

BIG DATA: TECHNOLOGIES



SAPIENZA
UNIVERSITÀ DI ROMA



Master in Data Intelligence e Strategie Decisionali
Dipartimento di Scienze Statistiche

Paolo Dell' Olmo

Outline

- 1. Trends**
- 2. Piattaforme**
- 3. Parallel and Distributed Computing**
- 4. Hadoop and MapReduce**
- 5. Esempi and Takeaways**

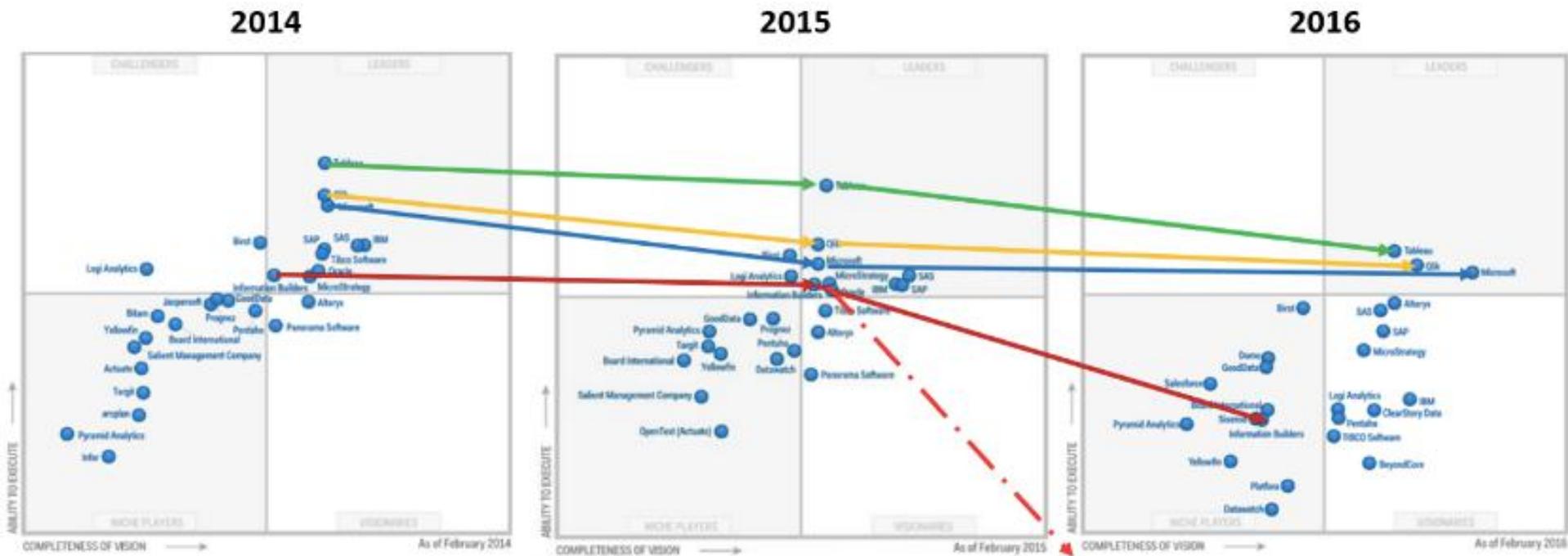
Trends

Gartner Magic Quadrant - 2014

2014 Gartner Magic Quadrant for BI & Analytics Platforms



Magic Quadrant Evolution



- Tableau
- Qlik
- Microsoft
- Information Builders
- - - → Oracle

Gartner Magic Quadrant - 2017

