

Statistical process control (SPC)

- Le carte di controllo sono uno strumento «tradizionale» per il raggiungere e mantenere obiettivi di qualità nell'industria

Idee di base:

Phase I: raccogliere campioni sufficientemente ampi dai processi

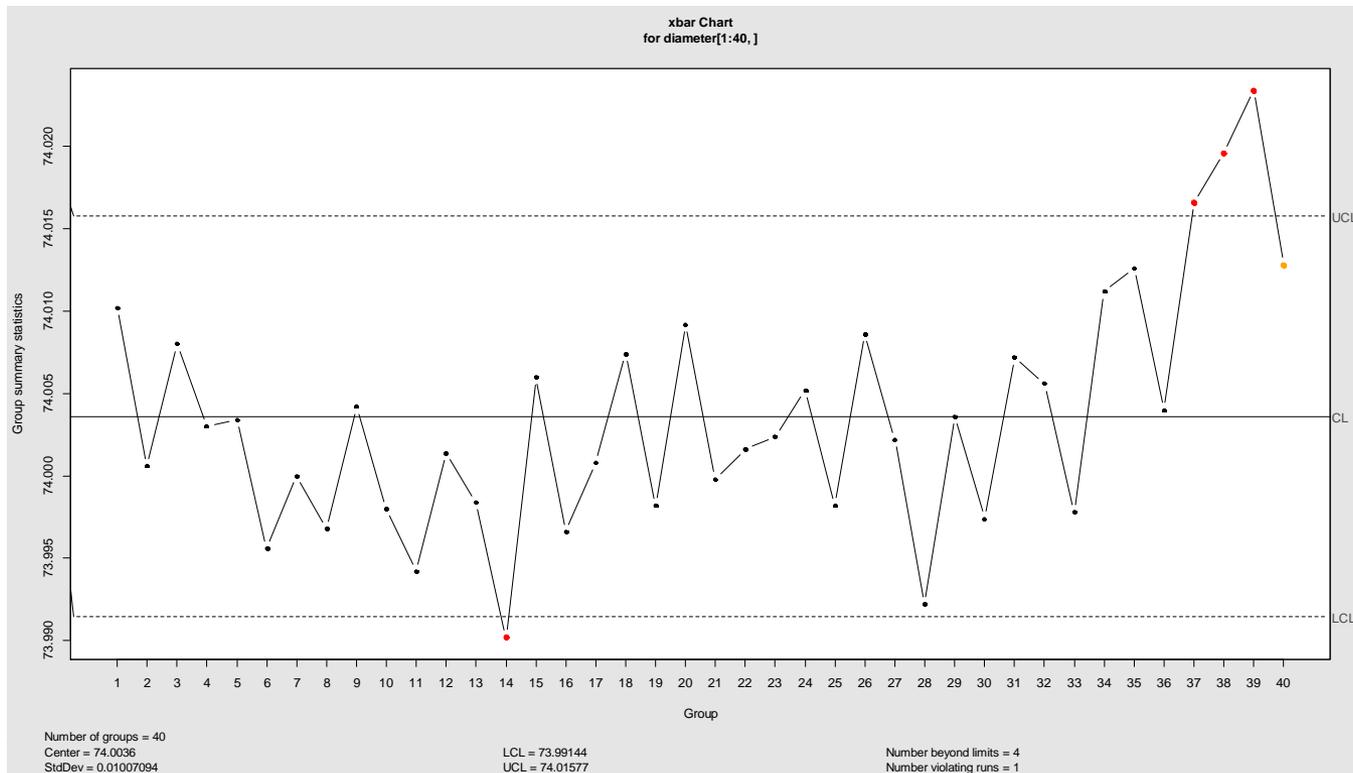
- per descrivere la loro variabilità naturale in condizioni operative normali (NOC)
- Per caratterizzare gli stati di cattivo funzionamento e decidere cosa deve essere monitorato

Phase II: controllo del processo attraverso l'analisi di campioni raccolti periodicamente dal processo.

