SISTEMI DI ANALISI DEI DATI ***(Gli 'Analytics')***

*Breve introduzione panoramica ai Sistemi di Analisi dei dati per le decisioni. I Componenti del processo decisionale***.**

1- Premessa.

Si tratta normalmente di componenti dei 'Sistemi Informativi gestionali e direzionali' delle Organizzazioni; e si osserva in essi anche il ruolo dei sistemi di 'Business Intelligence'; attraverso le loro Architetture specifiche, in continua evoluzione.
Si parla anche di 'Sistemi di supporto alle decisioni'. Che si avvalgono sempre di più dell'ausilio di Modelli e Metodi matematici per le decisioni. Ed il supporto fondamentale ed in notevole recente sviluppo della Information Technology.

Ora qui cercheremo di dare una breve panoramica introduttiva degli strumenti disponibili adesso per 'fare Analytics': metodologie, algoritmi matematici e statistici, hardware, software. La scienza relativa è molto ampia e si innova continuamente; vanno scelti i metodi e gli strumenti più adatti alla specifica problematica da risolvere.
L'argomento non è certo dei più attrattivi; però si tratta di un salto culturale di cui avremo sempre più bisogno.

Come abbiamo già riferito in precedenti articoli di questa rivista.
Dobbiamo prendere coscienza che tutto quello che oggi i nostri Manager stano imparando sarà senz'altro la loro Cultura tecnica/metodica di base. *Per poter incominciare ad agire nella loro professione.*
***MA NON BASTERÀ*. Occorrerà supportarla con più INFORMATICA e più MATEMATICA.**
Se vorranno 'sopravvivere' professionalmente. *E non essere superati ad es. dai tedeschi, dagli olandesi, e poi anche dai cinesi. Molti più dati ? Occorrerà molto maggiore capacità di elaborazione. Aumento di Volume, Varietà, Velocità, Variabilità, Veridicità e Decadibilità dei dati che saranno disponibili.* ***L' 'Analisi dei dati' odierna non sarà, o non lo è già più, sufficiente per prendere le migliori decisioni.*** *Prima che altri popoli più intraprendenti ci superino nelle nostre attività culturali ed economiche.*

Oggi si parla molto di IoT (Internet Of Things), Industria 4.0, Piani Industry 4.0, Networking 4.0, ecc.
Tutto è importante: i sensori, le tecnologie di fabbrica e di supply chain, i tecnologi; le facilitazioni economiche e fiscali, ecc…; i PID, i Digital Innovation Hub, i Competence Center, ecc…
Ma ***le cose più importanti di tutte*** *saranno* ***la gestione e l'utilizzo adeguati della enormemente maggiore entità di dati, di testi, di immagini,*** *ecc… provenienti dall'esterno e dall'interno; per l'innovazione e l'ottimizzazione dei processi aziendali.*
Molti più dati dai sensori ? L' 'Analisi dei dati' odierna non sarà, o non lo è già, più sufficiente per prendere le migliori decisioni.
Sarà necessaria una nuova 'Analitica' con e per i 'Big Data'; anzi un 'flusso di Analytics' lungo tutta l'organizzazione*.*
Per sapere, prevedere, ben operare. Le Analisi applicabili sui dati, strutturati e non strutturati, possono già essere descrittive, predittive, prescrittive; e ora potranno essere anche 'cognitive'.

Come noto, ***l'attuazione delle innovazioni del 4.0 impatterà soprattutto sui Manager aziendali****. I Manager sono i soli che hanno il know how adatto per le innovazioni: conoscono, impostano, controllano i Processi. E quindi dovranno conseguentemente innovare detti processi; ben supportati da una adeguata varietà di Analytics.*
Per fare esaminare, filtrare, normalizzare, analizzare, interpretare e comunicare i dati giusti; anche in sequenza per i livelli superiori dei processi in una struttura normalmente a piramide di una organizzazione.

Vari tipi di Analytics occorreranno *per i dati che scaturiscono ai vari livelli della struttura (piramidi di sistemi/analytics): per le tecnologie di comunicazione, attivazione e integrazione dei macchinari e degli impianti operativi di base:- per le attività operative dei processi gestionali ai vari livelli;- per il controllo e le interpretazioni del funzionamento di detti processi ai vari livelli;- per il supporto alle interpretazioni e le decisioni ai vari livelli manageriali;- per i Manager tecnici di sviluppi, di fabbrica e di supply chain;- per i Manager commerciali, amministrativi e finanziari;- per il Top Management. Ecc., ecc.*
E con quei sistemi informativi (Scada, Mes, Erp, Crm, ecc…) andranno integrati e il più possibile standardizzati.