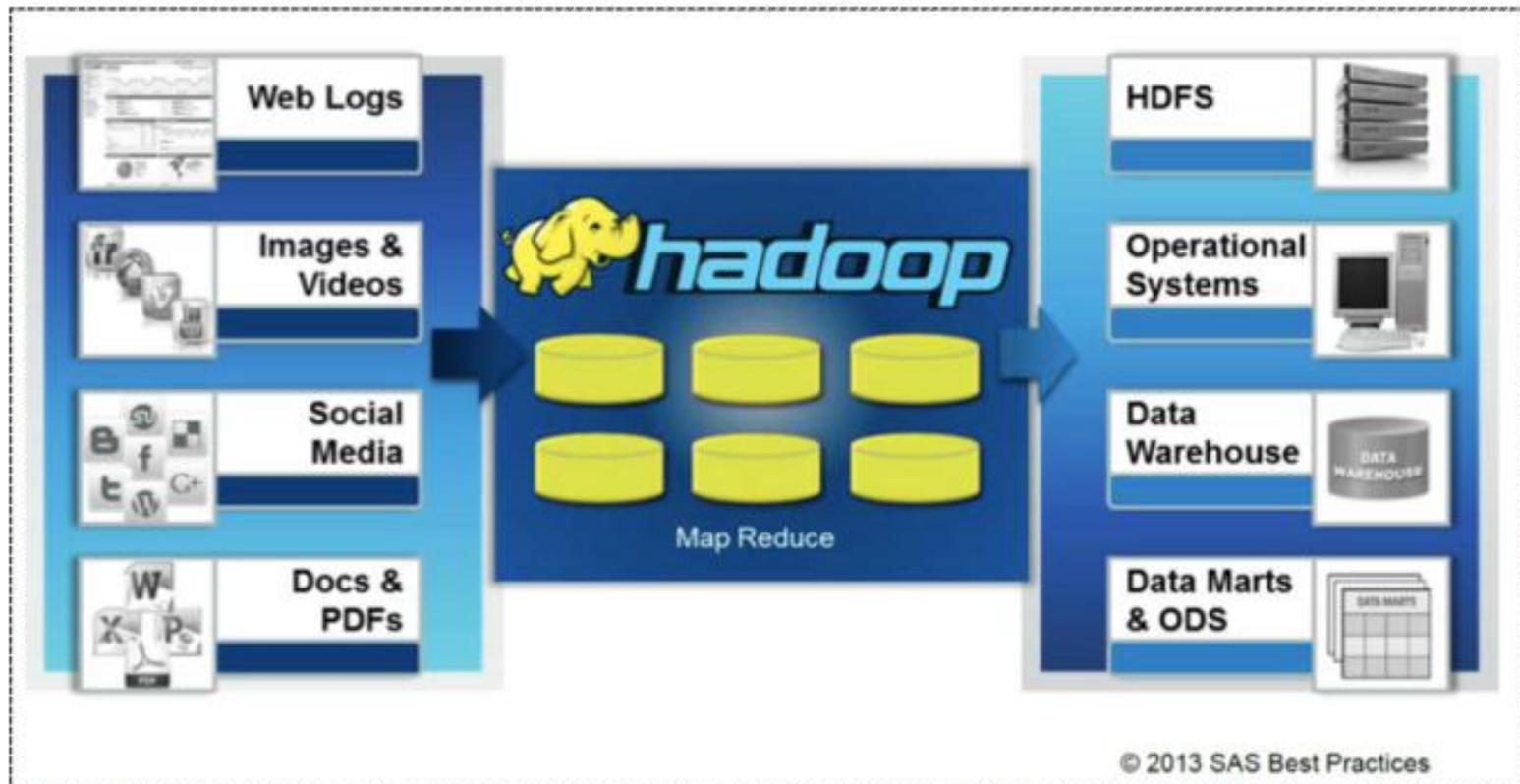


Gartner Magic Quadrant - 2017



Piattaforme

Piattaforme

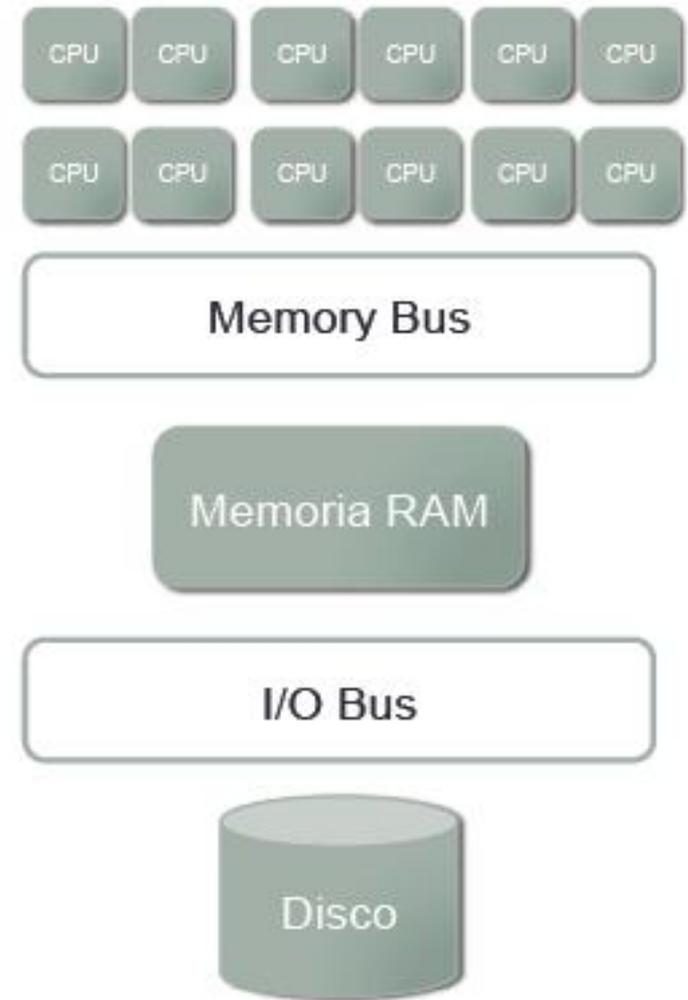
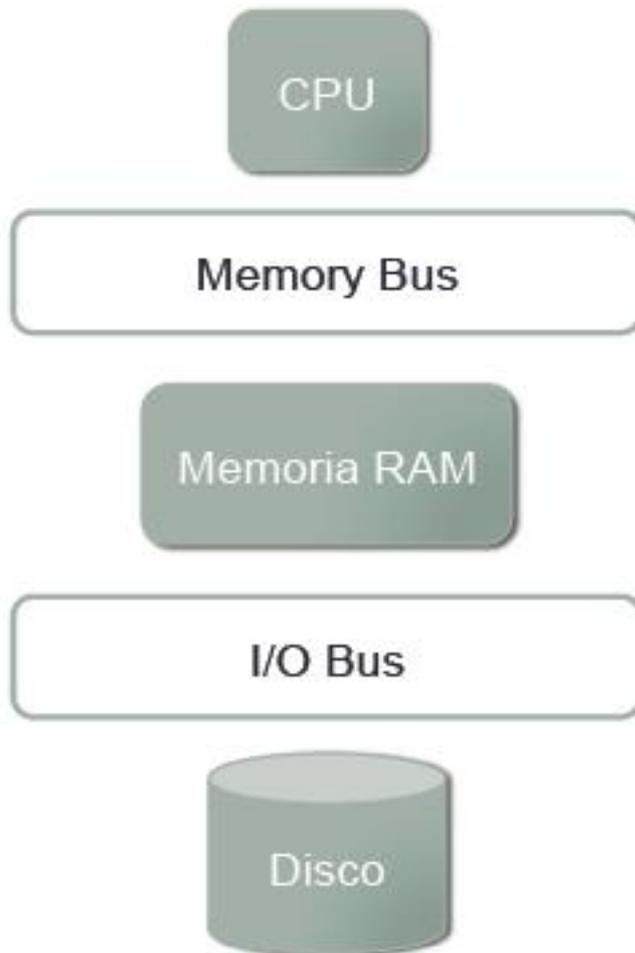


Architetture per le imprese



Calcolo Distribuito

Calcolo Parallelo



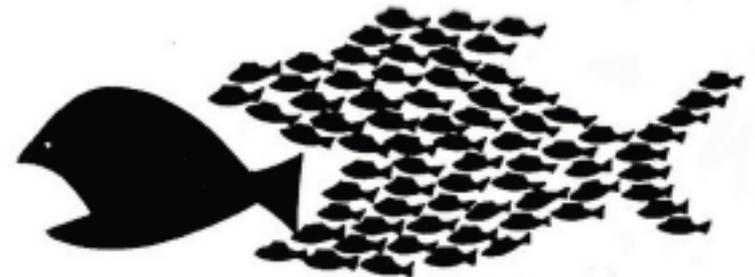
Scale-up

- Il numero di istruzioni necessarie per la risoluzione di un problema è proporzionale alla complessità (computazionale) dello stesso ed alla quantità di dati di input
- Fermo restando il numero di istruzioni necessarie per risolvere un certo problema, è possibile ridurre il tempo di esecuzione adoperando processori più potenti (**scale-up** o **scalabilità verticale**)



Scale-out

- Piuttosto che investire sul potenziamento di un singolo sistema, si allestisce un supercalcolatore virtuale aggregando tanti calcolatori distinti (**scale-out o scalabilità orizzontale**)
- Vantaggi:
 - Efficienza linearmente proporzionale al costo degli investimenti
 - Flessibilità nel dimensionamento del sistema (incrementando o decrementando le risorse del sistema a seconda delle necessità)
 - Flessibilità nell'aggiornamento del sistema
 - Possibilità di gestire i guasti



