

# Service Oriented BI

Alberto Abelló & Oscar Romero  
Dept. Enginyeria de Serveis i Sistemes  
d'Informació  
Universitat Politècnica de Catalunya

## Outline

- ◉ Services
- ◉ IaaS
  - > Cloud computing
- ◉ PaaS
  - > BigTable
  - > MapReduce
- ◉ SaaS
  - > CRM
  - > SCM
- ◉ BaaS
  - > SOA
  - > QoS
- ◉ BI on Services
  - > BPM
  - > KPI

## Service definition

"Services are economic activities offered by one party to another, most commonly employing time-based performances to bring about desired results in recipients themselves or in objects or other assets for which purchasers have responsibility. In exchange for their money, time and effort, service customers expect to obtain value from access to goods, labor, professional skills, facilities, networks, and systems; but they do not normally take ownership of any of the physical elements involved."

Lovelock & Wright

## Unified services theory

"All managerial themes unique to services are founded in customers providing significant inputs into the production process."

Sampson

## Benefits of outsourcing

- ◉ Focus on its core competence
- ◉ Decreases cost
- ◉ Access to latest technology
  - > Without investment
- ◉ Leverages benefits from a supplier
  - > Economies of scale

Alberto Abelló & Oscar Romero 05/07/2011

5

## Risks of outsourcing

- ◉ Loss of direct control over quality
- ◉ Exposure to data security issues
- ◉ Dependence on one supplier
- ◉ Coordination expense and delays
- ◉ Atrophy of in-house capacity

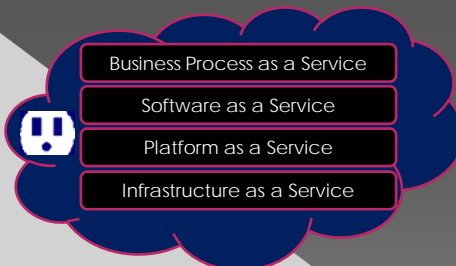
Alberto Abelló & Oscar Romero 05/07/2011

6

## Criteria to choose provider

- Convenience
  - > Locality issues
- Dependability
  - > Reliability
- Personalization
- Price
- Quality
  - > Expectations vs perception
- Reputation
- Speed
- Security
  - > Confidentiality
  - > Availability

## Service layers



## Business Process as a Service

- IBM WebSphere
- Oracle SOA suite
- webMethods
- Apache ServiceMix
- Microsoft Connected Services Framework
- Open ESB
- etc.



Alberto Abelló & Oscar Romero 05/07/2011

9

## Software as a Service

- Salesforce.com
- Cloud9
- Oco
- RightNow
- Microstrategy
- Quantivo
- Oracle on Demand
- etc.



Alberto Abelló & Oscar Romero 05/07/2011

10

## Platform as a Service



- Google BigTable
- Amazon SimpleDB
- Microsoft SDS
- FanthomDB
- Aster DB
- Vertica
- K2 Analytics
- etc.

Alberto Abelló & Oscar Romero 05/07/2011

11

## Infrastructure as a Service



- Amazon EC2
- IBM SmartCloud
- Google app engine
- Etc

Alberto Abelló & Oscar Romero 05/07/2011

12

# Infrastructure as a Service IaaS

Alberto Abelló & Oscar Romero 05/07/2011

13

## Electricity as a utility



Alberto Abelló & Oscar Romero 05/07/2011

14

## Computation as a utility



Alberto Abelló & Oscar Romero 05/07/2011

15

## Cloud computing definition

"Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction."

NIST (National Institute of Standards and Technology)

Alberto Abelló & Oscar Romero 05/07/2011

16



## Characteristics

- On-demand self-service
- Broad network access
- Resource pooling
- Rapid elasticity
- Measured service

Alberto Abelló & Oscar Romero 05/07/2011

17

## What's new

- Illusion of infinite resources
- Elimination of up-front commitment
- Pay-per-use

Alberto Abelló & Oscar Romero 05/07/2011

18

## A cow or bottled milk?

**Buy a cow**

- High upfront investment
- High maintenance cost
- Produces a fixed amount
- Not easy scaling

**Buy bottled milk**

- Pay-per-use
- Lower maintenance cost
- Linear scaling
- Fault tolerant

Daniel Abadi analogy

Alberto Abelló & Oscar Romero 05/07/2011 19

## Obstacles/Opportunities

- Availability of service
- Data lock-in
- Data confidentiality
- Data transfer bottlenecks
- Performance unpredictability
- Scalable storage
- Debugging
- Scaling quickly
- Reputation fate sharing
- Software licensing

