

Sistemi Informativi

Colognese, Rossini

Febbraio - Marzo 2018

Teoria dell'Organizzazione	2
Organizzazione Aziendale	2
Tecnologie Informatiche	2
Gestione dell'Informazione	3
Organizzazione come Sistema Aperto	3
Modello Gerarchico (Piramide di Anthony)	3
Progettazione del Sistema Informativo	4
Incertezza Comportamentale	4
Classificazione dei Sistemi Informativi	5
Segmentazione dei Bisogni: il Portafoglio Applicativo	5
Informatizzazione del Portafoglio Operativo nelle Imprese Manifatturiere	5
Enterprise Resource Planning (ERP)	6
SAP (multinazionale)	7
Il Portafoglio Operativo delle Società di Servizi	7
CRM e Aziende di Utility	8
CRM e Business Intelligence	8
Ingegneria Sociale	9
Fasi dell'Attacco	9
Fase Fisica	9
Fase Psicologica	10
Ingegneria dei Processi Gestionali	11
Griglia Metodologica	12
Variabili Organizzative	12
Fasi di Analisi	13
Trasformazione dei Processi	13
Business Process Model and Notation (BPMN)	14
Modellazione di processi e dati	14
Data Warehousing	15
Introduzione al Data Warehousing	15
Architettura a 2 Livelli	16
Architettura a 3 Livelli	16
Il modello Multidimensionale	17
Il Ciclo di Vita del Data Warehouse	18
Analisi e Riconciliazione delle Sorgenti Operazionali	19
Analisi dei Requisiti	19
Progettazione Concettuale	20
Carico di Lavoro e Volume Dati	22
Progettazione Logica	23
Sistemi MOLAP	23
Sistemi ROLAP: Schema a Stella	23
Snowflake Schema	23

Teoria dell'Organizzazione

Il **Sistema Informativo** è la componente di una *organizzazione* (persone e tecnologie) che *gestisce* (analizza, elabora, estrapola) le *informazioni di interesse* (utilizzate e prodotte durante l'esecuzione dei processi).

Organizzazione Aziendale

L'**Organizzazione** è il processo attraverso il quale un insieme di persone viene strutturato secondo i principi di divisione del lavoro e coordinamento, diventando un sistema (sinonimo di azienda).

Un'azienda è un'organizzazione di persone e tecnologie finalizzata alla soddisfazione di bisogni umani attraverso la produzione, la distribuzione o il consumo di beni economici.

- Dall'interazione tra risorse umane e tecnologie deriva il **comportamento aziendale**, rivolto al raggiungimento degli obiettivi, che produce dei risultati.

L'**Organizzazione Aziendale** è il processo attraverso il quale l'insieme di persone che partecipano direttamente allo svolgimento delle attività dell'azienda viene strutturato secondo i principi di divisione del lavoro e coordinamento (ruoli e funzioni diverse, seguendo scopi, metodi e regole). Contiene dei macroprocessi (operativo/produttivo, controllo e gestione) e dispone di risorse (umane, materiali, *informative*).

Teoria dell'Organizzazione: studio e progettazione scientifica dei compiti per migliorare le prestazioni.

Tecnologie Informatiche

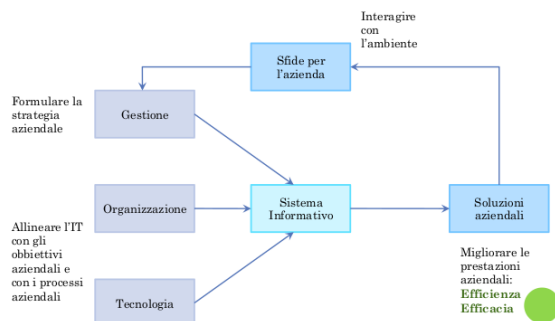
Le **Tecnologie Informatiche** sono un insieme di sistemi, strumenti e tecniche per automatizzare il trattamento delle informazioni. Portano innovazione scientifica e tecnica, riduzione dei tempi e costi di produzione. Supportano il ciclo di vita dell'informazione come risorsa aziendale (è oggetto di processi produttivi, operativi e attività gestionali).

Le **Infrastrutture di Tecnologie Informatiche** sono un insieme di risorse tecnologiche condivise che fornisce una piattaforma per le applicazioni informatiche di un'azienda. Include investimenti in hardware, software e servizi (consulenze, aggiornamenti e training del personale).

Un **Sistema Informativo Aziendale** è un insieme di elementi interconnessi che raccolgono, catalogano, ricercano, elaborano, memorizzano e distribuiscono dati trasformandoli in informazioni utili per supportare le attività decisionali di controllo di un'azienda.

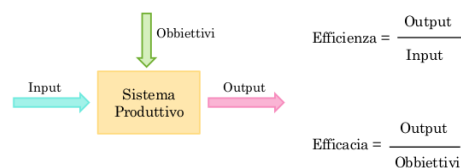
Deve essere progettato per svolgere tre macroprocessi: acquisizione dei dati (processo di Input), trasformazione dei dati (processo di Elaborazione) e restituzione di informazioni (processo di Output).

I **sistemi informativi** svolgono la funzione di automatizzare la gestione e l'elaborazione dei dati.



Il **sistema produttivo** si basa su: obiettivi (output atteso), input e output (effettivo).

- **Efficienza**: costo per il raggiungimento degli obiettivi.
- **Efficacia**: grado di raggiungimento degli obiettivi.



Le *tecnologie informatiche* hanno un impatto positivo sull'**efficienza** dei sistemi produttivi: riducono i costi unitari, aumentano la produzione a parità di risorse (aspetto quantitativo), incentiva la crescita delle dimensioni organizzative e cambia la struttura organizzativa.

Le *tecnologie informatiche* hanno aumentato l'**efficacia** dei sistemi produttivi: uso più efficiente dei fattori produttivi a parità di volumi di produzione, razionalizzazione dell'uso delle risorse (aspetto qualitativo), differenziazione dei prodotti e ampliamento della gamma.

A livello operativo le *tecnologie informatiche* hanno semplificato i compiti individuali, dato una maggiore specializzazione, aumentato l'interdipendenza e incrementato la complessità gestionale.

Gestione dell'Informazione

La *gestione dell'informazione* è quindi fondamentale per il raggiungimento degli scopi aziendali:

- **informazione** come **risorsa organizzativa** fondamentale.

L'**informazione** è la risorsa principale: viene scambiata, selezionata ed elaborata nelle attività gestionali di controllo e coordinamento.

Viene prodotta da qualunque attività (anche operativa) ma è una *risorsa immateriale* che sta alla base di ogni altra risorsa immateriale (conoscenza ed esperienza individuale ed organizzativa). Non viene distrutta dall'uso, può essere soggetta ad obsolescenza, non si esaurisce ma *si auto-rigenera* (cioè estraggo un'informazione e da essa ne posso dedurre altre, incrementando le prestazioni dei processi gestionali).

La quantità di informazione posseduta può causare:

- **Overload informativo**: aumento incontrollato dell'informazione disponibile, eccedendo le capacità di elaborazione individuale; causando rallentamento e peggioramento delle decisioni.
- **Underload informativo**: disponibilità di informazioni al di sotto delle capacità di elaborazione individuale; semplificazione delle decisioni che vengono prese in tempi più brevi.

Organizzazione come Sistema Aperto

In funzione delle opportunità fornite dall'*ambiente esterno* e tenendo conto dei vincoli da esso posti, l'azienda definisce i propri obiettivi.

Deve inoltre sapersi adattare ai cambiamenti dell'ambiente esterno (nuovi requisiti come ad esempio un cambio di mode), dunque all'**incertezza ambientale**. Esiste un **cruscotto aziendale** che prende decisioni strategiche in base ai dati che arrivano dall'esterno.

L'*informazione* gioca un ruolo fondamentale nel determinare la capacità di adattamento come risorsa delle attività gestionali di ripianificazione organizzativa.

Capacità elaborativa: adeguatezza di un'organizzazione rispetto alle necessità di elaborare informazioni a essa imposte dai propri obiettivi e dal contesto in cui opera.

Modello Gerarchico (Piramide di Anthony)



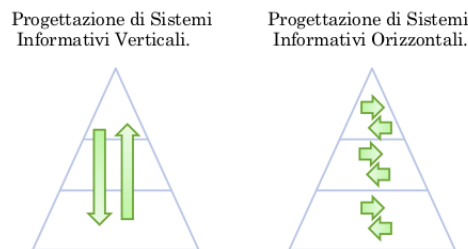
- I cambiamenti delle tendenze di mercato comportano una ripianificazione strategica degli obiettivi di produzione.
- Un malfunzionamento HW può essere gestito al livello di competenza senza far salire l'informazione al livello più alto (*se servono matite non si mandano richieste al livello strategico*).

La capacità elaborativa di un'organizzazione dovrebbe essere superiore a quella di ciascuno degli individui che la compongono. L'organizzazione supera i limiti individuali.

Progettazione del Sistema Informativo

La progettazione del Sistema Informativo è un'attività fortemente legata alla pianificazione della struttura organizzativa. Il Sistema Informativo massimizza le prestazioni dell'organizzazione.

Allineamento fra Sistema Informativo, struttura organizzativa e livello di incertezza: Come organizzare gli scambi informativi all'interno di una organizzazione a fronte di crescente incertezza ambientale.



Sistemi Informativi Verticali: storicamente i primi ad essere supportati da sistemi informatici. Accentramento delle decisioni. Al crescere dell'incertezza e del numero di eccezioni, i livelli gerarchici superiori vengono sovraccaricati dai compiti decisionali.

Limite: capacità elaborativa dell'organizzazione coincide con la capacità elaborativa dell'unità gerarchicamente più elevata (che potrebbe essere un individuo).

Sistemi Informativi Orizzontali: comunicazione fra unità organizzative allo stesso livello, anche se dipendenti gerarchicamente da unità distinte. Delega decisionale.

Collegamenti laterali a crescente capacità elaborativa:

- contatti diretti tra coloro che condividono lo stesso problema;
- gruppi temporanei di lavoro (task force) o permanenti (team);
- ruoli manageriali di integrazione per gestire le eccezioni relative ad un particolare prodotto (*product manager*), programma (*program manager*) o progetto (*project manager*);
- organizzazione a matrice, dunque ciascuna unità ha una doppia dipendenza gerarchica secondo due dimensioni rilevanti ai fini organizzativi (*ad esempio il docente sotto facoltà e dipartimento*).

Incertezza Comportamentale

Non esiste soltanto l'incertezza ambientale, ma anche l'incertezza legata ad un comportamento non prevedibile degli individui (*ad esempio un lavoratore che si sveglia male e rende di meno*).

Si deve considerare l'eventualità che un individuo decida di non scegliere la soluzione che massimizza l'efficienza e l'efficacia organizzativa (trattenendo/nascondendo informazione invece di scambiarla, non rilevando eccezioni o non eseguendo il proprio compito decisionale).

Si deve tener conto che gli individui sono mossi anche da interessi personali (spesso in conflitto con quelli organizzativi) e da comportamento opportunistico. Queste divergenze rendono necessarie delle forme di coordinamento che generano dei costi di controllo (per la verifica del comportamento) e di garanzia (produzione documentata). Diventa dunque necessario attuare una *minimizzazione dei costi* di produzione e coordinamento.

Classificazione dei Sistemi Informativi

Disposizione dei *Sistemi Informativi* lungo la piramide aziendale (definizione e funzioni attribuite dipendono dal livello al quale essi sono collocati):

- **Transaction Processing Systems (TPS):** alla base della piramide, destinati alla gestione delle transazioni, tengono traccia delle informazioni di routine nelle organizzazioni (ordini, spedizioni, ecc.);
- **Management Information Systems (MIS):** al livello superiore, destinato al management; acquisiscono dati dai TPS per fare una rappresentazione periodica della situazione delle operazioni aziendali;
- **Decision Support Systems (DSS):** allo stesso livello dei precedenti, affiancano il management nelle decisioni di non-routine per il supporto nelle decisioni aziendali; permettono di simulare ipotesi per testare la validità di una gestione, utilizzando i dati dai TPS e MIS e aggiungendo dati provenienti da fonti esterne (prezzi materie prime, tassi di mercato, ecc.);
- **Executive Information Systems (EES):** al vertice della gerarchia, sono riservati al senior manager dell'organizzazione per le decisioni strategiche di non-routine; sono spesso composti nel formato di un *cruscotto digitale* con il quale il senior manager riesce ad avere sotto controllo gli andamenti della gestione aggregando informazioni interne ed esterne, sintetizzandole con gli indicatori più importanti.

