


Data management in large-scale distributed systems

Claudia Lucia Roncancio

INP Grenoble
LIG – Sigma team
Claudia.Roncancio@imag.fr

Once upon a time...a DBMS!

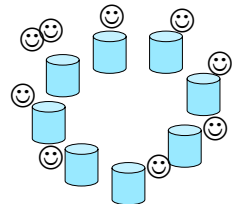
Models languages, DW, DM,...



data + R/W transactions

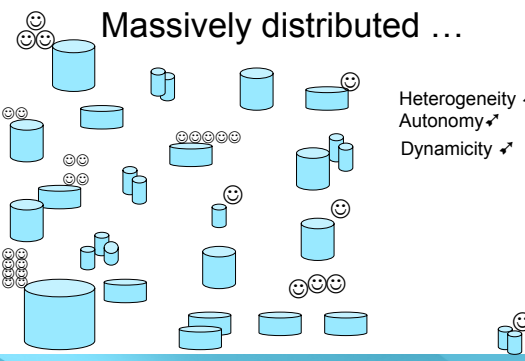
- Persistence
- Querying
- Transactions
- Concurrency control
- Fault tolerance
- Memory mgt. Caching
- Optimization

... distribution....



- Persistence
- Querying
- Transactions
- Concurrency control
- Fault tolerance
- Memory mgt. Caching
- Optimization

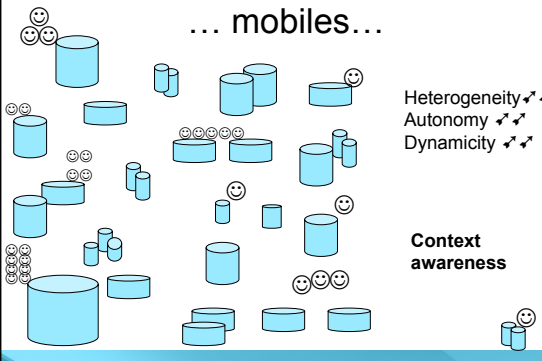
Massively distributed ...



- Heterogeneity ↗
- Autonomy ↗
- Dynamicity ↗

4

... mobiles...

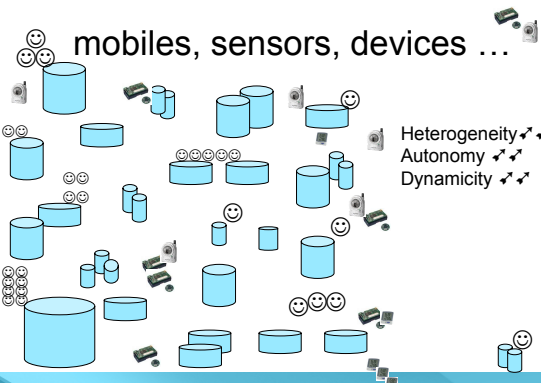


- Heterogeneity ↗
- Autonomy ↗
- Dynamicity ↗

Context awareness

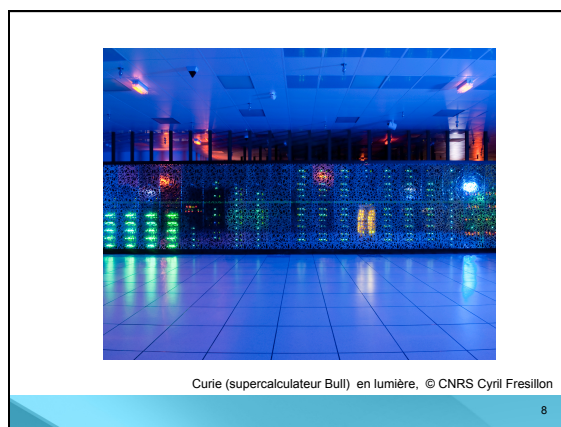
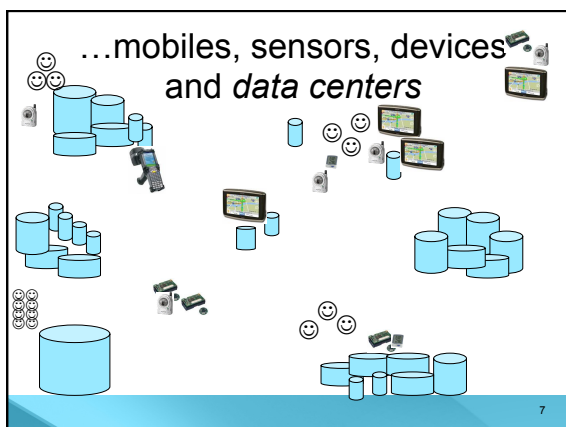
5

mobiles, sensors, devices ...



- Heterogeneity ↗
- Autonomy ↗
- Dynamicity ↗

6



Distributed and ubiquitous systems

- Data everywhere!!
- Data, information and knowledge management
 - very small → big data
 - mobile, elastic, ... !
- Extremely heterogeneous application domains and requirements

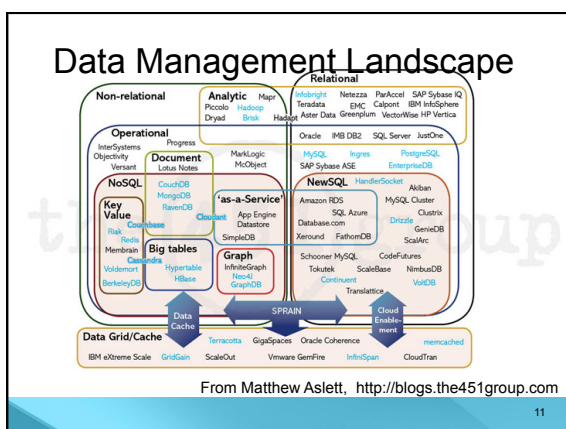
9

Application domains

Today the field covers all the largest sources of data ...

- Any company!
- Scientific and medical databases
- Integrating information
- Data mining
- Web search and data management
- Monitoring everywhere...

10

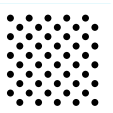
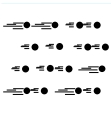




Many issues

- Relational, Key-Value, Graphs, Document
- Querying, analyzing, programming approaches
- Map-Reduce, Spark frameworks
- Scalability, storage, concurrency control, consistency, transactions


12

...Dimensions of Big Data

Volume	Velocity	Variety	Veracity*
 <p>Data at Rest</p> <p><small>Terabytes to exabytes of existing data to process</small></p>	 <p>Data in Motion</p> <p><small>Streaming data, milliseconds to seconds to respond</small></p>	 <p>Data in Many Forms</p> <p><small>Structured, unstructured, text, multimedia</small></p>	 <p>Data in Doubt</p> <p><small>Uncertainty due to data inconsistency & incompleteness, ambiguities, latency, deception, model approximations</small></p>

* Truthfulness, accuracy of precision, correctness


© 2012 IBM Corporation







Data is the new oil.

We see in data the same transformative, wealth-creating power that 19th-century visionaries once sensed in the crude black ooze trapped underground.

If "crude" data can be extracted, refined, and piped to where it can impact decisions in real time, its value will soar. And if data can be properly shared across an entire ecosystem and made accessible in the places where analytics are most useful, then it will become a true game changer, altering the way we live, work, learn, and play.



Source: Cisco IBSG, 2012 #DataInMotion