamo problemi di ridondanza, coerenza, correttezza. Il sistema di ERP invece ha un unico archivio con il sistema che vi interagisce.

Moduli

Amministrazione	<p>Gestisce i processi di contabilità generale È fortemente soggetto a vincoli di legge.</p> <ul style="list-style-type: none"> • Supporto alle attività amministrative, come ad esempio fatture, pagamenti, incassi • Produzione di informazioni di sintesi sull'andamento aziendale.
Logistica	<p>Gestione dei materiali Gestisce le materie prime, i materiali di funzionamento e i prodotti finiti. Il pacchetto si concentra soprattutto sulla gestione interna – per esempio, nelle versioni attuali abbiamo tutti gli aspetti di gestione delle informazioni spaziali (es. carico/scarico). Le informazioni gestite sono quelle necessarie alla gestione dei magazzini e dei vari materiali e prodotti: anagrafica dei fornitori, dove sono le merci, come sono organizzate, la movimentazione... Questo permette un'ottimizzazione dello spazio e dei tempi.</p>
Vendite	<p>Gestione dell'interazione con il cliente Spostano il CRM nell'ERP. Permettono anche analisi per la valutazione della collocazione dell'azienda sul mercato; è un'analisi strategica, che fa parte dei cruscotti aziendali. L'ERP cerca di immettere tutte le funzionalità necessarie nello stesso pacchetto.</p> <ul style="list-style-type: none"> • Ricezione ordini: spesso, se un cliente non è noto, prima ci si informa se è un cliente che paga... • Elaborazione ordini: attivare la priorità, coordinare i flussi informativi e valutare la priorità • Evasione ordini e emissione di documenti di trasporto: interagisce col magazzino • Fatturazione: interagisce con la contabilità <p>Le interazioni avvengono in maniera molto naturale grazie all'unico DB</p>
Acquisti	<p>Gestione dell'interazione con i fornitori Riguarda l'approvvigionamento di materiali, quindi ha uno stretto contatto con la logistica.</p> <ul style="list-style-type: none"> • Condizioni commerciali • Rapporto con i fornitori <p>Insomma è quello che farebbe un E-procurement, ma in maniera integrata.</p>
Produzione	Dipende dal settore.

Il leader di mercato: SAP

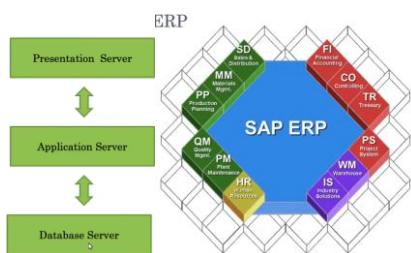
È una multinazionale europea, leader del settore degli ERP e in generale delle soluzioni enterprise. È del 72.

SAP = sistemi applicazioni e prodotti dell'elaborazione dati.

Negli anni 80 nasce **ABAP**, un linguaggio simile a cobol/fortran, e inizia a usarlo per la gestione dei processi informativi che permetteva la costruzione di rapporti informativi in maniera molto veloce.

Con l'avvento del relazionale permette di interfacciarsi con SQL. Negli anni 2000, le nuove tecnologie integrate sono la rete e la business intelligenze; poiché vi si basa SAP, è uno dei linguaggi più usati.

Architettura dei sistemi ERP



- **SD:** gestione commerciale
- **MM:** gestione materiali
- **PP:** pianificazione della produzione
- **QM:** gestione della qualità
- **PM:** manutenzione impianti
- **HR:** risorse umane
- **FI:** contabilità
- **CO:** controllo gestione
- **AM:** gestione cespiti
- **PS:** gestione commesse
- **WF:** flussi di lavoro; gestione processi che ci racconterà
- **IS:** industry solution specifici per industria

SISTEMI CRM - CLIENT RELATIONS MANAGEMENT

[Nice videino dallo internet](#)

Def: Processo CRM

E' un processo **integrato e strutturato di gestione della relazione con la clientela**, con lo scopo di **costruire relazioni personalizzate** di lungo periodo, capaci di aumentare la soddisfazione dei clienti e aumentare il valore della impresa per il cliente e del cliente per l'impresa.

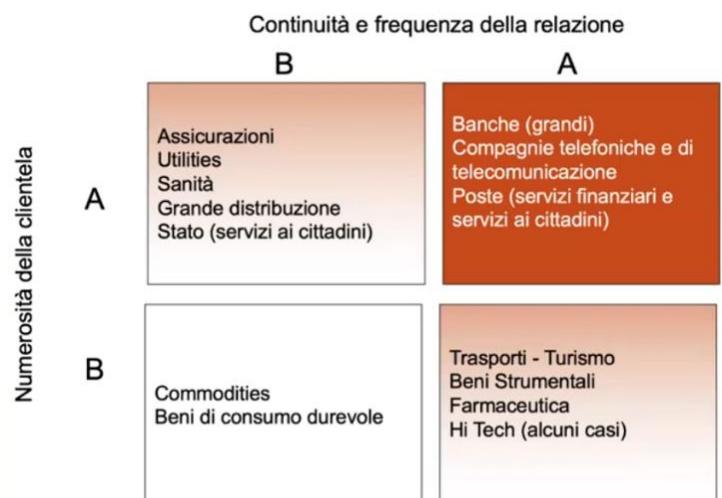
Aka, voglio fidelizzare il cliente.

Il cliente è l'elemento centrale della strategia commerciale: o rendo felice il cliente, o non campo 😊

Per fidalizzarlo, quindi, si attuano diverse strategie:

- + **Semplificazione dell'accesso**
- + **Offerte personalizzate**
- + **Servizio completo**
- + **Analisi dei reclami**

Ad oggi, ci riesco se ho un buon sistema CRM; le strategie CRM sono infatti senza sistemi CRM, che permettono di analizzare tutte le info e capire quale sia la proposta che soddisfa meglio il cliente. Chi ha più bisogno di un CRM? Chi ha più bisogno di info!



Schema

Ci concentriamo sul fatto che:

- + L'informazione passa da **tanti canali diversi**: live, telefono, web...
- + **Grande mole di dati**

Tanto più l'azienda ha un numero di clienti elevato e una frequenza elevata, tanto più sfrutterà questi canali; a maggior ragione dato che ciascun canale potrebbe avere preferenze diverse.

Uso

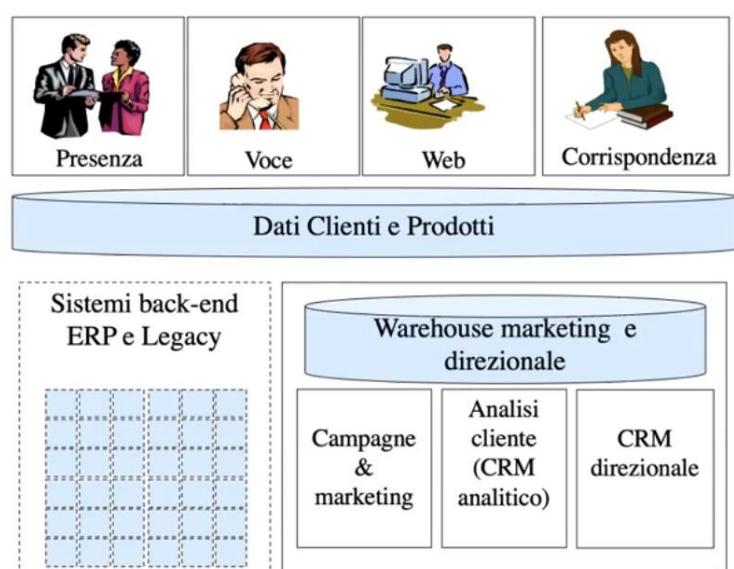
Tutti questi canali servono sia per **acquisire clienti**, sia per il **supporto**.

Questo tipo di interazione produce una **quantità esagerata di informazione**, che va travasata in sistemi informativi specifici (data warehouse). Il vantaggio è che così posso salvare separatamente tutti gli storici, per fare analisi.

Queste analisi potrebbero servire per agevolare il singolo cliente e fidelizzarlo con offerte specializzate, o creare offerte in base al mercato, o prendere decisioni strategiche.

Le aziende che più sfruttano queste tecnologie sono le aziende di utility.

Sono quelle che hanno più bisogno di un DW, perché è "scontato" che la fornitura del servizio in sé sia giusta; quello che conta davvero è in base al quale poi si valuta,

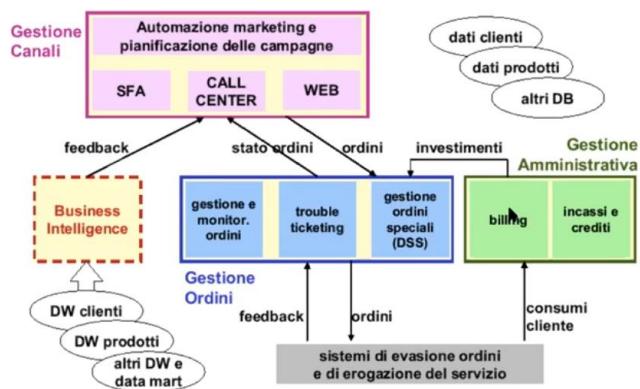


sono i servizi al contorno – come ad esempio il supporto.

Di conseguenza, è essenziale avere un buon CRM per:	Vantaggi
<ul style="list-style-type: none"> + Attività di marketing + Raccolta di ordini e fatturazione + Assistenza post-vendita. 	<ul style="list-style-type: none"> + Abbattimento di costi + Proposte di personalizzazioni facili facili

Fattori che spingono le aziende ad avere CRM:	Funzioni:
<ul style="list-style-type: none"> + Clientela massiccia + Rapporto continuativo + Concorrenza crescente + Canali virtuali che permettono un bacino più ampio ma anche più concorrenza 	<ul style="list-style-type: none"> + Sistemi di gestione degli ordini + Automazion delle forze di vendita; es. i venditori di medicinali fanno tutto da soli in loco + Sistemi di assistenza telefonica: call center + app + chatbot...

Moduli applicativi: esempio su azienda telefonica



Buisness intelligence

Il sistema CRM usa strumenti di **BI**, che sono applicazioni che **estraggono info dai sistemi di supporto alla produzione per fornire supporto alle decisioni**. Sono sistemi di tipo informazionale. Permettono di passare dai dati puri a rappresentazioni sintetiche e analitiche.

Rispondono a richieste complesse: siccome le operazioni decisionali sono fatte a livello alto, solitamente sono fatte da personale che non ha competenza informatica, e usa lo strumento informatico.

Focus sulle interrogazioni e non sugli aggiornamenti.

- + **DBMS** = supporto operativo e transizioni = OLTP
- + **BI** = supporto decisionale = OLAP

CRM e società di servizi

Il portafoglio applicativo cambia: la parte dell'azienda fatturiera cambia dato che non produco più un bene, e cambiano i processi aziendali. Qui, i processi aziendali sono ad alta intensità informativa e la maggior parte dei processi si occupa di informazione. Di conseguenza, **la catena del valore di Porter cambia**:

Di conseguenza, il CRM diventa fondamentale. In questa catena del valore ho:

- + **Predisposizione delle condizioni produttive**: mi metto in condizione di poter offrire il servizio
- + **Back office**: attività svolte in seguito a un ordine ma non in presenza del cliente, ad esempio, nel caso dell'assicurazione, la valutazione del rischio
- + **Front office**: erogazione del servizio di fronte al cietne (banca...)



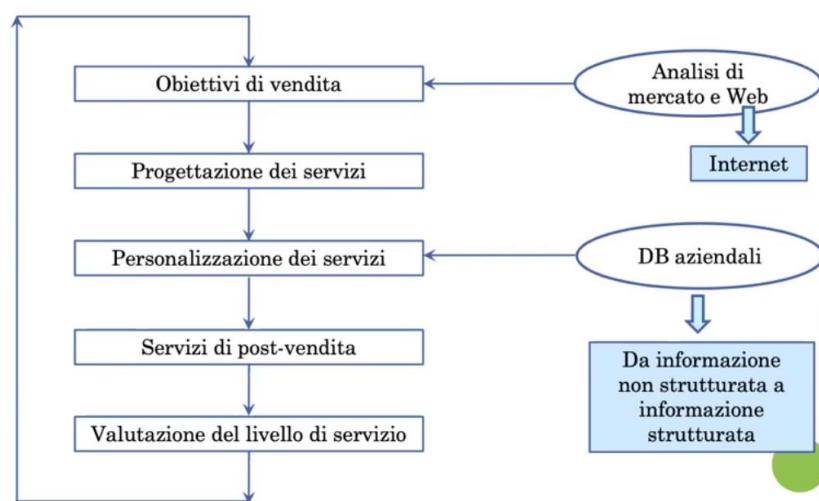
- + Pocacciamento della clientela: porto alla stipulazione dei contratti di servizio o all'ordine da parte dei clienti-

Il flusso degli ordini invece è analogo, manca solo la produzione materiale. È essenziale la gestione della conoscenza: es. se accendiamo un mutuo è importante gestire un flusso informativo relativo al cliente, perché ci vogliono delle garanzie e c'è della responsabilità importante.

Dunque, **la conoscenza è essenziale** perché fa la differenza fra dare i soldi a chi me li restituisce o meno. E quindi è un vantaggio competitivo. In questo contesto, spesso l'info è non strutturata, viene da posti diversi e ci vuole l'esperienza della persona che valuta... non è solo una formula matematica.

Knowledge holder

I CRM dunque devono prendere tutta questa informazione e renderla accessibile al knowledge worker, ovvero all'impiegato che svolge il lavoro intellettuale di natura decisionale. E' colui che partendo dalle info e dalla sua esperienza riesce a stipulare contatti vantaggiosi.



INGEGNERIA E TRASFORMAZIONE DEI PROCESSI GESTIONALI

L4 - 21/10/2020

I processi aziendali possono essere classificati e reingegnerizzati.

Processi

Def: Processi aziendali

I processi aziendali sono uno **sforzo coordinato per raggiungere un obiettivo**, e rappresentano il **modo di operare** di un'azienda.

L'avvento delle tecnologie informatiche permette di **ottimizzare le attività del processo stesso**, quindi diventa importantissimo coordinare e progettare sia la parte dei processi che la parte dei sistemi informatici.

Quando parliamo di gestione dei processi aziendali stiamo parlando di **sistemi orientati ai processi**; quando parliamo di basi di dati parliamo di sistemi informati ai dati. In questo corso impariamo a progettare i sistemi che organizzano le attività e i processi.

A livello di processi, analogamente al diagramma concettuale dei DB, avrà una rappresentazione grafica ad alto livello che permette di gestire i processi.

Anche in ambito processi l'innovazione tecnologica ha delle ricadute: nel BPR (Business Process Reengineering) abbiamo bisogno di correlare un'innovazione radicale di processi e organizzazione aziendale con un massiccio uso di informatica.

Classificazione ad alto livello dei processi

- + **Materiali:** flusso di materiali e attività
- + **Informativi:** flusso di informazioni; es. banche e assicurazioni
- + **Buisness process:** insieme strutturato di attività, finalizzato alla realizzazione di un ben definito risultato di interesse per l'organizzazione.

Buisness process (BP) / Processo aziendale

Def: Buisness process

In generale, un processo aziendale è formato da attività, realizzate come processi materiali o processi informativi, che sono collegati fra loro nel tempo e nello spazio (=coordinate) e che sono svolte dalle risorse di un'azienda.

Partendo da input definiti, le attività producono un output utilizzato dai clienti o da altre parti dell'azienda. I processi sono flussi di attività interfunzionali che concatenano le diverse competenze presenti in azienda:

Competenze **commerciali**
(marketing)

Competenze **tecniche**
(produzione)

Competenze **generali**
(approvvigionamento)

Formalmente

Un buisness process è definito come una tupla

$$BP = (A, I, O, C)$$

- + A = *attività*, ovvero serie di operazioni fisiche o di decisioni manageriali eseguite in maniera coordinata
- + I = *input*, ovvero le materie prime (informazioni, se il processo è decisionale) e le risorse aziendali
- + O = *output*, ovvero oggetti fisici, beni immateriali e servizi.
- + C = *clienti*, o più in generale i destinatari dell'output del processo.

Catena del valore di porter

È l'azienda vista secondo la prospettiva di processo. L'azienda è una successione di attività, finalizzate a produrre valore per il cliente. Il valore è misurato dal prezzo che il cliente è disposto a pagare.

Un esempio per le aziende manifatturiere:



Posso modellare le macroattività in processi. Ognuna delle macroattività sarà organizzata in processi.

Trasformazioni dei processi

Posso sempre decidere di reingegnerizzare il processo, ovvero passare da una forma di base a un modello diverso. Lo scopo è cercare di ottimizzare.

Rispetto alle strategie di reingegnerizzazione, possiamo vedere i processi come divisi nelle seguenti categorie:



Strategia	Benefici	How
Buy side Ha lo scopo di trasformare i processi che hanno a che fare con l'approvvigionamento delle materie prime, e con l'interazione con il mondo dei fornitori.	Ridurre i costi delle materie prime!	Sistemi di e-procurement
In-side Mira alla trasformazione dei processi interni all'azienda (es. produzione degli oggetti).	Ridurre i costi delle materie prime!	ERP
Sell-side Vuole trasformare i servizi di vendita, marketing, distribuzione e post vendita.	Ridurre i costi di transazione e aumentare il valore percepito dal cliente	CRM

Classificazione dei processi

- + **Processi intersettoriali:** processi generici che descrivono le pratiche di molteplici settori, aka cose che non cambiano in base al settore.
es. gestione materie, acquisto materie è uguale sia nelle banche che negli ospedali.
- + **Processi settoriali:** processi che si distinguono ne i vari settori.
es. settore sanitario (processi di ricovero, cura) != attivazione di un conto corrente
 - **Processi aziendali:** processi di una specifica azienda o di una sua parte; sono una specializzazione dei processi settoriali.
- + **Processi normativi e best practice:** sono i processi di riferimento. Derivano dal confronto tra processi diversi, e descrivono come sono o dovrebbero essere nelle migliori aziende del settore.

Scomposizione dei processi

Parlando di processi, posso dettagliare livelli di approfondimento successivi:

Macroprocesso	Processo	Fase	Attività	Operazione
primo livello; ad esempio, la catena del valore di Porter	illustrano il modo di operare dell'azienda	illustrano il modo in cui un processo è implementato. Non sono sempre definiti: sono utili se ho delle attività e parti di processo che voglio poter riutilizzare	livello minimo di analisi normalmente adottato	quasi mai usate; passi elementari

Griglia metodologica

È lo strumento usato per capire e studiare cosa cambiare di un processo; prevede di incrociare le **fasi di analisi** con le **variabili di analisi**

In ogni fase devo analizzare tutte le variabili organizzative descritte; in particolare, l'esempio per il bottom up:

Variabili	Fasi		
	Rilevazione della situazione esistente	Confronto e diagnosi	Riprogettazione
Flussi di attività	Macroprocessi. Processi. Attività. Fasi. Prestazioni.	Benchmarking con best practice. Analisi delle possibilità di integrazione.	Personalizzazione delle best practice. Prototipazione e simulazione.
Struttura aziendale	Struttura organizzativa. Mecanismi operativi. Stile di direzione.	Benchmarking. Analisi del ruolo delle diverse funzioni nell'integrazione dei processi.	Struttura organizzativa (cambiamento ruoli e/o nuova divisione delle attività). Mecanismi operativi e stile di direzione.
Risorse umane	Esame disponibilità al cambiamento. Inventario di professionalità e capacità esistenti.	Confronto con concorrenti o best in class. Diagnosi.	Definizione del nuovo sistema di valori. Predisposizione programmi di training e di assunzioni.
Misurazione delle prestazioni	Identificazione prestazioni strategiche.	Confronto con concorrenti o best in class. Diagnosi.	Definizione dei nuovi obiettivi e nuovi indici di prestazione.