LA STRATEGIA DIVENTA 'DIGITAL' E RICHIEDE COMPETENZE NUOVE.
I manager si trovano di colpo a dover affrontare e implementare strategie con una forte componente 'digital' e che comportano pure ***domande di business nuove*****,** la cui risposta richiede ***competenze e conoscenze che ora non sono esplicitamente presenti in azienda. Dal CIO (digital information officer) al … DIO (digital innovation officer).*** Ecco perché molte imprese cercano ***figure professionali nuove*** e affollano i social network con annunci di lavoro al limite del comprensibile. ***SEO SEM Specialist, Social Media Manager, Mobile Developer, Chief Data Officer*** sono solo alcune delle figure professionali più ricercate anche se, al momento, non è ancora chiaro quali leve avranno a disposizione per portare innovazione.

Quindi, per la sua pervasività la Digital Transformation non è più soltanto una questione tecnologica e neanche una questione solo di visione strategica, ma ***una sfida vera e profonda che coinvolge tutto il capitale umano*** e impone di sviluppare in ogni area aziendale ***nuove competenze e professionalità*** che siano in grado di interpretare al meglio le nuove opportunità e condurre il cambiamento.

2- Creare valore dai dati tradizionali e dai Big Data

Molti più dati ? I Dati sono un patrimonio sempre più importante da utilizzare. Occorre adeguare le competenze per poterli gestire al meglio ed in maniera innovativa**.**

Basi di dati. I dati vengono conservati e protetti su appositi archivi sempre più grandi e più complessi. I tradizionali sono di tipo OLTP (on line transaction processing). Si parla di Data Warehouse - Datamart – Cubi di dati - Qualità dei dati - Strumenti per la gestione dei Data Base.
I Big Data. Varie le Tipologie di Big Data (dati di enormi volumi, grandi velocità, varietà, ecc…), per le varie aree applicative.
Tecnologie per i Big Data. Per immagazzinamento e organizzazione. Ad esempio la Piattaforma Hadoop; o quella Spark. Ecc. Si parla di architettura "Data Lake" (raccolta e gestione); di Calcolo parallelo/distribuito; di Piattaforme e motori di calcolo - Di strumenti per la "data ingestion" (acquisizione): Etl, Apache Flume, Storm, Kafka, ecc…..

3- Tecniche per l'Analisi dei dati

Data mining (*ricerca, raccolta strutturata ed analitica dei dati****).*** Vari i modelli e i metodi di statistica classica e quelli OLAP (On-Line Analytical Processing). Tecniche di drill-down, roll-up, slicing/dicing, pivoting, drill-through. Rappresentazione dei dati in ingresso - Metodologie di analisi - Rappresentazione dei dati in output
Le tecniche di Analisi: come estrarre valore dai dati (*data monetization*). L'Analisi descrittiva (reportistica per mezzo di motori OLAP). L'Analisi predittiva (cosa è probabile che avvenga; tecniche di modellazione e preparazione dei dati, di analisi statistiche e di machime learning). L'Analisi prescrittiva (modelli cognitivi e predittivi, con anche tecniche per l'evidenza dei motivi dei probabili eventi; (alberi decisionali, Fuzzy Rule-Based System, Logic Learning Machine, ecc…).

4- Gli Strumenti

Strumenti software utilizzabili per le Elaborazioni e le Analisi dei dati *(esempi).*
Map Reduce, Pig,... Motori di acquisizione e di analisi di grandi dataset, di dati strutturati e non; con linguaggi per la gestione e la elaborazione.
SQL per l'analisi dei dati strutturati. SQL per la preparazione e l'analisi dei dataset - Utilizzo di Hive per interrogazione dei dati.- E poi Impala, U-SQL, Apache Drill.
Spark. Sistema indipendente ma integrabile in Hadoop. Motore di calcolo distribuito, fault tolerant, altamente scalabile.
L'ambiente R. Sistema software di elaborazione dati, che con un linguaggio specifico, basato sui vettori, consente di eseguire analisi statistiche di base ed avanzate. Utilizzo per calcoli statistici avanzati, machine learning, text mining, analisi dei grafi (reti, social network), applicazioni finanziarie, analisi econometri che, analisi di serie storiche, ecc….
L'ambiente Weka. Weka, acronimo di "Waikato Environment for Knowledge Analysis", è un software per l'apprendimento automatico sviluppato nell'università di Waikato in Nuova Zelanda. Weka è un ambiente software interamente scritto in Java. È open source. Normalmente applica dei metodi di apprendimento automatici (learning methods) ad un set di dati (dataset), per analizzarne il risultato. Attraverso questi metodi è possibile, avere quindi una previsione dei nuovi comportamenti dei dati. E' dotato di una buona interfaccia grafica.
Eccetera.
Dataset già esistenti utilizzabili free su Internet.

