

# Big OLAP Data Cubes in Multidimensional Big Data Analytics: A Shift-Paradigm for Next-Generation Big Data Analytics



**Alfredo Cuzzocrea**

iDEA Lab, University of Calabria, Italy  
& Dept. of CS, University of Paris City, France

*SEMINAR @ LIAS-ENSMA*

e-mail: [alfredo.cuzzocrea@unical.it](mailto:alfredo.cuzzocrea@unical.it)



# Outline

---

- Big Data Principles
- Context Example: Big Data Analytics on IoT
- OLAP: Problem Statement
- Big OLAP Data Cubes on the Cloud
- Multidimensional Big Data Analytics
- Case Study: Booking.com
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
- Conclusions



# Outline

---

- **Big Data Principles**
- Context Example: Big Data Analytics on IoT
- OLAP: Problem Statement
- Big OLAP Data Cubes on the Cloud
- Multidimensional Big Data Analytics
- Case Study: Booking.com
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
- Conclusions



# Big Data Principles

---

- Massive amounts of heterogeneous data
  - Relational Data (Tables/Transactions/Legacy Data)
  - Text Data (Web)
  - Semi-structured Data (XML)
  - Graph Data (Social Networks, Semantic Web)
- Large-scale data (distributed repositories, clouds)
- Scalability issues: running on very-large, growing data sets
- Elastic metaphors – Cloud Computing paradigms
- Database As A Service (DaaS)
- Easy and Interpretable Analytics
- Privacy-Preserving and Secure Data Management



# Big Data Principles

---

- Massive amounts of heterogeneous data
  - Relational Data (Tables/Transactions/Legacy Data)
  - Text Data (Web)
  - Semi-structured Data (XML)
  - Graph Data (Social Networks, Semantic Web)
- Large-scale data (distributed repositories, clouds)
- Scalability issues: running on very-large, growing data sets
- Elastic metaphors – Cloud Computing paradigms
- Database As A Service (DaaS)
- **Easy and Interpretable Analytics**
- Privacy-Preserving and Secure Data Management



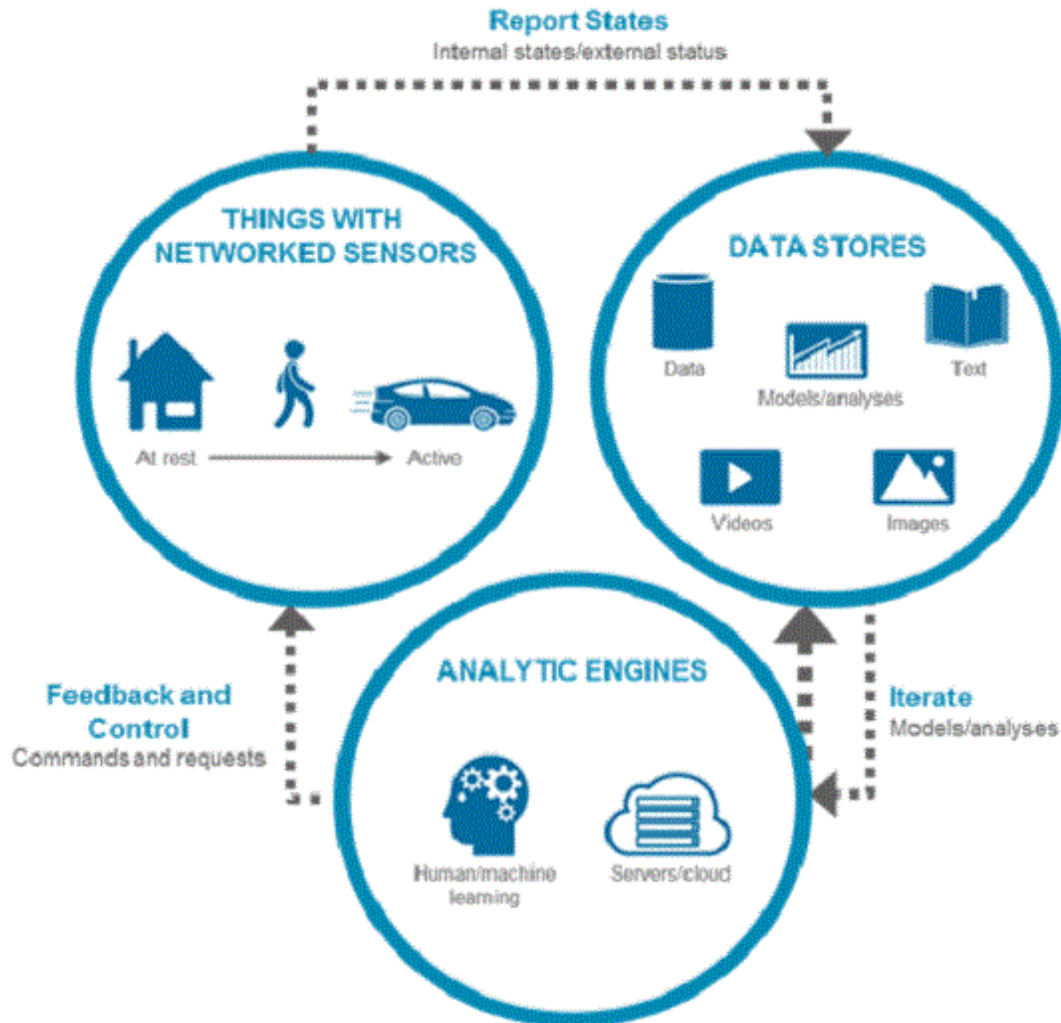
# Outline

---

- Big Data Principles
- **Context Example: Big Data Analytics on IoT**
- OLAP: Problem Statement
- Big OLAP Data Cubes on the Cloud
- Multidimensional Big Data Analytics
- Case Study: Booking.com
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
- Conclusions

# Context Example: Big Data Analytics on IoT/1

Interaction Between the Three Components of the Internet of Things



# Context Example: Big Data Analytics on IoT/2

Interaction Between the Three Components of the Internet of Things



**Data cubes** arise in several data layers of the reference framework:

- in the proper **data storage layer**;
- in the extended **analytical layer**.