Ci sono due tipi di approcci:

Bottom up	Top-down
<p>Volto al miglioramento del processo esistente.</p> <ol style="list-style-type: none"> 1. Rilevazione della situazione esistente 2. Confronto con altre imprese e diagnosi dei problemi 3. Ridisegno dei processi 	<p>Volto al disegno del processo dall'inizio, con criteri di ottimizzazione noti.</p>

La trasformazione efficace integra sia l'innovazione tecnologica che l'innovazione organizzativa.

A. Variabili organizzative

Le variabili organizzative sono quella cosa che posso modificare per aggiustare il processo.

1. Flusso delle attività

Sequenza delle attività attraverso cui è svolto il processo. A parità di fattori, il flusso di attività determina la durata del processo.

È la variabile che ci interessa di più a livello informatico. Determina:

- + **Durata del processo:** numero di attività previste
- + **Livello di servizio:** a seconda della flessibilità posso dare un servizio migliore (es. offre più modi per contattare)
- + **Qualità del prodotto**

Elementi da descrivere:

Natura del flusso	Attività	Sequenza attività	Attori	Eventi	Oggetti
<ul style="list-style-type: none"> + Fisico + Informativo + Controllo 	<ul style="list-style-type: none"> + Tipo (fisica/informativa) + Durata + Volumi + Tecnologie 	<ul style="list-style-type: none"> + Alternative 	<ul style="list-style-type: none"> + Tipologia + Azioni svolte sul flusso 	<ul style="list-style-type: none"> + Tipologia + Temporizzazione + Conseguenze 	<ul style="list-style-type: none"> + Natura + Profilo temporale

La modellazione può essere:

Schemi di sequenza: Indicano solo la sequenza di attività che formano un processo.

Schemi più ricchi Indicano sia la sequenza, sia gli attori coinvolti, gli eventi che determinano le attività e le informazioni utilizzate

2. Organizzazione

(=organizzazione gerarchica) divisione operativa del lavoro nell'ambito del processo. A parità di fattori, determina la parcellizzazione o l'accorpamento dell'attività.

Parliamo di in che modo userò le mie risorse per eseguire le attività del processo.

Qui è importante la descrizione dei ruoli e delle funzioni associate.

- + **Organigramma:** gerarchia delle responsabilità e delle autorità di una organizzazione a vari livelli di dettagli.
- + **Tabelle delle proprietà:** indicano le caratteristiche fondamentali delle strutture; es. scopo di una struttura, chi ci sta dentro, che compiti ha, che processi svolge.
- + **LRC (Linear responsibility charting):** sono tabelle che rappresentano le responsabilità organizzative dei vari settori.

- + D = decide, autorizza, ratifica
- + E = esegue
- + A = partecipa a tempo parziali, fornisce assistenza
- + I = sistematicamente informato

	Filiale vendita	Direzione commerciale	Direzione distribuzione	Magazzino Prodotti finiti	Spedizioni	Contabilità clienti
Ricezione ordini	E	D	I	I		A
Evasione degli ordini da magazzino	I		D	E	I	
Spedizione			I	A	E	I
Fatturazione			I			E

3. Risorse umane

L'adeguatezza delle competenze è condizione fonamentale per la trasformazione del processo a seguito di una innovazione tecnica.

Determinano la differenza tra il risultato effettivo di un processo e il massimo risultato teoricamente possibile in una data configurazione. Sono essenziali per il successo di un progetto correttamente ingegnerizzato.

L'innovazione tecnologica rende ruoli superati (perché lo fa in automatico il sistema), e causa la necessità di avere a disposizione figure professionali specializzate.

4. Misurazione delle prestazioni

Questa parte permette di valutare in che modo sono eseguiti i miei processi.

Unisco il controllo a un sistema di promozione o demozione.

B. Fasi di analisi

1. Identificazione della situazione esistente

- a. Identificazione dei macroprocessi
- b. Dettaglio dei processi
- c. Incrocio processi e unità organizzative
- d. Valuazione del process

2. Confronto con altre imprese e diagnosi dei problemi

- a. Confronto quantitativo e parametrizzazione; devo decidere a priori su cosa concentrarmi, dato che non posso ottimizzare tutto (es. località/tempo)
- b. Confronto qualitativo

3. Ridefinizione dei processi.

- a. Spesso si ridefinisce la *vision*, ovvero la rappresentazione sintetica degli elementi fondamentali della soluzione. La nuova vision, ad esempio, potrebbe essere “open space over spazi privati”.
- b. Analisi del cambiamento che mi aspetto di ottenere e in che tempo.

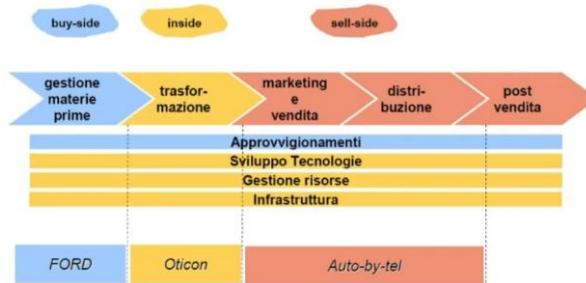


La trasformazione dei processi, dunque, consiste nell'incrociare le fasi per capire come ottimizzare i processi.

TRASFORMAZIONE DEI PROCESSI: CASI REALI

L5 – 26/10/2020

Posizionamento dei casi:



1. Trasformazione della supply chain (buy-side): caso Ford

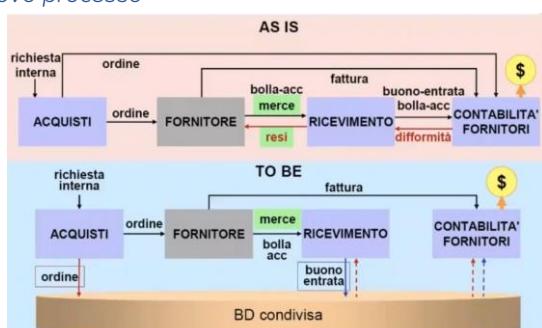
Nei primi anni 80 la ford stava cercando di ridurre le spese generali e amministrative; si riteneva che con l'introduzione dei computer si sarebbero potuti ridurre gli addetti del 20% rispetto ai loro 500 di partenza.

Il processo stava per partire, quando Ford acquistò parte di Mazda, che aveva ben 5 addetti! 😊 **Ford scema**. Allora ford decide di riesaminare il processo relativo alla contabilità.

Processo esistente

1. Arriva ordine da un reparto o stabilimento
2. Ufficio acquisti invia un ordine al fornitore e una copia alla contabilità
3. Alla consegna dei materiali l'addetto al ricevimento riempie un modulo da mandare alla contabilità
4. Il fornitore invia la fattura alla contabilità
5. La contabilità lavora su tre documenti: ordine di acquisto, modulo di ricevimento e fattura. Se sono appositi si paga. Nel 20% non erano d'accordo, e buona parte del tempo era data al gestire la non concordanza dei documenti.

Nuovo processo



Il nuovo processo si basa su una DB, senza bisogno di introdurre nuovi documenti e con informazioni automaticamente condivise e corrette.

Griglia metodologica

Leve	Processo esistente	Vision del nuovo processo	Gap e azioni da intraprendere
Flusso attività	Attività frammentate. Funzionamento <i>batch</i> . Isole di automazione. Informazione non condivisa.	Flusso unico. Funzionamento in tempo reale. BD condivisa con accesso immediato.	Riprogettazione flusso attività. Rifacimento SI aziendale.
Struttura Organizzativa	Struttura burocratica.	Struttura funzionale basata sul processo.	Revisione micro e macro struttura.
Risorse umane	Parcellizzazione. Addestramento sulle mansioni.	Separazione fra gestione di routine ed eccezioni.	Riaddestramento. Allargamento mansioni.
Misura prestazioni	Micro-efficienza.	Efficienza del processo.	Nuovo sistema di controllo.

2. Trasformazione di funzionamento (inside): caso Oticon

Oticon è un produttore di apparecchi acustici. Questo caso ci mostra una trasformazione completa del funzionamento interno, in risposta ad una crisi accaduta negli anni 80.

Nasce come azienda a conduzione familiare fondata ai primi del 900, partita da Hans Demant e poi passata al figlio.

- Negli anni 40, a causa della guerra e delle relative difficoltà nei trasporti il figlio William comincia a costruire apparecchi acustici.
- Negli anni 70 nasce una rete di società che permette un allargamento.
- Negli anni 80, Oticon entra in difficoltà perché il mercato inizia a prediligere apparecchi interni anziché esterni; arriva il primo bilancio passivo.

Nel 1988, quindi c'è un cambio di presidenza e un inizio della politica di risanamento. La nuova visione è:

- + **Diventare un'organizzazione che fornisca anche un servizio**, e non solo il prodotto fisico; tagliando fuori le cliniche specializzate.
- + **Organizzazione non più basata sulla gerarchia, ma su progetti** proposti ai quali ciascun dipendente può partecipare a seconda delle sue competenze.

Questo si realizza con un processo di informatizzazione e reingegnerizzazione:

- + **Assunzione di 2 manager per gestire i cambiamenti relativi all'informatica**
- + **Coinvolgimento di 2 società per la realizzazione dei sistemi informatici.**

Griglia metodologica

Tutto questo fa sì che già nel 92 ci siano dei risultati incoraggianti:

- + Time market ridotto di 6 mesi
- + Allocazione più rapida ed efficiente dei dipendenti
- + Maggiore soddisfazione personale
- + Progetti 9x rispetto al 1991.

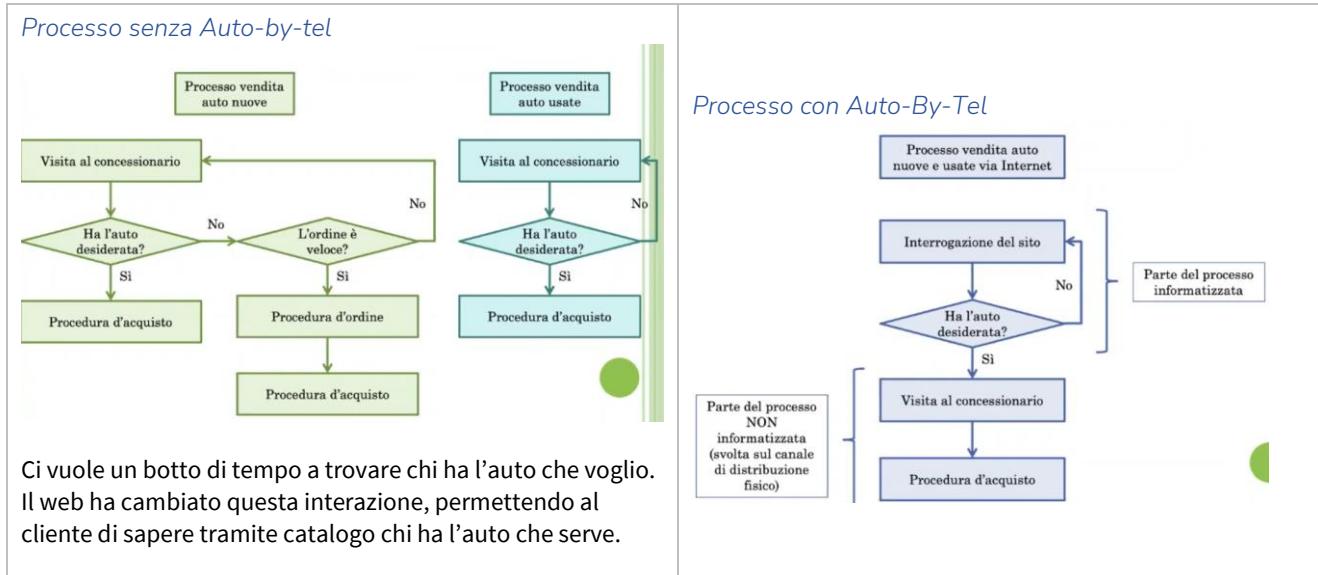
Leve	Processo esistente	Vision del nuovo processo	Gap e azioni da intraprendere
Flusso attività	Attività sequenziali. Isole di automazione. Archivi cartacei.	Attività parallele. Integrazione informazioni. Archivi elettronici.	Revisione flusso attività. Rifacimento sistemi per avere archivi condivisi accessibili da PC.
Struttura Organizzativa	Funzionale.	<i>Team-based.</i>	Revisione micro e macro struttura.
Risorse umane	Forte senso della gerarchia. Inamovibilità.	Cultura egualitaria. Posizione dinamica.	<i>Change management.</i> <i>Eliminazione status symbol.</i> <i>Open space</i> con postazioni di lavoro tutte uguali.
Misura prestazioni	Accuratezza. Micro-efficienza.	<i>Time to market.</i> Lavoro di gruppo.	Cambiare sistema di riferimento.

3. Trasformazione della gestione del cliente (sell-side): caso Auto-By-Tel.

È un esempio tipico di come il CRM trasforma completamente il processo di ricerca del bene, abbattendo i costi di transazione costenuti dal consumatore.

Auto-by-tel è un'azienda californiana fondata nel 1994. Nel 95 offre servizi via web, con accesso a un catalogo di auto; negli anni successivi aggiunge auto usate e assicurazione.

Auto-by-tel, quindi gestisce macchine nuove e usate. Senza l'intermediazione di auto-by-tel e senza internet, questi due processi sarebbero gestiti dai concessionari.



Griglia metodologica

Leve	Processo esistente	Vision del nuovo processo	Gap e azioni da intraprendere
Flusso attività	Attività svolte dal cliente interamente su canale fisico. Informazione non condivisa fra i concessionari.	Processo informatizzato. Risposta in tempo reale via Internet. BD condivisa con accesso via Internet.	Intermediazione flusso attività. Creazione da zero SI aziendale.
Struttura Organizzativa	Struttura distributiva delle auto decentralizzata.	Struttura a rete.	Creazione rete.
Risorse umane		Da acquisire.	Piano di assunzione e di inserimento.
Misura prestazioni	Numero auto vendute.	Efficienza del processo. Tempi di vendita.	Nuovo sistema monitoraggio (fornito da Auto-By-Tel ai concessionari appartenenti alla rete).

MODELLAZIONE DEI PROCESSI: BPMN

Fin'ora abbiamo parlato di processi aziendali composti da attività, organizzate per raggiungere un obiettivo. Abbiamo detto anche che il processo aziendale deve essere modellato così come già modelliamo l'informazione nei DB.

La catena del valore di Porter è già una modellazione di processi.

Quando penso alla modellazione dei DB, ho uno standard di riferimento: ER, UML. La stessa cosa mi serve in ambito processi; necessito di una notazione per la rappresentazione dei processi.

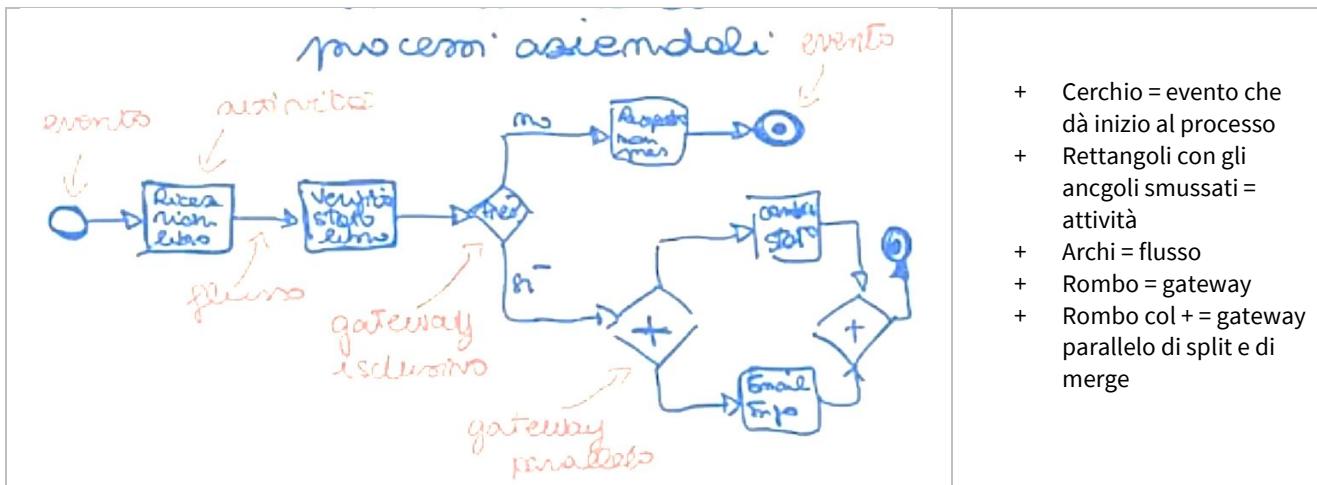
Def: Notazione grafica

Una notazione grafica (come BPMN) permette di differenziare i simboli che possono essere utilizzati per rappresentare i processi da un punto di vista grafico. Su questi simboli viene definito anche:

- Significato (semantica)
- Come posso combinarli

In questo contesto, quindi, una notazione è uno standard grafico per la descrizione di processi aziendali.

In BPMN:



PMS

Abbiamo anche già introdotto i PMS, **project management systems**, che sono **sistemi software che permettono di gestire i processi**. Sono l'altro lato della medaglia rispetto ai DBMS.

Sistemi che permettono di gestire i processi; come per le DB, gestire significa modellare, eseguire, monitorare.

A fronte di un modello di processo concettuale, poi posso arrivare a un processo eseguibile. I linguaggi per la descrizione delle operazioni sono

- + **XPDL** (basato su XML) Process Definition Language, permette di descrivere con una notazione
- + **BPEL** Business Process Execution Language:

Esiste BPMN 1.2, che però ha il problema di non essere mappabile fra BPEL. Questo è un bel problema se voglio automatizzare! Noi studieremo BPMN 2.0, che è direttamente e automaticamente traducibile. Si presta a diversi tipi di livelli di dettaglio.

Grammatica di BPMN

Processo strutturato

! Un processo BPMN può essere strutturato o meno (non è obbligatorio; BPMN non impone). Un processo strutturato è un processo dove **ogni gateway di split o ogni divisione** (es. eventi al bordo: **ha un corrispondente evento di merge**; questo è detto "single entry single exit"). In questo tipo di processi ho sempre un solo evento di end.

È stato dimostrato che qualunque processo può essere modellato così.. Costruire processi strutturati così diminuisce la possibilità di errore.

Eventi

- + Ogni processo deve avere almeno un elemento di start.
- + Un evento è qualcosa che accade istantaneamente durante l'esecuzione di un processo.
- + Un evento può influenzare il flusso del processo.