Predictive e Prescriptive analysis. Qui di seguito un dettaglio per le metodiche più innovative.
Esse fanno normalmente riferimento alla **CRISP-DM** (Cross Industry Standard Process for Data Mining): metodologia UE/IBM per processi di data analysis. Essa prevede le seguenti sei fasi: Business Understanding, Data Understanding, Data preparation, Modeling, Evaluation, Deployment. Con inoltre: Monitoring e feedback; Iterazioni.
E si esplicano dopo i processi propedeutici seguenti.
Esplorazione e Preparazione dei dati. Gli elementi, le variabili, i dati - Esplorazione dei dati: Analisi univariata, Analisi bivariata. Analisi multivariata.
Validazione dei dati. Dati incompleti - Dati soggetti a rumore - Trasformazione - Standardizzazione - Estrazione di attributi.
Riduzione dei dati. Errori comuni nella preparazione dei dati: variabili anacronistiche, campioni troppo piccoli, utilizzo dei campi ID dei database, casi errati per training e testing dei modelli, selezione delle variabili. Gestione degli outliers (dati anomali) e dei valori mancanti, degli utilizzi errati della distribuzione normale dei dati.

Gli Algoritmi. Qui di seguito elenchiamo principali esempi di classi di Algoritmi matematici, statistici, informatici utilizzati per i vari scopi analitici.
Si possono suddividere in classi di Algoritmi o supervisionati, o non supervisionati, o semi-supervisionati; a secondo del grado di interazione che si renda necessario nel loro utilizzo.
Classificazione dei dati. Classificazione per scopo - per modalità di apprendimento - per tipo di output - per similarità di funzionamento. Alberi di classificazione - Alberi decisionali (Decision Trees) – Algoritmi di classificazione con Metodi bayesiani: Naive Bayes: Classificatore bayesiano naive - Reti bayesiane - Regressione logistica.
Regole associative di dati. Struttura dei dati e valutazione - Regole a dimensione singola - Algoritmo 'A priori' - Generazione degli itemset (insiemi di dati frequenti) - Generazione delle regole – Ecc.
Algoritmi di regressione *(analisi di serie di dati)*- Struttura dei modelli di stima - Regressione lineare semplice - Retta di regressione - Regressione lineare multipla. Valutazione dei modelli di regressione - Selezione delle variabili predittive.
Serie storiche. Analisi delle componenti - Media mobile - Scomposizione di una serie storica - Modelli di smoothing esponenziale Modelli autoregressivi - Combinazione di modelli predittivi - Scelta di un modello di previsione.
Machine Learning (metodi che permettono a una 'macchina intelligente' di migliorare le proprie capacità e prestazioni nel tempo). Metodi di analisi che pure con vari appositi algoritmi automatizzano la costruzione di modelli analitici che possono imparare dai dati, identificare modelli autonomamente e prendere decisioni con un intervento umano ridotto al minimo.
Algoritmi di clustering *(raggruppamento).* Misure di affinità - Misure di distanza - Valutazioni preliminari - Metodi di partizione: Algoritmo.delle K-medie, K-means – Algoritmi dei K-medoidi, K-medoids (Partitioning Around Medoids) - Metodi gerarchici: di agglomerazione, di suddivisione.

Reti neurali *(reti di unità di calcolo adattive)***.** Preparazione della rete neurale artificiale (RNA) - Preparazione dei dati - Percettroni di Rosenblatt (classificatore binario)- Reti feed-forward a più livelli.
Deep learning *(apprendimento 'profondo')*. Strutture di base: auto-encoders (reti feed-forward a più livelli) e Restricted Boltzman Machines, in grado di imparare distribuzioni di probabilità - Strutture complesse di strati cognitivi: 'stacked auto-encoders' e 'deep belief networks'.
Support vector machines. L'algoritmo SVM effettua classificazioni tramite la costruzione di iperpiani, in grado di separare in modo ottimale classi di dati- Minimizzazione del rischio strutturale - Iperpiani di margine massimo per la separazione lineare – Separazione non lineare
Fuzzy rules based systems. Consistono di un'applicazione della 'logica fuzzy' a problemi di analisi predittiva. La logica fuzzy o 'logica sfumata' è una logica in cui si può attribuire a ciascuna proposizione un grado di verità diverso da 0 e 1 e compreso tra di loro. È un'estensione della logica booleana. È legata alla teoria degli 'insiemi sfocati'.
Logic Learning Machine. Applicazione di capacità predittiva, che si basa su apprendimento automatico (machine learning) e su programmazione logica (logic programming). Software Rulex.
Model ensembles *(insiemi di modelli di algoritmi*). Bagging (tecnica con estrazioni casuali)– Boosting (training sequenziale di modelli)- Gradient boosting (perfezionamento del boosting) - Random Forest (specifico per gli 'alberi decisionali') - Clustering ensembles (tecniche di ensembles per il 'clustering').