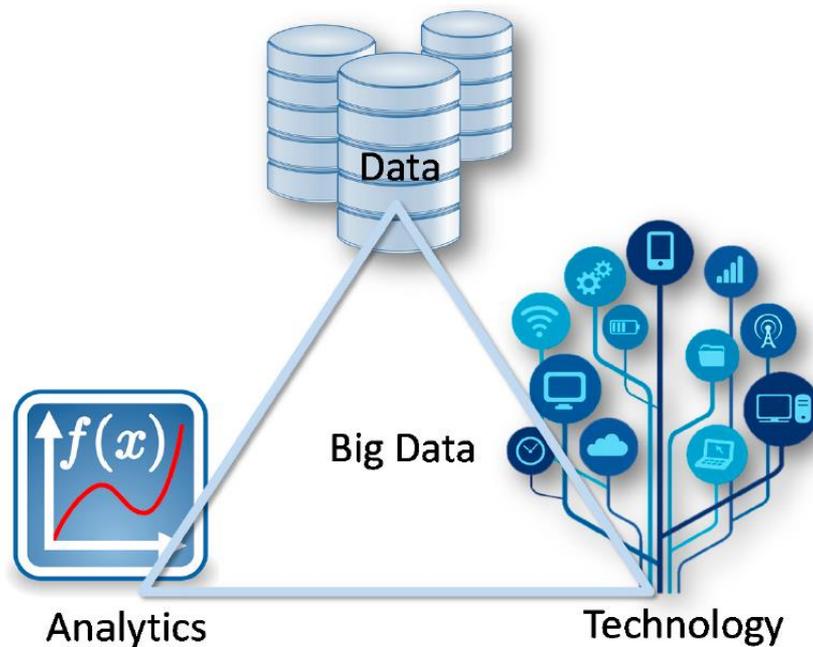
Nuove esigenze per SPC

Nel contesto «tradizionale»

- Campionamento periodico, $n_t \geq 2$ osservazioni per campione, analisi off-process, ruolo «attivo» dell'analista



Nuovo contesto



- I «big data» si presentano spesso nella forma di flussi di dati con osservazioni relativi a processi osservati sequenzialmente in tempi successivi

Nuove esigenze per SPC

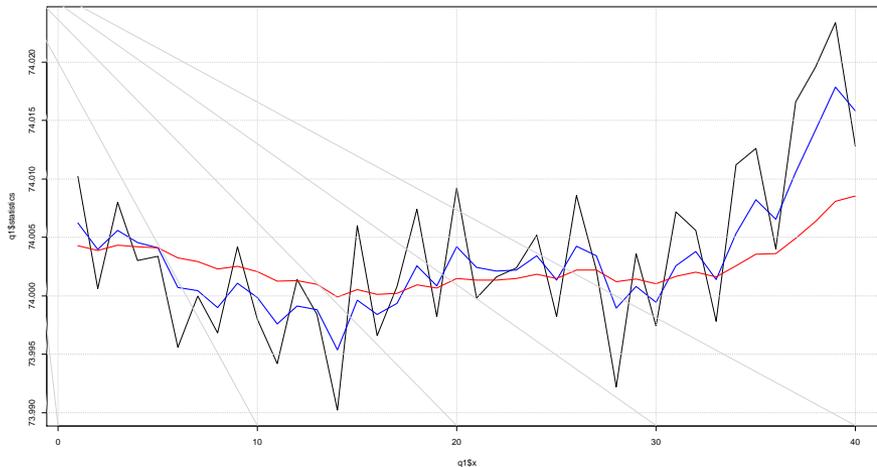
- Monitorare in parallelo un numero elevato di segnali, in modo il più possibile automatico;
- Aggiornamenti basati su osservazioni singole dai segnali;
- Elaborare i dati e fornire un feed-back in process e non post-process;
- Richiedere software facilmente reperibili, flessibili e di larga diffusione;
- Essere computazionalmente sostenibili, statisticamente «robuste»;

Alcune carte si adattano molto meglio di altre al contesto «nuovo»

EWMA (Exponentially weighted moving average) charts

- Sostituisco alla traiettoria del processo una traiettoria «smoothed», regolata da un parametro di memoria

$$Z_t = \lambda X_t + (1 - \lambda) Z_{t-1}$$



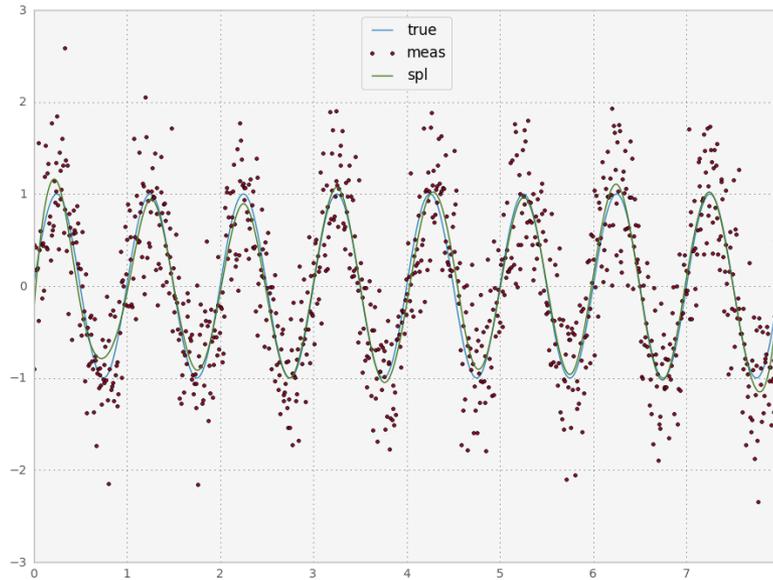
- Semplice da interpretare,
- Parametro λ regolabile (speed vs strength)
- Non richiede il ricorso a dati storici
- Implementazione elementare

Utilizzo:

- Dashborad di controllo remoto «facili» da leggere;
- early warning systems automatici;
- uso prognostico.