Ne esistono diversi tipi.

Evento di start

Mostra dove inizia il processo. È caratterizzato da un solo arco uscente e genera il token.



	None (generico)	
	Message	il processo inizia alla ricezione del messaggio; è il trigger.
	Timer	Il processo inizierà a fronte di una certa condizione temporale.
	Signal	Il processo inizia alla ricezione di un segnale. Il segnale (broadcast) è una comunicazione che può essere inviata da un partecipante al processo o da un altro processo.
	Multiple	
		Conditional
		Parallel multiple

Evento di end

A fronte della fine di un'attività e dell'arrivo a un evento di fine, l'end event consuma il token.

! Posso avere più di un evento finale ! Il processo è considerato attivo finché tutti i token raggiungono gli eventi di fine, a meno di usare l'evento terminate.

	None (generico)	
	Message	Alla fine del processo viene mandato un messaggio.
	Signal	
	Multiple	
	Terminate	Quando un token raggiunge questo evento, l'intero processo è terminato indipendentemente dal fatto che ci siano altri token ancora in viaggio.

Eventi intermedi

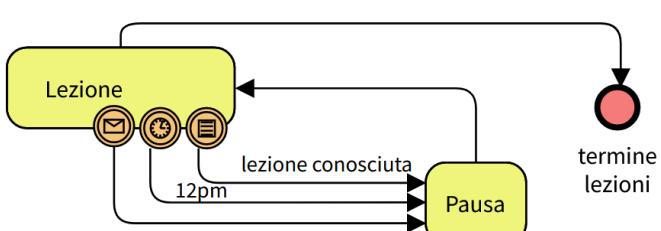
Ne esistono di due tipi:

- + Eventi di catching: aspettano un accadimento, messaggio, etc
- + Eventi di throwing: same ma lo lanciano.

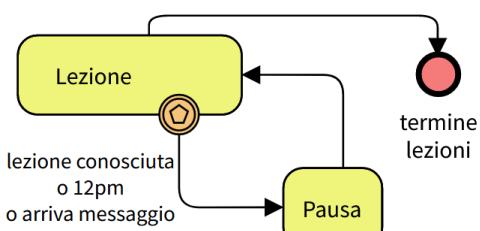
Volendo, gli eventi di start sono di tipo throwing e quelli di end sono di tipo catching.

	None (generico)	
	Timer	<p>Consiste nell'aspettare un istante di tempo.</p> <p>Se gli eventi sono di catching, si può anche porre al bordo:</p> <p>Il significato qui è che il token entra in A;</p> <ul style="list-style-type: none"> + Se l'attività termina entro 2 giorni, allora va in B + Se non termina in tempo va in C
	Message	<p>Può essere sia di tipo catching (aspetto una circostanza) che di tipo throwing (lancio un messaggio). Se di throwing è nero.</p> <p>Nel caso sopra, quando il token arriva e si attiva l'evento, allora viene mandato un messaggio e successivamente il token procede. Non ne esiste una versione attaccata al bordo!</p> <p>Posso anche bloccare e sbloccare flussi separati:</p>
	Signal	+ Same as above
	Conditional	
	Link.	<p>Crea delle connessioni tra parti del processo. Permette, a partire da un'attività, di specificare il fatto che il processo termina in un certo punto e riprende in un altro.</p> <p>Può essere comodo per avere porzioni di processo "combinabili"; il token viene "teletrasportato".</p>

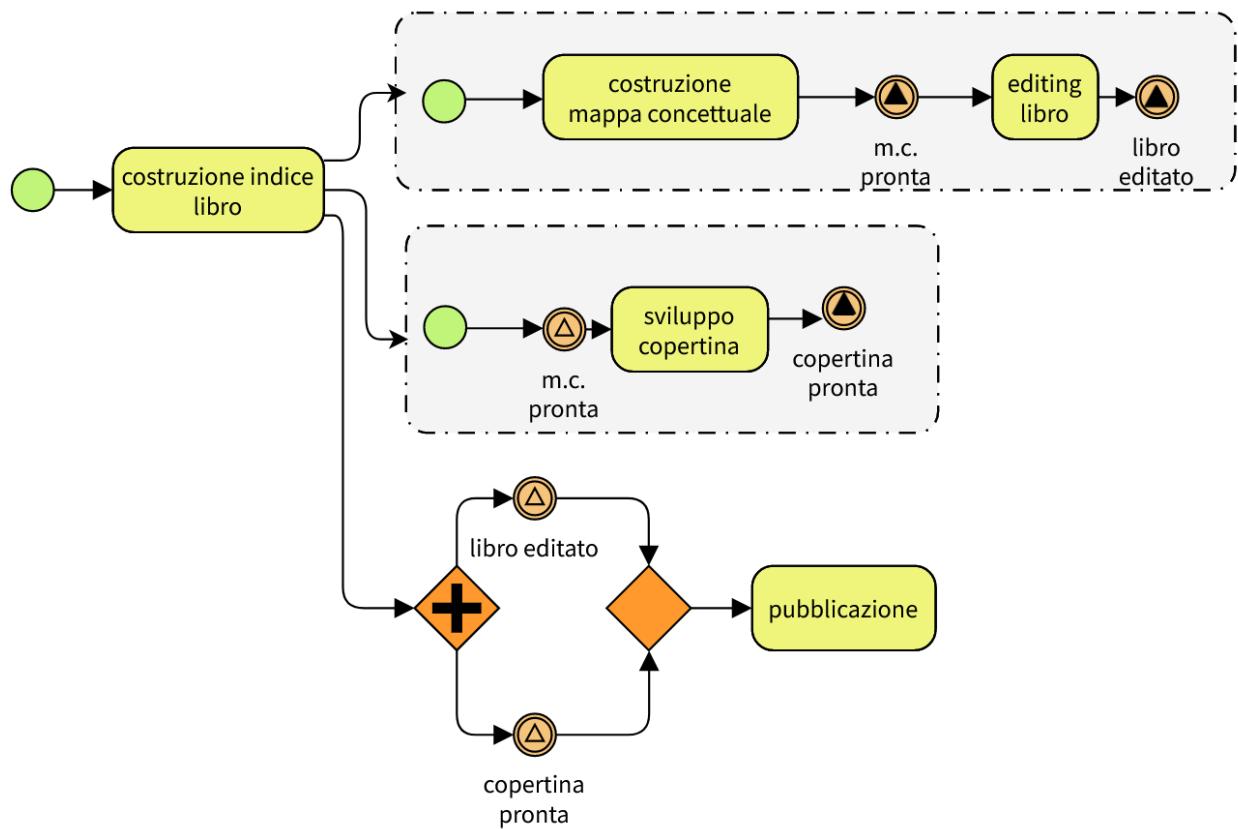
Posso specificare anche più di un evento al bordo:



Posso esprimere la stessa cosa con l'evento multiplo, scegliendo una condizione appropriata.



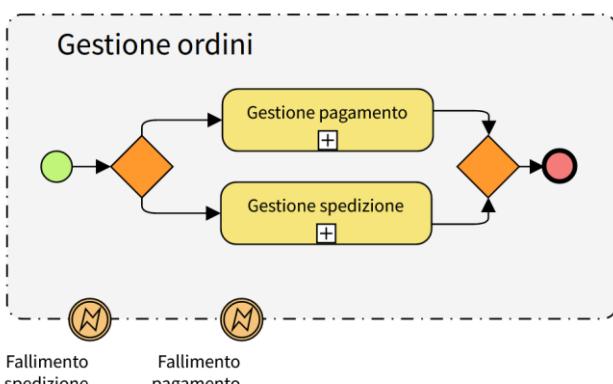
Esempiettino:



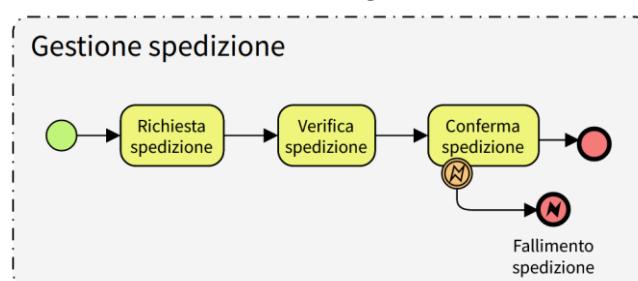
Evento di errore

	Error end event È usato per generare un errore
	Error intermediate event Può essere usato solo come evento al bordo di un'attività; non va mai sul flusso di un processo. È di tipo catching

Processo padre



Processo figlio

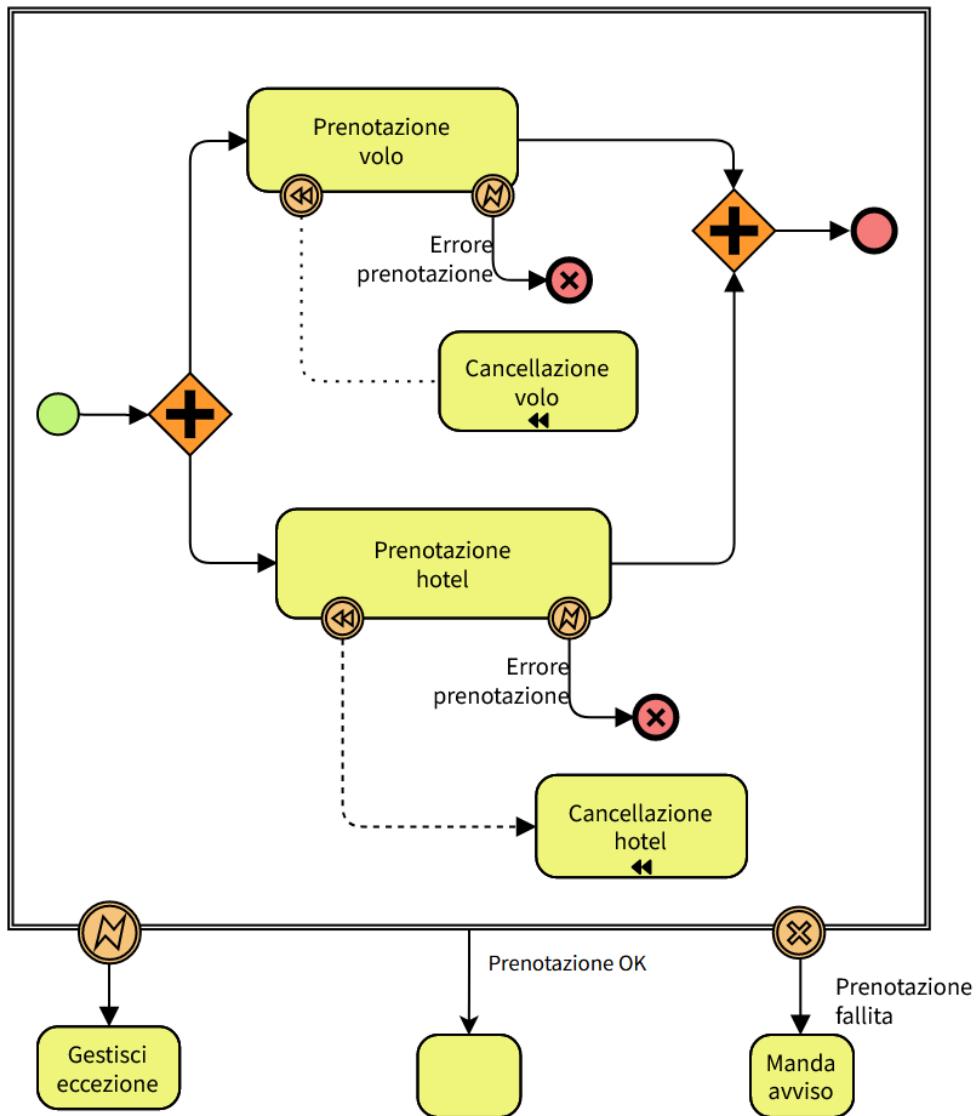


Da un sottoprocesso, lo scope dell'errore è il processo padre. Quando nel processo figlio si attiva il throwing fallimento spedizione, nel padre si attiva il fallimento spedizione catching.

Transazione

Una transazione è un insieme di operazioni che sono eseguite atomicamente, ovvero o tutto o niente. Si può rappresentare in BPMN attraverso un sottoprocesso con bordo doppio.

Devo prevedere la possibilità di fare rollback!



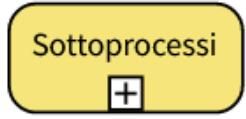
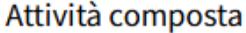
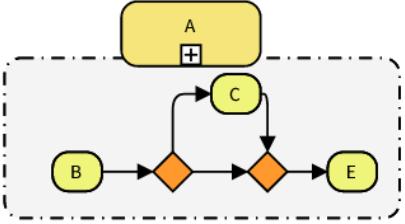
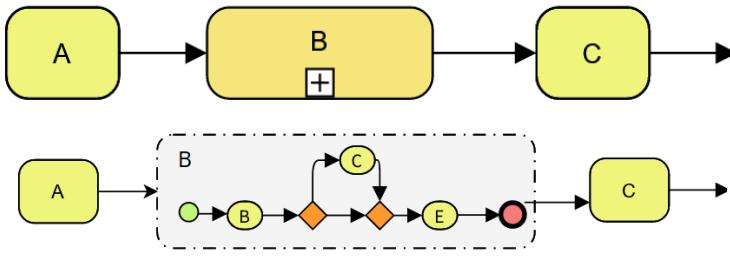
Costrutti tipici:

	Evento di compensazione È un evento intermedio di catching e può essere solo messo al bordo. Indica che bisognerà fare delle azioni di compensazioni (es. disdire) in caso di rollback
	Attività di compensazione Indica cosa va fatto in caso di rollback, MAI nel flusso normale del processo: infatti è collegata per associazione e non per flusso.
	Cancel event (throwing) Lancia un trigger all'evento al bordo
	Cancel event (catching) Dà inizio al rollback.
	Catching dell'eccezione al bordo della transazione

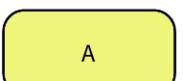
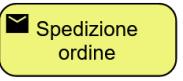
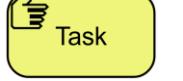
Attività/task

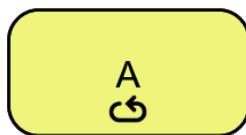
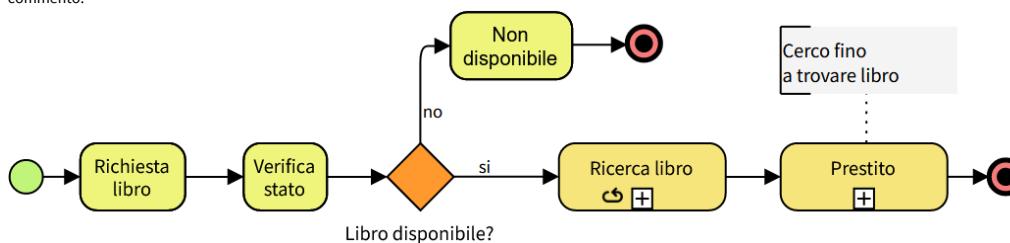
Describe un'unità di lavoro fatta in un processo aziendale.

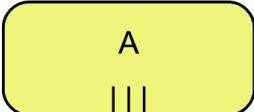
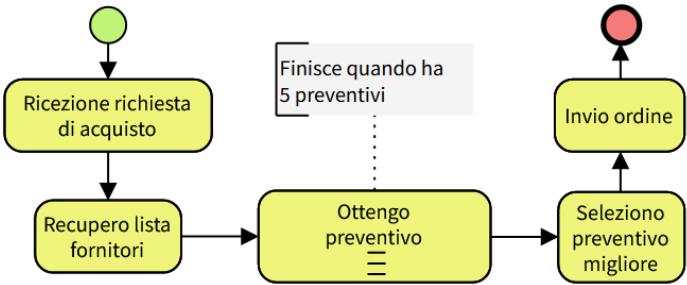
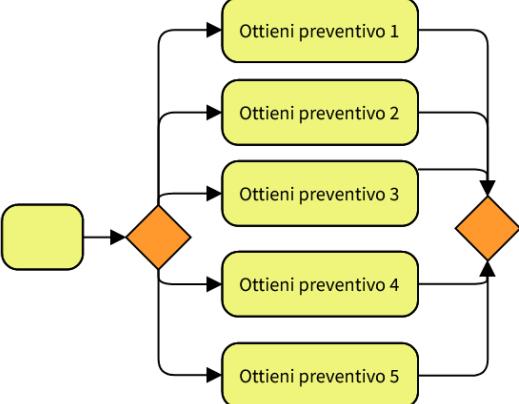
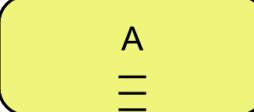
- + Utilizza del tempo per poter essere eseguita.
- + Involge le risorse aziendali.
- + Riceve un input e produce un output.

Attività  Attività atomica  Sottoprocessi  Attività composta 	<p>Quando un'attività non è atomica ma composta, allora sto rappresentando un'attività composta.</p>
	<p>Un sottoprocesso rappresenta un insieme di attività. Ne esistono due tipi:</p> <ul style="list-style-type: none"> + Embedded: sottoprocesso non riutilizzabile, che inizia con NONE start event  <ul style="list-style-type: none"> + Indipendenti: processo riutilizzabile + Nel caso dell'indipendente, i dati non sono passati in maniera automatica ma deve essere aggiunto un evento di passaggio dei dati (dato che posso riutilizzare la stessa procedura in più punto).

Tipi di task

 A	Attività generica. Eventualmente, il tipo è indicato in alto a sinistra.		
 Ricezione ordine	Attività di ricezione	 Spedizione ordine	Attività di spedizione
 Task	Attività manuale	 Task	Attività automatizzata

 A	Loop task È possibile specificare un certo numero di volte (fisso o condizionale) in cui deve essere eseguita un'attività. Indico la condizione di uscita con un commento.		
	Richiesta libro	Verifica stato	Non disponibile Libro disponibile? Cerco fino a trovare libro Ricerca libro Prestito

	<p>Attività multi-istanza parallela Sono attività svolte molte volte e in maniera concorrente. La condizione di loop dice quante volte deve essere eseguita.</p>  <p>A livello concettuale, corrisponde a</p>  <p>...che funziona, ma fa cageher in quanto a leggibilità.</p>
	<p>Attività multi-istanza sequenziale Same as above.</p>

Token

Il token è usato per descrivere il comportamento degli eventi d un processo, ovvero a livello di esecuzioni posso immaginare questa cosa che “viaggia” sul processo.



Flusso

→ Rappresenta vincoli temporali tra gli oggetti di flusso.

Gateway

Sono gli elementi che controllano il fusso, e possono essere di split o di merge.