Calcolo Parallelo e Calcolo Distribuito

- **Calcolo parallelo:** si intende generalmente la presenza di un unico calcolatore dotato di due o più CPU che cooperano per risolvere uno stesso problema condividendo o meno alcune risorse (disco, memoria)
- **Calcolo distribuito:** si intende generalmente la presenza di più calcolatori dotati di proprie CPU autonome. I calcolatori cooperano per risolvere uno stesso problema comunicando mediante una connessione di rete condivisa

Scalabilità

- **Misura la capacità di un sistema distribuito di rispondere efficientemente alle richieste esterne, indipendentemente dalla taglia del problema affrontato**
- Di norma, il tempo di risposta di un sistema è proporzionale alla taglia dei dati da elaborare ed al numero di richieste concorrenti ricevute
- E' possibile far fronte ad un maggior numero di richieste nello stesso tempo, allocando alla loro evasione un maggior numero di risorse
- Tuttavia, la scalabilità verticale di un sistema è intrinsecamente limitata da fattori fisici
- I sistemi distribuiti consentono una **scalabilità orizzontale** virtualmente illimitata

Scalabilità

- La scalabilità orizzontale di un sistema distribuito si ottiene:
 - ripartendo i dati del servizio tra i diversi nodi del sistema
 - dirottando le richieste del servizio ai diversi nodi del sistema
 - Adottando un protocollo di sincronizzazione che garantisca la consistenza dei dati e dello stato del servizio tra i diversi nodi del sistema
- Esempio: la rete di distribuzione di contenuti OVH

Disponibilità

- **Misura la capacità di un sistema distribuito di erogare più o meno ininterrottamente un servizio, e di farlo in un tempo accettabile**
- I sistemi di calcolo sono naturalmente soggetti a guasti o malfunzionamenti (i.e., faults), di natura hardware o software
- Questi possono compromettere la disponibilità di un servizio



Temporary Error (500)

We're sorry, but your Gmail account is temporarily unavailable. We apologize for the trying again in a few minutes. You can view the [Apps Status Dashboard](#) for the curre

If the issue persists, please visit the [Gmail Help Center](#).

[Try Again](#) [Sign Out](#)

[Show Detailed Technical Info](#)

Disponibilità

- La probabilità di tali eventi cresce con il numero di calcolatori coinvolti e con la durata di una elaborazione
- *“In each cluster’s first year, it’s typical that 1,000 individual machine failures will occur; thousands of hard drive failures will occur; one power distribution unit will fail, bringing down 500 to 1,000 machines for about 6 hours; 20 racks will fail, each time causing 40 to 80 machines to vanish from the network; 5 racks will “go wonky,” with half their network packets missing in action; and the cluster will have to be rewired once, affecting 5 percent of the machines at any given moment over a 2-day span, Dean said. And there’s about a 50 percent chance that the cluster will overheat, taking down most of the servers in less than 5 minutes and taking 1 to 2 days to recover.” – Jeff Dean (Google)*

Disponibilità

- L'approccio distribuito fornisce una soluzione naturale al problema tramite il concetto di *ridondanza*
 - *Ridondanza dei dati*
 - *Ridondanza dei servizi*

NoSQL

Criticità dei RDBMS: Gestione di grandi volumi di dati

- L'approccio relazionale rende complicato ed inefficiente il partizionamento di una base dati su un sistema distribuito
- Esempi:
 - L'interrogazione di una base dati relazionale normalizzata può richiedere il coinvolgimento di numerose tabelle. Per esigenze di spazio, queste potrebbero non trovarsi tutte sullo stesso server fisico. Per lo stesso motivo, non tutti i record di una stessa tabella potrebbero trovarsi sullo stesso server fisico, complicando quindi il piano di esecuzione di una eventuale query

Criticità dei RDBMS: Gestione di dati non strutturati

- I RDBMS impongono che i dati memorizzati obbediscano ad una struttura rigida e vincolante
- Nella realtà, i dati si presentano spesso in forma destrutturata o semi-strutturata, de-normalizzata e/o aggregata



Il movimento NO-SQL

- Termine genericamente usato per indicare DBMS non basati sull'approccio relazionale e/o non operanti con il linguaggio SQL
- Sono "sempre" esistiti
- Hanno visto la loro popolarità esplodere negli ultimi anni
- **Motivo:**
 - Le aziende protagoniste del Web 2.0 (Google, Yahoo, Amazon) necessitavano di tecnologie capaci di:
 - Gestire efficientemente grandi basi dati
 - Servire un numero di richieste potenzialmente molto alto e soggetto a picchi di carico
 - Erogare con continuità un servizio
 - Assecondare ed agevolare la rapida evoluzione dei servizi richiesta dal mercato e dalla competizione
 - I tradizionali RDBMS si dimostravano inadatti a tale scopo