Dynamic system modelling

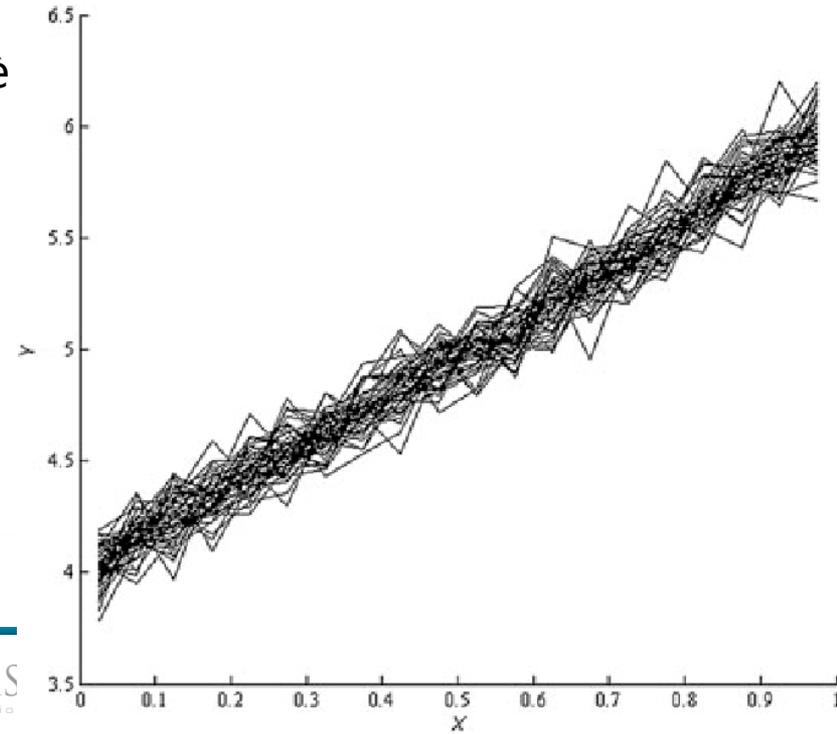
- Spesso i processi anche in controllo hanno una dinamica temporale complessa



- Phase I: modellazione statistica del processo in NOC +
- Phase II: monitoraggio automatizzato;

Profile monitoring

- Spesso tenere sotto controllo la correlazione tra segnali è più importante che tenere sotto controllo i segnali
 - Perché è l'obiettivo di qualità del processo ad essere definibile in termini di relazione «in equilibrio» tra segnali;
 - Perché singoli segnali fuori controllo possono essere corretti automaticamente ed è l'alterazione della struttura di relazione a rivelare l'anomalia.
 - Perché la mia modellazione dinamica è espressa in termini di relazione tra un segnale obiettivo e vari «predittori»



Linear profile monitoring

- Se le relazioni sono lineari possono essere descritte dall' indice di correlazione

$$r_{XY} = \frac{1}{T} \sum_{i=1}^T (X_i - \bar{X})(Y_i - \bar{Y})$$

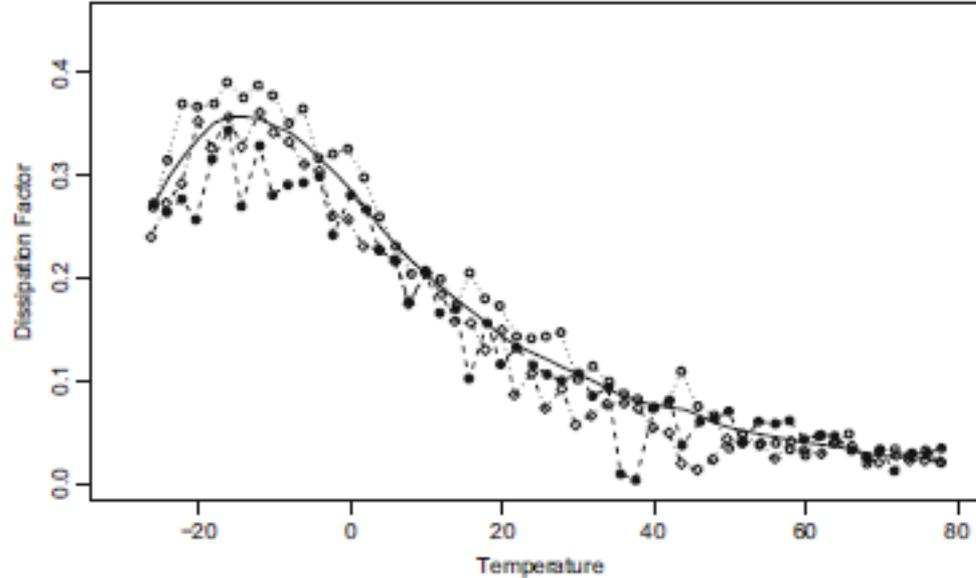
ed è possibile monitorare il processo «multivariato» con carte di tipo EWMA