La valutazione dei modelli ***I test per la valutazione dei modelli da adottare/adottati.***
Per considerare gli errori sistematici e la variabilità dei risultati nei modelli in adozione.
Valutazione dei modelli di classificazione. Hold-out: estrazione casuale di elementi e loro utilizzo per 'training' prima e 'test' dopo. Cross validation: suddivisione degli elementi in gruppi uguali e loro utilizzo alcuni per 'training' prima e altri per 'test' dopo – La matrice di confusione: tabulazione di classi di elementi reali e classi previste; in vere/false positive/negative; e le metriche che ne derivano per l'accuratezza, la precisione, la sensitività e la specificità –
Grafici per la valutazione: Gains Chart (per ogni livello percentuale degli elementi quanti veri positivi sono stati identificati, in percentuale sul totale); Lift Chart (indicazione della entità di volte che il modello adottato supera il modello casuale per sensitività); La curva ROC (mostra i possibili valori di falsi positivi e quelli veri, che è si possono ottenere variando la probabilità di appartenenza alla classe positiva).
Valutazione dei modelli di regressione. Tecniche, Metriche analoghe a quelle per i modelli di classificazione.
Valutazione dei modelli di clustering. Misure interne, per verificare la coesione tra gli elementi del cluster e la separazione tra i vari gruppi/cluster - Misure esterne, con utilizzo anche di un raggruppamento definito all'esterno e confronto con i cluster dell'algoritmo.
Valutazioni economiche per i modelli. Valutazioni dei costi fissi per i modelli da adottare, dei costi variabili per ogni azione, dei ricavi o eliminazione di costi per ogni azione.

5- Esempi di Applicazioni di Analysis/Business Intelligence

Molti degli algoritmi e delle tecniche sono già utilizzati ad esempio dai softwares/packages dei Sistemi Informativi a supporto della gestione. ***Sarebbe bene conoscerli*** per utilizzarli adeguatamente. O almeno 'capirli' per scegliere/verificare chi darà lo specifico supporto.
Il'Data Scientist', è un tipo di nuovo specialista molto importante. Ci sono **figure professionali nuove**, ancora 'tutte da costruire', ma saranno nuovi 'super manager'; in grado di lavorare sui dati per fornire risposte e suggerire strategie; affinché le aziende possano efficacemente muoversi, sviluppare nuove proposte e districarsi all'interno della crescente complessità globale. Però la loro formazione sarà molto impegnativa.
Un elenco dei **principali produttori mondiali** di Analytics comprende *ad esempio*: Business Objects, businessobjects.com, Cognos, cognos.com, Ibm, ibm.com, Microsoft, microsoft.com, Microstrategy, microstrategy.it, Oracle, oracle.com, Palisade, palisade.com, QlikView, qlik.com, SAS Institute, sas.com, Talend, talend.com, TARGIT, targit.com. ***Eccetera***.

Per l'Area della Domanda. Analisi di Serie storiche (Modelli diversi di analisi di serie. Modelli a media mobile). Analisi di Regressione (Regressione lineare semplice. Regressione lineare multipla). Modelli statistici predittivi (Modelli di smoothing esponenziale semplice, con correzione di tendenza, Modelli auto regressivi), Modelli di marketing.
Per l'Area dell'Offerta. Gestione/scelte di Marketing. Promotion relazionale, Ottimizzazione della Forza di vendita, Ottimizzazioni di Revenue Management, Simulazioni/analisi what-if, Analisi con 'Albero delle decisioni', Modelli statistici prescrittivi.
Per l'Area della Produzione e dei Materiali. Pianificazione a medio termine, Programmazione esecutiva, Ottimizzazione Capacità produttiva, Lottizzazioni e gestioni di Scorte, Calcolo Livelli Fisiologici di giacenze/stock materiali, Gestione code di servizio, Ottimizzazione gestione/picking di magazzini, Ottimizzazione trasporti/consegne, Modellizzazione/simulazione ed ottimizzazione di processi, Calcoli di affidabilità di processi, Ottimizzazione Impiantistica e layout, Ottimizzazione della supply chain, Processi decisionali di revenue management.
Per la Data Envelopment Analysis (*DEA****, per*** *unità/strutture a confronto)*. - Il modello CCR (Charnes, Cooper, Rhodes): definizione di Obiettivi target, Riferimenti eccellenti, Misura e Frontiera di efficienza. Individuazione di modi operativi efficienti: Analisi di cross-efficienza - Input e output virtuali - Restrizioni sui pesi.
Eccetera,

Riportiamo qualche breve descrizione di applicazione.