Fattori per il successo dei DB NO-SQL

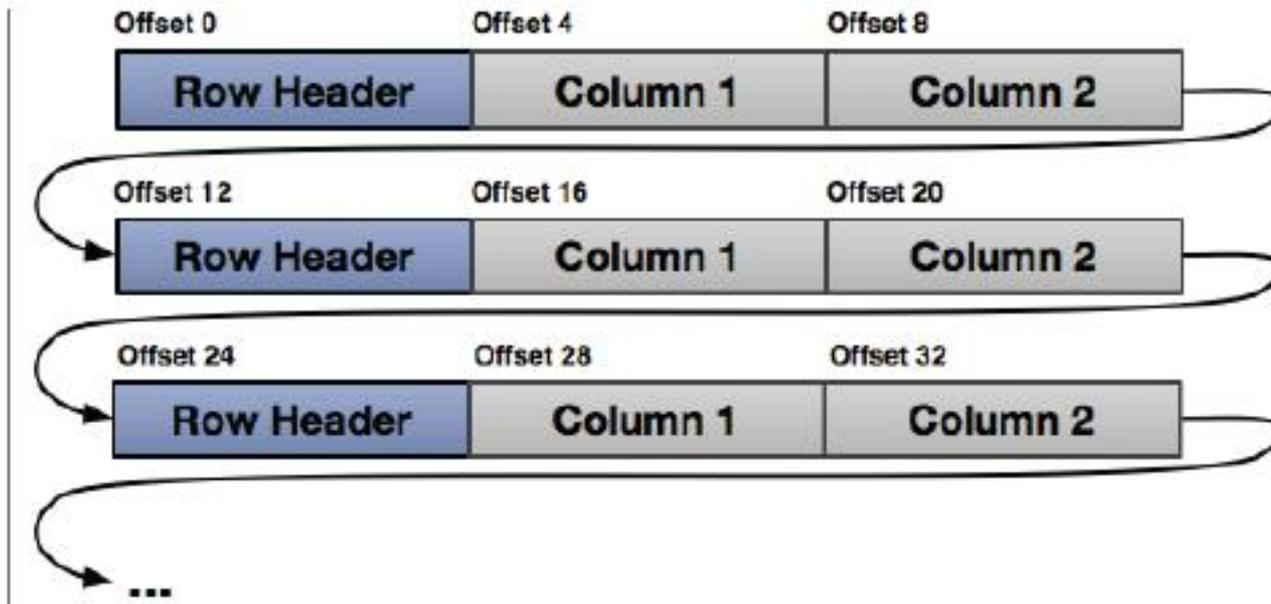
- **Flessibilità:** I DBMS NoSQL sono tipicamente Schema-Free. Questo implica la possibilità di modificare la struttura dei dati senza dover aggiornare le istanze già presenti (**schema on read vs. schema on write**)
- **Scalabilità:** Il modello dati adottato dai DBMS NoSQL semplifica il partizionamento di una base dati su più calcolatori di un sistema distribuito. La scalabilità viene spesso realizzata automaticamente e trasparentemente dal DBMS.
- **Tolleranza ai guasti/Disponibilità:** L'approccio distribuito agevola l'implementazione di soluzioni che consentono di mantenere in vita un servizio anche laddove una delle sue parti dovesse guastarsi

DB NO-SQL Caratteristiche

- Il termine NoSQL è generico ed include al suo interno tipologie di DBMS molto diverse tra loro
- Le principali sono:
 - Column based database
 - Key-Value database
 - Document based database
 - Graph database

Row Based

- **I RDBMS adottano un approccio row-based di memorizzazione dei dati**
 - Per conoscere il valore assunto da un certo attributo su un blocco di record, carichiamo in memoria l'intero blocco ed estrapoliamo il valore del solo attributo richiesto



Column Based

- **Scenario applicativo:**

- Analizzare il valore assunto da un certo dataset (potenzialmente di grandi dimensioni) relativamente ad un solo attributo
- Esempio: Conoscere l'importo di tutte le transazioni fatte su un sito di e-commerce