Alberto Abelló & Oscar Romero 05/07/2011 20

## Challenges

- ◉ Deployment
  - > Localization
  - > Routing
  - > Authentication
- ◉ Tuning
  - > Placement
  - > Resource partitioning
  - > Service level objectives
  - > Dynamically varying workloads

Alberto Abelló & Oscar Romero 05/07/2011

21

## Platform as a Service PaaS

Alberto Abelló & Oscar Romero 05/07/2011

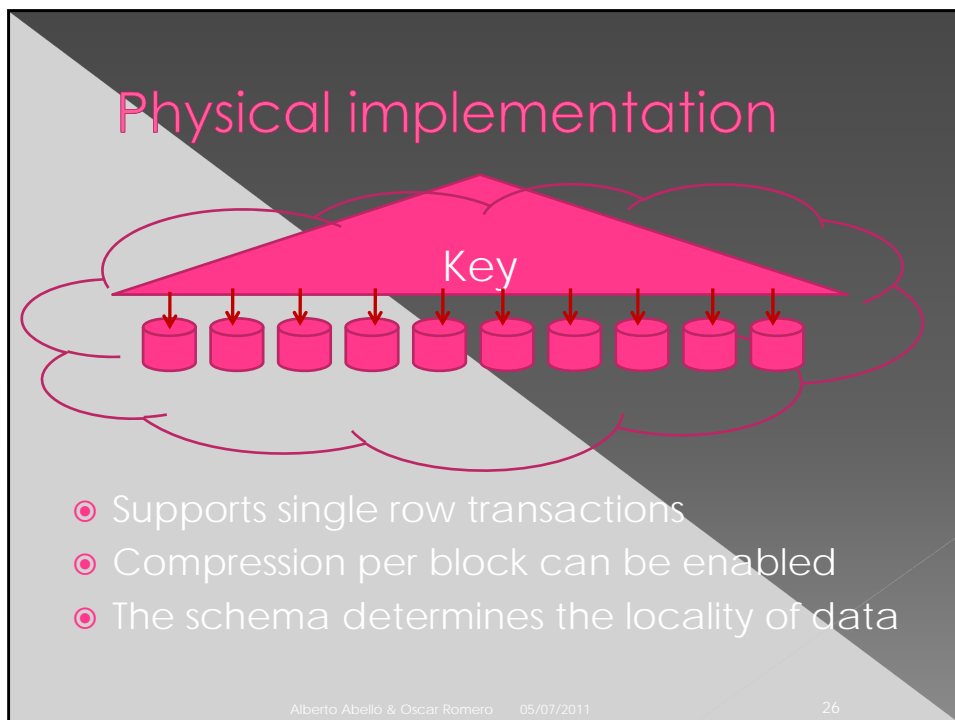
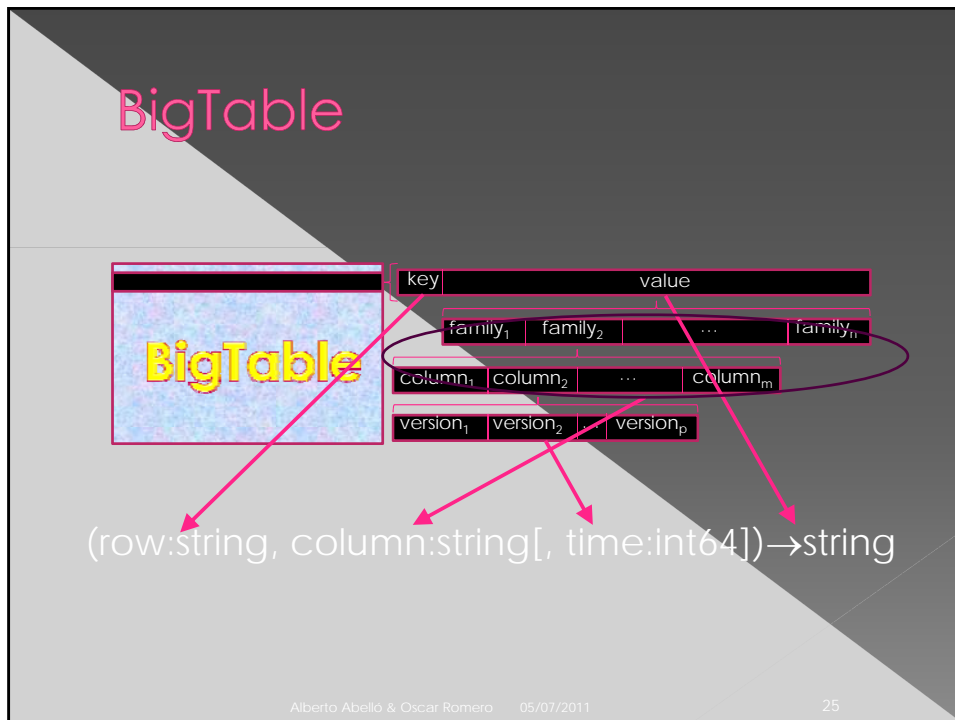
22