Segmentazione dei Bisogni: il Portafoglio Applicativo

Il **Portafoglio Applicativo** è l'insieme delle applicazioni utili in azienda e può essere diviso in tre segmenti principali:

- **Portafoglio Direzionale:** comprende le applicazioni informatiche a supporto dei cicli di pianificazione e controllo strategico e di pianificazione e controllo delle risorse aziendali (controllo budget, reporting, ecc.).
- **Portafoglio Istituzionale:** comprende le applicazioni informatiche per i processi di supporto all'amministrazione, alla gestione delle risorse umane alla contabilità:
 - processi che eseguono adempimenti di legge (contabilità, retribuzione, previdenza);
 - processi che supportano l'amministrazione di infrastrutture (immobili);
 - processi che supportano l'amministrazione di fattori di produzione (personale, scorte, denaro, impianti).

Riguarda aree con rilevante potenzialità di informatizzazione (forte proceduralità, ripetitività, periodicità).

L'informatizzazione nel portafoglio istituzionale inizia negli anni 90 con i sistemi *ERP (Enterprise Resource Planning)* riducendo i costi di elaborazione (aumentandone la capacità) e rendendo più efficaci i processi di pianificazione, controllo e ottimizzazione.

- **Portafoglio Operativo:** comprende le applicazioni informatiche per i processi primari della *Catena del Valore di Porter* (acquisizione materie prime, fabbricazione, marketing, vendita, distribuzione, post-vendita).

L'informatizzazione del portafoglio operativo porta i seguenti benefici:

- aumento della complessità gestibile nei processi aziendali;
- sincronizzazione dei dati all'interno dell'impresa (basi di dati uniche).

I portafogli *direzionale* e *istituzionale* due sono *orizzontali*: generali, per tutti i settori. Indipendenti dalle specifiche caratteristiche dei settori. Il portafoglio *operativo* è *verticale*: specifico di ciascun settore industriale poiché l'attrattiva informatica dei settori è variabile (molto alta nel settore bancario).

Informatizzazione del Portafoglio Operativo nelle Imprese Manifatturiere

1960-1990: Procedure per automazione di singole attività con soluzioni ad-hoc, linguaggi di programmazione classici (COBOL, C++) per migliorare l'efficienza nell'elaborazione dei dati e ottimizzare le risorse (meno scorte in magazzino).

Dal 1970 al 1995: Pacchetti MRP (*Manufacturing Resource Planning*). MRP è una tecnica che calcola i fabbisogni dei materiali e pianifica gli ordini di produzione e di acquisto, tenendo conto della domanda del mercato, della distinta base (materie prime per realizzare il prodotto), dei *lead time* di produzione e di acquisto e delle giacenze dei magazzini.

I pacchetti MRP erano inizialmente finalizzati alla sola informatizzazione del fabbisogno delle materie prime, produzione e gestione dei materiali (poi estesi a tutta la produzione). Fra le tecnologie chiave troviamo i database e pacchetti integrati. Tra i benefici troviamo: integrazione orizzontale e verticale dei processi intra-organizzativi e il bilanciamento dei fattori produttivi ed efficienza del processo di produzione.

Dal 1980: CIM (*Computer Integrated Manufacturing*) - Automazione di fabbrica con la gestione di parti di produzione affidate all'automazione computerizzata. Si ottiene l'eliminazione dei tempi morti e l'efficienza del processo di fabbricazione e qualità dei prodotti.

Dal 1990: Pacchetti ERP (*Enterprise Resource Planning*) con gestione produzione, integrata con applicazioni a supporto dei processi di vendita e distribuzione fisica (controllo degli inventari, tracciamento degli ordini, servizi per i clienti, finanza e risorse umane). Le

tecnologie chiave sono le architetture client-server e la rete (1997/98). I benefici sono rappresentati dalla trasformazione dei processi interni all'azienda e una maggiore efficienza dei fattori produttivi.

Dal 1995: Sistemi CRM (Customer Relationship Management) (per la fidelizzazione del cliente) con gestione della distribuzione, vendita e post-vendita. Le tecnologie chiave sono i pacchetti integrati per l'intero ciclo sul cliente e le architetture client-server e Web. I benefici sono rappresentati dall'abbattimento dei costi di transazione per il cliente e l'integrazione orizzontale e verticale dei processi di gestione dei clienti.

Dal 1995: E-Procurement con l'informatizzazione del buy-side delle imprese utilizzando pacchetti per l'intero ciclo di acquisto e architetture basate su tecnologie Internet (acquisizione di beni e servizi attraverso la rete). I benefici sono l'abbattimento dei costi di transazione per il compratore e per il venditore e l'integrazione orizzontale e verticale dei processi di gestione dei fornitori.

Enterprise Resource Planning (ERP)

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

Evoluzione dello schema MRP esteso ad attività amministrative e finanziarie, vendita e post-vendita, attività gestionali.

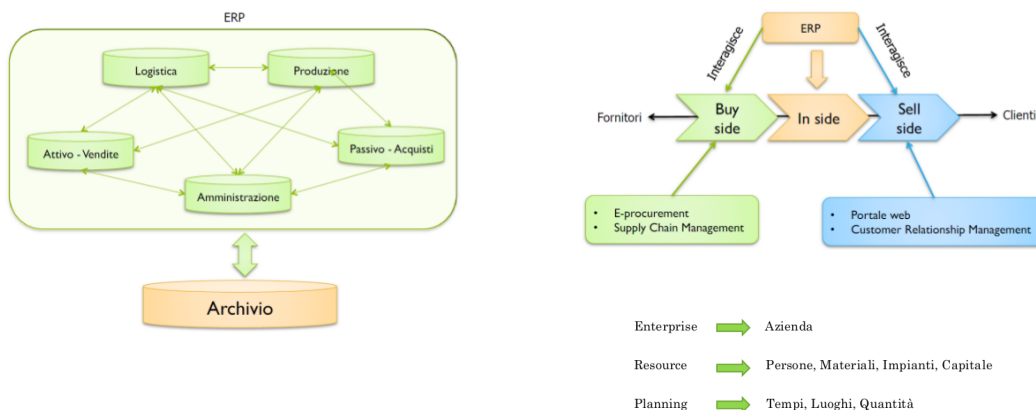
Con la seguente architettura troviamo:

- sistemi modulari composti da *database server* (gestione dati), *application server* (elaborazione dati), *presentation server* (presentazione dati);
- una base di dati condivisa per sincronizzare i processi interdipendenti, minimizzare i dati da immettere e garantire la coerenza delle informazioni.

Questi sistemi interagiscono tramite interfacce (con possibilità di estenderli), sono configurabili, serve periodo di avviamento (costoso e induce a ridefinizione dei processi aziendali) ed è presente sul mercato in modalità standard pre-configurata (inducono all'omologazione). Sono molto diffusi poiché i moduli sono venduti separatamente e parametrizzabili in base all'azienda che li usa.

Questi sistemi accelerano i cicli operativi e la condivisione dell'operazione (abbattendo i costi delle transazioni), però obbligano l'azienda ad adattare le procedure gestionali al sistema.

È applicabile a diversi settori mediante: la presenza di moduli indipendenti che coprono una specifica area aziendale, la struttura globale ERP unica in cui cambiano i moduli di gestione del *core-business* aziendale.



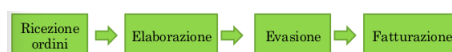
Amministrazione: modulo per le procedure di contabilità generale che supporta le attività amministrative (fatture, pagamenti e incassi) e produce informazioni di sintesi sull'andamento aziendale.

Logistica: procedure per la gestione dei materiali (materie prime, materiali di funzionamento, prodotti finiti). Si occupa della gestione interna della logistica e non del rapporto con clienti e fornitori.

- Articoli: descrizione, caratteristiche operative, anagrafe fornitori, lotti riordino;
- Layout depositi aziendali: dislocazione dei depositi (interni o esterni all'azienda) e spostamento dei beni tra i depositi (flussi di carico/scarico).

Vendite Flusso Attivo:

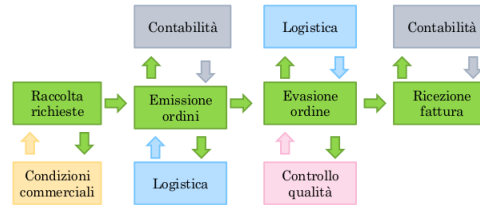
- processi che permettono all'azienda di interagire con il cliente (gestione ordini);
- trattamento delle condizioni commerciali (listini, sconti e analisi dei costi aziendali);
- analisi per la valutazione della collocazione dell'azienda sul mercato;
- trattamento dei processi di gestione del rapporto col cliente:
 - ricezione di ordini (valutazioni sul cliente prima di accettare, es. solvibilità);
 - elaborazione degli ordini (valutazione priorità);
 - evasione di ordini ed emissione di documenti di trasporto (interazione con il magazzino);
 - fatturazione (interagisce con contabilità).



Flusso Passivo: processi per l'interazione con i fornitori per approvvigionamento di materiali e lavorazioni/servizi esterni (molta correlazione con logistica e produzione).

Le principali funzioni di gestione riguardano condizioni commerciali (listini, sconti) e di rapporto con i fornitori.

Acquisti (gestione ordini): raccolta delle richieste da parte della produzione e degli altri reparti (materiale di funzionamento), emissione di ordini al fornitore, ricezione documenti di consegna ed evasione di ordini fornitore, ricezione di fatture fornitore.



Produzione: non esiste una soluzione unica per tutte le produzioni, ma diverse possibilità dipendenti. I produttori di ERP forniscono sottosistemi di produzione verticalizzati. ERP rispecchia il processo produttivo e viceversa (può aiutare ad ottimizzare il processo di produzione). Prima il processo produttivo influenzava il sistema informativo, adesso è il contrario.

Il modulo di produzione definisce dati tecnici: componenti per realizzare il prodotto (distinta base), risorse utilizzate (con cosa creo il prodotto) e processo produttivo (come creo il prodotto).

Il modulo di produzione svolge le seguenti funzioni:

- pianificazione della produzione (per ottimizzare le risorse): richiesta di produzione (vincola cosa produrre), disponibilità di magazzino (tempi di approvvigionamento), disponibilità risorse (interne ed esterne);
- avanzamento e controllo della produzione: modellata come flusso di fasi (senza entrare nei dettagli), informazione su prelievo/versamento di materiali/prodotti, informazioni di avanzamento (inizio e fine di una fase).

SAP (multinazionale)

SAP AG è una multinazionale europea per la produzione di software. È una delle principali aziende al mondo nel settore degli ERP e nelle soluzioni Enterprise.

Nel 1980 crea il linguaggio ABAP (simile a COBOL/Fortran) per interrogare dati e interfacciarsi al linguaggio SQL (con tecnologia client-server) dei maggiori database relazionali. Nel 2000 integra soluzioni internet e Business Intelligence.

I moduli dell'ERP di SAP sono:

- **Sales and Distribution (SD) - Gestione Commerciale:** colloca il prodotto/servizio sul mercato; comprende vendite spedizioni, trasporti, commercio estero, fatturazione.
- **Materials Management (MM) - Gestione Materiali:** effettua la movimentazione dei materiali (pianificazione dei consumi, acquisti, controllo delle fatture, gestione degli stock, controllo di qualità, gestione del magazzino).
- **Product Planning (PP) - Pianificazione della Produzione:** combina i fattori produttivi e li trasforma in beni e servizi (gestione della domanda, pianificazione delle capacità, della produzione e delle vendite).
- **Quality Management (QM) - Gestione della Qualità:** pianificazione, controllo, ispezioni, avvisi, certificati (per verificare le variazioni della qualità).
- **Plant Maintenance (PM) - Manutenzione Impianti:** pianificazione della manutenzione, la gestione degli oggetti tecnici e la gestione delle risorse.
- **Human Resources (HR) - Gestione delle Risorse Umane:** reclutamento, allocazione, formazione, rilevazione delle presenze, gestione dei turni, calcolo delle retribuzioni.
- **Financial Accounting (FI) - Contabilità:** rileva, elabora e comunica agli stakeholders i dati economici e quantitativi dell'azienda.
- **Controlling (CO) - Controllo di Gestione:** accerta che le risorse siano acquisite ed impiegate in modo efficiente ed efficace (controllo degli ordini, controllo dei progetti, contabilità dei costi, analisi di profittabilità).
- **Asset Manager (AM) - Gestione Cespiti:** gestisce i valori materiali e immateriali a utilità pluriennale.
- **Project System (PS) - Gestione Commesse:** pianifica i progetti di ricerca e di sviluppo (pianificazione dei costi, pianificazione dei tempi, pianificazione delle capacità).
- **Workflow (WF):** gestisce i flussi di lavoro (gestione delle autorizzazioni, dei livelli di accesso e delle password).
- **Industry Solution (IS):** fornisce soluzioni settoriali per imprese specifiche (gestione degli immobili, la gestione dell'auto-parco, la gestione del protocollo).