Hence, OLAP data cubes provide several functionalities for supporting big data analytics





# Outline

---

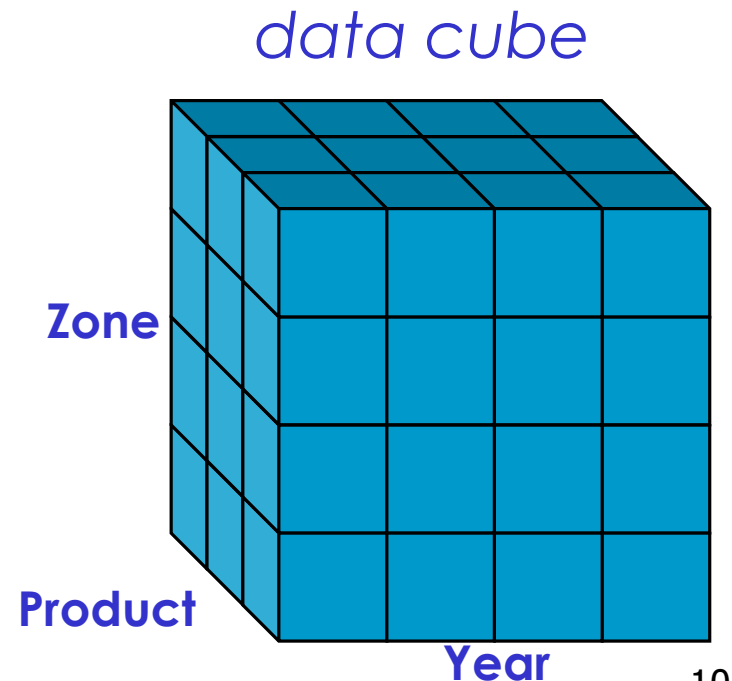
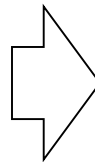
- Big Data Principles
- Context Example: Big Data Analytics on IoT
- **OLAP: Problem Statement**
- Big OLAP Data Cubes on the Cloud
- Multidimensional Big Data Analytics
- Case Study: Booking.com
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
- Conclusions

# OLAP: Problem Statement/1

- OLAP: Performing fast aggregations on huge amounts of data to support decision making processes

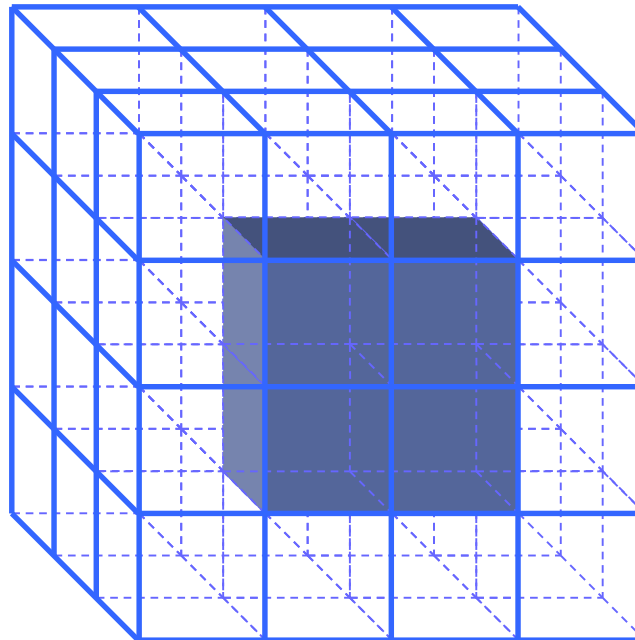
Dimensions			Measure
Product	Year	Zone	Sale

*multidimensional  
representation*



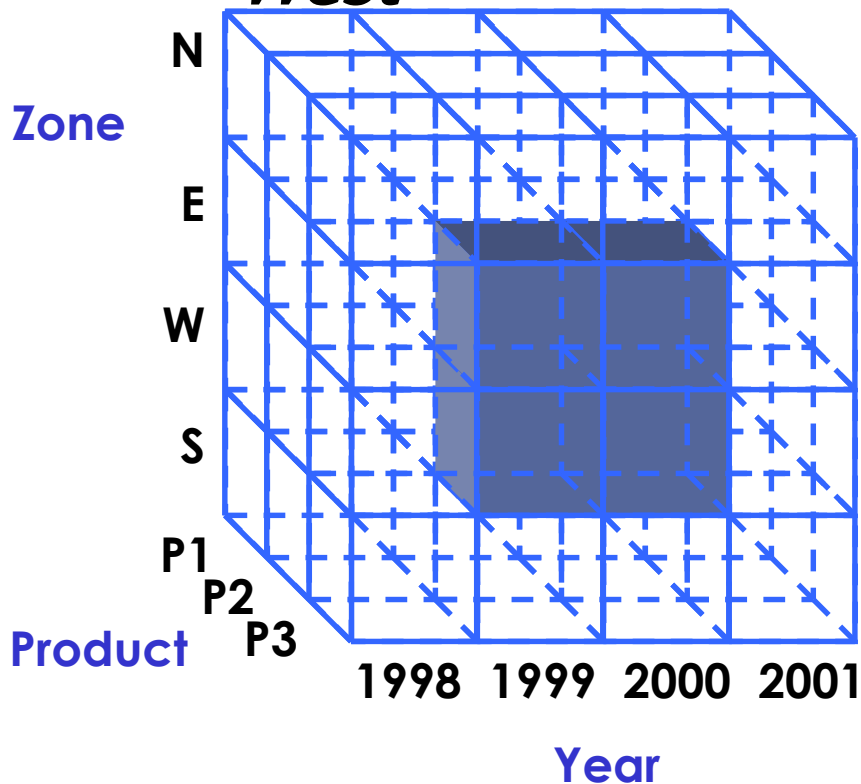
# OLAP: Problem Statement/2

- A range query over a data cube is defined as the application of a SQL aggregation operator (such as SUM, COUNT, AVG etc) on the subset of data which belong to a given range



# OLAP: Problem Statement/3

- Example: total amount of sales of the product *P3* between *1999* and *2000* in zones *East* and *West*



Product	Year	Zone	Sale
...	...	...	...
P3	1999	W	
...	...	...	...
P3	2000	E	
P3	2000	W	
...	...	...	...