- Esistono algoritmi per accomodare un buon numero di problemi statistici («big» data, outliers, altre «non normalità»)

Industria 4.0 e Big Data sono una grande opportunità per un controllo più efficace dei processi

Oltre il linear profile monitoring

- Non tutte le relazioni sono lineari



- Nei problemi dimensione elevate (molti segnali), il controllo di matrici di correlazione perde ogni potere diagnostico.

Oltre il profile monitoring

I passi oltre il monitoraggio sono rappresentati da «diagnosi» e «prognosi».

- La «**diagnosi**» richiede una rappresentazione coerente delle connessioni e delle direzionalità causali nei processi «multivariati»;
- La «**prognosi**» (= metodi predittivi) richiede informazioni aggiuntive sui faults, la loro classificazione;
- Metodi «knowledge based» e «data-driven»;

Tra automazione e analisi (conclusioni)

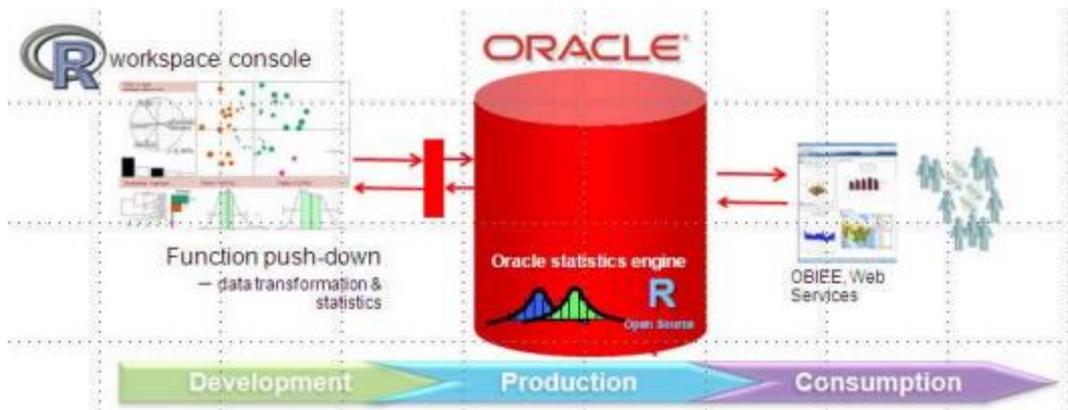
- Fino ad un certo livello di complessità è realistico puntare a controllo, diagnosi e prognosi dei processi multivariati

Per processi più complessi:

- Gli algoritmi di auto-apprendimento dai dati sono strumenti potenti, ma è bene essere scettici riguardo ad algoritmi «infallibili» o «universali» (rischio «caos»)
- Sfruttare la conoscenza fisica, causale, dei processi (e in generale l'informazione «non numerica») rende molto più efficiente gli algoritmi;
- Integrazione di metodi data-driven e modellazione rimane essenziale.

Softwares

- Un'analisi statistica «sostanziosa» richiede software specializzato (librerie);
- I software statistici tradizionalmente non attrezzati per gestione di basi di dati oltre il livello elementare;
- Negli ultimi anni molte proposte di integrazione di software statistico nelle piattaforme di gestione dati



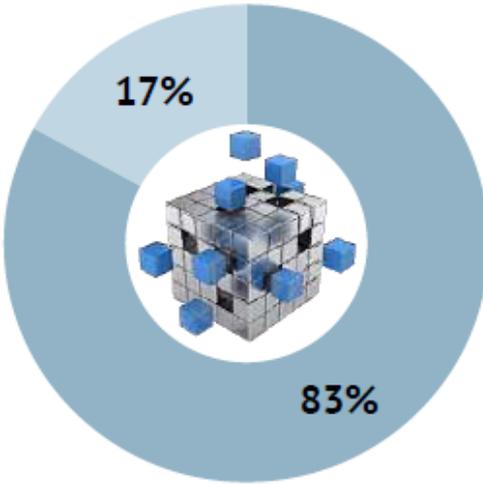
Il contesto: l'esplosione dei dati



- **Cause principali:** digitalizzazione, nuove tecnologie, interconnessione, ...
- L'azienda come nuova «**fabbrica di dati**»
- L'importanza strategica dei dati e delle informazioni
- Le **fonti di dati** continueranno a crescere ed evolversi
- **Nuovi paradigmi:**
 - Necessità di catturare, gestire e archiviare tutti i dati aziendali per preservare storia e contesto
 - Analizzare scientificamente i dati per 'arricchirli' di senso utile
 - Condividere i dati velocemente e liberamente a tutti coloro che hanno necessità