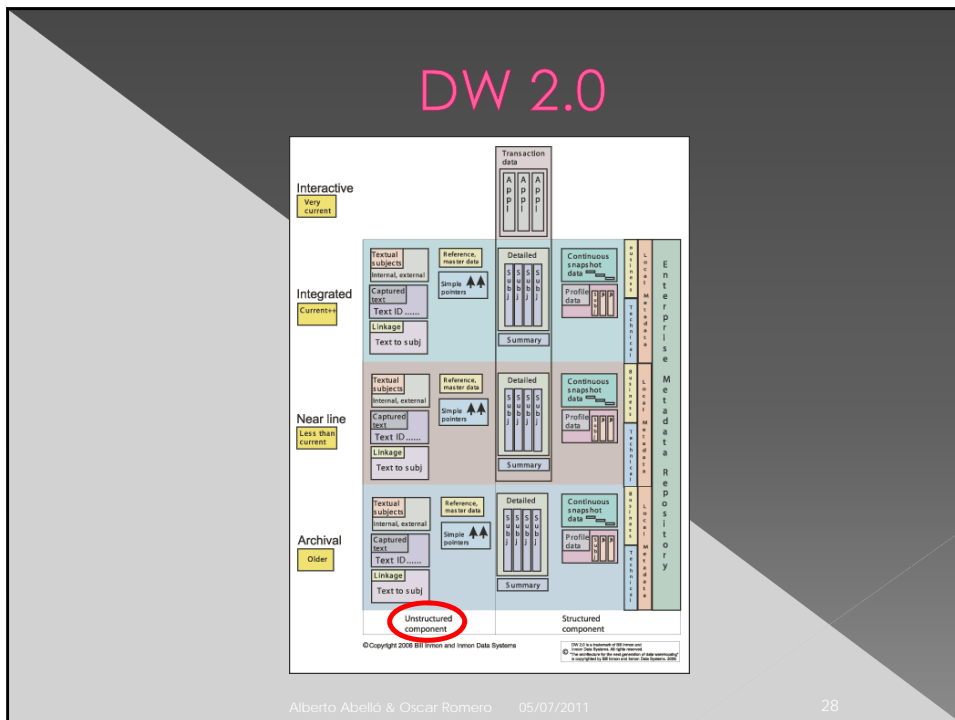
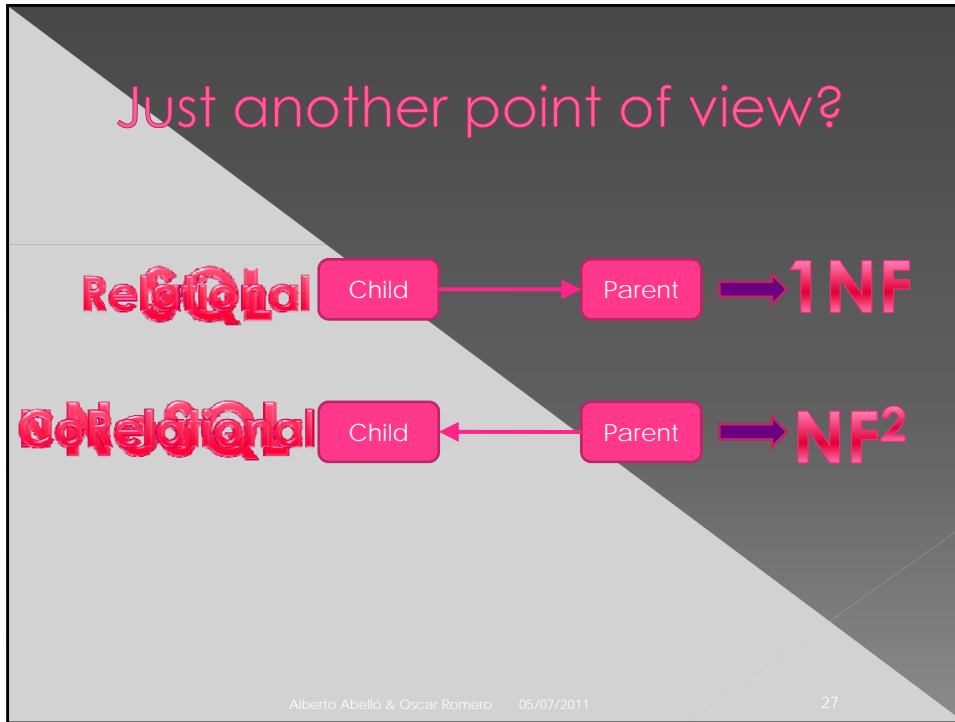
## Kinds of software