Il Demand Forecasting, o previsione della domanda, è l'insieme delle attività tese a prevedere quale sarà l'evoluzione, qualitativa e quantitativa, della domanda di un prodotto o servizio in un tempo che può variare da qualche anno per certi beni durevoli o industriali a pochi giorni, al limite un giorno per l'altro, per i prodotti deperibili.

Sentiment analysis.
Si tratta di un'applicazione di 'data mining' applicata soprattutto ai social network. Un metodo di analisi che raccoglie in tempo reale le reazioni degli utenti e/o i trend di comportamento per un qualsiasi evento, locale o globale. Grazie alle tante informazioni prodotte oggi dal popolo dei social network (una delle molte fonti dei Big Data), la Sentiment analysis rappresenta uno strumento accurato per individuare ed 'ascoltare' le conversazioni online fornendo alle aziende un'interpretazione del mercato molto realistica.

Manutenzione predittiva.
L'obiettivo di un'organizzazione in generale è di far avere sempre la disponibilità operativa dei sistemi; ossia di non avere, se possibile, interruzioni nella disponibilità di un sistema durante il periodo nella quale è richiesta. La Manutenzione può essere fatta preventiva, statistica, secondo condizione, incidentale, correttiva, migliorativa, opportunistica, *eccetera*. La Manutenzione Predittiva è un tipo di manutenzione preventiva; che viene organizzata con l'individuazione di parametri che vengono misurati ed i cui valori estrapolati utilizzando appropriati modelli matematici/fisici/informatici; allo scopo di individuare in tempo il tempo residuo prima di un possibile guasto. Una variazione delle misure effettuate rispetto allo stato di normale funzionamento indicherà l' eventuale aumentare del degrado e permetterà di prevedere il momento del guasto.

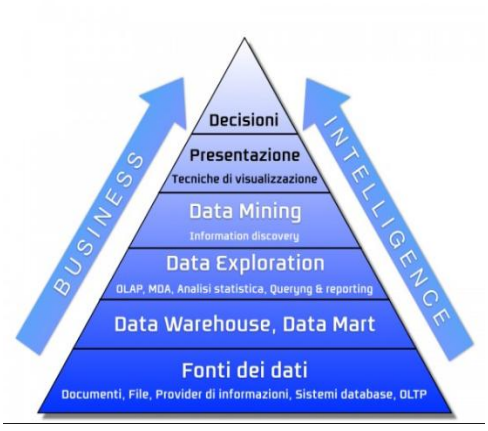
Apprendimento automatico. L'apprendimento automatico (noto anche come machine learning) rappresenta una delle aree fondamentali dell'intelligenza artificiale .E si occupa della realizzazione di sistemi, algoritmi, reti neurali , ecc… che si basano sulle osservazioni, trattandole come dati per la sintesi di nuova conoscenza. L'apprendimento può avvenire catturando caratteristiche di interesse provenienti da esempi concreti, da strutture di dati o da sensori, ecc…per analizzarle e valutarne le relazioni tra le variabili osservate.
L'apprendimento automatico è un campo multidisciplinare. Esso si basa sui risultati di intelligenza artificiale, probabilità e statistica, teoria della complessità computazionale, teoria di controllo, teoria dell'informazione; e anche altri campi.
L' Apprendimento può essere supervisionato, non supervisionato, con rinforzo, ecc…..
Ibm: "Il futuro prossimo è il Cognitive Business". Proliferazione dei dati e software economy stanno accelerando quella che si identifica come la nuova 'Cognitive Computing Era', dove Digital Business e Digital Intelligence confluiscono. La Digital Transformation sarà sempre più abilitata dalle tecnologie cognitive che supporteranno lo sviluppo di nuove applicazioni e servizi digitali aziendali. Il cardine tecnologico di questa importante evoluzione sarà sempre più l'integrazione. "Il cognitive business è qualcosa di completamente diverso dalla digitalizzazione". "I sistemi cognitivi hanno nella loro forza la comprensione e l'autoapprendimento degli eventi attraverso l'analisi dei dati non strutturati, la 'percezione' e l'interazione attraverso il linguaggio naturale dell'uomo; compiono ragionamenti generando ipotesi, considerazioni, argomentazioni e raccomandazioni; imparano dagli esperti (dall'uomo, quindi) e dalla continua 'acquisizione' e analisi di dati, ma con una velocità impensabile per una mente umana.