# Outline

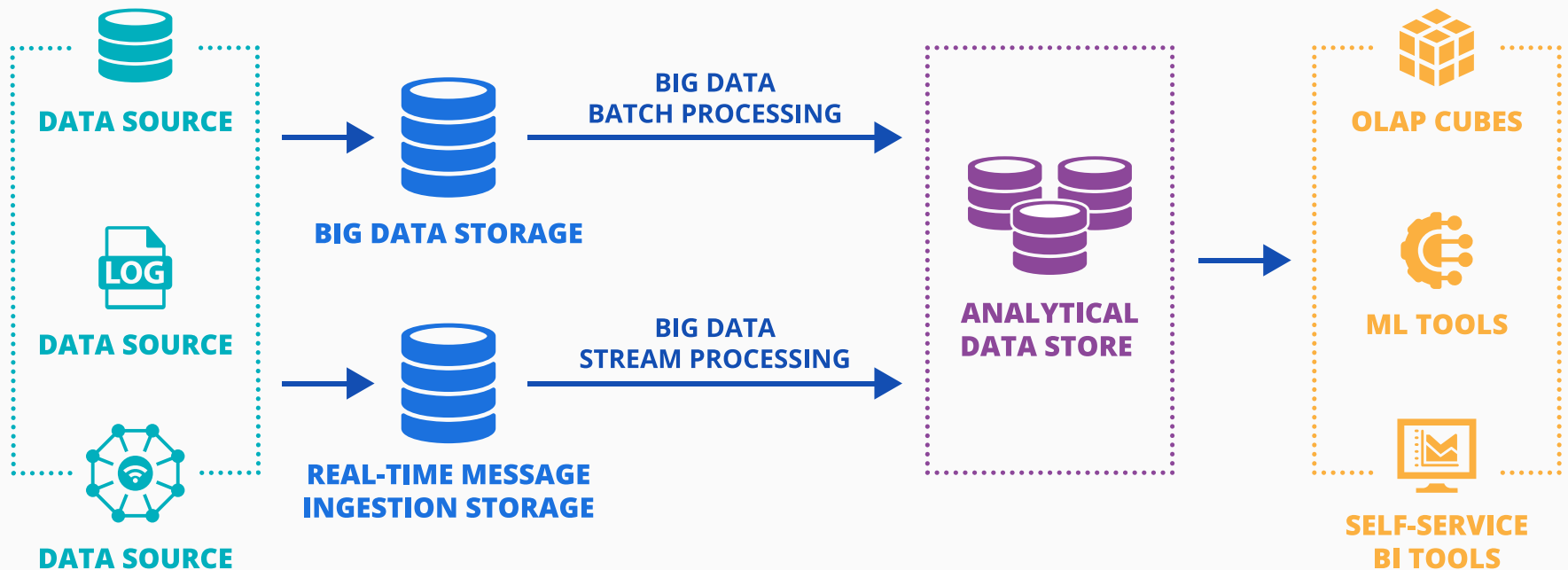
---

- Big Data Principles
- Context Example: Big Data Analytics on IoT
- OLAP: Problem Statement
- **Big OLAP Data Cubes on the Cloud**
- Multidimensional Big Data Analytics
- Case Study: Booking.com
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
- Conclusions

# Big OLAP Data Cubes on the Cloud

- What for Big OLAP Data Cubes on the Cloud?
- A powerful analytical model

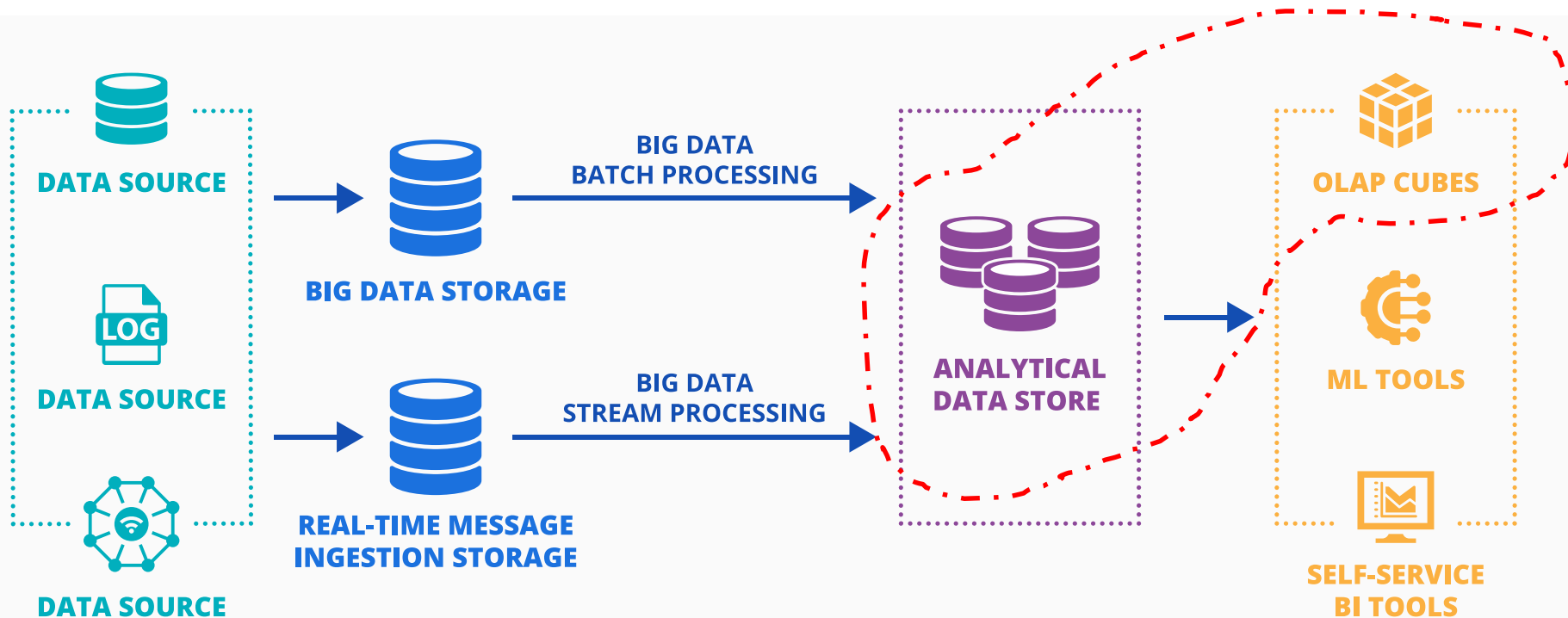
## BIG DATA ARCHITECTURE



# Big OLAP Data Cubes on the Cloud

- What for Big OLAP Data Cubes on the Cloud?
- A powerful analytical model

## BIG DATA ARCHITECTURE





# Outline

---

- Big Data Principles
- Context Example: Big Data Analytics on IoT
- OLAP: Problem Statement
- Big OLAP Data Cubes on the Cloud
- **Multidimensional Big Data Analytics**
- Case Study: Booking.com
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
- Conclusions



# Multidimensional Big Data Analytics/1

- New big data analytical metaphor proposed by me in 2011:



Alfredo Cuzzocrea

Professor of Computer Engineering, [University of Calabria](#), Italy

Email verificata su si.dimes.unical.it - [Home page](#)

[Big Data](#) [Database Systems](#) [Data Mining](#) [Data Warehousing](#) [Knowledge Discovery](#)

 SEGUI

TITOLO

CITATA DA

ANNO

[Analytics over large-scale multidimensional data: the big data revolution!](#)

480

2011

A Cuzzocrea, IY Song, KC Davis

Proceedings of the ACM 14th international workshop on Data Warehousing and ...