Il Portafoglio Operativo delle Società di Servizi

Le società di servizi sono rappresentate da: settore bancario, settore assicurativo, servizi finanziari. Sono caratterizzate da un'alta intensità informativa (sia dei processi produttivi che dei prodotti/servizi).

C'è sovrapposizione tra produzione e distribuzione:

- *processo produttivo*: predisposizione delle condizioni produttive (creare la capacità produttiva per erogare il servizio, anche con la stipulazione contratti con altri operatori, acquisto e apertura nuove filiali), attività di back-office (erogazione del servizio svolto in seguito ad un ordine ma non in presenza del cliente, come la valutazione del rischio di una polizza assicurativa);
- *processo distributivo*: attività di front office (erogazione del servizio in presenza del cliente come l'attività di sportello della banca), ricerca della clientela (per la stipulazione dei contratti di servizio da parte dei clienti).

Trattamento dell'informazione supportato dai sistemi informatici

- *Flusso degli ordini*: arrivo dell'ordine, produzione e distribuzione del servizio, post-vendita;
- *Flusso di gestione della conoscenza (knowledge management)*: elaborazione dell'informazione acquisita durante la produzione e l'erogazione del servizio; trasformazione in conoscenza accessibile alle attività operative e direzionali.

La gestione della conoscenza è vitale nelle società di servizi, dove essa rappresenta il vero capitale dell'azienda (*vantaggio competitivo*). La gestione della conoscenza ha lo scopo di trasformare l'informazione non strutturata in informazione comprensibile a qualunque esperto. L'informazione non è strutturata, ad esempio:

- la valutazione del rischio di una polizza assicurativa non è una formula (i fattori di rischio sono imprevedibili);
- la valutazione del premio è un'attività creativa basata su esperienza, intuito e sulle informazioni disponibili.

Knowledge Worker: impiegato che svolge un lavoro intellettuale di natura decisionale. È una figura tipica delle società di servizi con elevato grado di personalizzazione o di trattativa con il cliente.

CRM e Aziende di Utility

I sistemi di *Customer Relationship Management* supportano l'intero ciclo di rapporti con la clientela (marketing e pianificazione vendite, raccolta ordini e fatturazione, assistenza post-vendita). Questi sistemi informatizzano l'area *sell-side* e attraverso Internet trasformano l'interazione tra aziende e clienti.

Le aziende di **utility** (forniscono gas, energia, acqua, telefono e simili) sono spinte ad uno sviluppo massiccio di sistemi a supporto dei contatti con il cliente per i seguenti motivi: clientela numerosa, rapporto continuativo nel tempo, sviluppo di canali virtuali (Web e call center), concorrenza crescente, necessità di sviluppare e mantenere un'offerta personalizzata. I conseguenti **vantaggi** sono: abbattimento dei costi di relazione, grande impatto sui profitti e buon grado di personalizzazione mantenendo i costi contenuti.

Le funzioni principali dei sistemi CRM sono i sistemi di gestione degli ordini (es. Amazon), sistemi di automazione delle forze di vendita (venditori dotati di pc per gestione di offerte e ordini), sistemi di assistenza telefonica (call center).

CRM e Business Intelligence

Gli strumenti di *Business Intelligence* sono applicazioni che estraggono informazioni dai sistemi di supporto alla produzione per elaborarle e fornire supporto all'apparato decisionale (nel portafoglio decisionale servono ad analizzare lo stato dell'azienda e prendere decisioni). Si occupano di passare da dati a informazione sintetica.

L'architettura dei sistemi di *Business Intelligence* comprende: una serie di *data warehouse* e una serie di *data mart*.

Ingegneria Sociale

Ingegneria Sociale: Una manipolazione della naturale tendenza alla fiducia dell'essere umano, architettata e realizzata dall'ingegnere sociale con l'obiettivo *ottenere informazioni* che permettano di raggiungere l'obiettivo prefissato.

- **Sociale:** perché coinvolge e studia il comportamento umano.
- **Ingegneria:** è necessario pianificare la situazione per raggiungere i propri obiettivi (come nell'hacking, la maggior parte del lavoro è nella preparazione, piuttosto che nel tentativo stesso). L'ingegnere sociale programma l'attacco anticipando le domande che il bersaglio può porgli in modo da *avere sempre la risposta pronta*.

Ingegnere Sociale: sceglie la vittima, sviluppa una causa plausibile per ottenere subito confidenza, stabilisce un obiettivo minimo, costruisce il rapporto su simpatia o altre tecniche per influenzare la vittima, sviluppa risposte per superare le possibili obiezioni, prevede una via d'uscita per non bruciare l'origine del contatto.

FootPrinting: l'ingegnere sociale prima di compiere il gesto criminoso comincia a raccogliere informazioni sul sistema da colpire (può durare mesi). In questo tempo studia e impara: i sistemi di comunicazione aziendale, funzionamento della posta interna, organigramma aziendale, giorni e orari di pulizia, frequenta gli uffici (es. consegna buste, caffè, ecc.), conosce e dialoga col personale (addetti alla sicurezza, segretarie, sistemisti, ecc.), cerca di ottenere informazioni (manda e-mail, telefonate), finge di essere un utente inesperto che ha smarrito il password, manda un'offerta di un nuovo firewall per aumentare il sistema di sicurezza.

Phreaker: furono i primi utilizzatori dell'ingegneria sociale per ottenere informazioni vitali dalle compagnie telefoniche. Utilizzavano la rete telefonica sfruttando i sistemi e i dipendenti dell'azienda erogatrice del servizio.

Sicurezza: non è un prodotto acquistabile ma un processo evolutivo che coinvolge l'organizzazione nella sua interezza; però spesso è solo un'illusione poiché, se si rubano soldi o beni qualcuno si accorge che sono spariti. Ma quando si rubano informazioni nessuno lo nota, perché restano comunque in possesso della vittima.

Informazioni: sono memorizzate nei computer sono gestite da operatori umani che molto spesso ignorano le procedure di sicurezza. La sicurezza non è un problema di tecnologia, bensì di persone e gestione.

Le **Vittime Preferite** sono le persone che: non hanno niente da perdere nel fornire una informazione sensibile, sottostimano il valore delle informazioni, pensano che seguire le basilari regole per la sicurezza delle informazioni sia una perdita di tempo, non valutano le conseguenze delle loro azioni.

L'ingegneria sociale spesso funziona perché è più facile dell'hacking, non richiede dispositivi particolari, ha un costo basso e un basso rischio, lascia poche tracce, è efficace, non diventa obsoleta ed è poco conosciuta dalle vittime.

Questi attacchi possono avere successo perché di solito le persone non conoscono bene le pratiche di sicurezza.

Fasi dell'Attacco

- **Fase Fisica:** raccolta di informazioni attraverso persone (*shoulder surfing*), documenti (supporti digitali distrutti, stampe di dati sensibili), luoghi (*dumpster diving*).
- **Fase Psicologica:** impersonificazione, carpire le informazioni utili attraverso la persuasione.

Fase Fisica

Strumenti Essenziali: buona memoria per raccogliere le informazioni, telefoni, fax, scanner, stampanti, server di posta, cooperazione, argomenti e ragioni per giustificare l'azione (più la persona target è vicina alle informazioni e più forti devono essere le argomentazioni).

Obiettivi dell'Attacco: password, obiettivi fisici (modem, server), help desk (processo per fornire supporto riguardo prodotti/servizi informatici ed elettronici).

I **Metodi di Attacco** sono due (serve la giusta combinazione di entrambi):

- interazione con altre persone o azioni svolte senza l'ausilio di mezzi tecnologici;
- attacchi svolti con l'ausilio della tecnologia.

Attacchi mediante interazione umana: richiesta diretta di informazioni, truffe telefoniche, rovistare nella carta straccia (*dumpster diving*) che viene ancora stampata in grandi quantità, rovistare negli hard disk dismessi, richiesta di informazioni per un falso sondaggio, sbirciare alle spalle (*shoulder surfing*).

Attacchi Tecnologici: attacchi diretti al sistema (creazione di un nuovo account e aggiungere privilegi o diritti di accesso a un account esistente), esecuzione di un programma malintenzionato (trojan), attivazione di un piano di *phishing* (loghi contraffatti), intercettazioni telefoniche, invio di mail contenenti malware o false proposte (*Nigeriam Scam*, mail in cui si parla di somme di denaro da trasferire chiedendo informazioni sensibili), microfoni, webcam, tastiere (*KeyGhost*, un dispositivo che cattura qualsiasi cosa venga digitata sulla tastiera).

Altre tecniche utilizzate sono le finestre di pop-up (chiedere nuovamente username e password per qualche problema di rete), spam e catene di S. Antonio (non causano danni ma riducono la produttività), siti web (offrire qualcosa di gratis o la possibilità di vincere qualcosa), file scambiati tramite peer-to-peer (una volta scaricati, il sistema viene utilizzato come stazione di comunicazione e attacco per aggressioni verso l'esterno per comunicare dati sensibili).

Metodi Classici: fingersi un collega, fingersi un fornitore o un tutore dell'ordine, fingersi un nuovo assunto che chiede aiuto, fingersi un fabbricante di sistemi che offre aggiornamenti o patch al sistema, usare una falsa finestra pop-up per ottenere credenziali, inviare programmi o patch gratis da installare, inviare malware, far registrare ad un sito web fasullo, modificare intestazione dei fax (per far sembrare che provengano dall'interno), fingere di essere di un'altra sede dell'azienda per avere accessi a posta elettronica o altri dati.

Reverse Social Engineering: offrire aiuto in caso di problemi, poi fare in modo che il problema si presenti realmente, convincendo così la vittima a chiedere aiuto.

Trucchi Fondamentali:

- domanda "mina antiuomo": si pone la domanda "mina", se il bersaglio risponde senza cambiare tono di voce significa che non è scettico e si può porre la vera domanda senza insospettirlo.
- per non insospettire, non bisogna mai interrompere la conversazione una volta ottenuta l'informazione chiave (proseguire come se nulla fosse).

Ingegnerrizzare se stessi: diventare qualcun altro (meglio al telefono), vestirsi in base all'occasione, imparare a parlare in base al proprio ruolo, dire ciò che il bersaglio vuole sentirsi dire (imparare il più possibile su di esso), saper mentire.

Fase Psicologica

Sicurezza e Fiducia: il primo obiettivo dell'ingegnere sociale è creare fiducia nella vittima, a quel punto può scattare la trappola, che può sfruttare le conoscenze informatiche dell'aggressore (la sicurezza si basa sulla fiducia).

Peopleware: l'anello debole non è la persona ma la naturale tendenza a credere alla parola altrui. L'ingegnere sociale fa leva sui comportamenti (naturali, sollecitati, forzati) che tutti mostriamo.

Obiettivi della fase psicologica: osservare il comportamento delle persone, imparare come pensano, localizzare le debolezze umane, imparare quali parole usare quando si parla a certe persone.

Su cosa agire: l'ingegnere sociale fa leva sui bisogni primari dell'uomo per far sì che la propria richiesta venga accettata. I **bisogni primari** sono: fisiologici (funzionali al mantenimento psicofisico dell'uomo), sicurezza (essere fuori pericolo, stabilità, protezione), appartenenza (essere accettati da altri), stima (ottenere approvazione/riconoscimento).

I **bisogni secondari** sono: cognitivi (conoscere, comprendere, esplorare), estetici (simmetria, ordine, bellezza), autorealizzazione (trasformare in realtà il potenziale inespresso), trascendenza (aiutare altri a realizzare il loro potenziale).

Tendenze base della natura umana: autorevolezza (tendenza a cedere quando una persona autorevole fa una richiesta), simpatia (tendenza ad obbedire quando chi fa la richiesta suscita simpatia), ricambiare (tendenza ad obbedire automaticamente ad una richiesta quando viene dato o promesso qualcosa di valore), coerenza (tendenza a dire di sì dopo aver sposato pubblicamente una causa per evitare di apparire inaffidabili), convalida sociale (tendenza ad obbedire quando, in questo modo, si fa quello che fanno gli altri), scarsità (tendenza a dire di sì quando si crede che l'oggetto cercato stia scarseggiando e altri siano in competizione per averlo).