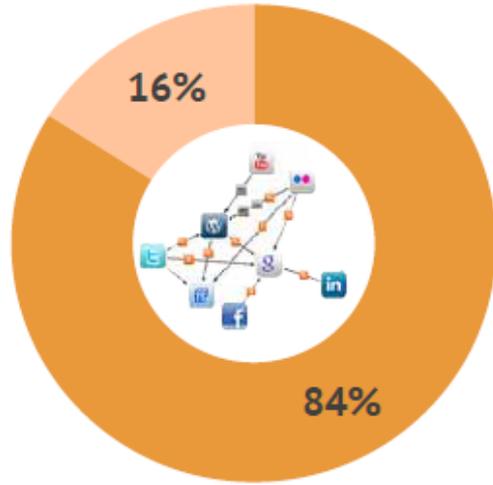
I dati nelle aziende: tipologia, origine e trend di crescita

Tipologia dei dati



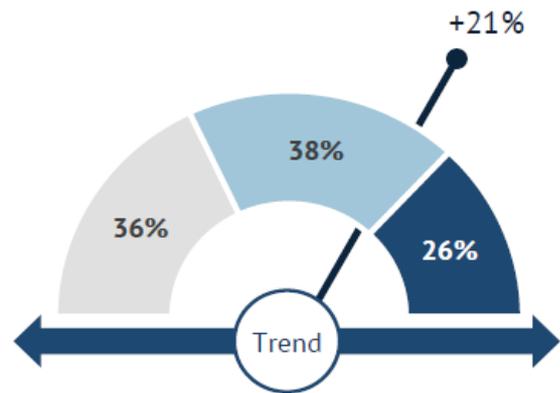
- Dati strutturati
- Dati destrutturati

Origine dei dati

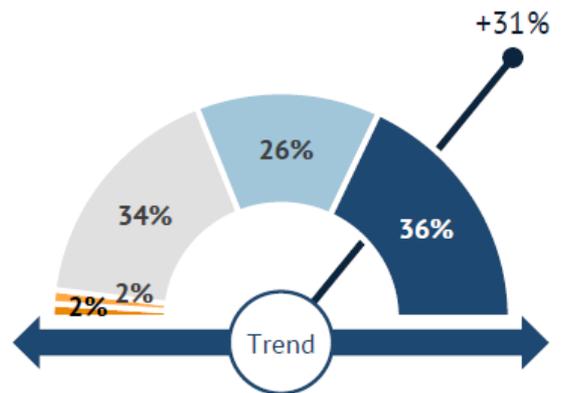


- Dati interni
- Dati esterni

Dati strutturati



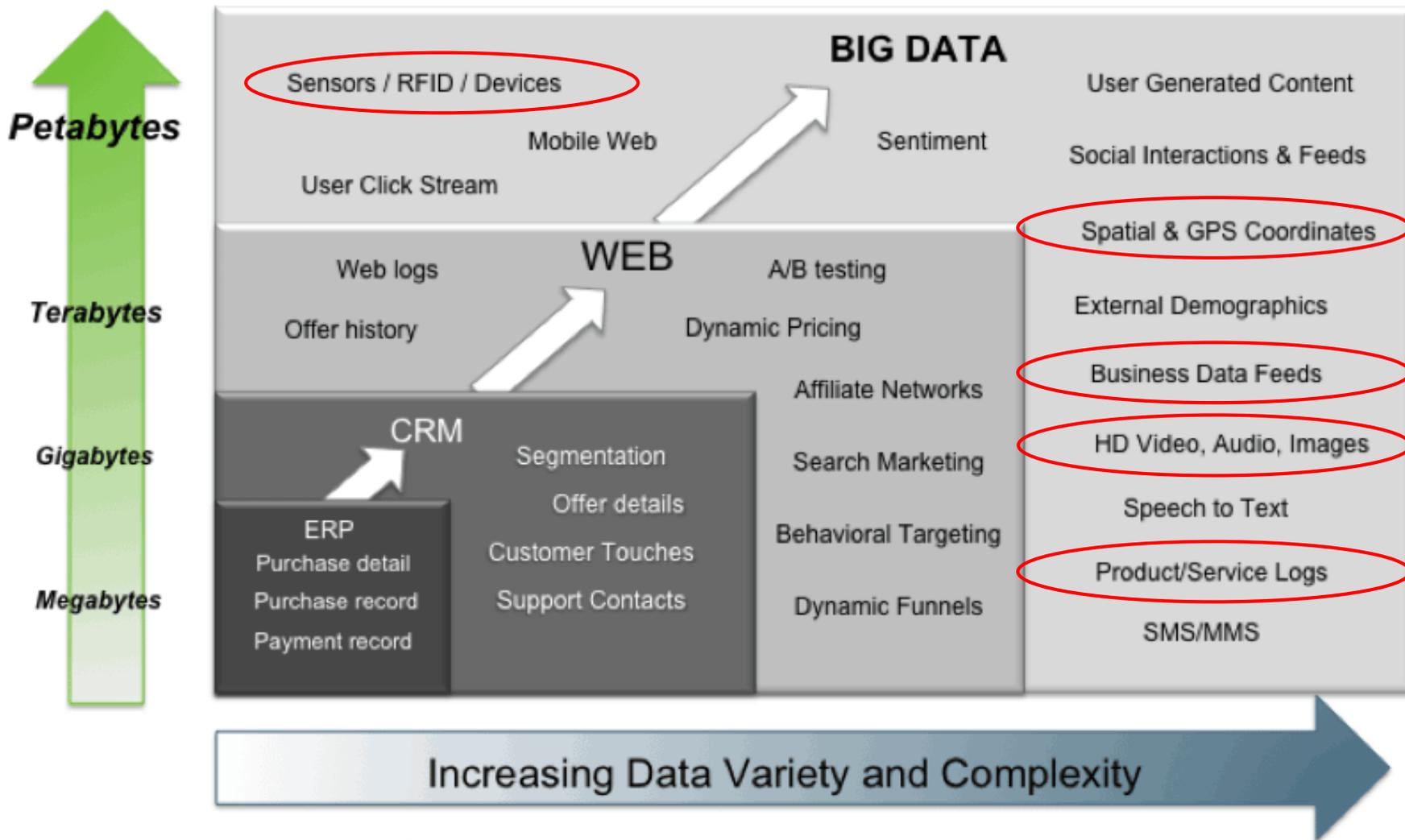
Dati destrutturati



- Crescita sostanziale (>25%)
- Crescita moderata (>25%)
- Stabilità (+/-3%)
- Diminuzione moderata (<25%)
- Diminuzione sostanziale (>25%)

Fonte Osservatori Digital Innovation – Rapporto I dati al centro dei processi aziendali 2014 – © DIG - Politecnico di Milano (www.osservatori.net)

Big Data = Transazioni + Interazioni + Osservazioni



Source: Contents of above graphic created in partnership with Teradata, Inc.

I Big Data si possono definire come l'insieme di:

- **Dati** che hanno caratteristiche tali da superare i limiti dei DataBase tradizionali
- **Tecnologie** nuove volte a estrarre conoscenza e valore da questa tipologia di dati

Dati che hanno caratteristiche tali da superare i limiti dei DataBase tradizionali

- **La crisi del modello Relazionale**

- **DBMS** relazionali classici (IBM DB2, SQL Server, Oracle, MySQL, ...): rigidi nella struttura, nati per i dati transazionali, non sono adatti a gestire tipologie di dati eterogenei e non strutturati

- Avvento di **nuovi modelli** per la gestione dei Big Data

- **NoSQL** DB: famiglia di sistemi («one size does not fit all») capaci di lavorare su dati distribuiti, schema non fisso, facile scalabilità
- **NewSQL** DB: modello relazionale con performance di scalabilità comparabili ai NoSQL

Dati che hanno caratteristiche tali da superare i limiti dei DataBase tradizionali