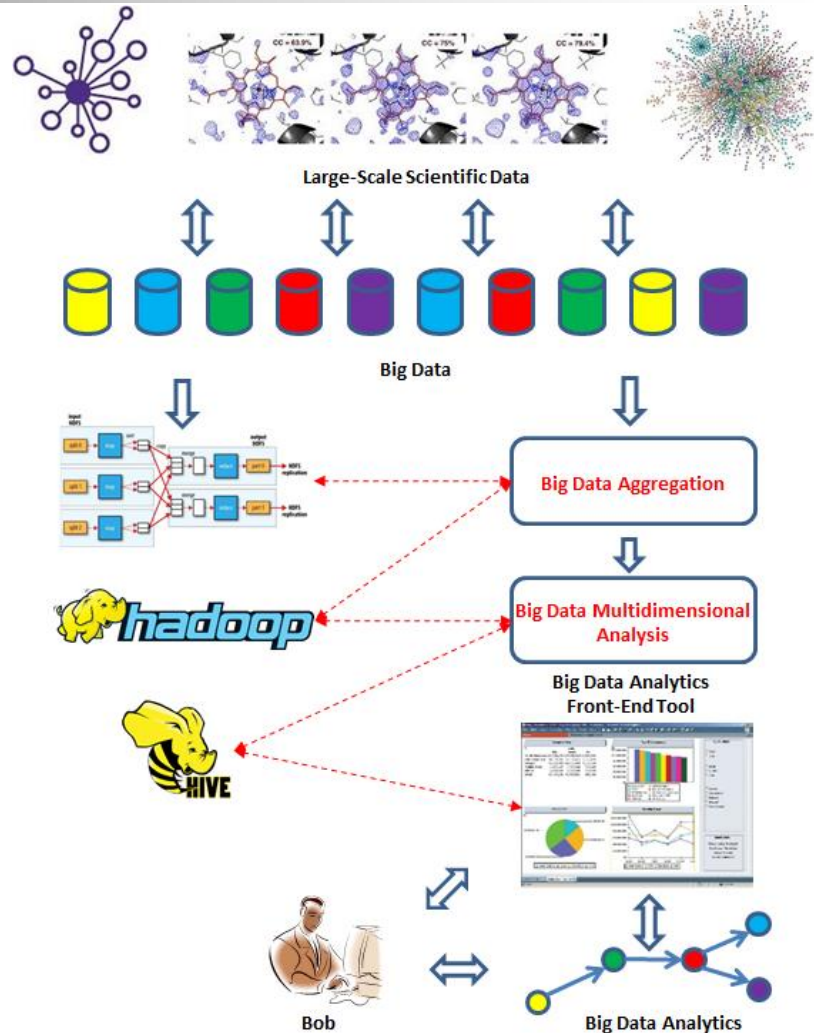
# Multidimensional Big Data Analytics/2

---

- Apply the principles of multidimensional data representation and analysis to the big data analytics process
- The target big data source is modeled in terms of **dimensions** and **measures**
- Aggregation procedures (e.g., pivoting) and machine learning algorithms (e.g., clustering) are applied on the derived multidimensional data structures
- A new paradigm that is overbearingly emerging

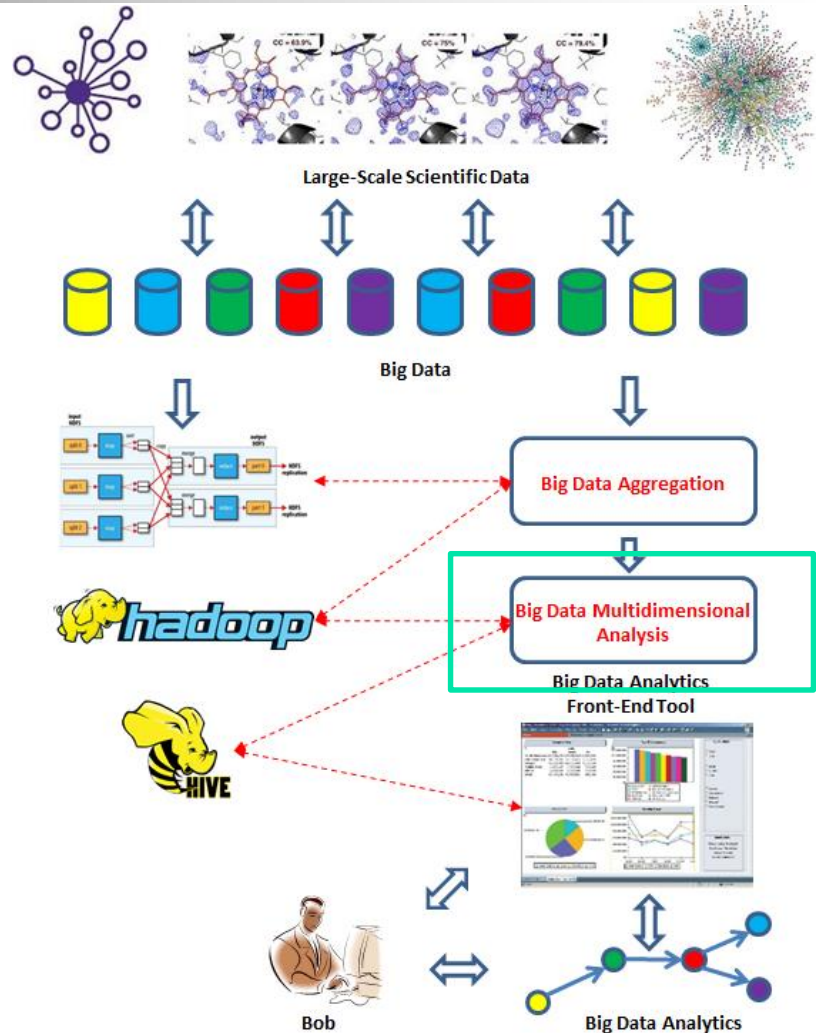
# Multidimensional Big Data Analytics/3

- A reference architecture for supporting multidimensional big data analytics (scientific settings)



# Multidimensional Big Data Analytics/3

- A reference architecture for supporting multidimensional big data analytics (scientific settings)






# Outline

---

- Big Data Principles
- Context Example: Big Data Analytics on IoT
- OLAP: Problem Statement
- Big OLAP Data Cubes on the Cloud
- Multidimensional Big Data Analytics
- **Case Study: Booking.com**
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
- Conclusions

# Case Study: Booking.com/1

Booking.com

EUR  ?

List your property

Register

Sign in

Stays

Flights

Flight + Hotel

Car rentals

Attractions

Airport taxis

Home > Italy > Lazio > Rome > Search results

Search

Destination/property name:

Check-in date

Check-out date

1 adult · 0 children · 1 room

☐ Entire homes & apartments

☐ I'm travelling for work

Search

Filter by:

Popular filters

☐ Spa and wellness centre

136

☐ Spa/wellness packages

94

☐ Private beach area

1

☐ Spa facilities

96

☐ Bed and breakfasts

1075

☐ Spa lounge/relaxation area

81

☐ Hot spring bath

7

Rome: 9,315 properties found

Show on map

Sort by: Our top picks

Commission paid and other benefits may affect an accommodation's ranking. [Find out more.](#)



RomeAsYouLike - Arancio Apartments N° 10

Spagna, Rome · [Show on map](#) · 1 km from centre

In the centre of Rome, located within a short distance of Trevi Fountain and Piazza Navona, RomeAsYouLike - Arancio Apartments N° 10 offers free WiFi, air conditioning and household amenities such as...