Sales			
Product	Customer	Date	Sale
Beer	Thomas	2011-11-25	2 GBP
Beer	Thomas	2011-11-25	2 GBP
Vodka	Thomas	2011-11-25	10 GBP
Whiskey	Christian	2011-11-25	5 GBP
Whiskey	Christian	2011-11-25	5 GBP
Vodka	Alexei	2011-11-25	10 GBP
Vodka	Alexei	2011-11-25	10 GBP

Column

- Dati memorizzati per colonne piuttosto che per righe
 - Ciascun attributo di una relazione viene memorizzato in una tabella dedicata e corredato di un id
 - Attributi diversi di una stessa occorrenza sono memorizzati usando lo stesso id

Product	
ID	Value
1	Beer
2	Beer
3	Vodka
4	Whiskey
5	Whiskey
6	Vodka
7	Vodka

Customer	
ID	Customer
1	Thomas
2	Thomas
3	Thomas
4	Christian
5	Christian
6	Alexei
7	Alexei

Date	
ID	Date
1	2011-11-25
2	2011-11-25
3	2011-11-25
4	2011-11-25
5	2011-11-25
6	2011-11-25
7	2011-11-25

Sale	
ID	Sale
1	2 GBP
2	2 GBP
3	10 GBP
4	5 GBP
5	5 GBP
6	10 GBP
7	10 GBP

Column

- **Aggregazione di tutti i record che presentano uno stesso valore in corrispondenza di uno stesso attributo**
 - Spesso, occorrenze diverse di uno stesso attributo presentano gli stessi valori
 - E' possibile comprimere significativamente la taglia delle tabelle-attributo raggruppando i valori identici ed indicando il range di id per cui si applicano