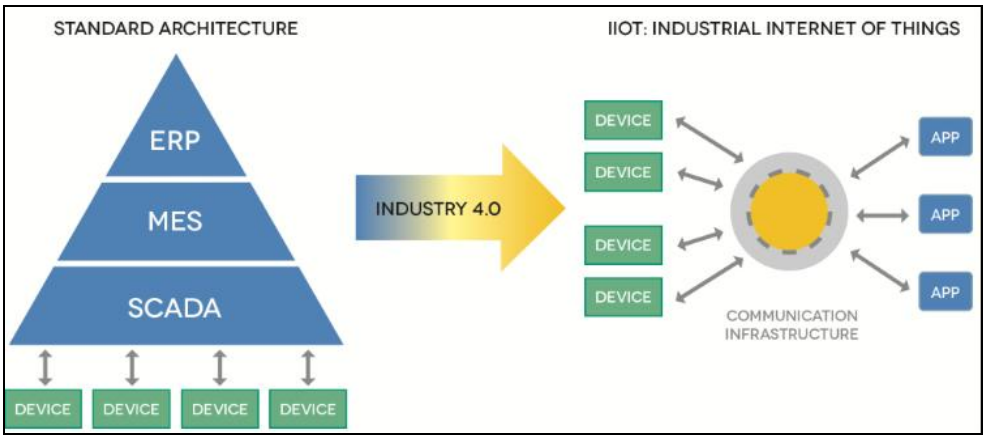
Franco Boccia. Bologna, ottobre 2018. www.b-it.it
(Riferimento fatto a nostri precedenti articoli già pubblicati; e ai testi di Rezzani e Vercellis).

---------- ooo0ooo ----------

Figure tratte da Google Immagini: esempi di Architetture.

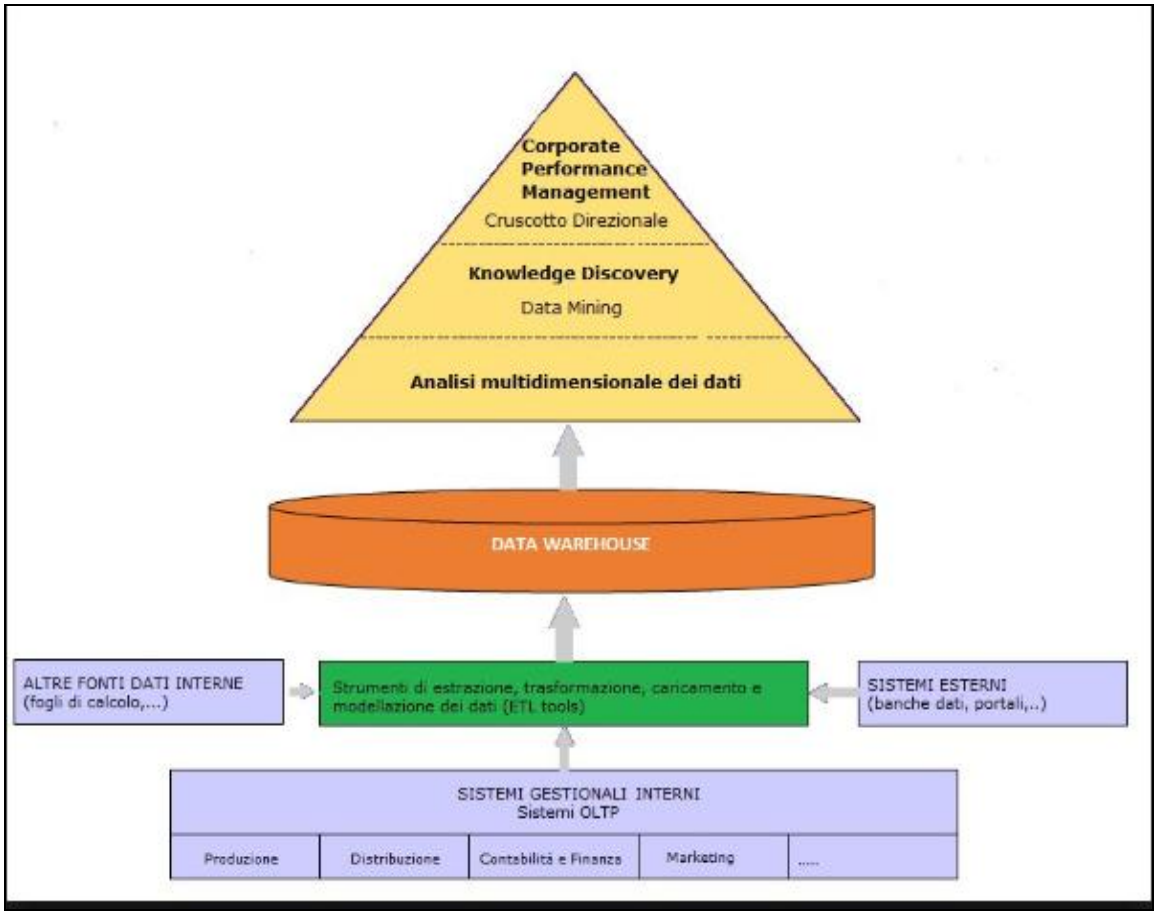
Piramidi nelle Organizzazioni.