- Software in the cloud
  - > DBMS
  - > Workflow management
  - > Versioning
- Cloud software

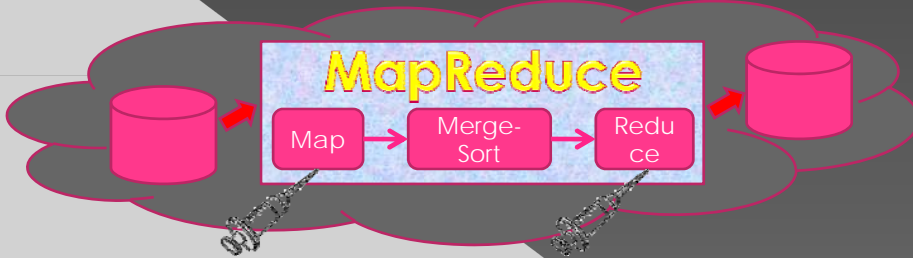
## Open source platform

- HBase
  - > Based on Google BigTable (2006)
- Hadoop
  - > Based on Google MapReduce (2004)





## MapReduce



- Processes pairs [key, value]
- Hides parallelization, fault-tolerance, data distribution and load balancing

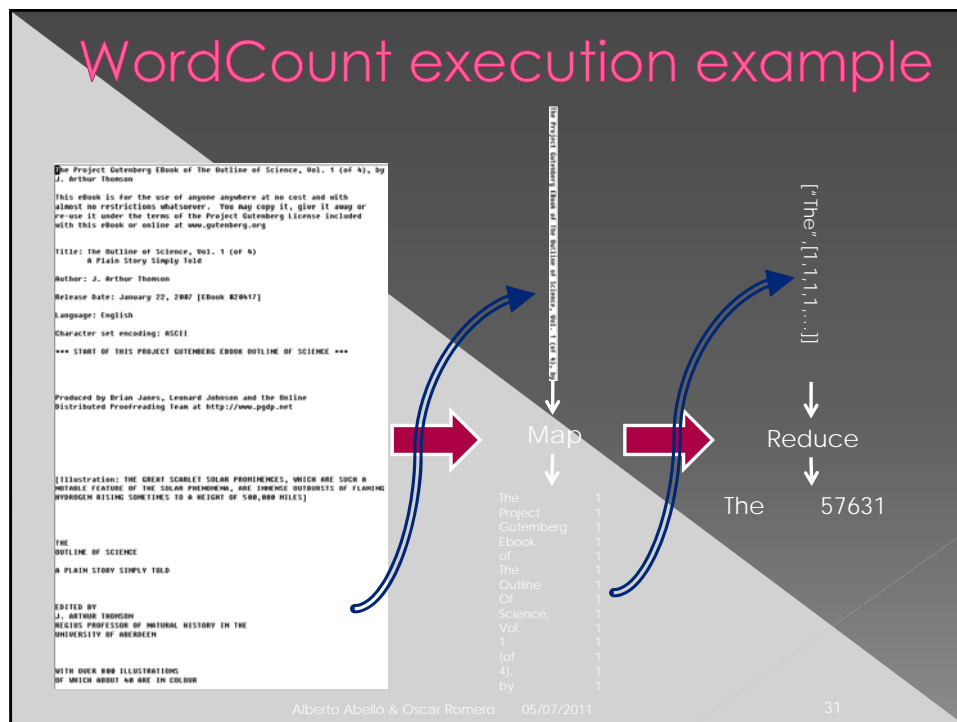
Alberto Abelló & Oscar Romero 05/07/2011 29

## WordCount code example

```
public void map(LongWritable key, Text value) {
    String line = value.toString();
    StringTokenizer tokenizer = new StringTokenizer(line);
    while (tokenizer.hasMoreTokens()) {
        write(new Text(tokenizer.nextToken()), new IntWritable(1));
    }
}

public void reduce(Text key, Iterable<IntWritable> values) {
    int sum = 0;
    for (IntWritable val : values) {
        sum += val.get();
    }
    write(key, new IntWritable(sum));
}
```