Product'	
ID	Value
1-2	Beer
3	Vodka
4-5	Whiskey
6-7	Vodka

Customer'	
ID	Customer
1-3	Thomas
4-5	Christian
6-7	Alexei

Date'	
ID	Date
1-7	2011-11-25

Sale'	
ID	Sale
1-2	2 GBP
3	10 GBP
4-5	5 GBP
6-7	10 GBP

Hadoop

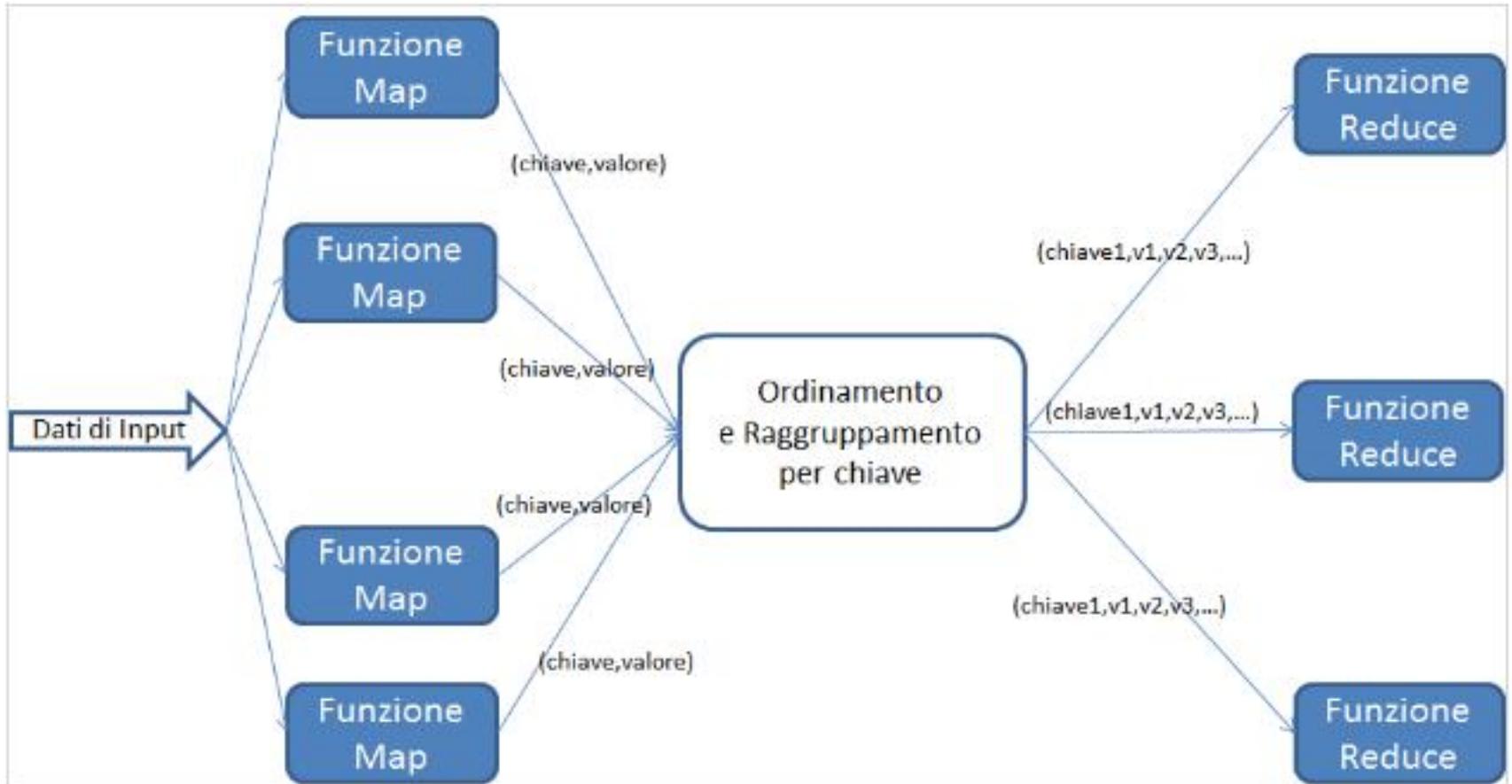
RDBMS - Hadoop

RDBMS	Hadoop
Schema <i>on Write</i> : lo schema dei dati deve essere creato prima che i dati stessi vengano caricati	Schema <i>on Read</i> : i dati sono semplicemente copiati nel file system, nessuna trasformazione è richiesta
Ogni dato da caricare deve essere trasformato nella struttura interna del database	I dati delle colonne sono estratte durante la fase di lettura
Nuove colonne devono essere aggiunte esplicitamente prima che i nuovi dati per tali colonne siano caricate nel database	I nuovi dati possono essere aggiunti ed estratti in qualsiasi momento