Anche qui ne abbiamo diversi tipi:

Gateway esclusivo	Solo uno dei cammini può essere attraversato dal token. La X dentro è facoltativa.
--------------------------	--

Gateway parallelo 	<p>Le attività V,B,C vengono eseguite TUTTE e in maniera concorrente.</p> <p>Attenzione! Sono svolte in maniera concorrente, quindi non posso sapere quale è eseguita prima o dopo!</p>
Condizione di default 	<p>Indica il path seguito di default.</p>
Gateway inclusivo 	<p>Permette di prendere uno o più cammini. Il gateway di merge rilascerà il token solo quando è arrivato un numero di token pari al numero di cammini che si sono attivati.</p>
Gateway (esclusivo) basato su eventi 	<p>Il path che viene eseguito è il primo dei tre eventi che si verifica.</p>

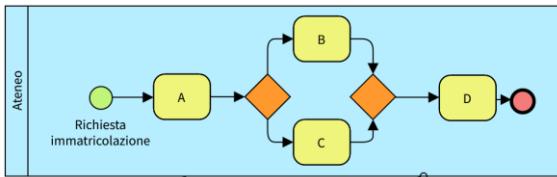
All'esame chiede differenza fra token e evento e se posso rappresentare la cosa in entrambi i modi. La risposta è dipende: se l'operazione implica di occupare del tempo è un'attività, mentre se è istantanea è un evento.

Pools e lanes

Message flow

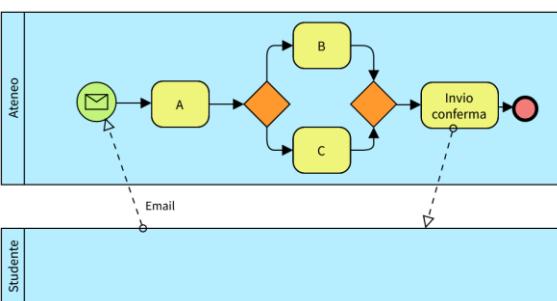
Rappresenta la comunicazione fra due partecipanti indipendenti (pool). È sempre fra pool, mai fra oggetti dello stesso processo!

Pool



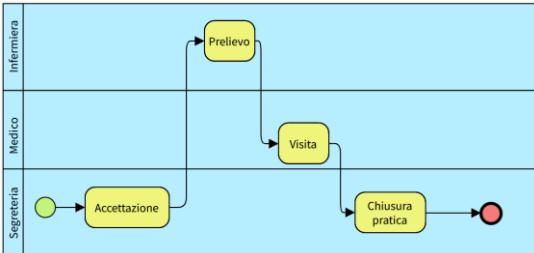
Una **pool** permette di rappresentare un'entità organizzativa indipendente che esegue il processo.

Le due pool sono processi separati e non condividono dati, quindi la comunicazione / scambio di messaggi deve avvenire in maniera esplicita con le frecce tratteggiate.



Posso anche rappresentare il fatto che c'è un partecipante allo studente, ma senza specificare il suo processo se non mi interessa.

Lane



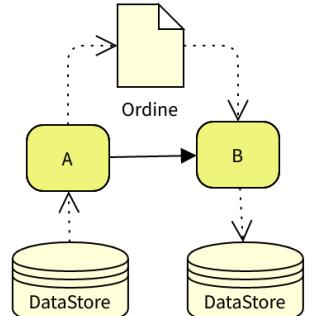
Una lane è rappresentata in una pool, e specifica altre entità organizzative all'interno della stessa organizzazione. Di fatto, serve a specificare chi esegue cosa.

Nella pool, il passaggio di messaggi è implicito.

Artifacts

Permettono di rappresentare informazioni addizionali sul processo, che però non impattano direttamente il processo. Ovviamente sono facoltativi.

	Annotazione testuale È un commento. La uso per specificare condizioni o aggiungere info utili.
	Data object Rappresenta eventuali dati utilizzati durante l'esecuzione del processo. Può rappresentare dati in unput o output delle attività. La caratteristica è che non sono memorizzati da nessuna parte.
	Data store Same as above, ma i dati sono memorizzati in una base di dati.
	Associazioni dirette o indirette Servono a collegare i dati alle attività.



SISTEMI DI SUPPORTO ALLE DECISIONI

Lezione 11 - 16/11/2020

Nascono negli anni 80 con lo scopo di **sfruttare i cumuli di dati**. Inizialmente, lo scopo era solo di descrivere il passato, evidenziare problemi per risolverli. Adesso, invece, ci si spinge oltre: si cerca di **anticipare il futuro**, usando il sistema per prendere decisioni.

Abbiamo già parlato di portafoglio direzionale, che è l'insieme di applicativi che permette ai manager di gestire tutte le situazioni che hanno a che fare col prendere decisioni. È volto a analizzare lo stato dell'azienda e prendere decisioni.

Def. Piattaforma di business intelligence

Correlato al portafoglio direzionale, si parla anche di **piattaforma per la business intelligence**: la business intelligence è quella disciplina che consente a chi deve prendere decisioni di avere a disposizione soluzioni SW per capire i fattori chiave del business e prendere decisioni.

Quando parliamo di piattaforma è perché di solito serve HW dedicato, infrastruttura di rete e tutti i SW appropriati.

Disponibilità e valore delle informazioni

Il ruolo chiave di queste piattaforme è la **trasformazione dei dati aziendali in informazioni**.

Sappiamo già che dati != informazioni, ovvero il dato è quello che memorizziamo e l'informazione è ciò che estraiamo. Abbiamo bisogno di informazioni, che sono risorsa organizzativa, come materia prima delle decisioni.

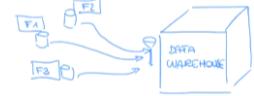


I dati transazionali sono in **enorme quantità, ma hanno poco valore**. Da questi dati però possiamo trarre fuori delle informazioni selezionate (es. report) per capire meglio. **Le informazioni di valore sono sempre meno**.

Scenario tipico

Lo scenario tipico è quello di una grande azienda con numerose filiali, e in ognuna di queste filiali possiamo avere soggetti informativi diversi – che eventualmente non condividono DB. Vogliamo capire quanto il contributo di ogni filiale contribuisce al rendimento globale.

Per far questo c'è bisogno del cosiddetto **data warehouse**; posso immaginare che nelle varie filiali ci sono a disposizione delle sorgenti informative; voglio far sì che **tutte queste informazioni siano raccolte in un unico DW**, che è un raccoglitore di informazioni che integra e riorganizza i dati provenienti da soggetti di varia natura allo scopo di poter analizzare l'informazione a supporto del processo decisionale.



In pratica, avere a disposizione un DW fa sì che si possano **tenere separate le elaborazioni transazionali dall'insieme di elaborazioni analitiche** che faccio per elaborare le transazioni.

Data warehousing

Def: Data warehousing

È una collezione di **metodi, tecnologie e strumenti** di ausilio al **knowledge worker** per produrre analisi dei dati finalizzate all'esecuzione dei processi di tipo decisionale.

Il DW, assieme a metodi e strumenti, si inserisce perfettamente in tutti i contesti in cui ho tantissime informazioni (al punto che sono "inutili", o non riesco a capire se sono importanti).

Caratteristiche:

- + **Accessibilità:** le informazioni devono essere accessibili anche a utenti con conoscenze limitate di informatica
- + **Integrazione dei dati:** devo gestire basi di dati diverse che contengono info simili, e che in qualche modo vanno travasate nel DW
- + **Flessibilità dell'interrogazione:** stiamo parlando di tantissimi dati, strumenti messi a disposizione di persone che vogliono realizzare query complesse. Quindi servono metodologie che permettano di passare da visualizzazioni e metodi diversi e complessi.
- + **Sintesi:** se voglio analizzare in modo efficace i dati voglio poterli visualizzati in qualche modo
- + **Dati corretti e completi:** altrimenti non posso prendere decisioni appropriate
- + **Rappresentazione multidimensionale:** per offrire all'utente una visione intuitiva delle informazioni devo mettere a disposizione una rappresentazione multidimensionale, vedremo meglio poi.

Def Data warehouse

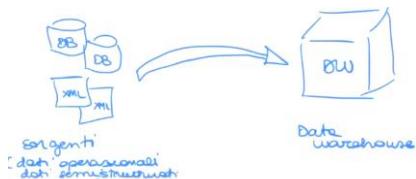
È una collezione di dati di supporto per il processo decisionale.

Caratteristiche:

- + **Collezione orientata ai soggetti di interesse:** ovvero il DB è generico e ho tante applicazioni diverse che lo usano. IL DW invece è orientato al soggetto di interesse, oper esempio solo al marketing/logistica e per ogni area ho delle informazioni relative.



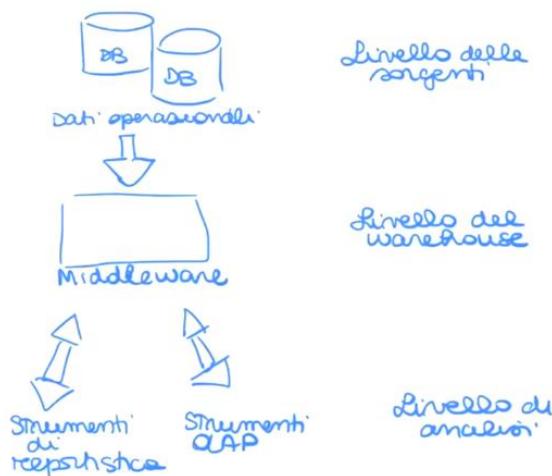
- + **Informazione integrata e consistente:** il DW parte da sorgenti informative diverse, che finiscono nella stessa DW.
- + **L'informazione è rappresentativa dell'evoluzione temporale:** è difficile fare analisi senza evoluzione... si fa un archivio temporale in cui il tempo è parte della chiave, e tengo traccia della storia andando a vedere come i dati sono cambiati rispetto al tempo.
- + **Non è volatile.**



Interrogazioni

	DB = OLTP	DW = OLAP
Caratteristiche	<ul style="list-style-type: none"> + Eseguiamo transazioni + Leggono e scrivono un numero ridotto di tuple, che coinvolgono tabelle tra loro relazionate da semplici relazioni 	<ul style="list-style-type: none"> + Analisi dinamica: il DW è usato per fare analisi che si possono concatenare. + Leggono e scrivono grandi quantità di dati per trovare dati di sintesi. + Interattività: possibilità di fare delle sessioni di analisi che concatenano analisi diverse.
Utenti	Migliaia 🌟🌟	Centinaia 💯
Accesso	Lettura e scrittura 🖊️📘	Solo lettura 📖📖📖
Tempo	Dati correnti 🕒	Dati storicizzati 🕰️
Modello dei dati	Normalizzato 😊	Denormalizzato e multidimensionale 🎩🌐
Orientamento	Applicazione	Soggetto di interesse.

Architetture

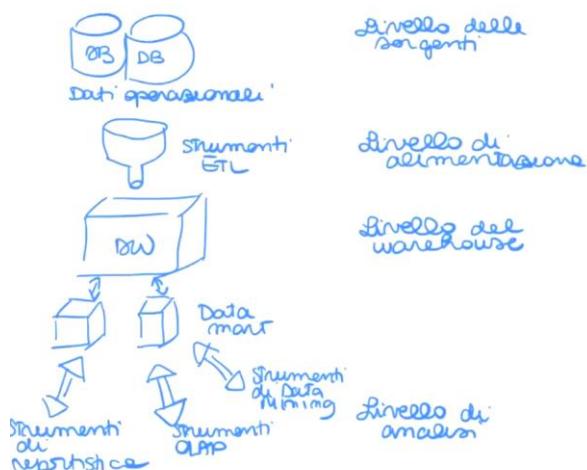


Ad 1 livello

Si parte dalle DB, che sono il livello delle sorgenti. Il livello del warehouse è implementato per mezzo di un middleware, che simula la presenza del DW

Svantaggio:

Non c'è una vera separazione fra OLAP e OLTP, perché in pratica la presenza del DW è simulata dal middleware ma di fatto restano solo nel DB. Questo rallenta e genera confusione



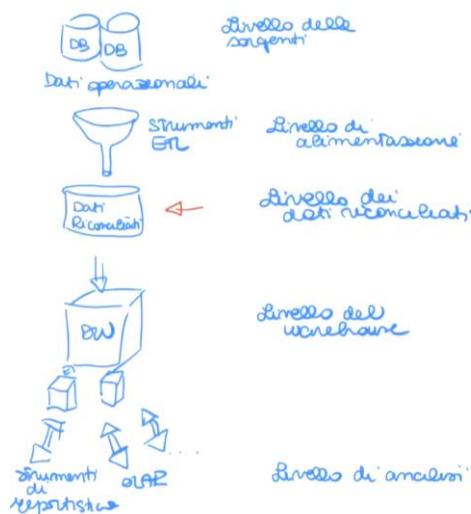
A due livelli

Def: Data mart

Sottoinsieme o aggregazione dei dati presenti nel DW principale, che contiene le informazioni rilevanti per una particolare area del business.

Vantaggio:

Ho finalmente la separazione fra DB/DW, quindi le due interazioni non possono mai interferire a vicenda.



A tre livelli

La cosa in più sono i **dati riconciliati**. I dati riconciliati sono dei dati operazionali risultato del processo di integrazione e ripulitura.

Vantaggio:

Ho un modello comune di tutti i dati per l'azienda.

Svantaggio:

Ridondanza.

ETLs



Def: ETL

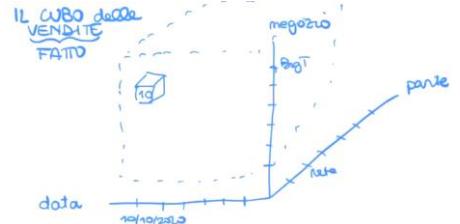
Sono un insieme di strumenti di **Extraction, Transformation and Loading**. Il ruolo di questi strumenti è di estrarre trasformare e ripulire i dati del DW.

Modello del DW: modello multidimensionale

È quello che usiamo per rappresentare i dati in DW. Si basa sul concetto di fatto.

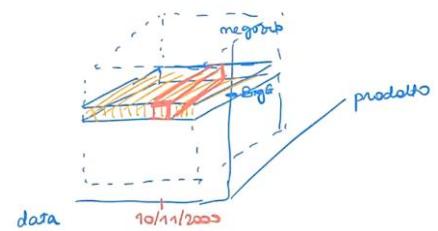
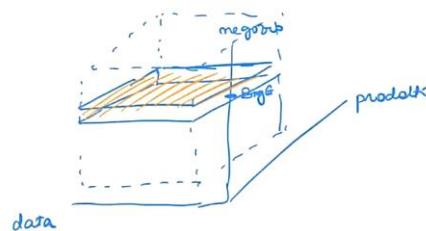
I fatti di interesse sono rappresentati in **cubi** in cui:

- + **Ogni cella contiene MISURE numeriche** che quantificano il fatto da diversi punti di vista.
- + **Ogni asse rappresenta una dimensione** di interesse per l'analisi.
- + **Ogni dimensione può essere la radice di una gerarchia** di attributi, usati per aggregare i dati.



Ad esempio:

- Se prendo una **fetta** del cubo, la fetta mi dà tutto quello che ho venduto nel tempo di un certo negozio.
- Se fisso anche una certa data, ottengo una **strisciolina**.



Gerarchie

Ad esempio:

Prodotto	Tipo	Categoria
brullo		
sbianco	detersivo	pulizia cose
lucido		
lattel		
lattell	latticini	
cola		
aranciata	bibite	alimentari

Negozio	Città	Regione

Data	Mese	Anno

Approcci al OLAP

Parlando di modello multidimensionale e di olap, esistono due diversi approcci:

<i>ROLAP (relational OLAP)</i>	<i>MOLAP (Multidimensional OLAP)</i>
<ul style="list-style-type: none">+ Ci permette di usare un gestore di dati relazionali per fare le analisi di cui abbiamo parlato. La possibilità di avere il ROLAP viene dal fatto che i db relazionali sono super diffusi. → Ovviamente, nel momento in cui usiamo una base di date relazionale abbiamo un problema di prestazioni poiché i dati sono denormalizzati! Fare una query relazionale costerà parecchio.	<ul style="list-style-type: none">+ Ha il vantaggio di usare una struttura dati più adatta alla gestione dei dati, ma posso avere problemi relativi alla scarsità dei dati. → Ad esempio, non è detto che in un negozio si vendano sempre tutti i dati: rischio di avere tante celle vuote e informazioni sparse.

Operatori OLAP

Sono gli operatori più utilizzati per accedere alle informazioni di un DW. Questo perché **visualizzano l'informazione del cubo** e permettono all'utente di **costruire sessioni di analisi complessa**, e ottenere le informazioni ottenute a ciascun passo precedente.

Di solito, gli operatori olap mettono a disposizione **un'interfaccia flessibile e facile** da usare.

Sessione OLAP

Indichiamo con sessione OLAP una **sessione di navigazione** in cui posso analizzare uno o più fatti di interesse. Ogni passo della sessione prevede l'applicazione di un operatore olap. Il risultato di una sessione olap è di tipo multidimensionale.

Data mining

È uno fra gli strumenti che posso usare; permette di **scoprire informazioni nascoste**. È una tecnica che è stata proposta per i carrelli delle spese: serve a capire i comportamenti ricorrenti.

+ Scoprire informazioni nascoste nei dati

Viene fatta anche adesso; le tessere raccoglipunti sono fatte per quello :P È interessante sapere cosa comprano le persone per fare ricerche di mercato, oppure per esempio per studiare tecniche di esposizione della merce (es. si preferisce mettere lontana la roba che normalmente viene acquistata insieme). Serve anche in medicina, per provare a trovare relazioni fra sintomi e genetica...

+ Determinare regole di implicazione logica

presenti nelle basi di dati, individuando gruppi di affinità tra oggetti

PROGETTAZIONE DI UN DATA WAREHOUSE

Tipi di approcci

Top down

Analizza i bisogni globali dell'intera azienda e pianifica/progetta il DW nella sua interezza, ovvero progettando tutti i data mart per gli usi. Devo avere un'idea generale dell'azienda.

Vantaggi	Svantaggi
<ul style="list-style-type: none"> + Avendo una visione globale dell'azienda probabilmente permette di produrre un DW consistente e ben integrato. 	<ul style="list-style-type: none"> - Tempi lunghi: non ho un primo risultato che mi dà l'idea; far vedere subito le potenzialità spesso è utile anche per recepire più requisiti, quindi avere tutto insieme porta ad avere tempi molto lunghi e potrebbe scoraggiare l'azienda. - Progettare tutto assieme è complesso.

Bottom up

Il DW viene costruito in maniera incrementale; progetto i singoli datamart, poi li collego costruendo il tutto un pezzettino alla volta.

Vantaggi	Svantaggio
<ul style="list-style-type: none"> + Risultati concreti in tempi brevi, per avere un riscontro immediato sull'utilità. + Ci possiamo concentrare su un singolo settore, e questo permette di avere un dominio ristretto. Questo permette di restringere le persone coinvolte e dati da integrare. 	<ul style="list-style-type: none"> - Visione parziale del dominio di interesse

Ciclo di sviluppo di un data warehouse

1. Definizione degli obiettivi e pianificazione

- Individuazione degli obiettivi: bisogna capire quali sono gli obiettivi del sistema; è la raccolta di tutti i requisiti, di cui abbiamo già parlato in ingegneria del software.
- Stima delle dimensioni
- Valutazione dei costi
- ... (tutto ciò che ha a che fare con la progettazione).