- Le **nuove soluzioni tecnologiche** consentono:
 - di garantire **buone prestazioni su volumi di dati molto superiori** a quelli tradizionali
 - la **scalabilità** delle applicazioni alla crescita delle dimensioni dei dati
 - di bilanciare **tempi di risposta, numero di operazioni** svolte e **tolleranza ai guasti** hw e sw
- Alcune **architetture di riferimento**:
 - **dati distribuiti** e replicati su Cluster di Computer (es: Hadoop)
 - **elaborazione distribuita** (es: MapReduce, Apache Spark)
 - **file system distribuiti** (es: HDFS) per dati memorizzati direttamente nei formati nativi
 - **cloud computing** per virtualizzare risorse hardware e software

Tecnologie nuove per l'analisi dei Big Data: gli Analytics

- Gli Analytics sono **nuovi strumenti** di trattamento e analisi sui Big Data superando i limiti dei tradizionali strumenti analitici e di Data Warehousing
- Sistemi realizzati allo scopo di effettuare **analisi (anche tradizionali)** che incontrano limiti tecnici a causa della natura dei nuovi dati
- L'ambito degli Analytics è quello di **estrarre valore dal dato** per mezzo di tecniche statistiche avanzate che portano ad un utilizzo più proficuo dei **numerosi dataset disponibili**
- I Data Analytics sono modellati su **dati né statici né prevedibili**, con un livello di agilità che ne permetta l'utilizzo in tempo reale
- Esempio: una estensione di analisi aggregate a livelli di **granularità più fine** o l'estensione della finestra temporale utilizzabile, **raggiungendo profondità storiche precluse ai sistemi tradizionali**
- Esempio: un sistema di Analytics basata su tecnologie «Big Data» consente una storicizzazione adeguata dei dati e **tempi di recupero delle informazioni drasticamente ridotti** rispetto a sistemi «legacy»