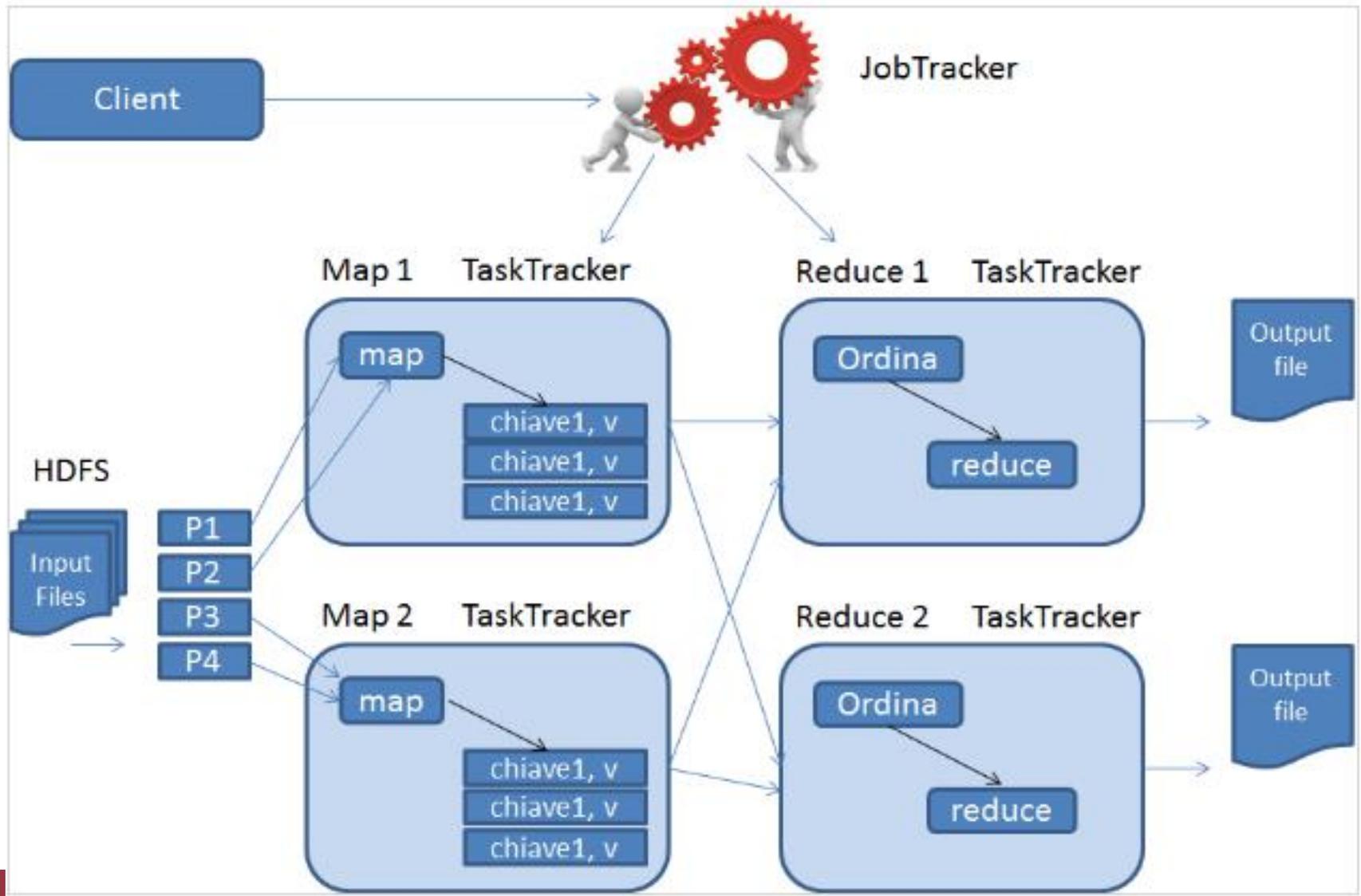
Framework MapReduce

Componente	Descrizione
Hadoop common	Uno strato software comune che fornisce funzioni di supporto agli altri moduli.
HDFS	File system distribuito che fornisce un'efficace modalità di accesso ai dati. Garantisce che i dati siano ridondanti nel cluster rendendo le operazioni sui dati stessi immuni dall'eventuale guasto di un nodo. HDFS accetta dati in qualsiasi formato, strutturati e non strutturati. Di HDFS parleremo a breve in maniera dettagliata.
Map-Reduce	Un pattern che implementato permette di realizzare sistemi di computazione parallela e distribuita di grandi quantità di dati lavorando secondo il principio del " <i>divide-et-impera</i> ".
YARN	Un framework che consente di creare applicazioni o infrastrutture per il calcolo distribuito (sulla base di MapReduce). Esso si occupa della gestione delle risorse del cluster (memoria/CPU/storage).

Funzione Map-Reduce



Framework MapReduce



Terastat Sapienza

TeraStat in sintesi

- 336 core di calcolo
- ca. 1,8 Terabyte di RAM
- 24 Terabyte di spazio disco locale
- 72 Terabyte di spazio disco condiviso
- Interconnessione su fibra ottica ad 8 Gbit/s
- Nuova Richiesta per Progetto TeraDrive

Ciclo di vita delle componenti di TeraStat

- **nodi n3-n6:** attivi da dicembre 2012, da dismettere entro quest'anno
- **nodi n7-n10:** attivi da luglio 2015, operativi sino a luglio 2020
- **nodi n11-n18:** attivi da luglio 2015, operativi sino a luglio 2020
- **sistema storage e sistema gestione code:** attivo da gennaio 2012, da dismettere entro quest'anno

Scenario Attuale

- **Obiettivo:** Rendere il servizio stabile e finanziariamente autosufficiente
- **Orizzonte temporale:** virtualmente illimitato
- **Azioni:**
 - Ampliamento bacino di utenza includendo altri gruppi/dipartimenti
 - Aggiornamento sistema di storage e di gestione code ed ampliamento sistema di calcolo da finanziarsi mediante un grande progetto 'grandi attrezzature' (ordine di 100k euro)
 - Individuazione risorsa interna Sapienza per affiancare/sostituire Cineca
 - Rinnovo contratto Cineca sino a quando necessario

Comparazione costi Cluster in cloud (TeraStat)

- Profili tipici di uso di Terastat:
 - **Profilo 1:** Job di 100 ore su unico core
 - **Profilo 2:** Job di 100 ore su 16 core
 - **Profilo 3:** Job di 100 ore su 4 nodi con 16 core ciascuno
 - **Profilo 4:** Job di 100 ore su 1 nodo con 16 core e 756 Giga di RAM

Comparazione costi Cluster in cloud (Amazon Web Services)

Profilo	Costo 100h
Profilo 1 (1 nodo, 1 core)	≈7 €
Profilo 2 (1 nodo, 16 core)	≈89 €
Profilo 3 (4 nodo, 64 core)	≈357 €
Profilo 4 (1 nodo, 16 core, 976 GB RAM)	≈1877 €

Grazie per l'attenzione !

paolo.delloolmo@uniroma1.it