Superb

2 reviews

9.0

Location 10

Show prices



Casa Dania

Managed by a private host

Vaticano Prati, Rome · [Show on map](#) · 1.9 km from centre · Metro access

Located in an historic building, Casa Dania offers air-conditioned accommodation in Rome, a 15-minute walk from St Peter's Square.

Fabulous

52 reviews

8.9

Location 9.5

Show prices

# Case Study: Booking.com/2

**Booking.com**

EUR🇬🇧?

List your property

Register

Sign in

Stays

Flights

Flight + Hotel

Car rentals

Attractions

TRE Airport taxis

Home > Italy > Lazio > Rome > Search results

Search

Destination/property name:  
🔍 Rome

Check-in date  
📅 Check-in date ▾

Check-out date  
📅 Check-out date ▾

1 adult · 0 children · 1 room ▾

☐ Entire homes & apartments ⓘ

☐ I'm travelling for work ⓘ

Search

Filter by:

Popular filters

☐ Spa and wellness centre136

☐ Spa/wellness packages94

☐ Private beach area1

☐ Spa facilities96

☐ Bed and breakfasts1075

☐ Spa lounge/relaxation area81

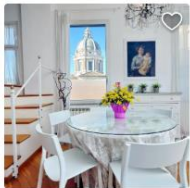
☐ Hot spring bath7

Rome: 9,315 properties found

Show on map

Sort by: Our top picks ▾

Commission paid and other benefits may affect an accommodation's ranking. Find out more. ✕



RomeAsYouLike - Arancio Apartments N° 10

👍👍👍👍

Spagna, Rome · Show on map · 1 km from centre

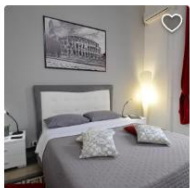
In the centre of Rome, located within a short distance of Trevi Fountain and Piazza Navona, RomeAsYouLike - Arancio Apartments N° 10 offers free WiFi, air conditioning and household amenities such as...

Superb9.0

2 reviews

Location 10

Show prices



Casa Dania👍👍👍

Managed by a private host

Vaticano Prati, Rome · Show on map · 1.9 km from centre · Metro access

Located in an historic building, Casa Dania offers air-conditioned accommodation in Rome, a 15-minute walk from St Peter's Square.