I modelli e le soluzioni disponibili

Performance management & Basic Analytics

- **Descriptive Analytics (Analisi Descrittiva)**

- insieme di strumenti orientati a descrivere la situazione attuale e passata dei processi aziendali e/o aree funzionali. Tali strumenti permettono la visualizzazione grafica dei livelli di performance

Advanced Analytics

- **Predictive Analytics (Analisi Predittiva)**

- strumenti avanzati che effettuano l'analisi dei dati per rispondere a domande relative a cosa potrebbe accadere nel futuro (regressione, forecasting, modelli predittivi, e altri)

- **Prescriptive Analytics (Analisi Prescrittiva)**

- applicazioni Big Data avanzate che, insieme all'analisi dei dati hanno la capacità di assumere e gestire processi decisionali

- **Automated Analytics (Analisi Automatica)**

- capaci di implementare autonomamente l'azione proposta secondo il risultato delle analisi svolte

Le funzionalità dei Big Data Analytics

Performance management & Basic Analytics

- **Strumenti di query e reporting**
 - risposta a domande specifiche, accesso ai DataBase in modo flessibile e tempestivo
- **Cubi multidimensionali e analisi OLAP**
 - navigazione dei dati secondo viste logiche dinamiche
- **Dashboard, Scorecard, KPI**
 - cruscotti, visione grafica di una funzione o processo, controllo prestazioni, indici di performance
- **Meccanismi di Alerting**
 - visualizzazione di segnali di allarme per il superamento di valori di soglia assegnati su indicatori specifici

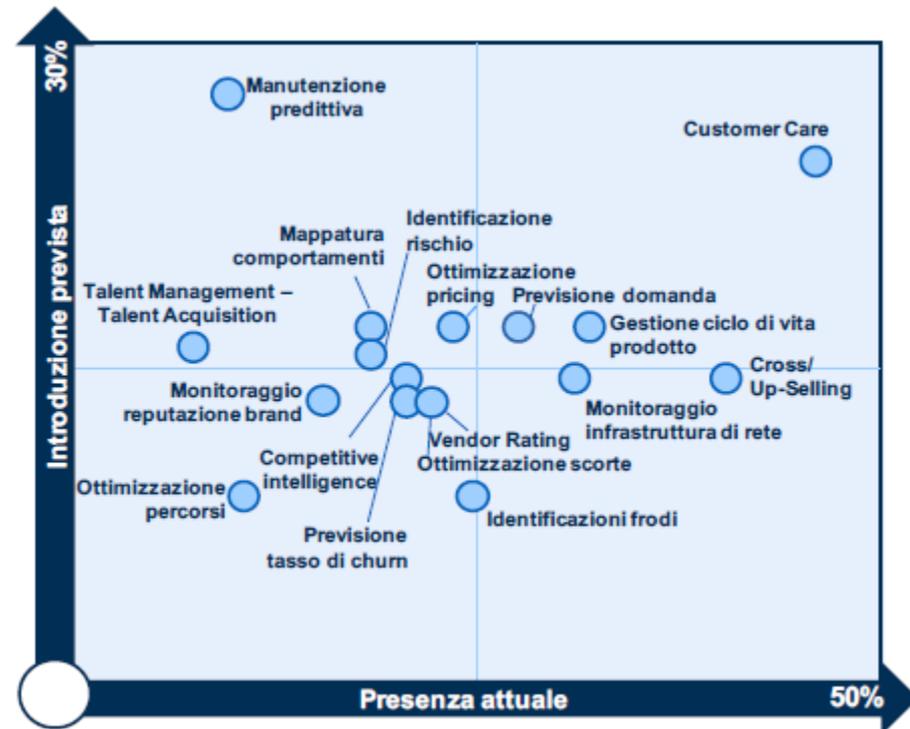
Le funzionalità dei Big Data Analytics

Advanced Analytics

- **Metodi estrapolativi di forecasting**
 - cogliere tendenza e stagionalità di una serie storica di valori
- **Metodi predittivi di Data Mining**
 - identificare le relazioni tra variabili, con diverse tecniche (es: classificazione, regressione, clustering, link analysis, web mining, ...)
- **Modelli di ottimizzazione**
 - identificare la decisione migliore tra un insieme di azioni possibili