Caratteristiche delle Vittime: paura delle conseguenze, senso di colpa, sentimentalismo, ritenere che l'azione possa dare loro un vantaggio, tendenza a dare fiducia agli altri, dovere morale identificazione con l'ingegnere sociale, attitudine all'aiuto, convinzione della propria non vulnerabilità (i nuovi assunti sono una preda facile).

Persuasione: la finalità non è quella di forzare un individuo a fare qualcosa ma quella di aumentare la sua volontaria condiscendenza alle richieste altrui. La vittima viene guidata sul percorso scelto dall'attaccante, lasciandole credere di avere il controllo totale della situazione.

Strumenti di Persuasione: impersonificazione, conformità (conformarsi al giudizio degli altri), diffusione di responsabilità (ritengono che l'azione riduca le loro responsabilità), accattivarsi la fiducia altrui e usare la cortesia.

Ingegneria dei Processi Gestionali

I **processi** (*business process*) rappresentano il modo di operare di un'azienda. Le **tecnologie informatiche** e di rete trasformano il modo di operare di un'azienda e quindi i processi. Si ha la necessità di integrare l'innovazione informatica con l'innovazione organizzativa per migliorare costi, tempi o qualità (l'innovazione tecnologica ha ricadute sull'organizzazione aziendale).

I processi possono essere:

- **materiali** (flusso di materiali e attività);
- **informativi** (flusso di informazioni input/output);
- **Business Process o Processo Aziendale** (un insieme strutturato di attività, finalizzato alla realizzazione di un risultato di interesse per l'organizzazione).

Business Process Reengineering (BPR): necessità di correlare un'innovazione radicale di processi e organizzazione aziendale con un massiccio uso di informatica.

Business Process (BP): un processo aziendale è formato da attività, realizzate come processi materiali o processi informativi, che sono collegate fra loro nel tempo e nello spazio e sono svolte dalle risorse di un'azienda. Partendo da input definiti, le attività producono un output utilizzato dai clienti.

Sono flussi di attività interfunzionali che concatenano: competenze commerciali (marketing), competenze tecniche (produzione), competenze generali (approvvigionamento).

Un **Business Process** è definito come una tupla $BP = (A, I, O, C)$, dove:

- **A** = attività (operazioni fisiche o di decisioni manageriali con esecuzione coordinata);
- **I** = input (materie prime e risorse aziendali oppure informazioni se è un processo decisionale);
- **O** = output (oggetti fisici, beni immateriali, servizi);
- **C** = clienti (destinatari dell'output del processo).

Catena del Valore di Porter: l'azienda è vista come una successione di attività (un processo) finalizzate a produrre valore per il cliente che viene misurato dal prezzo che il cliente è disposto a pagare.



Trasformazione dei Processi - Strategia Buy-Side: trasforma il processo nelle varie fasi di ricerca del fornitore, definizione di prezzi e condizioni, ordinazione, ricezione dei beni e servizi ordinati. L'obiettivo è ridurre il costo di transazione e del materiale stesso. La trasformazione si appoggia a sistemi di *E-procurement* e infrastrutture internet.

Trasformazione dei Processi - Strategia In-Side: trasformazione dei processi interni all'impresa per ridurre i costi di funzionamento e della durata dei processi e migliorare la qualità e il servizio al cliente. La trasformazione si appoggia a sistemi *ERP* e infrastrutture internet.

Trasformazione dei Processi - Strategia Sell-Side: orientata ai processi di marketing, vendita, distribuzione e servizio post-vendita/assistenza al cliente. Si ha un maggiore valore del prodotto percepito dal cliente e un abbattimento dei costi di transazione. La trasformazione si appoggia a sistemi *CRM* e infrastrutture internet.

Classificazione dei processi

- **Processi intersettoriali**: sono generici e descrivono le pratiche di vari settori (gestione/acquisto materie ...).
- **Processi settoriali**: distinguono i vari settori (nel settore sanitario ci sono processi di ricovero, di cura ...);
- **Processi aziendali**: processi di una specifica azienda o di una sua parte (specializzazione dei processi settoriali);
- **Processi normativi e best practice**: processi di riferimento; Best practice derivano dal confronto tra processi diversi e descrivono come dovrebbero essere nelle migliori aziende del settore.

Scomposizione dei processi (dettaglia i processi per livelli di approfondimento successivi)

- **Macroprocessi**: primo livello di segmentazione di un'azienda (es. la *Catena del Valore di Porter*);
- **Processi**: illustrano il modo di operare dell'azienda;
- **Fasi (Sottoprocessi)**: illustrano il modo in cui un processo è implementato (una fase è una tappa del processo);
- **Attività**: livello minimo di analisi normalmente adottato nello studio dei processi; le attività sono determinate scomponendo i processi secondo una logica sequenziale;
- **Operazioni**: passi elementari per eseguire una certa attività (quasi mai usate).

Griglia Metodologica

La *griglia metodologica* è orientata a una strategia integrata di progettazione dei processi. Essa comprende:

- descrizione delle variabili di analisi (*variabili organizzative*);
- descrizione delle *fasi di analisi*;
- identificazione degli strumenti a supporto delle attività di analisi (es. diagrammi di flusso).

Variabili Organizzative

La trasformazione dei processi deve integrare innovazione tecnologica e organizzativa. L'azione è necessaria su tutte le variabili organizzative che, insieme alla tecnologia informatica, sono le leve attraverso i quali sono trasformati i progetti.

Le variabili di progettazione integrata dei processi sono:

- *Flusso delle attività*: sequenza delle attività attraverso cui è svolto il processo;
- *Organizzazione del processo*: divisione operativa del lavoro nell'ambito del processo, cioè come i processi sono mappati sui ruoli (può indicare parallelizzazione o accorpamento delle attività);
- *Competenza delle risorse umane che operano nel processo*: fondamentale per la trasformazione del processo a seguito di un'innovazione tecnica;
- *Sistema di misurazione e controllo delle prestazioni usato per governare il processo e valutare gli attori aziendali che lo eseguono*.

Variabile 1 - Flussi delle Attività

Il flusso delle attività determina: la durata del processo (dipende dal numero di attività), il livello di servizio (dipende dal grado di flessibilità), la qualità del prodotto (insieme a tecnologia e risorse umane).

La modellazione dei flussi può essere condotta a diversi livelli: *schemi di sequenza* (indicano solo la sequenza delle attività che formano un processo) e *schemi più ricchi* (indicano gli attori coinvolti, gli eventi che determinano le attività e le informazioni utilizzate).

Gli elementi della modellazione dei flussi sono:

- natura del flusso: fisico, informativo, controllo;
- attività: tipo, durata, volumi, tecnologie;
- sequenza attività: alternative nella sequenza (eventuale struttura di controllo);
- attori: tipologia, azioni svolte sul flusso;
- eventi: tipologia, temporizzazione, conseguenza dell'evento sull'attività (avvia, ferma, modifica);
- oggetti: natura (fisico, informativo), profilo temporale (informazioni permanenti/temporanee);

Variabile 2 - Organizzazione

L'organizzazione è una determinante strutturale della prestazione del processo e si dividono in due tipi:

- organizzazioni parcellizzate (per compiti ripetitivi e semplici con alti volumi);
- organizzazioni piatte con mansioni articolate (per lavori complessi e volumi limitati).

Per descrivere l'organizzazione si possono usare: organigrammi, tabelle di proprietà, Linear Responsibility Charting (*LRC*).

Organigramma: Gerarchia delle responsabilità e delle autorità di un'organizzazione con vari livelli di dettaglio.

Tabella delle Proprietà (*logiche/quantitative*): indica le proprietà *logiche* (descrizione del mandato, elenco dei compiti assegnati e dei processi svolti) e *quantitative* (organici e volumi del lavoro) delle strutture.

Linear Responsibility Charting (LRC): offre una visione tabellare della responsabilità organizzativa che integra quella dell'organigramma. Per ciascun processo, specifica il ruolo svolto da ogni struttura nel processo.

Il ruolo può essere: **D** (decide, autorizza), **E** (esegue le attività), **A** (dà assistenza operativa e supporto), **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

Variabile 3 - Risorse Umane

Le risorse umane determinano la differenza tra il risultato effettivo di un processo e il massimo risultato teoricamente possibile. L'innovazione tecnologica causa la necessità di avere a disposizione figure professionali specializzate:

- creandole internamente riconvertendo il personale si possono avere problemi gestionali (superamento dell'avversione al cambiamento e accettazione delle nuove tecnologie);
- acquisite sul mercato (reperimento complesso e costoso).

Variabile 4 - Sistema di Misurazione delle Prestazioni

Ogni processo è governato da un sistema di misurazione delle prestazioni che comprende:

- *sistema di pianificazione e controllo*: fissa gli obiettivi di efficienza e efficacia del processo e li controlla periodicamente;
- *sistema di promozione e incentivazione*: fissa gli obiettivi del processo come vendite e soddisfazione del cliente;
- *sistema dei valori*: posiziona gli obiettivi dell'azienda e decide i valori rispetto a cui incentivare (in un processo orientato all'utente, è prioritaria la soddisfazione del cliente).

Fasi di Analisi

Rappresentano i passi logici attraverso cui si svolge l'analisi dei processi.

- Fasi (approccio *bottom-up* al miglioramento del processo esistente): rilevazione della situazione esistente, confronto con altre imprese e diagnosi dei problemi, ridisegno dei processi.
- Fasi (approccio *top-down*): disegno del processo con applicazione di criteri di ottimizzazione noti, analisi della situazione esistente per capire cos'è necessario per attuare il progetto.

In ogni fase vengono analizzate le variabili organizzative appena descritte.

Fase 1 - Rilevazione della Situazione Esistente

- **Passo 1 - Identificazione dei macroprocessi:** identificazione dei processi usando un opportuno modello (catena del valore, checklist, best practice) e rilevazione delle proprietà fondamentali:
 - quali sono i clienti del processo (esterni / interni);
 - tipo di processo (inside/buy-side/sell-side);
 - input del processo (materie prime, competenze);
 - output del processo (prodotto/servizio fornito al cliente).
- **Passo 2 - Dettaglio dei processi:** composizione dei processi in fasi e attività, documentata da:
 - diagrammi gerarchici (processo → fasi → attività);
 - diagrammi di flusso (flussi fisici e/o informativi);
 - schede che descrivono le proprietà di processi, fasi, attività, operazioni.
- **Passo 3 - Incrocio processi/unità organizzative:** analisi della relazione fra strutture organizzative e processi da tre punti di vista:
 - rilevazione delle strutture organizzative (organigrammi);
 - definizione dei ruoli delle strutture nei processi (*LRC*);
 - mappatura delle attività delle strutture nel flusso dei processi (opportuni diagrammi di flusso chiamati *Responsibility Activity Diagram – RAD*).
- **Passo 4 - Valutazione del processo:** definizione dei parametri di funzionamento e giudizio sul valore dei prodotti. Comprende:
 - giudizio sul processo da parte degli esecutori (costo del processo in termini di tempo usato dalle risorse e tempo di completamento del processo);
 - giudizio sul processo da parte dei clienti interni o esterni in termini di:
 - * utilità/valore (prezzo che il cliente è disposto a pagare);
 - * livello del servizio (puntualità e velocità);
 - * qualità (contenuto del prodotto e assenza di difetti).

Fase 2 - Confronto e Diagnosi

- **Passo 1 - Confronto quantitativo e parametrizzazione:** per il confronto tra prestazioni di aziende concorrenti secondo alcuni parametri (produttività, livello di servizio, durata attività);
- **Passo 2 - Confronto qualitativo:** misurare le cause della diversità rispetto a valori di mercato (fase decisiva per la scelta di eventuali cambiamenti organizzativi). Si tratta di esaminare le variabili organizzative per individuare le aree in cui una diversità può indicare un legame causa-effetto (evidenziazione delle differenze di modello organizzativo e tecnologico).

Fase 3 - Ridefinizione

- Definizione della *vision* che dà una rappresentazione sintetica degli elementi fondamentali della soluzione: illustra la nuova struttura del flusso di attività e gli aspetti più significativi della configurazione delle variabili gestionali.
- Analisi del cambiamento: specificazione dell'impatto organizzativo del progetto (incrocia le variabili organizzative con gli stakeholder) e valutazione di rischi, benefici, tempi e costi del cambiamento.