Fabulous8.9

52 reviews

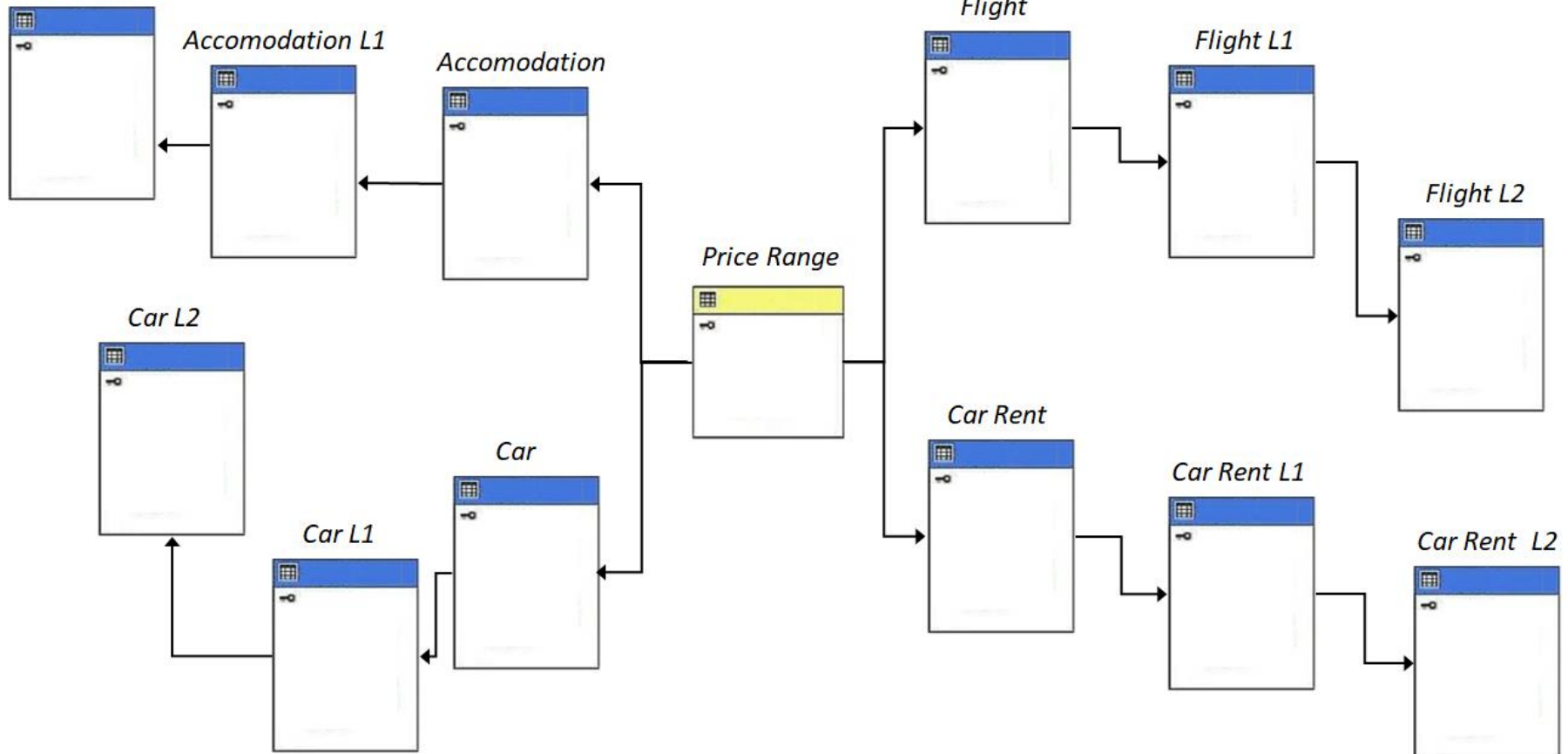
Location 9.5

Show prices

23

# Case Study: Booking.com/3

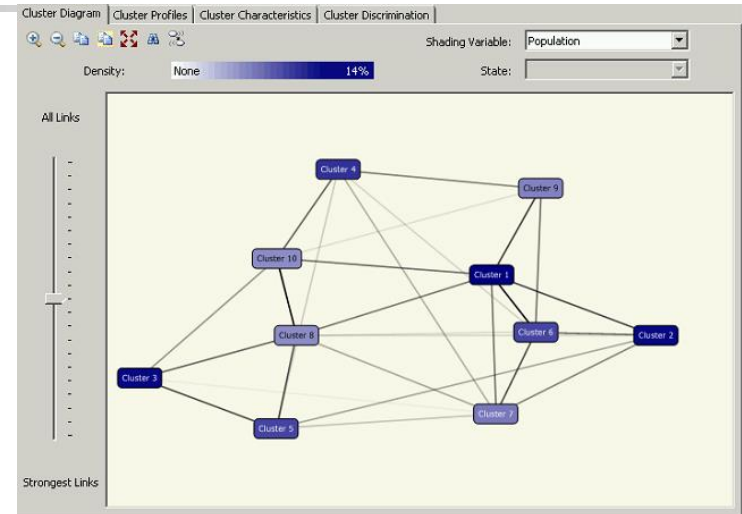
Accommodation L2





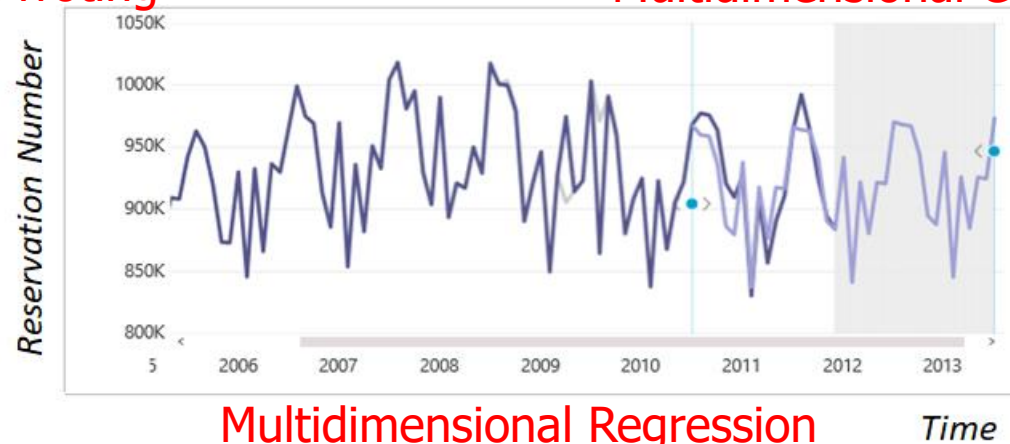
# Case Study: Booking.com/4

Reservation Number						
Zone						
		Calabria			Sicily	
Time	RC	CZ		ME	CT	
2009	3,084	3,120	6,204	3,090	3,126	6,216
January	1,024	1,036	2,060	1,026	1,038	2,064
February	1,028	1,040	2,068	1,030	1,042	2,072
March	1,032	1,044	2,076	1,034	1,046	2,080
2010	3,012	3,048	6,060	3,018	3,054	6,072
January	1,000	1,012	2,012	1,002	1,014	2,016
February	1,004	1,016	2,020	1,006	1,018	2,024
March	1,008	1,020	2,028	1,010	1,022	2,032
Grand Total	6,096	6,168	12,264	6,108	6,180	12,288



Multidimensional Pivoting

Multidimensional Clustering



Multidimensional Regression

Time



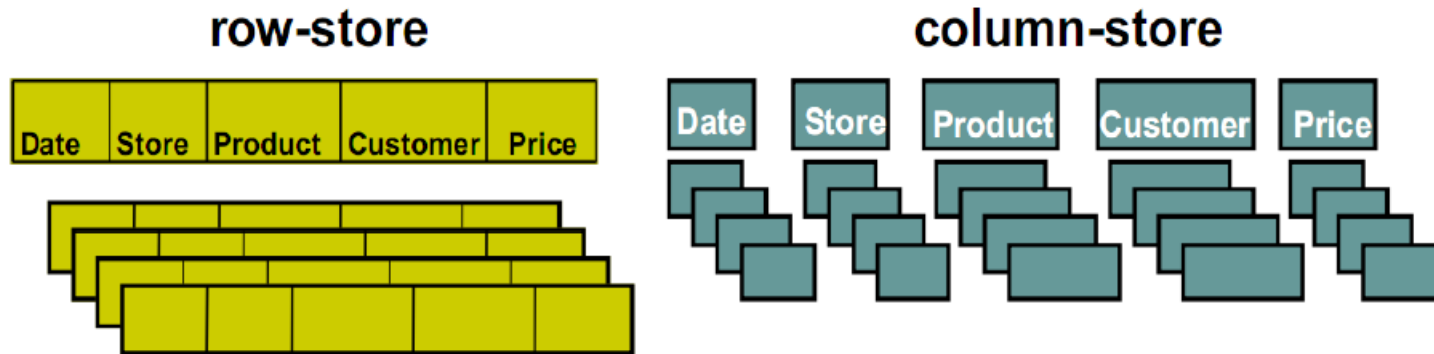
# Outline

---

- Big Data Principles
  - Context Example: Big Data Analytics on IoT
  - OLAP: Problem Statement
  - Big OLAP Data Cubes on the Cloud
  - Multidimensional Big Data Analytics
  - Case Study: Booking.com
  - **Column-Oriented Representation of Big OLAP Data Cubes on the Cloud**
  - Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
  - Conclusions

# Column-Oriented Representation of Big OLAP Data Cubes on the Cloud/1

- Idea: using a column-store representation

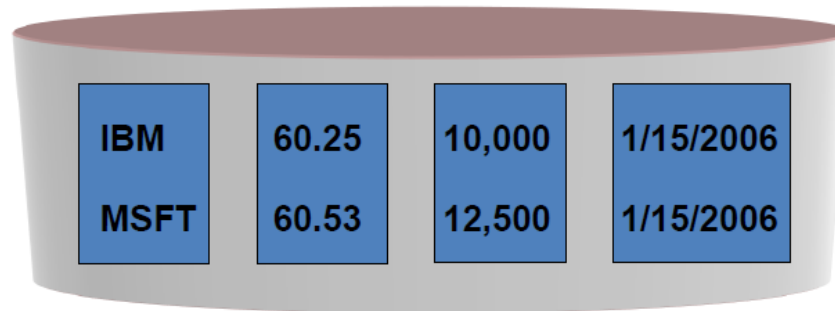


- In row store data are stored in the disk tuple by tuple.
- Where in column store data are stored in the disk column by column

# Column-Oriented Representation of Big OLAP Data Cubes on the Cloud/2

- Example

*Column Store:*

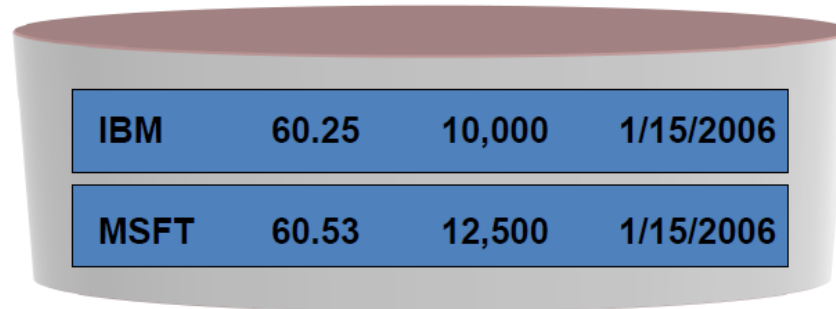
A 3D cylinder representing a data cube, divided into four vertical columns. Each column contains two blue rectangular cells, one for 'IBM' and one for 'MSFT', containing their respective values.