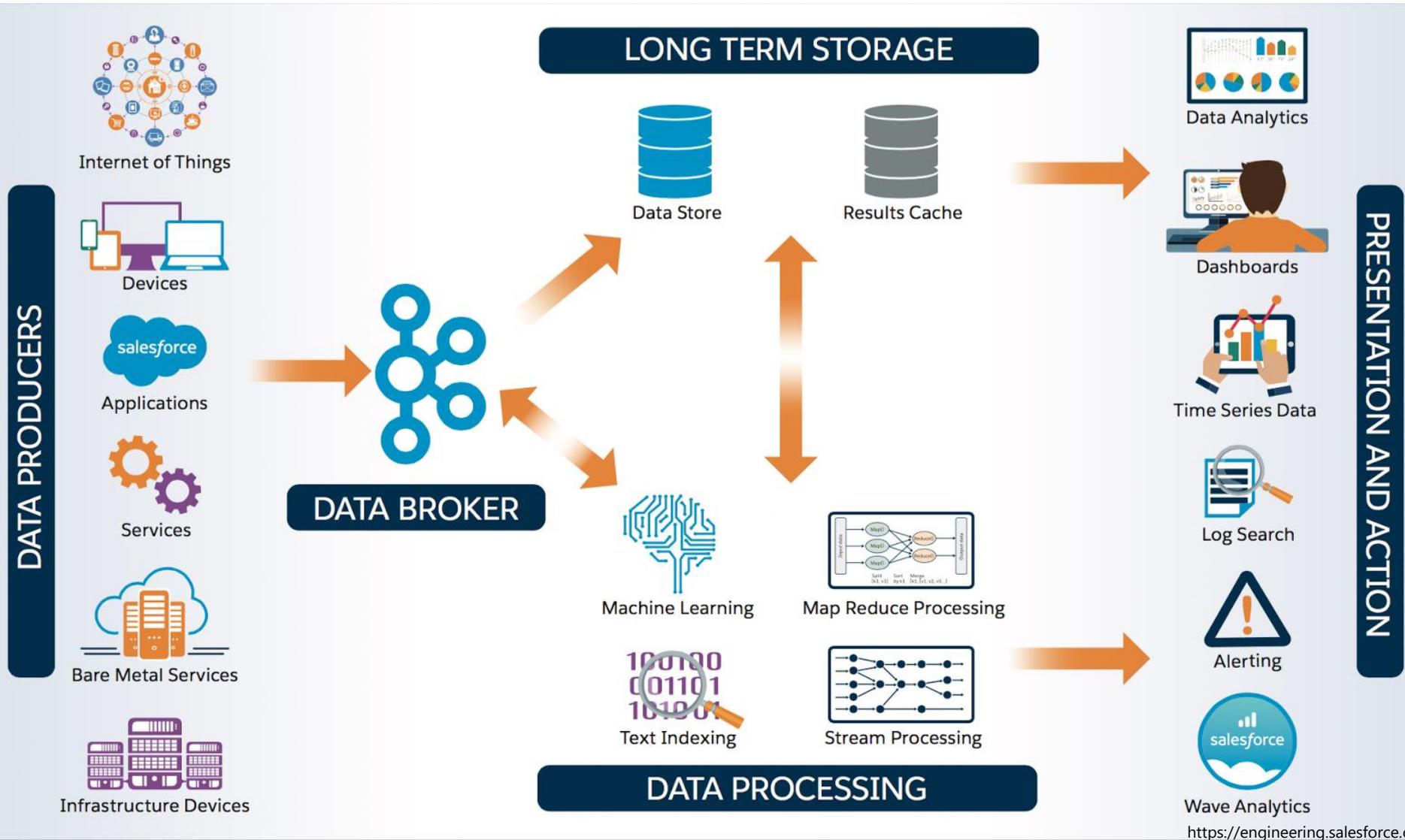
Diffusione delle principali iniziative di Analytics

- Le **PMI** rappresentano ancora un **settore marginale** (13% del mercato Analytics)
- Campi di utilizzo della Big Data Analysis:**
 - Efficienza e rischi Operativi**
 - analisi finanziarie, asset management, gestione personale, supply chain, manutenzione preventiva, ...
 - Sicurezza e performance applicative**
 - funzionamento e servizi legati al mondo IT, ...
 - Conoscenza e servizio ai clienti**
 - progetti di marketing e vendite, sviluppo prodotti, ottimizzazione della digital experience, ...



Fonte Osservatori Digital Innovation – Rapporto Big Data: lo scenario di mercato 2016 – © DIG - Politecnico di Milano (www.osservatori.net)

Big Data life cycle: una possibile architettura di riferimento



Conclusioni

- **L'onda dei Big Data per le PMI non è ancora arrivata**, nessun settore ne sarà immune
- Forse già oggi **posso sfruttare al meglio i dati che ho già a disposizione** dai sistemi informatici di **IT** (Information Teconology) o di **OP** (Operation Technology)
- **Non sarà solo un cambiamento tecnologico** ma culturale, organizzativo, strategico, ...
- Focalizzare con chiarezza **cosa si pensa di ottenere dai Big Data**
- Nell'attuale clima Nazionale ed Internazionale la possibilità di **prendere decisioni** basandoci su informazioni e analisi nuove ed aggiornate può portare **grandi benefici** e vantaggi competitivi alle imprese