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. Meccanismi 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). Meccanismi 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. Predispersione 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 indicatori di prestazione.

Trasformazione dei Processi

L'applicazione delle tecnologie dell'informazione ha come effetto l'integrazione inter-funzionale o inter-organizzativa dei processi che offre la possibilità di condividere informazioni prima non disponibili e di eseguire congiuntamente attività prima svolte separatamente dalle singole unità (le organizzazioni possono ottenere benefici consistenti).

L'informatica consente una maggiore integrazione inter-funzionale e inter-organizzativa per due ragioni fondamentali: aumenta la disponibilità di informazioni e supporta l'esecuzione dei compiti individuali di natura decisionale.

Business Process Model and Notation (BPMN)

Modellazione di processi e dati

Data Warehousing

L'informatica in azienda era vista come una scienza di supporto che rende più rapide ed economiche le operazioni di gestione delle informazioni (ma non crea ricchezza). Adesso invece i sistemi informatici si sono trasformati da semplici strumenti per migliorare l'efficienza dei processi a elementi centrali dell'organizzazione aziendale (rivoluziona i processi).

L'informatica ha un duplice ruolo:

- tecnologia di supporto alla gestione del Sistema Informativo;
- disciplina organizzativa che influenza processi, servizi e struttura aziendale.

I motivi principali per migliorare il Sistema Informativo sono tre:

- **Trasformazione dell'Economia:** l'economia moderna è basata sulle conoscenze e sull'informazione ed è caratterizzata da una breve vita dei prodotti che richiede decisioni tempestive;
- **Trasformazione dell'Impresa:** per operare con profitto in un sistema economico altamente competitivo le aziende hanno bisogno di una struttura flessibile in grado di reagire rapidamente alle mutate situazioni esterne/interne;
- **Globalizzazione:** con l'allargamento dei mercati a livello mondiale nasce l'esigenza del controllo di mercati a larga scala.

Per il portafoglio direzionale si parla anche di **Business Intelligence**: disciplina che consente a chi deve decidere in azienda di capire, attraverso soluzioni software, i fattori chiave del business e conseguentemente di prendere le migliori decisioni in quel momento. Si parla di piattaforma perchè necessita di un'apposita infrastruttura hardware e software di supporto.

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

L'informazione è un bene a valore crescente e costituisce la materia prima trasformata dai sistemi informativi. Spesso la disponibilità di troppi dati rende difficile (se non impossibile) estrapolare le informazioni veramente importanti (**dati \neq informazione**).

Per ogni azienda è fondamentale poter disporre in maniera rapida e completa delle informazioni necessarie al processo decisionale; l'aumento esponenziale del volume dei dati operazionali ha reso il calcolatore l'unico supporto adatto e il ruolo dell'informatica è passato da passivo strumento per la registrazione delle operazioni a fattore decisivo per l'individuazione di elementi critici dell'organizzazione e potenziali aree di business.

Negli anni '80 nascono i sistemi di supporto alle decisioni (**DSS, Decision Support System**): l'insieme delle tecniche e degli strumenti informatici atti a estrapolare informazioni da un insieme di dati memorizzati su supporti elettronici.

Introduzione al Data Warehousing

Uno scenario tipico è quello di una grande azienda, con tante filiali, i cui dirigenti vogliono monitorare il contributo dato da ciascuna di esse al rendimento dell'impresa. Un *raccoglitore di informazioni (Data Warehouse)* che integra e riorganizza i dati provenienti da sorgenti distinte e li rende disponibili per analisi e valutazioni finalizzate alla pianificazione e al processo decisionale.

Mescolare interrogazioni *analitiche* e *transazionali* di routine porta a inevitabili rallentamenti che rendono insoddisfatti entrambi gli utenti. È necessario separare l'elaborazione di tipo analitico (**OLAP, On-Line Analytical Processing**) da quella legata alle transazioni (**OLTP, On-Line Transactional Processing**), costruendo un raccoglitore di informazione che integri i dati provenienti da sorgenti di varia natura. Alcune aree di utilità sono: commercio, manifattura, servizi finanziari, trasporti, telecomunicazioni, sanità ...

Data Warehousing: collezione di metodi, tecnologie e strumenti di ausilio al *knowledge worker* (dirigente, amministratore, gestore, analista) per condurre analisi dei dati finalizzate all'attuazione di processi decisionali e al miglioramento del patrimonio informativo. Il processo di *warehousing* possiede le seguenti caratteristiche:

- *accessibilità* a utenti con conoscenze limitate di informatica e strutture dati;
- *integrazione dei dati* sulla base di un modello standard dell'impresa;
- *flessibilità di interrogazione* per trarre il massimo vantaggio dal patrimonio informativo esistente;
- *sintesi* per permettere analisi mirate ed efficaci;
- *rappresentazione multidimensionale* per offrire all'utente una visione intuitiva e manipolabile delle informazioni;
- *correttezza e completezza* dei dati integrati.

Un **Data Warehouse** è una collezione di dati di supporto per il processo decisionale che presenta le seguenti caratteristiche: orientata ai soggetti di interesse, integrata e consistente, rappresentativa dell'evoluzione temporale (c'è un ricco contenuto storico a differenza dei DB operazionali, la "fotografia" dei dati ad un certo istante di tempo non può essere aggiornata), non volatile.

Le interrogazioni sono di due tipi:

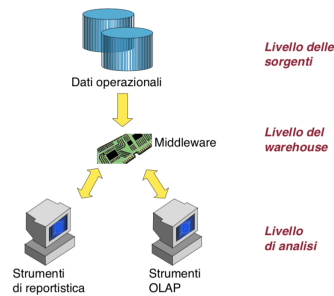
- **OLTP:** le interrogazioni eseguono transazioni che leggono e scrivono un ridotto numero di record da diverse tabelle;
- **OLAP:** le interrogazioni effettuano un'analisi dinamica e multidimensionale che richiede la scansione di un'enorme quantità di record per calcolare un insieme di dati numerici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

Le **architetture** devono avere i seguenti requisiti:

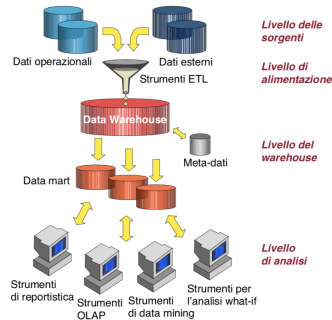
- *separazione*: l'elaborazione analitica e quella transazionale devono essere mantenute il più possibile separate;
- *scalabilità*: l'architettura hardware e software deve essere facilmente ridimensionabile a fronte della crescita nel tempo dei volumi di dati da gestire degli utenti da soddisfare;
- *estendibilità*: poter accogliere nuove applicazioni e tecnologie senza riprogettare integralmente il sistema;
- *sicurezza*: il controllo sugli accessi è essenziale a causa della natura strategica dei dati memorizzati;
- *amministrabilità*: la complessità dell'attività di amministrazione non deve risultare eccessiva.

Architettura a 2 Livelli

Architetture a 1 livello



Architetture a 2 livelli



Data Mart: un sottoinsieme o un'aggregazione dei dati presenti nel DW primario, contenente l'insieme delle informazioni rilevanti per una particolare area del business, divisione dell'azienda o categoria di soggetti. Ci sono due tipi di *data mart*:

- *dipendenti*: sono quelli alimentati dal DW primario; utili in aziende medio-grandi poiché delineano i contorni delle informazioni necessario ad un certo tipo di utenti e permettono di raggiungere migliori prestazioni (dimensioni inferiori);
- *indipendenti*: alimentati direttamente dalle sorgenti; l'assenza di un DW primario snellisce le fasi progettuali ma determina uno schema complesso di accesso ai dati (rischio di inconsistenze tra i data mart).

L'*architettura a 2 livelli* presenta i seguenti **vantaggi**:

- informazione di buona qualità sempre disponibile anche quando, per motivi tecnici, non è disponibile l'accesso alle sorgenti;
- l'interrogazione analitica non interferisce con la gestione delle transazioni
- l'organizzazione logica del DW si basa sul modello relazionale (le sorgenti offrono modelli relazionali o semi-strutturali);
- c'è una distanza temporale tra sistemi OLTP (trattano dati correnti e al massimo livello di dettagli) e sistemi OLAP (operano su dati storici e di sintesi);
- è possibile impiegare tecniche per ottimizzare le prestazioni per applicazioni di analisi e reportistica.

Architettura a 3 Livelli

L'*architettura a 3 livelli* ha il livello dei dati riconciliati, il cui vantaggio principale è che crea un modello di dati comune e di riferimento per l'intera azienda. Introduce una separazione netta tra le problematiche legate all'estrazione e integrazione dei dati e quelle inerenti l'alimentazione del DW. I dati riconciliati, però, introducono un'ulteriore ridondanza rispetto ai dati sorgente.

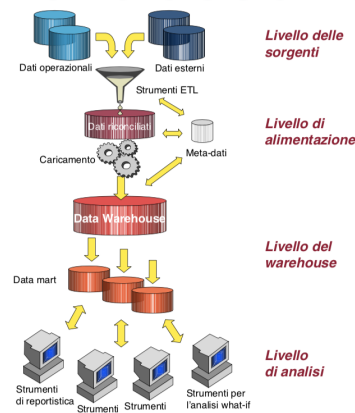
Dati Riconciliati: dati operazionali ottenuti dopo il processo di integrazione e ripulitura dei dati sorgente, quindi dati integrati, consistenti, corretti, volatili, correnti e dettagliati.

Il ruolo degli strumenti di *Extraction, Transformation and Loading* è quello di alimentare una sorgente dati singola, dettagliata, esauriente e di alta qualità che possa a sua volta alimentare il DW (*riconciliazione*).

La riconciliazione avviene due volte durante il processo di alimentazione del DW: quando viene popolato per la prima volta, e periodicamente quando viene aggiornato (estrazione, pulitura, trasformazione e caricamento).

Estrazione: i dati rilevanti vengono estratti dalle sorgenti e vengono scelti principalmente in base alla loro qualità. L'*estrazione statica* avviene quando il DW deve essere popolato per la prima volta (fotografia dei dati operazionali). L'*estrazione incrementale* viene usata per l'aggiornamento periodico del DW (cattura i cambiamenti avvenuti nelle sorgenti dall'ultima estrazione).

Architetture a 3 livelli



Pulitura: si incarica di migliorare la qualità dei dati delle sorgenti (dati duplicati, incosistenza tra valori, dati mancanti, valori impossibili o errati, valori inconsistenti per le diverse convenzioni o errori di battitura).

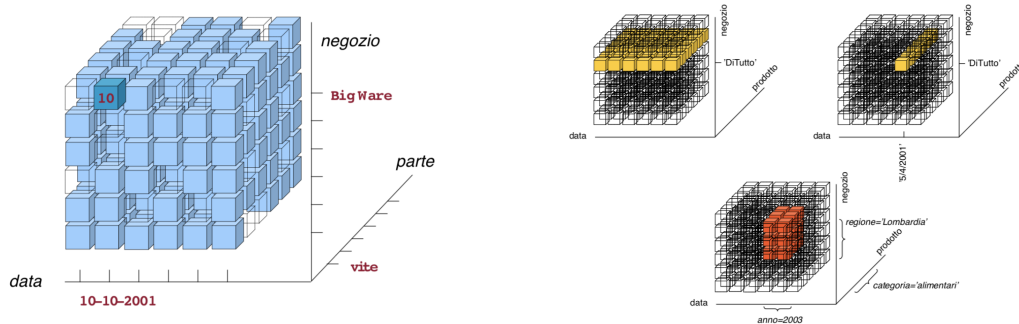
Trasformazione: converte i dati dal formato operativo sorgente a quello del DW. La corrispondenza è complicata dalla presenza di fonti eterogenee, che richiede una complessa fase di integrazione (es. formati differenti per lo stesso dato).

Caricamento: i dati vengono caricati nel DW con *Refresh* (i dati vengono riscritti integralmente, sostituendo quelli precedenti, come nel caso della prima popolazione del DW) e con *Update* (i soli cambiamenti occorsi nei dati sorgente vengono aggiunti nel DW).