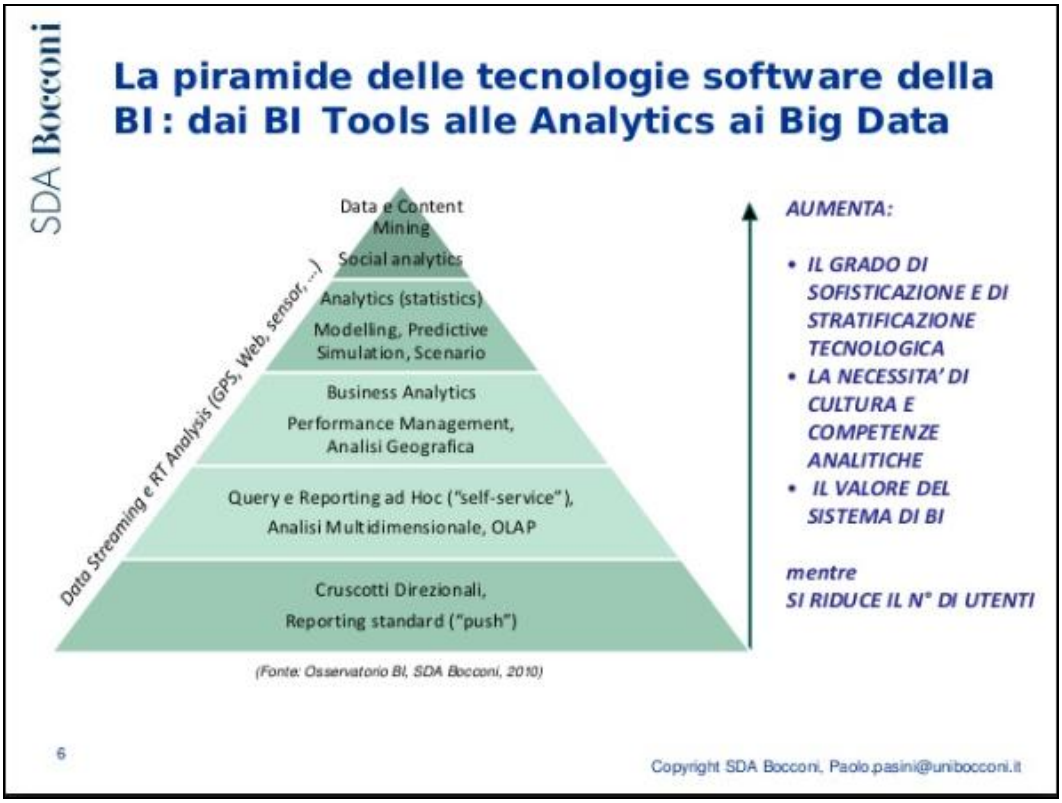




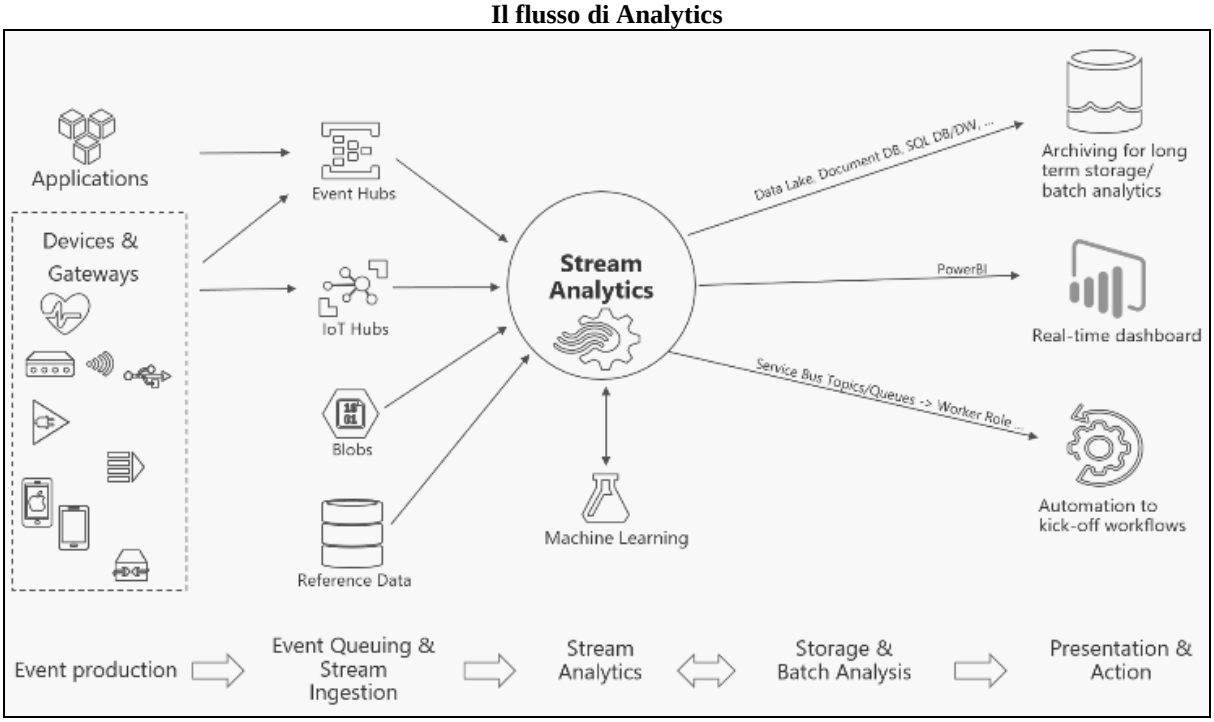
Le capacità occorrenti

Una macrostruttura sistemica