2. Progettazione dell'infrastruttura

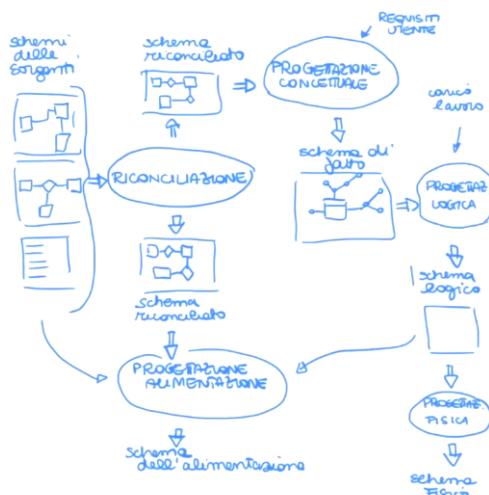
Consiste nell'analisi e comparazione di possibili soluzioni architettoniche.

3. Progettazione e sviluppo del data mart vero e proprio

È qualcosa che si ripete: il data mart è un pezzo del DW, quindi ripeto questa progettazione per ogni data mart.

L12 – 18/11/2020

Progettazione del data mart



La progettazione di un data mart è abbastanza simile alla progettazione di un database.

- Analisi e riconciliazione delle sorgenti
- Analisi dei requisiti
- Progettazione concettuale
- Raffinamento del carico di lavoro
- Progettazione logica
- Progettazione dell'alimentazione del data warehouse
- Progettazione fisica

Partiamo dagli schemi delle sorgenti, che possono essere molto eterogenei, abbiamo una fase di riconciliazione. La fase di riconciliazione permette di ottenere uno schema dei dati riconciliato. Questo servirà per 2 passi successivi:

- + **Procedo alla progettazione concettuale** del data warehouse / data mart, attraverso la quale produco lo schema di fatto. Da qui si procede con la progettazione logica e la fisica.
- + **Procedo alla progettazione dell'alimentazione**, di cui però non ci occupiamo. Insomma, permette di mappare le sorgenti dei dati nel formato dei dati che ci serve.

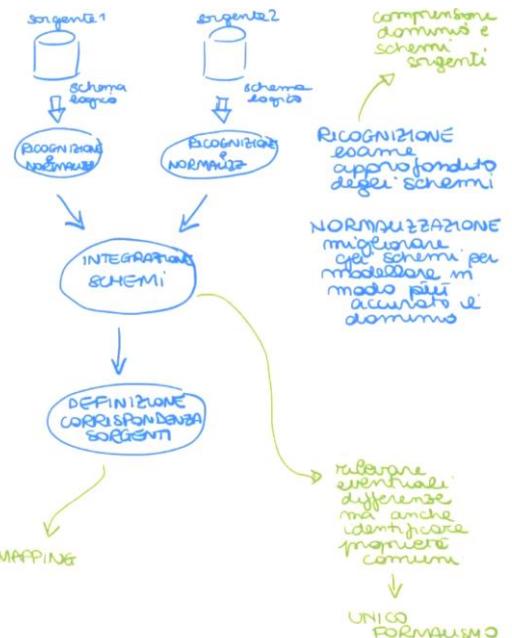
1. Riconciliazione delle sorgenti

Partiamo da un **insieme di sorgenti eterogenee**, quindi quello che facciamo è **capire dagli schemi delle sorgenti come sono rappresentati i dati** del dominio di interesse. Questi passaggi vanno fatti anche nel caso in cui la sorgente sia unica!

- La **riconoscione** è un esame approfondito delle sorgenti e degli schemi.
- La **normalizzazione** ha l'obiettivo di correggere gli schemi per modellare in modo più accurato il dominio applicativo.

Il frutto è uno **schema trasformato**, ovvero **l'integrazione degli schemi**. Si cercano di **evidenziare le proprietà interschema**: vogliamo identificare eventuali differenze e proprietà condivise, per usare un unico formalismo.

Ottenuto lo schema integrato, si definisce la corrispondenza con le sorgenti; vogliamo ottenere un **mapping** che permetta di correlare gli elementi degli schemi locali con quello degli schemi riconciliati. Ovviamente, nella fase di integrazione potremo incontrare situazioni diverse... per esempio, potrei avere che:



Prospettive diverse per rappresentare gli stessi dati	Stessa informazione rappresentata con costrutti equivalenti	Incompatibilità
<pre> graph TD subgraph S1 [Sorgente 1] D1[DIPENDENTE] --> P1[PERSONA] P1 --> M1[PRODIGO] M1 --> D1 D1 --> H1[HIPERDIPENDENTE] H1 --> D1 end subgraph S2 [Sorgente 2] D2[DIPENDENTE] --> P2[PERSONA] P2 --> M2[ASSISTENTE] M2 --> D2 D2 --> H2[HIPERDIPENDENTE] H2 --> D2 end </pre>	<pre> graph TD LIBRO[LIBRO] --> ISBN[ISBN] LIBRO --> TITOLO[TITOLO] LIBRO --> EDITORIO[Editorio] LIBRO --> OBA[OBA] LIBRO --> EDITRICE[EDITRICE] LIBRO --> INDUSTRIA[Industria] </pre>	<pre> graph TD PROFESSORE[PROFESSORE] --> M1[maestro] PROFESSORE --> C1[CORSO] PROFESSORE --> I1[Insegnante] PROFESSORE --> T1[Titolo] PROFESSORE --> N1[Nome] PROFESSORE --> V1[Vnd. corsa ed.] PROFESSORE --> C1 PROFESSORE --> I1 PROFESSORE --> T1 PROFESSORE --> N1 PROFESSORE --> V1 </pre>

Le possibili relazioni fra costrutti usati sono:

- + **Identità**: stesso punto di vista, stessi costrutti. Non vengono commessi errori.
- + **Equivalenza**: costrutti diversi ma equivalenti, e non ci sono errori di specifica (esempio del libro)
- + **Comparabilità**: costrutti e punti di vista non sono in contrasto fra loro (primo esempio)
- + **Incompatibilità**: gli schemi sono in contrasto.

Fasi di integrazione

1. Pre-integrazione

Vado a definire la strategia di integrazione. (per esempio, integro le cose una alla volta? Oppure analizzo once e metto tutto assieme?)

2. Comparazione degli schemi

Faccio una analisi comparativa che mira a identificare i conflitti e le correlazioni

3. Allineamento degli schemi

Cerco dirisolvere eventuali conflitti, applicando delle primitive che permettono di sistemare gli schemi. (es. rinomino le cose nominate come diverse):

Attenzione; non tutti i conflitti possono essere risolti automaticamente!

4. Fusione e ristrutturazione degli schemi

Durante la fusione cerchiamo di allineare gli schemi per ottenere lo schema conciliato.

Lo schema conciliato deve essere ottimale, ovvero avere:

- Completezza
- Minimalità (minima ridondanza)
- Leggibilità

1. Analisi dei requisiti

Va i pari passo con l'analisi delle sorgenti. È importante perché per riuscire a capire in che modo costruire lo schema dei dati, l'alimentazione, etc. è importante in maniera **indipendente** dall'analisi delle sorgenti.

Lo scopo è **raccogliere le esigenze di utilizzo del data mart**; non solo vogliamo capire quali sono i dati a disposizione, ma anche che cosa mettere nel data warehouse per poter soddisfare le esigenze dei clienti. L'obiettivo, dunque, è **raccogliere le esigenze degli utenti**.

Fonti

- + Da questo punto di vista le fonti da utilizzare e considerare sono i **futuri utenti del data mart**; potenzialmente si tratta di persone non informatiche, dato che viene usato sul top della piramide, e che di conseguenza potrebbe avere delle difficoltà a interagire col datamart; sarà importante mediare sul linguaggio da utilizzare.
- + Per quanto riguarda gli aspetti più tecnici, eventualmente, si può interagire anche con gli **amministratori tecnici**; comunque non sono gli utenti principali.

Interviste

Ci sono avanti due approcci possibili:

+ **A piramide**

Approccio induttivo, in cui partiamo da domande molto dettagliate e poi andiamo via via ad ampliare le informazioni da chiedere, portando l'intervistato a dare risposte generali a domande aperte.
Questo tipo si presta molto bene a soggetti un po' riluttanti, che si trovano bene a rispondere a domande più precise.

+ **A imbuto**

Approccio opposto; si parte da domande molto generali per poi andare a restringere su argomenti più specifici. Si presta meglio con intervistati che vivono l'intervista come "pesante", e per i quali la domanda aperta viene vata come una rassicurazione sul fatto che non ci sono risposte giuste o sbagliate.

Ovviamente le domande dipendono fortemente dal ruolo; dovranno avere a che fare con le mansioni dell'intervistato:

- Dirigente = obiettivi dell'azienda, come misura il successo, analisi a cui sono interessati
- Direttore di reparto = competenza sul reparto, quindi a lui ha più senso specializzarsi sui suoi margini di manovra, analisi a cui sono interessati
- Amministratore del sistema informatico = domande tecniche, relative a come vengono gestiti i dati e le richieste

Ovviamente, all'inizio non è necessario stabilire esattamente quali siano gli strumenti che verranno messi a disposizione; ma sarebbe importante rendere chiaro quali siano **capacità e vantaggi dei sistemi OLAP**.

Fatti

I fatti sono i **concetti** che gli utenti finali del data mart utilizzeranno per fare analisi e prendere decisioni.

L'analisi dei requisiti ha lo scopo di andare a capire quali siano i fatti di interesse per l'azienda. Li scopro partendo da **sorgenti informative e interviste**.

Quando riusciamo ad identificare il fatto abbiamo già capito l'interesse, ovvero il concetto che sarà analizzato per prendere decisioni.

- + Inoltre, oltre ai fatti di interesse, dobbiamo capire quali siano le **dimensioni di analisi**. Fissare le dimensioni di un fatto è importante poiché ci permette di specificare quali siano le granularità a cui analizzare i dati, e con quali gerarchie.
- + È importante anche l'**evoluzione temporale / storizziazione**: il fatto di riuscire a memorizzare eventi nel tempo permette di fare analisi.

Esempi di fatti possono essere:

- Ambito sanitario e pronto soccorso = accessi fatti al pronto soccorso
- Turismo = prenotazione soggiorno, noleggio dell'auto

DFM - DIMENSIONAL FACT MODEL

Oonestamente, non mi sono trovata molto bene con la spiegazione. [Queste slides](#) però sono godlike

Per modellare i fatti non esiste ancora un accordo su un unico modello di progettazione concettuale, e l'ER non si presta molto bene.

Quello più usato è il dimensional fact model (DFM), modello concettuale grafico che consiste in un insieme di schemi di fatto. Ciascuno schema di fatto permette di modellare un fatto, le sue misure, le dimensioni di analisi e le gerarchie.

Vantaggi

- Supporta efficacemente il progetto concettuale
- Crea un ambiente su cui formulare in modo intuitivo le interrogazioni dell'utente
- Permette il dialogo tra progettista e utente finale, essendo intuitivo
- Crea una piattaforma stabile per il progetto logico, indipendentemente dal modello logico target
- È una documentazione espressiva e non ambigua.

Costrutti di base

Fatto

Concetto di interesse per il processo decisionale. Modella un insieme di eventi che accadono nel contesto, e deve avere aspetti dinamici che evolvono nel tempo.
(vendita)

Misura

Proprietà numerica di un fatto, che descrive un aspetto quantitativo di interesse per l'analisi.
(incasso)

Dimensione

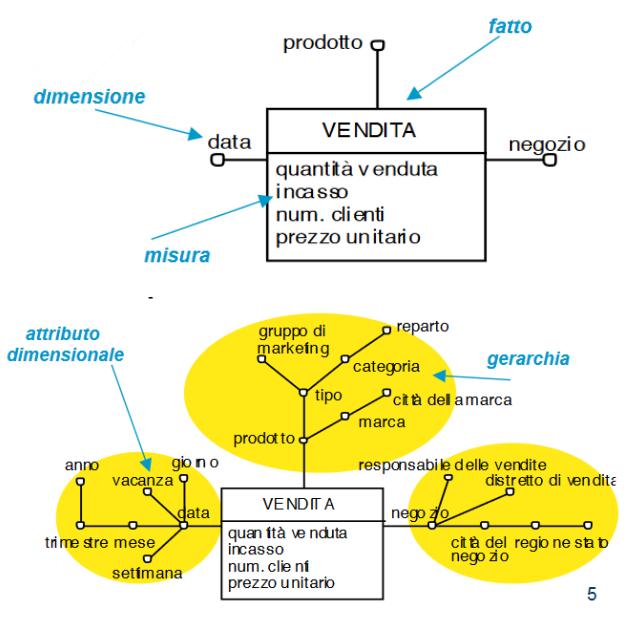
Proprietà di un fatto che ne descrive una coordinata di analisi.

Gerarchie

Albero direzionale i cui nodi sono attributi dimensionali, e i cui archi modellano associazioni molti a uno. L'arco da X a Y rappresenta la dipendenza funzionale X → Y.

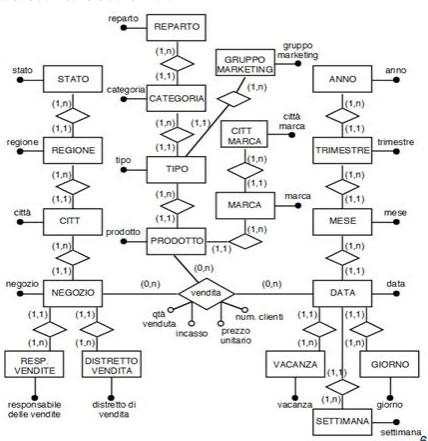
Attributo dimensionale

Le dimensioni e gli altri attributi che le descrivono (ad esempio un prodotto è descritto dal suo tipo, dalla marca,...)



Un fatto esprime un'associazione molti a molti fra le dimensioni.

Corrispondenza con l'ER



Attenzione: una settimana non sempre è in relazione uno a molti con un mese, potrebbe essere spezzettata :') quindi la settimana va a parte.

L'ER non si presta: lo schema di fatto è molto più comprensibile anche per chi non ha dimestichezza con 'sta roba.

Additività e questioni relative all'aggregazione di dati

Una misura è additiva quando può essere aggregata secondo l'operatore somma.

!!! L'aggregazione ci richiede di definire un operatore adatto per associare i valore degli eventi primari con gli eventi secondari. Ci permetterà di riassumere.

Ci sono tre tipi possibili di operatori:

- **Misure di flusso:** si riferiscono ad un certo periodo, al cui termine vengono valutate in maniera cumulativa
 - Vendite mensili, nascite annuali, ...
- **Misure di livello:** vengono valutate in particolari istanti di tempo
 - Quanti items in inventario al momento, quanti dipendenti...
- **Misure unitarie:** valutate in particolari istanti di tempo, ma espresse in senso relativo
 - Prezzo unitario di un prodotto, percentuale di sconto, ...

	GERARCHIE TEMPORALI	GERARCHIE NON TEMPORALI
MISURE DI FLUSSO	SUM, AVG, MIN, MAX	SUM, AVG, MIN, MAX
MISURE DI LIVELLO	Avg, MIN, MAX	SUM, AVG, MIN, MAX
MISURE UNITARIE	Avg, MIN, MAX	Avg MIN, MAX

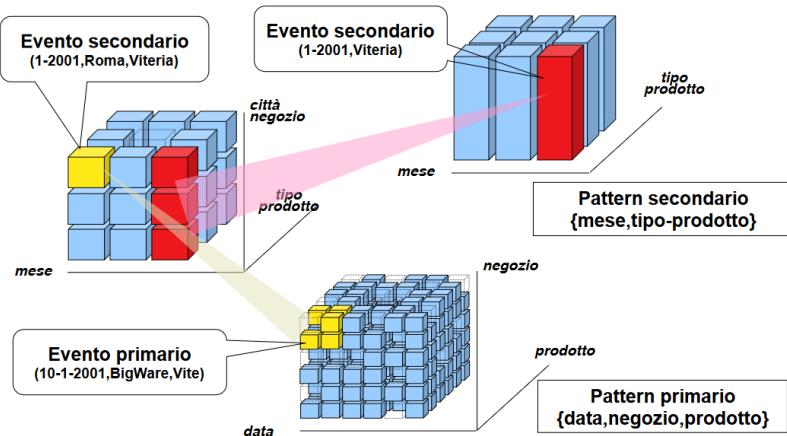
Gli operatori che posso usare dipenderanno dal tipo di gerarchia:

- **Gerarchie temporali:** il fatto cambia nel tempo
- **Gerarchie non temporali:** il fatto non cambia nel tempo; città, negozio...

Abbiamo già visto le misure non additive. Se non si possono usare neanche gli altri operatori si parla di misura non aggregabile.

Evento primario Un evento primario è una particolare occorrenza di un fatto, individuata da una ennupla costituita da un valore per ciascuna dimensione. A ciascun evento primario è associato un valore per ciascuna misura.	Evento secondario Dato un insieme di attributi dimensionali, detto pattern, ciascuna ennupla di loro valori individua un evento secondario , che aggrega tutti gli eventi primari corrispondenti. A ciascun evento secondario è associato un valore per ciascuna misura, che riassume in sé tutti i valori della stessa misura negli eventi primari corrispondenti.
Pattern primario Un pattern primario è il pattern formato dall'insieme delle dimensioni.	Pattern secondario Un pattern secondario è un qualsiasi altro pattern diverso dal primario.

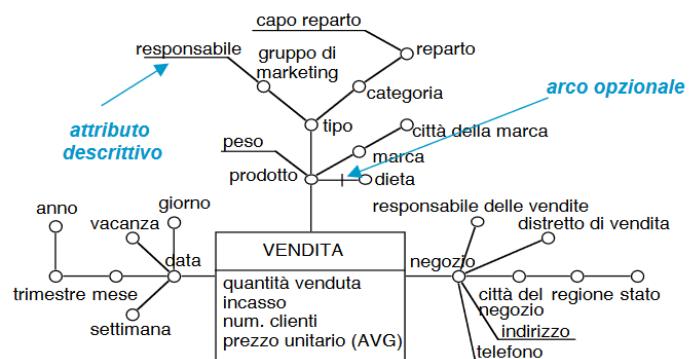
Questa immagine spiega tutto in maniera cristallina.
Grazie mille professore Beneventano, la amo.



Costrutti avanzati

Attributo descrittivo

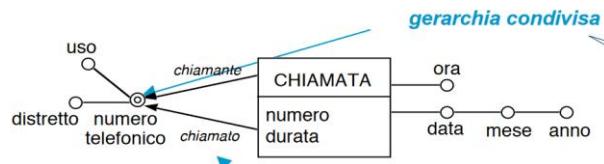
Contengono delle **informazioni aggiuntive su un attributo dimensionale** della gerarchia, connessi con una relazione a 1 (es. a un negozio è associato un indirizzo). **Non vengono usati per l'aggregazione**; servono a completare le informazioni.



Arco opzionale

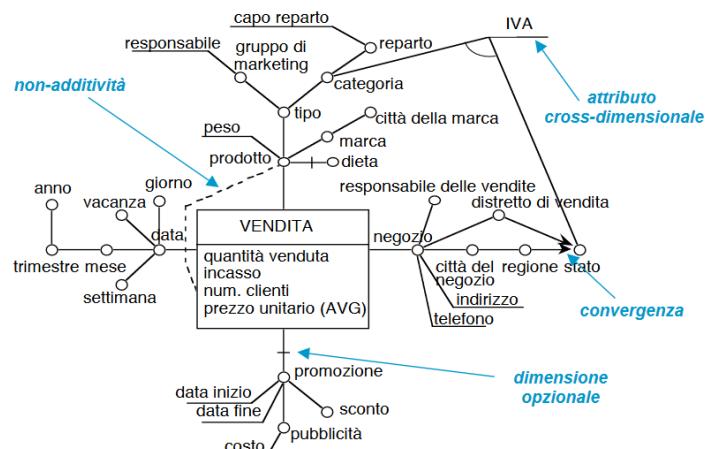
Gerarchia condivisa

Rappresenta una gerarchia condivisa. Questo tipo di rappresentazione **equivale ad avere la gerarchia ripetuta due volte**.