Il modello Multidimensionale

È il fondamento per la rappresentazione e l'interrogazione dei dati nei DW. I **fatti di interesse** sono rappresentati in *cubi* in cui:

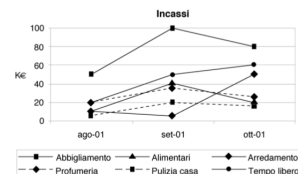
- ogni *cella* contiene *misure* numeriche che quantificano il fatto da diversi punti di vista;
- ogni *asse* rappresenta una *dimensione* di interesse per l'analisi;
- ogni *dimensione* può essere la radice di una *gerarchia* di attributi usati per aggregare i dati memorizzati nei cubi base.



Una volta che i dati sono stati ripuliti, integrati e trasformati, occorre capire come trarne il massimo vantaggio informativo. Esistono in sostanza tre approcci per l'interrogazione di un DW da parte degli utenti finali:

Reportistica: non richiede conoscenze informatiche; orientato agli utenti che hanno necessità di accedere, a intervalli di tempo predefiniti, a informazioni strutturate in modo pressoché invariabile.

incassi (K€)	Ottobre 2001	Settembre 2001	Agosto 2001
Abbigliamento	80	100	50
Alimentari	20	40	10
Arredamento	50	5	10
Profumeria	25	35	20
Pulizia casa	15	20	5
Tempo libero	60	50	20



OLAP: l'utente deve ragionare in modo multidimensionale e deve conoscere l'interfaccia dello strumento grafico utilizzato (orientamento verso utenti non esperti di informatica); È il metodo principale per ottenere le informazioni contenute in un DW. Consente di analizzare ed esplorare interattivamente i dati sulla base del modello multidimensionale. Gli utenti OLAP sono in grado di costruire attivamente una sessione di analisi complessa in cui ciascun passo effettuato è conseguenza dei risultati ottenuti al passo precedente. Una sessione OLAP consiste in un *percorso di navigazione* che rappresenta l'analisi di uno o più fatti di interesse sotto diversi aspetti e a diversi livelli di dettaglio. È sequenza di interrogazioni spesso formulate non direttamente, ma per differenza rispetto all'interrogazione precedente. Ogni passo della sessione di analisi è scandito dall'applicazione di un *operatore OLAP* che trasforma l'ultima interrogazione formulata in una nuova interrogazione. Il risultato è di tipo multidimensionale (rappresenta i dati in modo tabellare).

Data Mining: richiede all'utente la conoscenza dei principi che stanno alla base degli strumenti utilizzati. È un'attività orientata a scoprire informazioni nascoste nei 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 si vuole ricercare (ricerche di mercato, studio dell'efficacia del marketing, segmentazione di mercato, analisi delle abitudini di acquisto, pianificazione aziendale, modellazione degli investimenti, rilevamento di attività fraudolente, valutazione delle categorie di rischio).

- **Regole Associative:** consentono di individuare i gruppi di affinità tra oggetti attraverso lo studio delle abitudini di acquisto per la pubblicità mirata e l'organizzazione della merce sugli scaffali (*market-basket analysis*).
- **Clustering:** significa raggruppare gli oggetti in un ridotto numero di insiemi (*cluster*) che caratterizzino al meglio la popolazione (es. segmentazione della clientela in categorie).
- **Alberi Decisionali:** usati per la comprensione di un particolare fenomeno poiché permettono di classificare, in ordine di importanza, le cause che portano al verificarsi di un evento (es. valutazione delle categorie di rischio dei clienti per le società che concedono mutui e prestiti).
- **Serie Temporal:** individuazione di pattern ricorrenti o atipici in sequenze di dati complesse (identificazione di schemi associati alla crescita dei titoli di borsa, rilevazione anomalie nel sistema di monitoraggio, analisi percorsi di navigazione in siti web).

ROLAP (Relational OLAP): giustificato dall'enorme lavoro svolto sul modello relazionale, dalla diffusa esperienza aziendale sull'utilizzo di basi di dati relazionali e dall'elevato livello di prestazioni e flessibilità dei DBMS relazionali.

Si presenta la necessità di elaborare tipologie specifiche di schemi che permettano di traslare il modello multidimensionale sul modello relazionale. C'è il problema delle prestazioni (costose operazioni di join tra tabelle di elevate dimensioni).

MOLAP (Multidimensional OLAP): è basato su un modello logico ad hoc sul quale i dati e le operazioni multidimensionali possono essere direttamente rappresentati. I dati vengono fisicamente memorizzati in vettori e l'accesso è di tipo posizionale.

Il **vantaggio** dell'approccio **MOLAP** rispetto a quello **ROLAP** è che le operazioni multidimensionali sono realizzabili in modo semplice e naturale, senza necessità di ricorrere a join (prestazioni ottime). Non esistendo ancora uno standard per il modello logico multidimensionale, le diverse implementazioni **MOLAP** hanno poco in comune (solo l'utilizzo di tecnologie ottimizzate per trattare il problema della sparsità).

La **qualità di un processo** misura la sua aderenza agli obiettivi degli utenti ed è caratterizzata da: *accuratezza* (conformità tra valore memorizzato e quello reale), *attualità* (dato memorizzato non obsoleto), *completezza* (non mancano informazioni), *consistenza* (rappresentazione uniforme dei dati), *disponibilità* (dati facilmente disponibili all'utente), *tracciabilità* (possibilità di risalire alla fonte di ciascun dato), *chiarezza* (dati facilmente interpretabili).

Il Ciclo di Vita del Data Warehouse

Molte organizzazioni mancano di esperienza e capacità per affrontare le sfide implicite nei progetti di data warehousing. Una delle cause principali è la mancata adozione di un *approccio metodologico*, che minimizza i rischi di insuccesso essendo basato su un'analisi costruttiva degli errori commessi. Il rischio di ottenere un risultato insoddisfacente è particolarmente alto a causa delle elevatissime aspettative degli utenti. Una larga parte della responsabilità della riuscita del progetto ricade sulla qualità dei dati sorgente e sulla lungimiranza, disponibilità e dinamismo del personale dell'azienda.

Approccio Top-Down

Analizza i bisogni globali dell'intera azienda e pianifica lo sviluppo del DW per poi progettare e realizzarlo nella sua interezza.

Il **vantaggio** di questo approccio è il seguente:

- promette ottimi risultati poiché si basa su una visione globale dell'obiettivo e garantisce di produrre un DW consistente e ben integrato.
-
-
-
-
-

Gli **svantaggi** di questo approccio sono i seguenti:

- il preventivo con costi onerosi e lunghi tempi di realizzazione scoraggiano la direzione dall'intraprendere il progetto;
- affrontare contemporaneamente l'analisi e la riconciliazione di tutte le sorgenti di interesse è complesso;
- prevedere a priori le esigenze delle diverse aree aziendali è pressoché impossibile, e il processo di analisi rischia di subire una paralisi;
- il fatto di non prevedere la consegna a breve termine di un prototipo non permette agli utenti di verificare l'utilità del progetto e ne fa scemare l'interesse e la fiducia.

Approccio Bottom-Up

Il DW viene costruito in modo incrementale, assemblando iterativamente più data mart, ciascuno dei quali incentrato su un insieme di fatti collegati a uno specifico settore aziendale e di interesse per una certa categoria di utenti.

I **vantaggi** di questo approccio sono i seguenti:

- determina risultati concreti in tempi brevi;
- non richiede elevati investimenti finanziari;
- permette di studiare solo le problematiche relative al data mart in oggetto;
- fornisce alla dirigenza un riscontro immediato sull'effettiva utilità del sistema in via di realizzazione;
- mantiene costantemente elevata l'attenzione sul progetto

Lo **svantaggio** di questo approccio è il seguente:

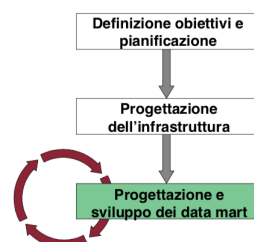
- determina una visione parziale del dominio di interesse.
-
-
-
-
-

Il primo *data mart* da prototipare deve essere quello che gioca il ruolo più strategico per l'azienda, deve ricoprire un ruolo centrale per l'intero DW, si deve appoggiare su fonti dati già disponibili e consistenti. Il ciclo di sviluppo è il seguente:

Definizione obiettivi e pianificazione: individuazione degli obiettivi, stima delle dimensioni del sistema, scelta dell'approccio per la costruzione, valutazione dei costi e del valore aggiunto, analisi dei rischi e delle aspettative.

Progettazione dell'infrastruttura: si analizzano e si comparano le possibili soluzioni architetturali valutando le tecnologie e gli strumenti disponibili.

Progettazione e sviluppo dei data mart: ogni iterazione comporta la creazione di un nuovo data mart e nuove applicazioni, che vengono integrate nel sistema di data warehousing.



Analisi e Riconciliazione delle Sorgenti Operazionali

Il progettista, confrontandosi con gli esperti del dominio applicativo, acquisisce una buona conoscenza delle sorgenti operazionali:

- **Ricognizione:** esame approfondito degli schemi locali per comprendere appieno il dominio applicativo;
- **Normalizzazione:** per correggere gli schemi locali al fine di modellare in modo più accurato il dominio applicativo.

Ricognizione e normalizzazione devono essere svolte anche qualora sia presente una sola sorgente dati; se esistono più sorgenti, l'operazione dovrà essere ripetuta per ogni singolo schema.

L'**integrazione** di un insieme di sorgenti dati eterogenee (basi di dati relazionali, file dati, sorgenti legacy) consiste nell'individuazione delle corrispondenze tra i concetti rappresentati e nella risoluzione dei conflitti evidenziati, per creare un unico schema globale i cui elementi possano essere correlati con i corrispondenti elementi degli schemi locali (*mapping*).

Deve anche identificare l'insieme di concetti distinti e memorizzati in schemi differenti che sono correlati attraverso proprietà semantiche (*interschema*). Per ragionare sui concetti negli schemi delle diverse sorgenti è necessario utilizzare un unico formalismo. Ci possono essere diverse relazioni tra concetti comuni:

- **identità:** utilizzati stessi costrutti, concetto modellato dallo stesso punto di vista e non vengono commessi errori di specifica;
- **equivalenza:** sono stati utilizzati costrutti diversi (ma equivalenti) e non ci sono errori di specifica o diversità di percezione;
- **comparabilità:** i costrutti utilizzati e i punti di vista dei progettisti non sono in contrasto tra loro;
- **incompatibilità:** gli schemi sono in contrasto a causa dell'incoerenza nelle specifiche;

Ci sono quattro **fasi dell'integrazione**:

- **preintegrazione:** viene definita la strategia di integrazione;
- **Comparazione degli schemi:** analisi comparativa dei diversi schemi per identificare le correlazioni e i conflitti tra i concetti.
- **Allineamento degli schemi:** risoluzione dei conflitti evidenziatisi al passo precedente, che si ottiene applicando primitive di trasformazione agli schemi sorgenti o allo schema riconciliato temporaneamente definito (cambio di nomi e tipi degli attributi, modifica delle dipendenze). Non sempre i conflitti possono essere risolti, poiché derivano da inconsistenze di base del sistema informativo; in questo caso la soluzione deve essere discussa con gli utenti che dovranno fornire indicazioni su qual è la più fedele interpretazione del mondo reale.
- **Fusione e ristrutturazione degli schemi:** gli schemi allineati vengono fusi a formare un unico schema riconciliato (sovrapposizione dei concetti comuni a cui saranno collegati i rimanenti concetti degli schemi locali). Poi si applicano altre trasformazioni per migliorare la struttura dello schema riconciliato rispetto a completezza, minimalità e leggibilità.

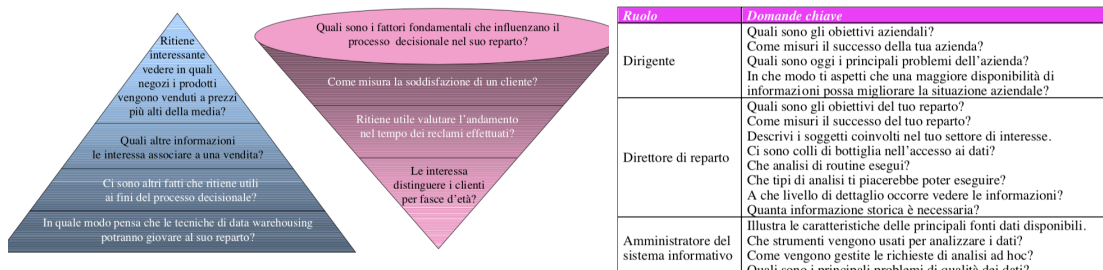
Analisi dei Requisiti

La fase di analisi dei requisiti ha l'obiettivo di raccogliere le esigenze di utilizzo del data mart espresse dagli utenti finali. Ha un'importanza strategica poiché influenza le decisioni da prendere riguardo: schema concettuale dei dati, specifiche delle applicazioni per l'analisi dei dati, architettura del sistema, linee guida per la manutenzione e l'evoluzione del sistema.