IBM	60.25	10,000	1/15/2006
MSFT	60.53	12,500	1/15/2006

Used in: Sybase IQ, Vertica

---

*Row Store:*

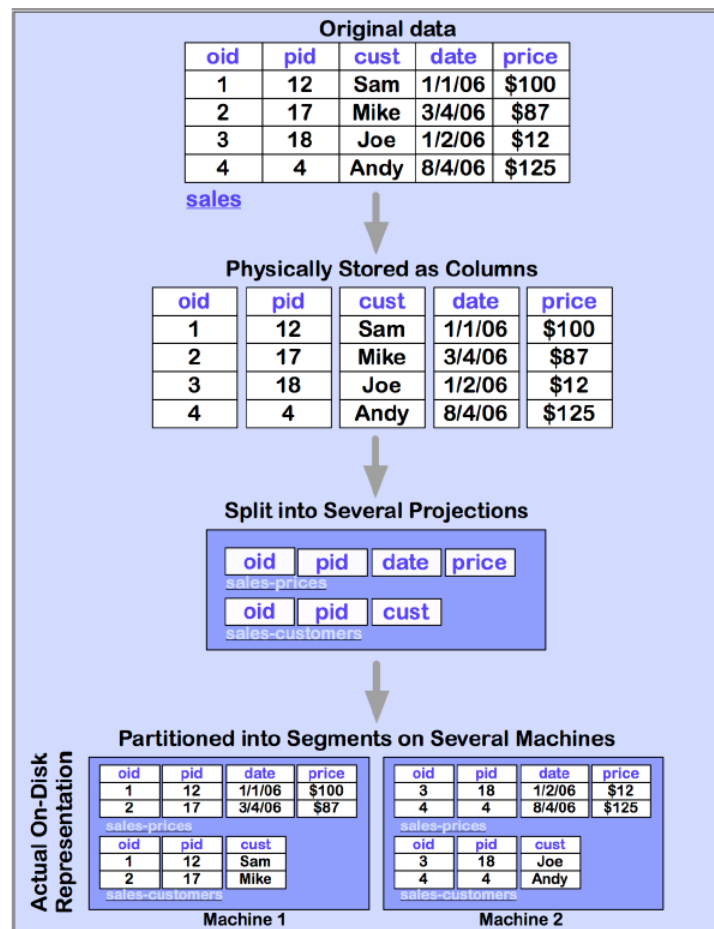
A 3D cylinder representing a data cube, divided into two horizontal rows. Each row contains four blue rectangular cells for 'IBM', a price, a volume, and a date.

IBM	60.25	10,000	1/15/2006
MSFT	60.53	12,500	1/15/2006

Used in: Oracle, SQL Server, DB2, Netezza,...

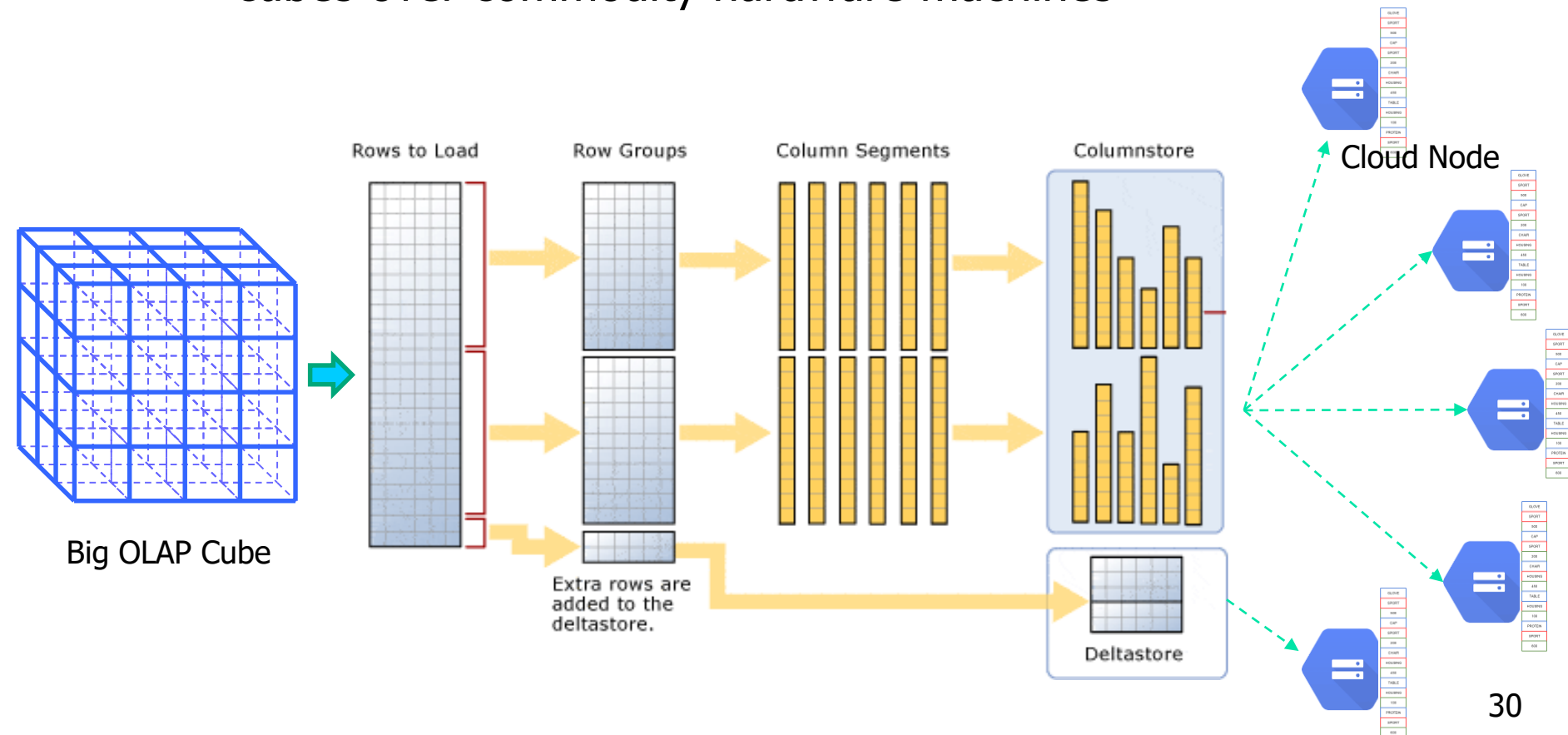
# Column-Oriented Representation of Big OLAP Data Cubes on the Cloud/3