Convergenza

Posso avere che le misure **convergono**; ad esempio ha senso che il distretto di vendita non sia incluso nella regione ma sia incluso nello stato. Implica un vincolo di integrità del tipo "lo stato della città del negozio deve essere lo stesso di quello del distretto del negozio"



Attributo crossdimensionale

Ad esempio, l'IVA dipende sia dalla categoria che dallo stato in cui mi trovo.

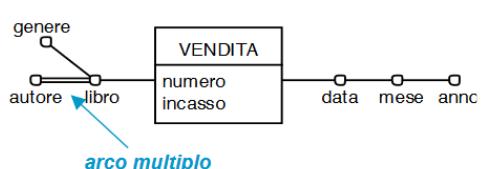
Non additività

Una misura è additiva su una dimensione quando i suoi valori possono essere aggregati lungo l'operatore **somma**; ad esempio

- L'incasso è additivo rispetto al prodotto; posso sommare l'incasso rispetto alla dimensione del prodotto
- Il numero di clienti non è additivo rispetto al prodotto.

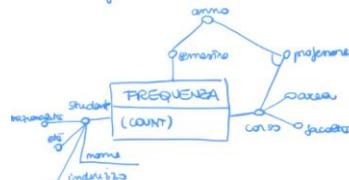
Archi multipli

Possiamo anche rappresentare degli archi multipli; permette di **rappresentare una relazione molti a molti**.



Schemi di fatto vuoti

Uno schema di fatto si dice vuoto se non ha misure. Serve solo a **registrare il verificarsi di un evento**.



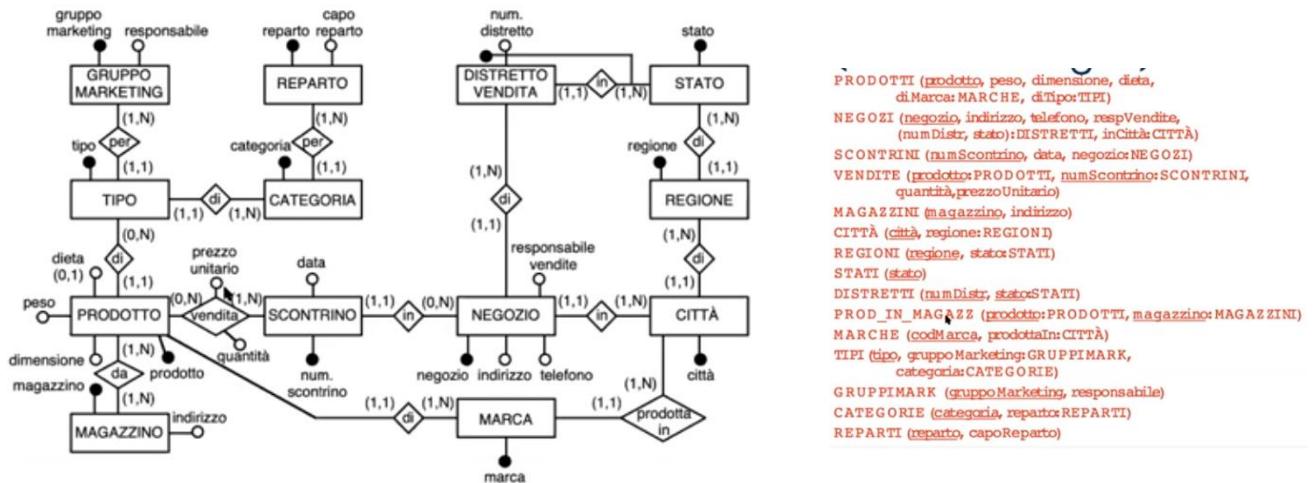
PROGETTAZIONE CONCETTUALE

Ci sono due approcci:

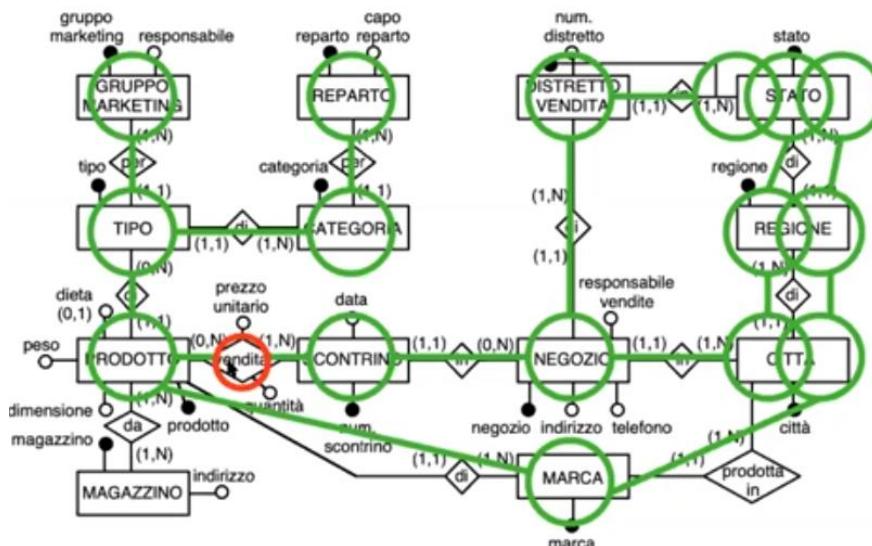
<i>Approccio basato sui requisiti</i>	<i>Approccio basato sulle sorgenti</i>
<p>Il progettista deve sfruttare le interviste e usarle per capire i fatti, le misure e le gerarchie. Questo approccio è difficoltoso, perché NON tiene conto delle sorgenti (spesso eterogenee)</p>	<p>È possibile, in maniera pseudoautomatica, creare degli schemi. Noi vediamo questo.</p> <p>Per costruire gli schemi di fatto:</p> <ol style="list-style-type: none"> 1. Definizione dei fatti. Serve perché il DW si basa sui fatti! Qui potrebbe essere utile usare i requisiti. 2. Per ogni fatto: <ol style="list-style-type: none"> a. Costruire l'albero degli attributi b. Definizione delle dimensioni c. Definizione delle misure d. Creazione dello schema di fatto con il DFM.

L'approccio funzionale è quello che unisce entrambi.

Esempio: le vendite



1. Riconosco il fatto d'interesse – per noi, la vendita – e le possibili dimensioni.



2. Costruisco l'albero degli attributi

<pre> graph TD A((prodotto + num scontum)) --- B[quantità] B --- C["prodotto + num scontum"] C --- D["prezzo unitario"] </pre>	La radice dell'albero corrisponde all'identificatore del fatto
<pre> graph TD A((prodotto + num scontum)) --- B[regione] B --- C[stato] C --- D[dati] D --- E[prodotti] E --- F[prodotto + num scontum] F --- G["prezzo unitario"] G --- H[gruppo marketing] H --- I[categoria] I --- J[tipi] J --- K[gruppo marketing] K --- L[reparto] L --- M[neopartito] M --- N[adattamento] N --- O[adattamento] O --- P[adattamento] P --- Q[adattamento] Q --- R[adattamento] R --- S[adattamento] S --- T[adattamento] T --- U[adattamento] U --- V[adattamento] V --- W[adattamento] W --- X[adattamento] X --- Y[adattamento] Y --- Z[adattamento] Z --- AA[adattamento] AA --- BB[adattamento] BB --- CC[adattamento] CC --- DD[adattamento] DD --- EE[adattamento] EE --- FF[adattamento] FF --- GG[adattamento] GG --- HH[adattamento] HH --- II[adattamento] II --- JJ[adattamento] JJ --- KK[adattamento] KK --- LL[adattamento] LL --- MM[adattamento] MM --- NN[adattamento] NN --- OO[adattamento] OO --- PP[adattamento] PP --- QQ[adattamento] QQ --- RR[adattamento] RR --- SS[adattamento] SS --- TT[adattamento] TT --- UU[adattamento] UU --- VV[adattamento] VV --- WW[adattamento] WW --- XX[adattamento] XX --- YY[adattamento] YY --- ZZ[adattamento] ZZ --- AAA[adattamento] AAA --- BBB[adattamento] BBB --- CCC[adattamento] CCC --- DDD[adattamento] DDD --- EEE[adattamento] EEE --- FFF[adattamento] FFF --- GGG[adattamento] GGG --- HHH[adattamento] HHH --- III[adattamento] III --- JJJ[adattamento] JJJ --- KKK[adattamento] KKK --- LLL[adattamento] LLL --- MLL[adattamento] MLL --- NLL[adattamento] NLL --- OLL[adattamento] OLL --- PLL[adattamento] PLL --- QLL[adattamento] QLL --- RLL[adattamento] RLL --- SLL[adattamento] SLL --- TLL[adattamento] TLL --- ULL[adattamento] ULL --- VLL[adattamento] VLL --- WLL[adattamento] WLL --- XLL[adattamento] XLL --- YLL[adattamento] YLL --- ZLL[adattamento] ZLL --- AAAA[adattamento] AAAA --- BBBB[adattamento] BBBB --- CCCC[adattamento] CCCC --- DDDD[adattamento] DDDD --- EEEE[adattamento] EEEE --- FFFF[adattamento] FFFF --- GGGG[adattamento] GGGG --- HHHH[adattamento] HHHH --- IIII[adattamento] IIII --- JJJJ[adattamento] JJJJ --- KKKK[adattamento] KKKK --- LLLL[adattamento] LLLL --- MLLL[adattamento] MLLL --- NLLL[adattamento] NLLL --- OLLL[adattamento] OLLL --- PLLL[adattamento] PLLL --- QLLL[adattamento] QLLL --- RLLL[adattamento] RLLL --- SLLL[adattamento] SLLL --- TLLL[adattamento] TLLL --- ULLL[adattamento] ULLL --- VLLL[adattamento] VLLL --- WLLL[adattamento] WLLL --- XLLL[adattamento] XLLL --- YLLL[adattamento] YLLL --- ZLLL[adattamento] ZLLL --- AAAAA[adattamento] AAAAA --- BBBBB[adattamento] BBBBB --- CCCCC[adattamento] CCCCC --- DDDDD[adattamento] DDDDD --- EEEEE[adattamento] EEEEE --- FFFFF[adattamento] FFFFF --- GGGGG[adattamento] GGGGG --- HHHHH[adattamento] HHHHH --- IIIII[adattamento] IIIII --- JJJJJ[adattamento] JJJJJ --- KKKKK[adattamento] KKKKK --- LLLLL[adattamento] LLLLL --- MLLLL[adattamento] MLLLL --- NLLLL[adattamento] NLLLL --- OLLLL[adattamento] OLLLL --- PLLLL[adattamento] PLLLL --- QLLLL[adattamento] QLLLL --- RLLLL[adattamento] RLLLL --- SLLLL[adattamento] SLLLL --- TLLLL[adattamento] TLLLL --- ULLLL[adattamento] ULLLL --- VLLLL[adattamento] VLLLL --- WLLLL[adattamento] WLLLL --- XLLLL[adattamento] XLLLL --- YLLLL[adattamento] YLLLL --- ZLLLL[adattamento] ZLLLL --- AAAAA </pre>	Aggiungo eventuali attributi che trovo sulla radice nell'ER
<pre> graph TD A((prodotto + num scontum)) --- B[regione] B --- C[stato] C --- D[dati] D --- E[prodotti] E --- F[prodotto + num scontum] F --- G["prezzo unitario"] G --- H[gruppo marketing] H --- I[categoria] I --- J[tipi] J --- K[gruppo marketing] K --- L[reparto] L --- M[neopartito] M --- N[adattamento] N --- O[adattamento] O --- P[adattamento] P --- Q[adattamento] Q --- R[adattamento] R --- S[adattamento] S --- T[adattamento] T --- U[adattamento] U --- V[adattamento] V --- W[adattamento] W --- X[adattamento] X --- Y[adattamento] Y --- Z[adattamento] Z --- AA[adattamento] AA --- BB[adattamento] BB --- CC[adattamento] CC --- DDD[adattamento] DDD --- EEE[adattamento] EEE --- FFF[adattamento] FFF --- GGG[adattamento] GGG --- HHH[adattamento] HHH --- III[adattamento] III --- JJJ[adattamento] JJJ --- KKK[adattamento] KKK --- LLL[adattamento] LLL --- MLL[adattamento] MLL --- NLL[adattamento] NLL --- OLL[adattamento] OLL --- PLL[adattamento] PLL --- QLL[adattamento] QLL --- RLL[adattamento] RLL --- SLL[adattamento] SLL --- TLL[adattamento] TLL --- ULL[adattamento] ULL --- VLL[adattamento] VLL --- WLL[adattamento] WLL --- XLL[adattamento] XLL --- YLL[adattamento] YLL --- ZLL[adattamento] ZLL --- AAAA[adattamento] AAAA --- BBBB[adattamento] BBBB --- CCCC[adattamento] CCCC --- DDDD[adattamento] DDDD --- EEEE[adattamento] EEEE --- FFFFF[adattamento] FFFFF --- GGGGG[adattamento] GGGGG --- HHHHH[adattamento] HHHHH --- IIIII[adattamento] IIIII --- JJJJJ[adattamento] JJJJJ --- KKKKK[adattamento] KKKKK --- LLLLL[adattamento] LLLLL --- MLLLL[adattamento] MLLLL --- NLLLL[adattamento] NLLLL --- OLLLL[adattamento] OLLLL --- PLLLL[adattamento] PLLLL --- QLLLL[adattamento] QLLLL --- RLLLL[adattamento] RLLLL --- SLLLL[adattamento] SLLLL --- TLLLL[adattamento] TLLLL --- ULLLL[adattamento] ULLLL --- VLLLL[adattamento] VLLLL --- WLLLL[adattamento] WLLLL --- XLLLL[adattamento] XLLLL --- YLLLL[adattamento] YLLLL --- ZLLLL[adattamento] ZLLLL --- AAAAA </pre>	Ogni vertice corrisponde a un attributo dello schema sorgente. Collego tutto, navigando l'ER e senza davver pensare a come poi rappresenterò il tutto in uno schema di fatto

3. Editing dell'albero

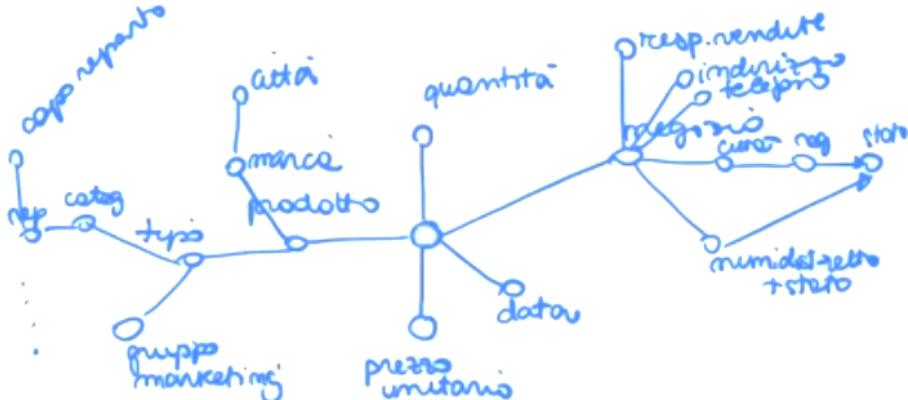
Le operazioni partono dal presupposto che non tutti gli attributi sono interessanti per il data mart. Possiamo quindi manipolare l'albero per eliminare i livelli di dettaglio non necessari.

Potatura di un vertice v	Innesto
Consiste nell'eliminazione dell'intero sottoalbero che ha v per radice.	Viene utilizzato per mantenere il sottoalbero di un vertice v che è stato eliminato.

Esempio

POTATURA di Regione, Dimentezione, Num distretto

POTATURA num.scontum con innesto



4. Definizione di dimensioni

Le dimensioni devono essere scelte tra i vertici figli della radice. È importante capire che scegliere la dimensione implica rispetto a quali punti di vista potrò fare l'analisi.

Il tempo dovrebbe sempre essere una dimensione: una delle caratteristiche del DW è che vogliamo rappresentare lo storico, e che i fatti sono dinamici.

Nell'esempio delle vendite, alcune dimensioni papabili sono prodotto, data e negozio.

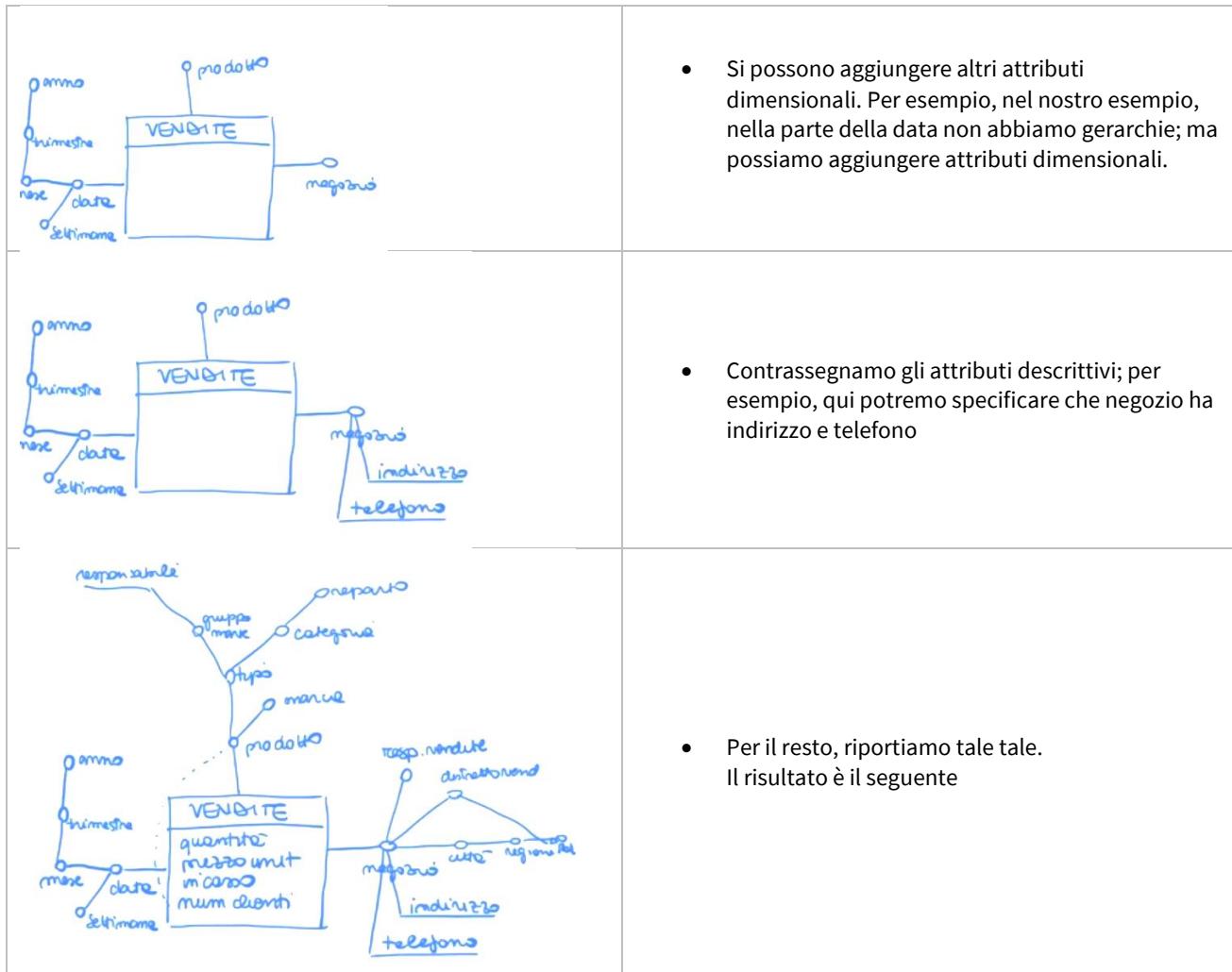
5. Definizione delle misure

Se tra le dimensioni compaiono tutti gli attributi che costituiscono un identificatore di fatto, allora le misure corrispondono agli attributi numerici che sono figli della radice dell'albero.