La "fonte" principale da cui attingere i requisiti sono i futuri utenti del data mart (*business users*). La differenza nel linguaggio usato da progettisti e utenti rende il dialogo difficile e a volte infruttuoso. Per gli aspetti più tecnici, saranno gli amministratori del sistema informativo a fungere da riferimento per il progettista (in questo caso, i requisiti che dovranno essere catturati riguardano principalmente vari vincoli sul sistema di data warehousing).

Esistono due tipi di **Interviste**:

- a **piramide** (*approccio induttivo*): si parte da domande dettagliate per poi ampliare l'argomento con domande aperte che richiedono risposte più generali (per superare lo scetticismo dell'intervistato, poiché non richiede un forte coinvolgimento).
- a **imbuto** (*approccio deduttivo*): si parte da domande molto generali per poi restringere l'argomento a temi specifici. Utile nel caso in cui l'intervistato sia emozionato (le domande generali non prevedono risposta sbagliata e alleviano la tensione).



I **fatti** sono i concetti su cui gli utenti finali del data mart baseranno il processo decisionale; ogni *fatto* descrive una categoria di eventi che si verificano in azienda. Fissare le dimensioni di un fatto è importante poiché significa determinarne la *granularità* (il più fine livello di dettaglio a cui i dati saranno rappresentati). La scelta della granularità di un fatto nasce da un compromesso tra due esigenze contrapposte: raggiungere un'elevata flessibilità d'utilizzo e conseguire buone prestazioni. Per ogni fatto occorre definire l'*intervallo di storicizzazione* (l'arco temporale che gli eventi memorizzati dovranno coprire).

Il riconoscimento di fatti (dimensioni e misure) è strettamente collegato all'identificazione di un *carico di lavoro preliminare* che può essere espresso in linguaggio naturale (utile per valutare la granularità dei fatti e iniziare ad affrontare il problema dell'aggregazione).

	Data mart	Fatti
commerciale/ manfatturiero	approvvigionamenti	acquisti, inventario di magazzino, distribuzione
	produzione	confezionamento, inventario, consegna, manifattura
	gestione domanda	vendite, fatturazione, ordini, spedizioni, reclami
	marketing	promozioni, fidelizzazione, campagne pubblicitarie
finanziario	bancario	conti correnti, bonifici, prestiti ipotecari, mutui
	investimenti	acquisto titoli, transazioni di borsa
	servizi	carte di credito, domiciliazioni bollette
sanitario	scheda di ricovero	ricoveri, dimissioni, interventi chirurgici, diagnosi
	pronto soccorso	accessi, esami, dimissioni
trasporti	medicina di base	scelte, revoche, prescrizioni
	merci	domanda, offerta, trasporti
telecomunicazioni	passenger	domanda, offerta, trasporti
	manutenzione	interventi
turismo	traffico	traffico in rete, chiamate
	CRM	fidelizzazione, reclami, servizi
gestionale	gestione domanda	biglietteria, noleggi auto, soggiorni
	CRM	frequent-flyers, reclami
	logistica	trasporti, scorte, movimentazione
	risorse umane	assunzioni, dimissioni, promozioni, incentivi
	budgeting	budget commerciale, budget di marketing
	infrastrutture	acquisti, opere

Fatto	Possibili dimensioni	Possibili misure	Storicità
inventario di magazzino	prodotto, data, magazzino	quantità in magazzino	1 anno
vendite	prodotto, data, negozio	quantità venduta, importo, sconto	5 anni
linee d'ordine	prodotto, data, fornitore	quantità ordinata, importo, sconto	3 anni

Fatto	Interrogazione
inventario di magazzino	Quantità media di ciascun prodotto presente mensilmente in tutti i magazzini. Prodotti per i quali è stata esaurita la scorta contemporaneamente in tutti i magazzini in almeno un'occasione durante la settimana passata. Andamento giornaliero delle scorte complessive per ciascun tipo di prodotto.
vendite	Quantità totali di ciascun tipo di prodotto vendute durante l'ultimo mese. Incasso totale giornaliero di ciascun negozio. Per un dato negozio, incassi relativi alle diverse categorie di prodotti durante un certo giorno. Riepilogo annuale degli incassi per regione relativamente a un dato prodotto.
linee d'ordine	Quantità totale ordinata annualmente presso un certo fornitore. Importo giornaliero ordinato nell'ultimo mese per un certo tipo di prodotto. Sconto massimo applicato da ciascun fornitore durante l'ultimo anno per ciascuna categoria di prodotto.

Altri requisiti sono: vincoli di progettazione logica e fisica (spazio disponibile), progetto dell'alimentazione (periodicità dell'alimentazione), architettura del sistema di data warehousing (architettura da implementare, numero di livelli, presenza di data mart dipendenti o indipendenti, materializzazione del livello riconciliato), applicazioni per l'analisi dei dati (analisi delle tipologie di interrogazioni e dei rapporti analitici normalmente richiesti), piano di avviamento e piano di formazione.

Progettazione Concettuale

È universalmente riconosciuto che un DW si appoggia sul modello multidimensionale, ma non c'è accordo sulla metodologia di progetto concettuale (il modello entità/relazione non può essere usato per modellare i DW).

Il **Dimensional Fact Model (DFM)** è un modello concettuale grafico per data mart pensato per: supportare il progetto concettuale, creare un ambiente su cui formulare in modo intuitivo le interrogazioni dell'utente, permettere il dialogo tra progettista e utente finale per raffinare le specifiche dei requisiti, creare una piattaforma stabile da cui partire per il progetto logico (indipendentemente dal modello logico), restituire una documentazione espressiva e non ambigua.

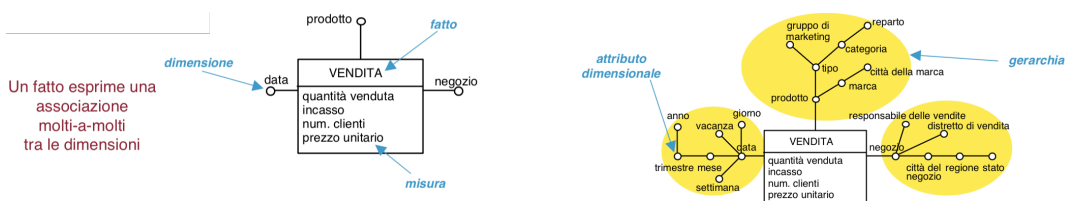
La rappresentazione concettuale del DFM consiste in un insieme di **schemi di fatto**.

Gli **elementi di base** da modellare sono:

- **fatti**: modella un insieme di eventi che accadono nell'impresa (vendite, spedizioni, acquisti...). È essenziale che un fatto abbia aspetti dinamici, ovvero evolva nel tempo;
- **misure**: proprietà numerica di un fatto e ne descrive un aspetto quantitativo per l'analisi (vendita misurata dall'incasso); uno **schema di fatto** si dice **vuoto** se non ha misure (il fatto registra solo il verificarsi di un evento).
- **dimensione**: proprietà di un fatto e ne descrive una coordinata di analisi (per le vendite sono prodotto, negozio, data).

I **costrutti di base** del DFM sono:

- **attributi dimensionali**: dimensioni ed eventuali altri attributi che le descrivono (un prodotto è descritto dal suo tipo, dalla categoria cui appartiene, dalla sua marca, dal reparto in cui è venduto);
- **gerarchia**: albero direzionato i cui nodi sono attributi dimensionali e i cui archi modellano associazioni molti-a-uno tra coppie di attributi dimensionali. Racchiude una dimensione (alla radice dell'albero) e gli attributi dimensionali che la descrivono.



Le **Naming Conventions** prevedono: attributi dimensionali con nomi diversi (nomi uguali devono essere differenziati con il nome di un attributo dimensionale che li precede nella gerarchia), i nomi degli attributi non dovrebbero riferirsi al fatto a cui appartengono (evitare *shipped product* e *shipment date*), attributi con stesso significato in schemi diversi devono avere stesso nome.

Costrutti avanzati: un *attributo descrittivo* contiene informazioni aggiuntive su un attributo dimensionale di una gerarchia. Non viene usato per l'aggregazione poiché ha valori continui e/o poiché deriva da un'associazione uno-a-uno (alcuni archi dello schema possono essere *opzionali*).

Addittività: l'aggregazione richiede di definire un operatore adatto (SUM, AVG, MIN, MAX) per comporre i valori delle misure che caratterizzano gli eventi primari in valori da abbinare a ciascun evento secondario. Ci sono tre categorie di misure:

- **Misure di Flusso**: si riferiscono a un periodo, al cui termine vengono valutate in modo cumulativo (il numero di prodotti venduti in un giorno, l'incasso mensile, il numero di nati in un anno);
- **Misure di Livello**: vengono valutate in particolari istanti di tempo (n° prodotti in inventario, n° abitanti di una città);

- **Misure Unitarie:** vengono valutate in particolari istanti di tempo, ma sono espresse in termini relativi (prezzo unitario di un prodotto, percentuale di sconto, cambio di una valuta)

Una misura è detta *additiva* su una dimensione se i suoi valori possono essere aggregati lungo la corrispondente gerarchia tramite l'operatore di somma, altrimenti è *non-additiva*. Una misura **non-additiva** è *non-aggregabile* se nessun operatore di aggregazione può essere usato su di essa.

Ci sono due diversi approcci alla progettazione concettuale:

- **basata sui requisiti:** ricavare dalle interviste condotte presso l'utente, un'indicazione precisa sui fatti da rappresentare, le misure che li descrivono e le gerarchie attraverso cui aggregarli utilmente. Il problema del collegamento tra lo schema concettuale e le sorgenti operazionali viene affrontato in seguito;
- **basata sulle sorgenti:** si può definire lo schema concettuale in funzione della struttura delle sorgenti, evitando il complesso compito di stabilire il legame con esse a posteriori. Inoltre, è possibile derivare uno schema concettuale dagli schemi operazionali in modo pressoché automatico.

La progettazione concettuale viene effettuata a partire dalla documentazione relativa al database riconciliato (schemi E/R, schemi relazionali, schemi XML . . .). I passi di progettazione consistono nella definizione dei fatti e per ogni passo: costruzione albero degli attributi, editing dell'albero, definizione delle dimensioni, definizione delle dimensioni, creazione dello schema di fatto.

Definizione dei Fatti: entità o relazioni che rappresentano archivi frequentemente modificati (come VENDITA) possono definire *fatti*; non è così per quelli che rappresentano archivi quasi-statici. Ogni fatto identificato diventa la radice di un nuovo schema.

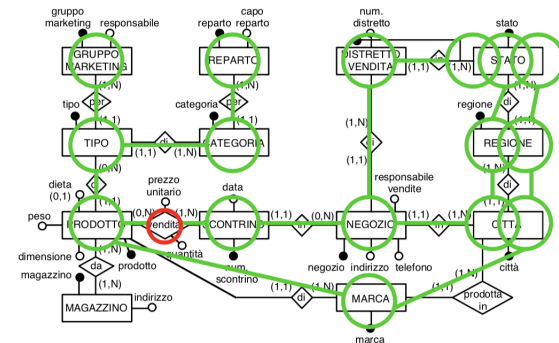
Costruzione dell'Albero degli Attributi: è un albero in cui ogni vertice corrisponde a un attributo dello schema sorgente, la radice corrisponde all'identificatore (chiave primaria), per ogni vertice v , l'attributo corrispondente determina tutti gli attributi corrispondenti ai discendenti di v . L'albero degli attributi può essere costruito in modo automatico applicando una procedura che naviga ricorsivamente le dipendenze espresse, nello schema sorgente, dagli identificatori e dalle associazioni a-uno.

```

root=nuovoVertice(ident(F)); // ident(F) è l'identificatore di F
// la radice dell'albero è etichettata con l'identificatore dell'entità scelta come fatto
traduci(F, root);

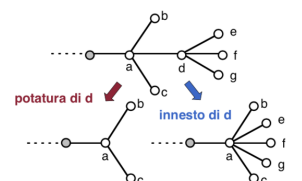
procedura traduci(E, v):
// E è l'entità corrente dello schema sorgente, v il vertice corrente dell'albero
{
  per ogni attributo a di E tale che a <= ident(E)
  {
    aggiungiFiglio(v, nuovoVertice(a));
    // aggiunge al vertice v un figlio a
  }
  per ogni entità G connessa a E da un'associazione R tale che max(E, R)=1
  {
    per ogni attributo b di R
    {
      aggiungiFiglio(v, nuovoVertice(b));
      // aggiunge al vertice v un figlio b
    }
    prossimo=nuovoVertice(ident(G));
    // crea un nuovo vertice con il nome dell'identificatore di G ...
    aggiungiFiglio(v, prossimo);
    // ... lo aggiunge a v come figlio ...
    traduci(G, prossimo);
    // ... e innesca la ricorsione
  }
}

```

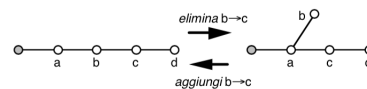


Editing dell'Albero degli Attributi: non tutti gli attributi dell'albero sono d'interesse per il data mart; quindi, l'albero può essere manipolato per eliminare i livelli di dettaglio non necessari.