- Cloud-based representation over multiple nodes



# Column-Oriented Representation of Big OLAP Data Cubes on the Cloud/4

- Cloud-based partitioned representation of big OLAP data cubes over commodity hardware machines





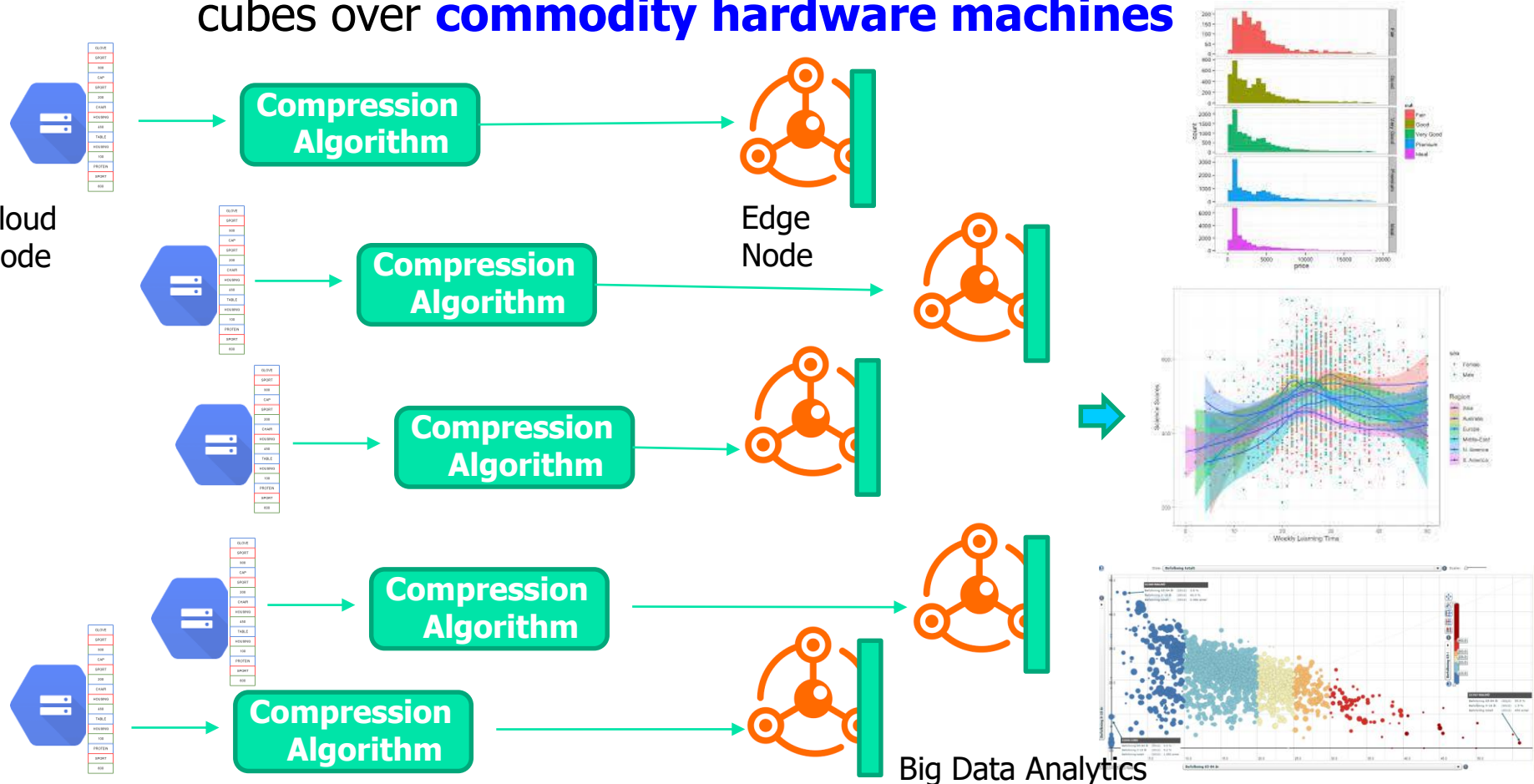
# Outline

---

- Big Data Principles
- Context Example: Big Data Analytics on IoT
- OLAP: Problem Statement
- Big OLAP Data Cubes on the Cloud
- Multidimensional Big Data Analytics
- Case Study: Booking.com
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- **Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud**
- Conclusions

# Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud

- Cloud-based partitioned representation of big OLAP data cubes over **commodity hardware machines**







# Compression Algorithms for Multidimensional Data

---

- Across 10 years, we propose a set of compression techniques for OLAP data cubes:
  - $\Delta$ -Syn, an analytical synopsis data structure that introduces a polynomial approximation technique for OLAP data cubes;
  - $K_{LSA}$ , which further extends the proposal in order to provide accuracy control over compressed OLAP data cubes;
  - LCS-Hist, a histogram-based complex methodology for compressing massive-in-size high-dimensional OLAP data cubes.



# Outline

---

- Big Data Principles
- Context Example: Big Data Analytics on IoT
- OLAP: Problem Statement
- Big OLAP Data Cubes on the Cloud
- Multidimensional Big Data Analytics
- Case Study: Booking.com
- Column-Oriented Representation of Big OLAP Data Cubes on the Cloud
- Innovation: Column-Based Big OLAP Data Cube Compression on the Cloud
- **Conclusions**



# Conclusions

---

- An innovative paradigm for supporting multidimensional big data analytics over big OLAP cubes, deployed on Clouds
- Scalable compression of multidimensional data cubes for supporting multidimensional big data analytics
- Comprehensive approach that can be integrated in next-generation big data analytics systems
- Nice integration with the wide family of open-source tools (e.g., Apache)

# Big OLAP Data Cubes in Multidimensional Big Data Analytics: A Shift-Paradigm for Next-Generation Big Data Analytics



**Alfredo Cuzzocrea**

iDEA Lab, University of Calabria, Italy  
& Dept. of CS, University of Paris City, France

*SEMINAR @ LIAS-ENSMA*

**Thanks for your attention!**