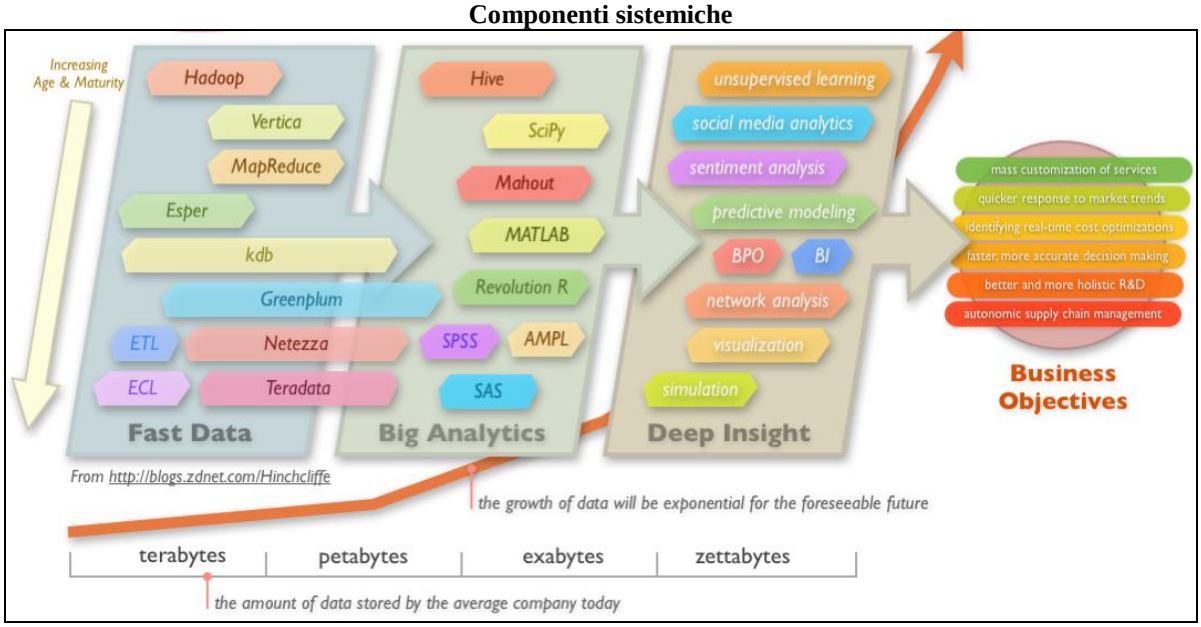




Il flusso di Analytics



Componenti sistemiche



---------- ooo0ooo ----------