La *potatura* di un vertice v si effettua eliminando l'intero sottoalbero con radice in v . L'*innesto* viene utilizzato quando, sebbene un vertice esprima un'informazione non interessante, è necessario mantenere nell'albero i suoi discendenti. L'innesto del vertice v , con padre v' , viene effettuato collegando tutti i figli di v direttamente a v' ed eliminando v . Quando un vertice opzionale viene innestato, tutti i suoi figli ereditano il trattino di opzionalità.



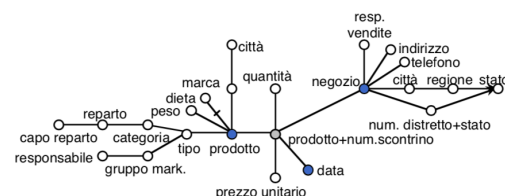
Può anche essere necessario modificarne radicalmente la struttura sostituendo il padre di un certo nodo: ciò corrisponde ad aggiungere o eliminare una dipendenza.



Definizione delle Dimensioni: devono essere scelte nell'albero degli attributi tra i vertici figli della radice; possono corrispondere ad attributi discreti o a intervalli di valori di attributi discreti o continui. La loro scelta definisce la granularità degli eventi primari.

Il tempo dovrebbe sempre essere una dimensione se:

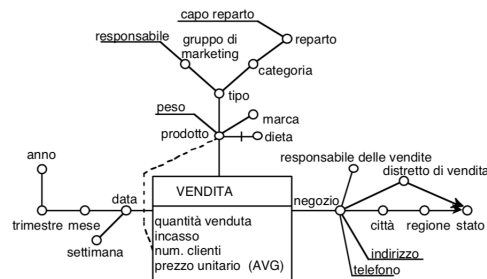
- la sorgente è uno schema storico e il tempo è rappresentato esplicitamente come un attributo; se appare nell'albero degli attributi come figlio di un vertice diverso dalla radice, si può effettuare un innesto o eliminare una dipendenza per farlo diventare un figlio diretto della radice e quindi una dimensione.
- nelle sorgenti snapshot il tempo non è rappresentato; viene aggiunto manualmente allo schema di fatto.



GLOSSARIO

Creazione dello Schema di Fatto: l'albero degli attributi si traduce in uno schema di fatto che include dimensioni e misure definite.

- è possibile aggiungere attributi dimensionali definendo opportuni intervalli per attributi numerici (es. sulla dimensione tempo);
- gli attributi che non verranno usati per l'aggregazione possono essere contrassegnati come descrittivi; tra questi compariranno anche gli attributi determinati da associazioni uno-a-uno e privi di discendenti;
- attributi alfanumerici figli della radice ma non prescelti né come dimensioni né come misure, se la granularità degli eventi primari corrisponde a quella dell'entità verranno rappresentati come attributi descrittivi (altrimenti devono essere potati).



Il diagramma illustra la struttura di un database relazionale per un negozio. Le entità e i loro attributi sono:

- NEGOZIO**: data, incasso, quantità
- STATO**: data, incasso, quantità
- VENDITA**: data, incasso, quantità
- CODVENDITA**: data, incasso, quantità
- PRODOTTI**: data, incasso, quantità
- CATEGORIA**: data, incasso, quantità
- REPARTO**: data, incasso, quantità
- RICOVERO**: data, incasso, quantità
- PAZIENTE**: data, incasso, quantità
- DIAGNOSI**: data, incasso, quantità

Le relazioni e le cardinalità sono:

- NEGOZIO** (1) a **STATO** (N): relazione 1:N.
- STATO** (1) a **VENDITA** (N): relazione 1:N.
- VENDITA** (1) a **CODVENDITA** (1): relazione 1:1.
- CODVENDITA** (1) a **PRODOTTI** (N): relazione 1:N.
- PRODOTTI** (1) a **CATEGORIA** (N): relazione 1:N.
- REPARTO** (1) a **RICOVERO** (N): relazione 1:N.
- RICOVERO** (1) a **PAZIENTE** (N): relazione 1:N.
- RICOVERO** (1) a **DIAGNOSI** (N): relazione 1:N.
- DIAGNOSI** (1) a **CATEGORIA** (N): relazione 1:N.

È necessario identificare in fase di progettazione un carico di lavoro di riferimento (reportistica, standard, colloqui con gli utenti). Le interrogazioni OLAP sono facilmente caratterizzabili: pattern di aggregazione, misure richieste, clausole di selezione.

Il Volume di Dati: consiste nelle informazioni necessarie a determinare/stimare la dimensione del data mart (numero di valori distinti degli attributi nelle gerarchie, lunghezza degli attributi, numero di eventi di ogni fatto). È utilizzato sia durante la progettazione logica sia durante la progettazione fisica per determinare: dimensione delle tabelle, dimensione degli indici, costi di accesso. La bontà delle stime è spesso compromessa a causa del problema della sparsità.

22

Progettazione Logica

La modellazione concettuale è indipendente dal modello logico, non è così per i temi legati alla modellazione logica. La struttura multidimensionale dei dati può essere rappresentata utilizzando due distinti modelli logici:

- **MOLAP** (*Multidimensional On-Line Analytical Processing*): memorizzano i dati utilizzando strutture multidimensionali (es. vettori multidimensionali);
- **ROLAP** (*Relational On-Line Analytical Processing*): utilizza modello relazionale per rappresentare i dati multidimensionali.

Sistemi MOLAP

L'utilizzo di soluzioni MOLAP:

- può fornire ottime prestazioni poiché le operazioni non devono essere "simulate" mediante complesse istruzioni SQL;
- pone il problema della *sparsità*: in media solo il 20% delle celle dei cubi contiene effettivamente informazioni, mentre le restanti celle corrispondono a fatti non accaduti;
- è frenato dalla mancanza di strutture dati standard: i diversi produttori di software utilizzano strutture proprietarie che li rendono difficilmente sostituibili e accessibili mediante strumenti di terze parti;
- progettisti e sistemisti non vogliono rinunciare alla loro lunga esperienza sui sistemi relazionali.

Le tecniche di **gestione della sparsità** sono basate sui seguenti principi:

- **suddivisione delle dimensioni**: consiste nel partizionare un cubo n-dimensionale in più sottocubi n-dimensionali (*chunk*). I singoli chunk potranno essere caricati più agevolmente in memoria e saranno gestiti in modo differente a seconda che siano *densi* (la maggior parte delle celle contiene informazioni) oppure *sparsi* (la maggior parte non contiene informazioni).
- **compressione dei chunk sparsi**: essi vengono rappresentati in forma compressa al fine di evitare lo spreco di spazio dovuto alla rappresentazione di celle che non contengono informazioni. Una struttura dati comunemente usata prevede un indice che riporti il solo offset delle celle che effettivamente contengono informazioni.

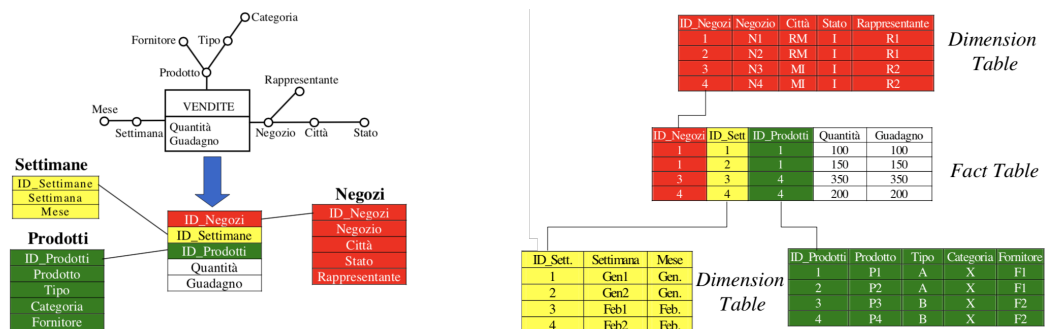
Sistemi ROLAP: Schema a Stella

La modellazione multidimensionale su sistemi relazionali è basata sullo schema a stella (*star schema*) e sulle sue varianti.

Uno **schema a stella** è composto da:

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

Non si hanno problemi di *sparsità* in quanto vengono memorizzate soltanto le tuple corrispondenti a punti dello spazio multidimensionale per cui esistono eventi



Le *Dimension Table* sono completamente denormalizzate (es. Prodotto \rightarrow Tipo); è sufficiente un join per recuperare tutti i dati relativi a una dimensione. La denormalizzazione introduce una forte ridondanza nei dati.

La *Fact Table* contiene tuple relative a diversi livelli di aggregazione (l'elevata dimensione incide sui tempi di accesso ai dati).

Snowflake Schema

Lo **schema a fiocco di neve** riduce la denormalizzazione delle *dimension table* DT_i degli schemi a stella eliminando alcune delle dipendenze transitive che le caratterizzano. Le *dimension table* $DT_{i,j}$ di questo schema sono caratterizzate da:

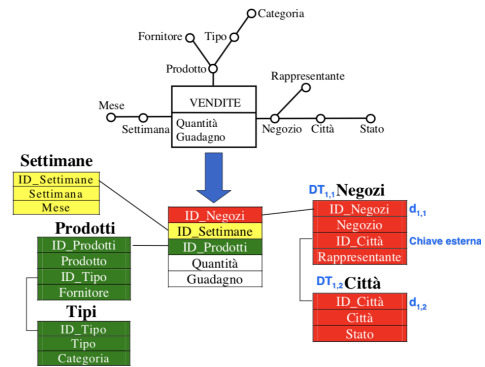
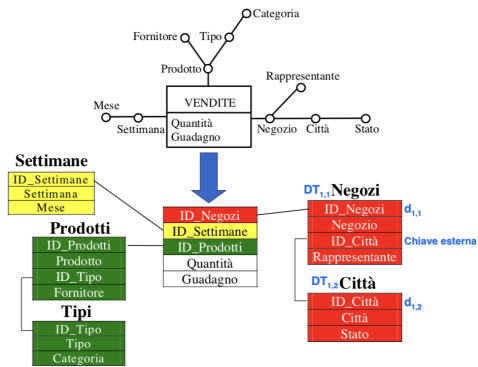
- una chiave primaria $d_{i,j}$;
- il sottoinsieme degli attributi di DT_i che dipendono funzionalmente da $d_{i,j}$;
- zero o più chiavi esterne importate da altre $DT_{i,k}$ necessarie a garantire la ricostruibilità del contenuto informativo di DT_i .

Denominiamo **primarie** le *dimension table* le cui chiavi sono importate nella *fact table*, **secondarie** le rimanenti.

Lo spazio richiesto per la memorizzazione dei dati si riduce grazie alla normalizzazione. È necessario inserire nuove chiavi surrogate per determinare le corrispondenze tra *dimension table* primarie e secondarie.

L'esecuzione di interrogazioni che coinvolgono solo gli attributi contenuti nella *fact table* e nelle *dimension table* primarie è avvantaggiata. Il tempo di esecuzione delle interrogazioni che coinvolgono attributi delle *dimension table* secondarie aumenta.

Lo **snowflake schema** è particolarmente utile in presenza di dati aggregati.



Normalizzazione con lo Snowflake Schema: le caratteristiche degli schemi a stella richiedono particolare attenzione affinché nella nuova relazione sia spostato il corretto insieme di attributi. La presenza di più dipendenze funzionali transitive in cascata fa sì che, affinché la decomposizione sia efficace, tutti gli attributi che dipendono dall'attributo che ha determinato lo snowflaking siano posti nella nuova relazione.