Nell'esempio, quantità e prezzo unitario

6. Creazione dello schema di fatto

Le gerarchie delle dimensioni corrispondono ai sottoalberi degli attributi, con radice nelle diverse dimensioni. Devo aggiungere il sottoalbero delle varie dimensioni all'opportuna dimensione.



Negli esercizi proveremo anche a costruire lo schema di fatto dall'analisi dei requisiti.

PROGETTAZIONE LOGICA

Come per le DB, la progettazione logica è fatta a valle di quella concettuale; quindi ci domandiamo quale modello dei dati scegliere per rappresentare le informazioni che abbiamo in un DW. Per la concettuale abbiamo usato il DFM, che però non si occupa assolutamente dell'implementazione, e abbiamo costruito uno schema concettuale direttamente dagli schemi del DB (e not yet dai requisiti).

Abbiamo due possibili modelli logici:

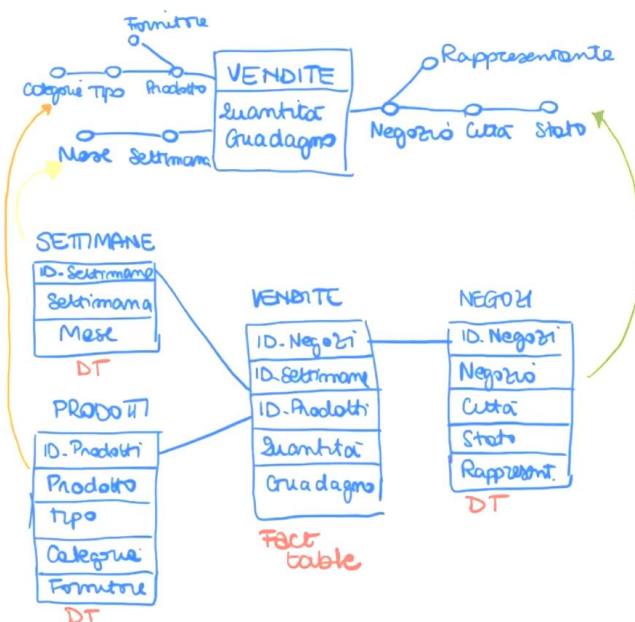
MOLAP		ROLAP	
Multidimensional on-line analytical processing; usa strutture multidimensionali. Non sono utilizzati.		Relational on-line analytic processing Modella lo schema per mezzo del relazionale.	
<ul style="list-style-type: none"> + Rappresentazione naturale per i dati: la definizione si basa su una serie di dimensioni, come il concettuale. + Ottime prestazioni anche dal punto di vista delle interrogazioni, perché non deve simulare operazioni sui cubi via SQL. 	<ul style="list-style-type: none"> - Problema della sparsità: mediamente abbiamo solo il 20% delle celle che ha un valore, perché non + per nulla detto che ogni giorno ogni prodotto sia venduto. Questo significa memorizzare celle per fatti non accaduti. - Non ci sono strutture dati standard come per il relazionale. - Mancanza di esperienza rispetto al relazionale, e quindi i progettisti sono riluttanti. 	<ul style="list-style-type: none"> + Risolve la sparsità, perché solo le tuple esistenti sono memorizzate + Usiamo il relazionale e quindi strutture ben conosciute 	<ul style="list-style-type: none"> - Sfida del dover costruire query complesse per simulare operazioni sul cubo.

Schema a stella

Procediamo quindi col ROLAP. La modellazione dei dati nel relazionale avviene attraverso il cosiddetto schema a stella, e sulle sue varianti. Uno schema a stella è uno schema relazionale composto da:

- Un insieme di relazioni DT_1, \dots, DT_n chiamate **dimension table**, ciascuna delle quali corrisponde a un dimensione.
- Ciascuna dimension table DT_i è caratterizzata da una chiave primaria d_i , tipicamente chiave surrogata, e da un insieme di attributi che descrivono le dimensioni. Di analisi a diversi livelli di aggregazione.
- Una relazione FT , chiamata **fact table**, che importa le chiavi d tutte le dimension table. La chiave primaria di FT è data dall'insieme delle chiavi esterne delle dimension table d_1, \dots, d_n . FT contiene inoltre un attributo per ogni misura.

Esempio



VENDITE (ID.Negozio, ID.Settimana, ID.Prodotto, Quantità, Guadagno)
 SETTIMANE (ID.Settimana, Settimana, Mese)
 NEGOZI (ID.Negozio, Negozio, Città, Stato, Rappresent.)
 PRODOTTI (ID.Prodotto, Prodotto, tipo, Categorie, Fornitore)

Esempio di istanza:

ID.Negozi	Negozi	Città	Stato	Rappresentante
1	N1	RM	I	R1
2	N2	RM	I	R1
3	N3	M1	I	R2
4	N4	M1	I	R2

ID.Negozi	ID.Sett	ID.Prodotti	Quantità	Guadagno
1	1	1	100	100
1	2	1	150	150
3	3	4	350	350
4	4	4	200	200

ID.Sett	Settimana	Mese
1	Gen1	Gen
2	Gen2	Gen
3	Feb1	Feb
4	Feb2	Feb

ID.Prod	Prodotti	Tipo	Categoria	Forn
1	P1	A	X	F1
2	P2	A	X	F1
3	P3	B	X	F2
4	P4	B	X	F2

! le DT sono completamente denormalizzate; abbiamo una denormalizzazione che porta ad una forte ridondanza dei dati, ma che permette di fare delle interrogazioni decenti usando una join.
! No sparsità dei dati.

Esempio interrogazione

Ovviamente sono in SQL. Per rappresentare le interrogazioni OLAP su schemi a stella, dato lo schema, posso esprimere una query del tipo

```
SELECT città, settimana, tipo,
SUM(Quantità)
FROM settimane, negozi, prodotti,
vendite
WHERE settimane.id_settimane =
vendite.id_settimane
AND negozi.id_negozi =
vendite.id_negozi
AND prodotti.id_prodotti =
vendite.id_prodotti // ha fatto il join fra FT e DT con le chiavi
AND prodotti.categoria =
'alimentari'
GROUP BY città, settimane, tipo
```

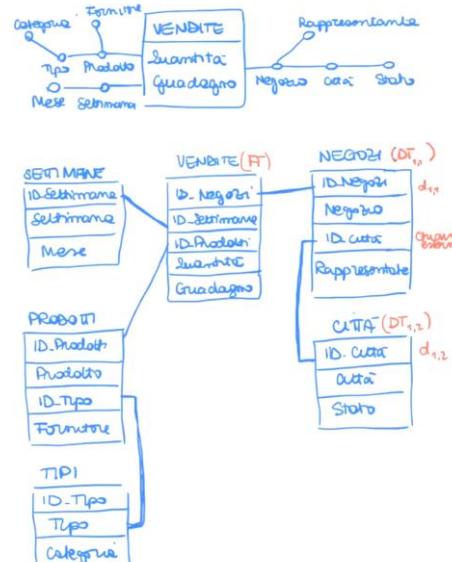
Schema a fiocco di neve

Permette di ridurre la denormalizzazione dello schema a stella. Lo schema a fiocco di neve riduce la denormalizzazione delle dimension table schemi a stella. Lo fa eliminando alcune delle dipendenze transitive.

Le dimension table $DT_{i,j}$ sono caratterizzate da :

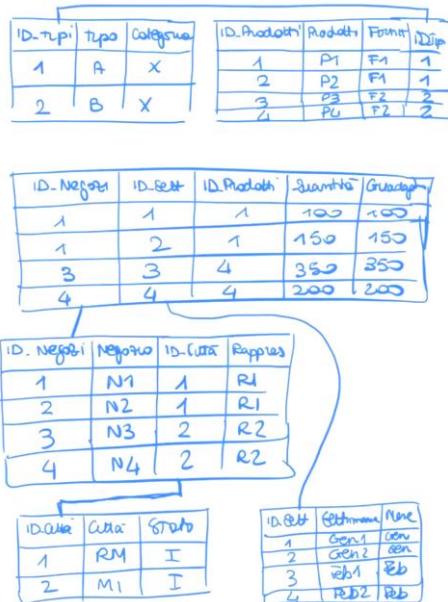
- Una chiave primaria surrogata $d_{i,j}$, come prima
- Il sottoinsieme degli attributi di DT_i che dipendono funzionalmente dalla chiave $d_{i,j}$
- Zero o più chiavi esterne importate dalle altre dimension table $DT_{i,k}$, che servono a ricostruire la dimension table che ho "spartito".

Definiamo primarie le dimension table le cui chiavi sono importate nella fact table, e secondarie le rimanenti.



Esempio dell'istanza: meno ridondante 😊

Ma poi se da una parte riduco la ridondanza, dall'altra le query saranno complesse se coinvolgono roba dalle secondarie. Aka, se nelle query non uso le secondarie questa soluzione è vantaggiosa perché riduco la ridondanza; altrimenti poi mi aumenta il tempo di esecuzione delle interrogazioni a causa delle join.



Esempio interrogazione

```

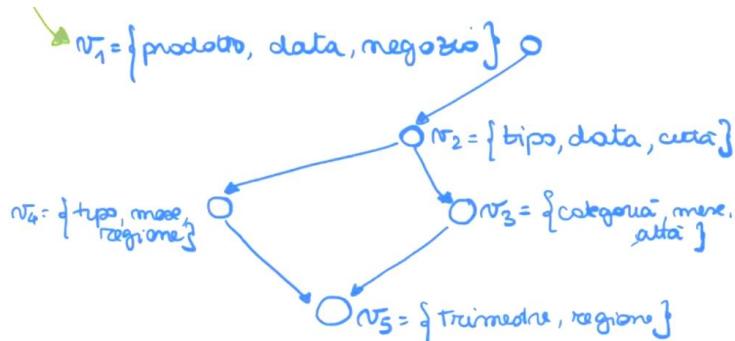
SELECT città, settimana, tipo,
SUM(Quantità)
FROM settimane, negozi, prodotti,
vendite, città, tipi
WHERE settimane.id_settimane =
vendite.id_negozi
AND negozi.id_negozi =
vendite.id_negozi
AND negozi.id_città =
città.id_città
AND prodotti.id_prodotti =
vendite.id_prodotti
AND prodotti.ID_tipo =
tipi.id_tipo
AND prodotti.categoria =
'Alimentari'
GROUP BY città, settimane, tipo
    
```

Viste

Difficilmente le interrogazioni di analisi coinvolgeranno cose nel dettaglio, perché l'obiettivo è aggregare. Di conseguenza nasce il concetto di vista: sono delle fact table contenenti dati aggregati.

Le viste possono essere identificate in base al livello (pattern) di aggregazione che le caratterizza.

- **Vista primaria.** Vista che corrisponde al pattern di aggregazione non aggregato, ovvero quello di cui abbiamo parlato fino ad ora. È il più fine e definito dalle dimensioni
- **Viste secondarie.** Pattern di aggregazioni secondari, ovvero aggregati.



La soluzione più semplice per memorizzare le viste è usare lo schema a stella, ovvero mettere tutti i dati in una sola fact table.

La dimensione di questa fact table, ovviamente, andrà crescendo; le dimension table conterranno le tuple relative ai vari livelli di aggregazione.

Il valore null viene usato per identificare l'origine delle tuple, ovvero nei campi in cui il livello di aggregazione è più fine rispetto a quello su cui si sta operando.

Esempio

ID-Negozio	ID-Sett	ID-Prodotti	Quantità	Guadagno
1	1	1	100	100
1	2	1	150	150
3	3	4	350	350
4	4	4	200	200
2	1	5	3600	3600
1	3	6	2400	2400
1	1	7	1000	1000

ID-Prodotti	Prodotto	Tipo	Categ	Fornito
1	P1	A	X	F1
2	P2	A	X	F1
3	P3	B	X	F2
4	P4	B	X	F2
5	NULL	A	X	F1
6	NULL	B	X	F2
7	NULL	NULL	NULL	F2

Pattern { Negozio, Settimane, Prodotti }

Pattern { Negozio, Settimane, Tipo }

Pattern { Negozio, Settimane, Fornitore }

ANALISI IN UN DATA WAREHOUSE

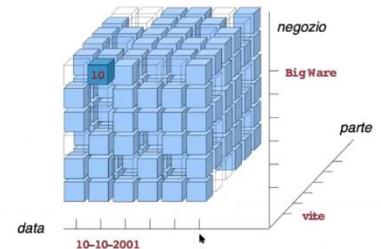
Recap

Il modello multidimensionale

Quando parliamo di un data warehouse abbiamo a che fare col modello multidimensionale, dove il fatto di interesse può essere rappresentato per mezzo di un cubo in cui:

- Ogni cella contiene misure numeriche che quantificano il fatto
- Una dimensione è un punto di vista dell'analisi
- Ogni dimensione può essere la radice di una gerarchia.

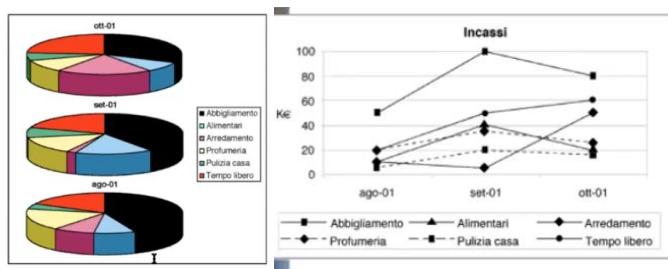
Attenzione: in realtà, possiamo avere anche più di 3 dimensioni! Ci fermiamo a 3 per comodità grafica.



Tecniche di analisi dei dati

Reportistica

Costruzione di report a partire dai dati nel DW, solitamente costruiti ad intervalli di tempo fissi. È un tipo di analisi fattibile anche con DB classici.



OLAP

Abbiamo a che fare con utenti che analizzano in maniera attiva, ovvero l'utente ha delle sessioni OLAP in cui ogni passo è l'input del passo successivo. L'utente del DW di solito non ha competenze informatiche ma di marketing.

Sessione olap

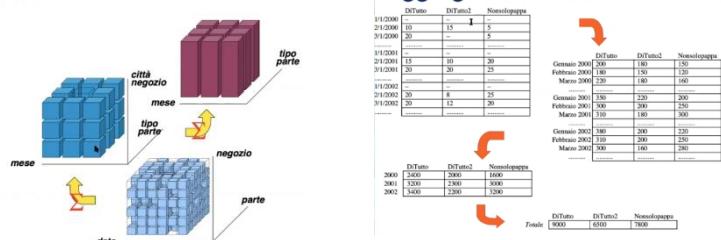
È il percorso di navigazione fatto dall'utente per analizzare uno o più fatti di interesse, su più dimensioni e su diversi livelli di dettaglio. Questo tipo di percorso, in pratica, è una sequenza di interrogazioni; ogni interrogazione si basa sulla precedente. L'interrogazione è analoga all'affettamento del cubo. La sessione di interrogazione di analisi olap, in pratica, è una sequenza di interrogazioni SQL su un diagramma a stella.

Il risultato della sessione olap è di tipo multidimensionale; se la guardo sul cubo ho una porzione del cubo.

Aggregazione

La logica dell'aggregazione è di partire dal cubo di partenza e, come una sommatoria, riunire i cubetti in cubi maggiori.

È importante avere la possibilità di fare proprio questi passaggi in una sessione olap, per la logica del DW secondo cui voglio seguire la logica del pensiero step by step.



Operatori

Roll-up

Elimina un livello di dettaglio, aumentando l'aggregazione dei dati. Posso eventualmente anche tornare indietro.

OLAP: operatori

Incassi Mensili Anni 97 e 98 per Regione del Cliente

Metrics	Dollar Sales											
	Customer	Region	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany
Month												
Jan 97	\$ 420		\$ 750	\$ 30	\$ 460	\$ 240	\$ 1.312	\$ 440	\$ 1.020	\$ 1.020	\$ 393	\$ 210
Feb 97	\$ 258		\$ 252	\$ 800	\$ 975	\$ 360	\$ 780	\$ 744	\$ 310	\$ 799	\$ 118	\$ 357
Mar 97	\$ 244		\$ 244	\$ 140	\$ 250	\$ 1.05	\$ 2.061	\$ 650	\$ 1.240	\$ 119	\$ 142	\$ 94
Apr 97	\$ 250		\$ 250	\$ 250	\$ 250	\$ 250	\$ 250	\$ 250	\$ 250	\$ 250	\$ 250	\$ 250
May 97	\$ 1.350		\$ 245	\$ 936	\$ 159	\$ 664	\$ 626	\$ 107	\$ 125	\$ 125	\$ 177	\$ 230
Jun 97	\$ 1.350		\$ 245	\$ 936	\$ 159	\$ 664	\$ 626	\$ 107	\$ 125	\$ 125	\$ 177	\$ 230
Jul 97	\$ 652		\$ 652	\$ 652	\$ 481	\$ 1.293	\$ 605	\$ 303	\$ 618	\$ 103	\$ 124	\$ 173
Aug 97	\$ 1.350		\$ 245	\$ 936	\$ 159	\$ 664	\$ 626	\$ 107	\$ 125	\$ 125	\$ 177	
Sep 97	\$ 2.291		\$ 2.291	\$ 1.032	\$ 1.170	\$ 398	\$ 396	\$ 452	\$ 390	\$ 241	\$ 407	\$ 258
Oct 97	\$ 1.350		\$ 245	\$ 936	\$ 159	\$ 664	\$ 626	\$ 107	\$ 125	\$ 125	\$ 177	
Nov 97	\$ 261		\$ 1.588	\$ 343	\$ 118	\$ 1.459	\$ 635	\$ 201	\$ 250	\$ 230	\$ 119	\$ 389
Dec 97	\$ 2.291		\$ 2.291	\$ 1.032	\$ 1.170	\$ 398	\$ 396	\$ 452	\$ 390	\$ 241	\$ 407	
Jan 98	\$ 2.510		\$ 762	\$ 1.123	\$ 1.236	\$ 2.027	\$ 2.087	\$ 546	\$ 266	\$ 277	\$ 292	
Feb 98	\$ 2.459		\$ 1.023	\$ 1.178	\$ 708	\$ 1.420	\$ 3.314	\$ 1.940	\$ 1.705	\$ 276	\$ 1.246	
Mar 98	\$ 667		\$ 667	\$ 1.721	\$ 440	\$ 1.148	\$ 80	\$ 1.310	\$ 303	\$ 154	\$ 457	\$ 45
Apr 98	\$ 667		\$ 667	\$ 1.721	\$ 440	\$ 1.148	\$ 80	\$ 1.310	\$ 303	\$ 154	\$ 457	\$ 45
May 98	\$ 556		\$ 1.097	\$ 412	\$ 226	\$ 406	\$ 361	\$ 1.628	\$ 247	\$ 1.011	\$ 41	
Jun 98	\$ 556		\$ 1.097	\$ 412	\$ 226	\$ 406	\$ 361	\$ 1.628	\$ 247	\$ 1.011	\$ 41	
Jul 98	\$ 556		\$ 1.097	\$ 412	\$ 226	\$ 406	\$ 361	\$ 1.628	\$ 247	\$ 1.011	\$ 41	
Aug 98	\$ 556		\$ 1.097	\$ 412	\$ 226	\$ 406	\$ 361	\$ 1.628	\$ 247	\$ 1.011	\$ 41	
Sep 98	\$ 544		\$ 412	\$ 1.487	\$ 299	\$ 679	\$ 706	\$ 356	\$ 440	\$ 405	\$ 99	\$ 385
Oct 98	\$ 544		\$ 412	\$ 1.487	\$ 299	\$ 679	\$ 706	\$ 356	\$ 440	\$ 405	\$ 99	\$ 385
Nov 98	\$ 898		\$ 2.096	\$ 1.726	\$ 642	\$ 395	\$ 1.740	\$ 1.943	\$ 1.142	\$ 366	\$ 207	
Dec 98	\$ 2.510		\$ 762	\$ 1.123	\$ 1.236	\$ 2.027	\$ 2.087	\$ 546	\$ 266	\$ 277		

roll-up

Incassi Trimestrali per Regione del Cliente

Incassi Annuali Anni 97 e 98 per CATEGORIA di prodotto e per ciascuna Regione

Incassi Annuali per CATEGORIA di prodotto

55

Drill down

Diminuisce l'aggregazione dei dati, introducendo un ulteriore livello di dettaglio in una gerarchia.

drill-down

Incassi Trimestrali per Città del Cliente

Ci sono tanti campi vuoti perché sono le città in cui non avevo incassi.

Incassi Annuali per CATEGORIA di prodotto

drill-down

Incassi Annuali per CATEGORIA di prodotto e per Regione dei clienti

Slice and dice

Riduce la dimensionalità del cubo; passa ad avere una singola fetta o una sottoparte del cubo.

OLAP: operatori

In cassi Annuali Anni 97 e 98 per Categorie di prodotto e per ciascuna Regione

Category	Year	Metrics	Dollar Sales	Customer Region	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany
Electronics	1997		\$ 130		\$ 1.774	\$ 384	\$ 130	\$ 2.346	\$ 2.254	\$ 2.104	\$ 566	\$ 199	\$ 1	
	1998		\$ 1.750		\$ 4.529	\$ 1.092	\$ 2.232	\$ 651	\$ 4.498	\$ 4.767	\$ 1.269	\$ 462	\$ 1.7	
Food	1997		\$ 538		\$ 925	\$ 399	\$ 677	\$ 213	\$ 1.503	\$ 261	\$ 165	\$ 156	\$ 615	\$ 1
	1998		\$ 5.376		\$ 1.325	\$ 1.054	\$ 413	\$ 2.077	\$ 1.342	\$ 1.296	\$ 408	\$ 175	\$ 1.0	
Gifts	1997		\$ 424		\$ 640	\$ 1.137	\$ 647	\$ 588	\$ 1.754	\$ 654	\$ 143	\$ 101	\$ 1	
	1998		\$ 611		\$ 887	\$ 1.110	\$ 560	\$ 499	\$ 1.182	\$ 644	\$ 273	\$ 177	\$ 273	\$ 1
Household	1997		\$ 5.787		\$ 5.320	\$ 5.416	\$ 6.812	\$ 4.334	\$ 5.008	\$ 7.598	\$ 2.239	\$ 1.000	\$ 1.0	
	1998		\$ 201		\$ 399	\$ 405	\$ 185	\$ 409	\$ 323	\$ 296	\$ 105	\$ 105	\$ 105	\$ 1
Kid's Korner	1997		\$ 247		\$ 422	\$ 441	\$ 412	\$ 424	\$ 446	\$ 446	\$ 140	\$ 139	\$ 139	\$ 1
	1998		\$ 624		\$ 595	\$ 564	\$ 385	\$ 200	\$ 978	\$ 416	\$ 48	\$ 38	\$ 38	\$ 1
Travel	1997		\$ 608		\$ 559	\$ 1.098	\$ 611	\$ 464	\$ 316	\$ 973	\$ 257	\$ 190	\$ 1	

slice-and-dice

In cassi Annuali Anno 98 per Categorie di prodotto e per ciascuna Regione

Category	Year	Metrics	Dollar Sales	Customer Region	North-East	Mid-Atlantic	South-East	Central	South	North-West	South-West	England	France	Germany
Electronics	1997		\$ 1.184		\$ 4.529	\$ 1.892	\$ 7.222	\$ 651	\$ 9.489	\$ 476	\$ 2.683	\$ 462	\$ 702	\$ 1
	1998		\$ 1.005		\$ 2.785	\$ 2.405	\$ 1.675	\$ 1.813	\$ 2.844	\$ 1.778	\$ 1.195	\$ 717	\$ 686	\$ 1
Food	1997		\$ 431		\$ 887	\$ 566	\$ 382	\$ 499	\$ 1.182	\$ 1.644	\$ 273	\$ 72	\$ 1	
	1998		\$ 5.787		\$ 5.320	\$ 5.416	\$ 6.812	\$ 4.334	\$ 5.008	\$ 7.598	\$ 2.239	\$ 1.000	\$ 1.0	
Gifts	1997		\$ 247		\$ 422	\$ 441	\$ 412	\$ 424	\$ 446	\$ 446	\$ 140	\$ 139	\$ 139	\$ 1
	1998		\$ 624		\$ 595	\$ 564	\$ 385	\$ 200	\$ 978	\$ 416	\$ 48	\$ 38	\$ 38	\$ 1
Household	1997		\$ 38.383		\$ 50.391	\$ 54.616	\$ 61.412	\$ 4.334	\$ 5.008	\$ 7.598	\$ 2.239	\$ 1.000	\$ 1.0	
	1998		\$ 2.559		\$ 2.943	\$ 2.943	\$ 2.943	\$ 2.943	\$ 2.943	\$ 2.943	\$ 2.943	\$ 2.943	\$ 2.943	\$ 1
Kid's Korner	1997		\$ 4.497		\$ 4.792	\$ 4.792	\$ 4.792	\$ 4.792	\$ 4.792	\$ 4.792	\$ 4.792	\$ 4.792	\$ 4.792	\$ 1
Travel	1997		\$ 608		\$ 559	\$ 1.098	\$ 611	\$ 464	\$ 316	\$ 973	\$ 257	\$ 190	\$ 1	

Pivoting

Semplicemente cambia la modalità di presentazione / il sistema di riferimento.

Category	Year	Metrics	Dollar Sales
Electronics	1997		\$ 10.616
	1998		\$ 29.299
Food	1997		\$ 5.300
	1998		\$ 5.638
Gifts	1997		\$ 16.315
	1998		\$ 20.047
Health & Beauty	1997		\$ 6.042
	1998		\$ 5.665
Household	1997		\$ 38.383
	1998		\$ 50.391
Kid's Korner	1997		\$ 2.559
	1998		\$ 2.943
Travel	1997		\$ 4.497
	1998		\$ 4.792

Category	Year	Metrics	Dollar Sales
Electronics	1997		\$ 10.616
	1998		\$ 29.299
Food	1997		\$ 5.300
	1998		\$ 5.638
Gifts	1997		\$ 16.315
	1998		\$ 20.047
Health & Beauty	1997		\$ 6.042
	1998		\$ 5.665
Household	1997		\$ 38.383
	1998		\$ 50.391
Kid's Korner	1997		\$ 2.559
	1998		\$ 2.943
Travel	1997		\$ 4.497
	1998		\$ 4.792

Drill across

Mette insieme due o più cubi correlati per comparare i dati.

Category	Quarter	Metrics	Dollar Sales
Electronics	Q1 1997		\$ 4.883
	Q2 1997		\$ 817
Food	Q1 1997		\$ 1.546
	Q2 1997		\$ 1.910
Gifts	Q1 1997		\$ 3.598
	Q2 1997		\$ 3.893
Health & Beauty	Q1 1997		\$ 0
	Q2 1997		\$ 1.826
Household	Q1 1997		\$ 9.314
	Q2 1997		\$ 8.124
Kid's Korner	Q1 1997		\$ 685
	Q2 1997		\$ 531
Travel	Q1 1997		\$ 603

Category	Quarter	Metrics	Dollar Sales
Electronics	Q3 1997		\$ 4.883
	Q4 1997		\$ 827
Food	Q3 1997		\$ 1.546
	Q4 1997		\$ 1.268
Gifts	Q3 1997		\$ 3.598
	Q4 1997		\$ 4.582
Health & Beauty	Q3 1997		\$ 0
	Q4 1997		\$ 0
Household	Q3 1997		\$ 9.314
	Q4 1997		\$ 9.331
Kid's Korner	Q3 1997		\$ 685
	Q4 1997		\$ 531
Travel	Q3 1997		\$ 603

Data mining

E'Ha un approccio diverso, ed è un tipo di analisi orientato a scoprire informazioni nascoste; un esempio è l'analisi sui carrelli delle persone, per vedere se ci sono pattern significativi presenti non sempre visibili in presenza di mole elevata di dati.

Il data mining raccoglie tecniche di intelligenza artificiale e pattern recognition, per aiutare l'utente nella ricerca di pattern; è sufficiente indicare cosa e dove lo si vuole cercare.

In ambito clinico permette di valutare casi clinici

Regole associative

Le regole di associative permettono di determinare regole di implicazione logica presenti nella base di dati, quindi di individuare gruppi di affinità tra oggetti.

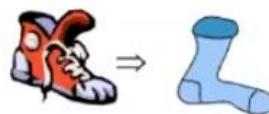
Applicazioni:

- Studio delle abitudini di acquisto (es. pubblicità mirata, organizzazione della merce sugli scaffali)
- Studio della variabilità delle vendite in assenza di un certo prodotto.

$\{scarpe\} \Rightarrow \{calze\}$

Chi compra
scarpe compra
anche calze

supporto = 70 %



confidenza = 85 %

La regola $X \Rightarrow Y$ ha supporto s se una frazione pari a s delle transazioni contengono tutti gli item in $X \cup Y$
es.: (il 70% delle transazioni include scarpe e calze)

La regola $X \Rightarrow Y$ ha confidenza c se una frazione pari a c delle transazioni in cui compare X contiene Y
es.: (l'85% delle persone che comprano scarpe comprano anche calze)

Supporto: % delle transazioni che ha entrambi gli item

Confidenza: % che nelle transazioni dove compare X prende anche Y

La regola $A \Rightarrow D$ ha

- SUPPORTO = 60% perché $\{A, D\}$ compare in 6 transazioni su 10
- CONFIDENZA = 75% perché su 8 transazioni in cui compare A, in 6 compare anche D

La regola $D \Rightarrow A$ ha

- SUPPORTO = 60%
- CONFIDENZA = 86% su 7 transazioni in cui compare D, in 6 compare anche A

TRANSAZIONE	ITEMS
1	A B C D
2	A B C
3	A D
4	B C D
5	A B D
6	B C
7	A C D
8	A B D
9	A B
10	A D

PROBLEMA: determinare tutte le regole associative che abbiano
SUPPORTO \geq minS
e CONFIDENZA \geq minC

QUALITÀ DEI DATI

È molto importante perché avere dati scadenti può avere conseguenze importanti; ad esempio, una lettera sbagliata in un indirizzo. Questo errore, che denota una scadente qualità dei dati, causa un ritardo; oppure, nei sistemi di invio automatico di mail, se ho record duplicati potrei inviare la mail più volte alla stessa persona.

Insomma, le conseguenze di una qualità scadente dei dati possono essere gravi. Il funzionamento efficiente e efficace si basa anche sulla qualità dei dati, che può avere una ricaduta anche sui processi aziendali.

Il problema della qualità dei dati è stato affrontato da diversi punti di vista, e in svariati diversi.

- **Statistica** (fine anni 60)

Alla fine degli anni 60, nel settore della statistica, gli esperti sono stati i primi a dedicarsi ai problemi della qualità dei dati quanto si resero conto dell'importanza del problema. Proposero una prima teoria matematica per verificare l'esistenza di duplicati.

- **Scienze gestionali** (inizio anni 80)

Nella gestione delle info aziendali e nei sistemi di manifattura voleva dire avere problemi in fase di produzione.

- **Informatica** (primi anni 90)

Gli esperti informatici iniziano a parlarne con la digitalizzazione, e si occupano del definire, misurare e migliorare dati prima nei DB, poi nei DW, ora nei sistemi legacy e nelle big data. Insomma, viaggia insieme alle tecnologie. Per esempio, la fase di pulitura della costruzione DW è uno dei tipi di azione sulla qualità dei dati.

Dimensioni

La qualità dei dati non è un problema semplice; non dobbiamo pensare che si riferisca solo alla accuratezza dei dati. I vari punti di vista possono essere visti come "dimensioni".

- **Accuratezza:**

In alcuni casi è facile fixare (battitura), ma in altri casi di accuratezza è più difficile (valori sbagliati)

- Errori di battitura (me ne rendo conto guardando il dato)
- Valori sbagliati su un certo attributo (non me ne rendo conto guardando i dati).

- **Completezza:**

È difficile da valutare: se manca del tutto una tupla è difficile accorgersene. È invece più facile se alla tupla manca un attributo

- **Consistenza:**

Se ho inconsistenze, potrebbe succedere che il dato non è aggiornato in uno dei posti. È difficile, perché non so quale sia il dato sbagliato fra i molteplici -> Livello di aggiornamento

ESEMPIO

FILM

ID	Titolo	Regista	Anno	#Rental	AnnoVR
1	Casablanca	Weisz	1942	3	1940
2	L'ultimo giuggente	Curtiz	1939	0	NULL
3	Vacanze Romane	Wyler	1953	0	NULL
4	Sabrina	NULL	1954	0	1985

Qualità dei dati e tipi di dati

Anche dai dati può dipendere la qualità dei dati (per esempio se so lo schema è più facile).

È importante parlare dei tipi di dati, perché le dimensioni considerate e le tecniche per il miglioramento della qualità devono essere adattate ai tipi di dati che ho a disposizione

Una prima classificazione è

- **Dati strutturati**

Struttura fissa e definita a priori : BD relazionali

- **Dati semistrutturati**

C'è una qualche struttura, ma questa è abbastanza flessibile : XML

- **Dati non strutturati**

Non hanno neanche quasi struttura flessibile; non c'è una definizione. : Linguaggio naturale

Frequenza di aggiornamento

Posso classificare in

- **Dati stabili:** il cambiamento è improbabile; ad esempio, per le pubblicazioni scientifiche ho una parte di info che non si modifica, al più aggiungo altre
- **Dati che variano nel lungo termine:** frequenza di aggiornamento molto bassa, ad esempio listini dei prezzi, delle palestre... frequenza tipo 1 volta l'anno.
! Il concetto di frequenza di aggiornamento dipende dal contesto.
- **Dati che cambiano di frequente**
Ho una frequenza di aggiornamento elevata. Ad esempio sensori termici, traffico, sensori vitali umani
Avere dati di buona qualità significa poter prendere decisioni migliori.

Scarsa qualità dei dati

- **Cambiamenti storici:** l'importanza di determinata info può cambiare rispetto al tempo. Ad esempio, info importanti ora non lo erano nel passato, e quindi cambia anche la qualità.
- **Modo di utilizzo:** a seconda del tipo di processo che metto in atto saranno importanti dati diversi (es. operazionale != processo decisionale)
- **Privacy:** la privacy sta prendendo sempre più piede, quindi quello che posso fare coi dati è sempre più ristretto. Anche la gestione corretta fa cambiare la valutazione di qualità.

Metodologia

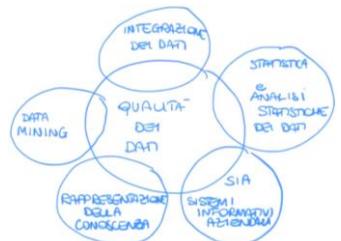
Per migliorare la qualità dei dati esiste una metodologia che

- **Parte dalla definizione :** rispetto a quale dimensione voglio migliorare? identifico le dimensioni che voglio considerare per valutare la qualità dei dati
- **Misura :** identificate le dimensioni, vogliamo valutare la qualità dei dati.
- **Analisi:** evidenziare eventuali problemi
- **Miglioramento:** metto in atto azioni per migliorare la qualità.

Ricerca

Gli aspetti relativi alla qualità di dati e il suo studio è molto recente.

- **Statistica:** lo studio dei problemi comprende dei metodi usati per raccogliere, analizzare e presentare i dati dal punto di vista statistico. Sono stati sviluppati metodi e modelli per esprimere predizioni e formulare decisioni.
Le problematiche si occupano soprattutto di:
 - Riassumere, descrivere ed esplorare i dati
 - Dedurre la natura del processo che li ha prodotti.
- **Rappresentazione della conoscenza**
Riguarda lo studio di come può essere rappresentata la conoscenza su un certo dominio applicativo, e quali tipi di ragionamento possono essere fatti.
Ovviamente la qualità dipende dal fatto di rappresentare in maniera corretta la qualità dei dati. Alcuni esempi di rappresentazioni sono formule logiche, attivazione di nodi neurali...
Le tecniche di ragionamento automatiche sono uno strumento importante per il miglioramento della qualità dei dati.
- **Data mining**
Abbiamo già accennato per i DW; è un processo analitico che permette di esplorare grandi insiemi di dati, alla ricerca di pattern. Esistono anche tipi di DM che permettono di estrarre la struttura dei dati.
Una buona qualità dei dati è fondamentale per il DM. Se non ho dati di buona qualità, la scoperta dei pattern viene gravemente compromessa.
- **Sistemi informativi aziendali/gestionali**
Abbiamo parlato dell'informazione come risorsa aziendale, e dati/conoscenza sono fondamentali. Dati di qualità scadente implicano processi e risultati scadenti.



I SIA gestiscono l'informazione; dati e conoscenza sono fondamentali risorse dei processi aziendali, e di conseguenza dati di qualità scadente hanno ricadute sulla esecuzione dei processi aziendali.

- **Processi decisionali**