Alberto Abelló & Oscar Romero 05/07/2011 30



## Benefits of MapReduce

- ◉ Programming model simple yet expressive

$$\text{map}(\text{key } k, \text{value } v) \mapsto [(ik_1, iv_1), \dots, (ik_{m(k,v)}, iv_{m(k,v)})]$$

$$\text{reduce}(\text{key } ik, \text{vset } ivs) \mapsto [ov_1, \dots, ov_{r(ik, ivs)}]$$

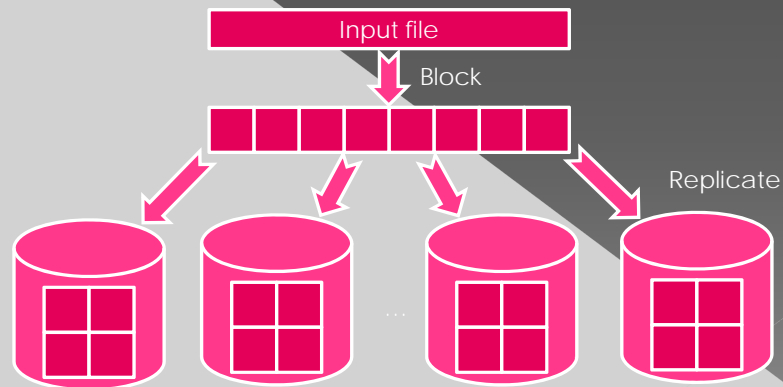
- ◉ Able to process structured or unstructured
- ◉ Elastically scalable
- ◉ Fine grained fault tolerance

Alberto Abelló & Oscar Romero 05/07/2011 32



## Algorithm: Data load

1. The input data is partitioned into blocks
2. Replicate them in different nodes

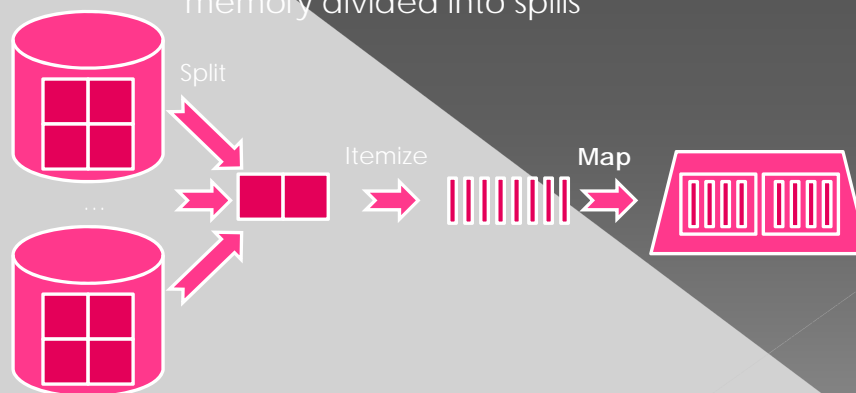


Alberto Abelló & Oscar Romero 05/07/2011

33

## Algorithm: Map phase (I)

3. Each map subplan reads a subset of blocks (i.e., split)
4. Divides it into records
5. Executes the map for each record and leaves them in memory divided into spills

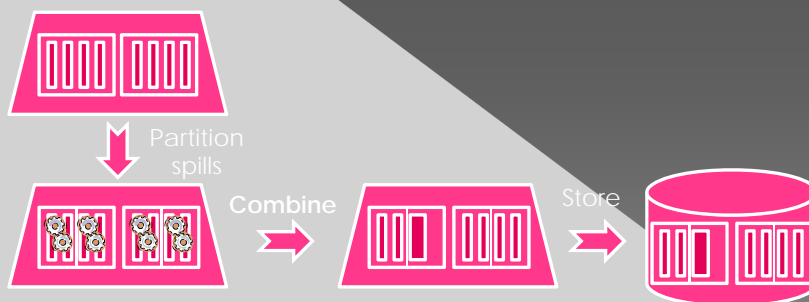


Alberto Abelló & Oscar Romero 05/07/2011

34

## Algorithm: Map phase (II)

6. Each spill is then partitioned per reducers
7. Each partition is sorted independently
8. Store the spills into disk



Alberto Abelló & Oscar Romero 05/07/2011

35

## Algorithm: Map phase (III)

9. Spill partitions are merged
10. Each merge is sorted independently
11. Store the result into disk

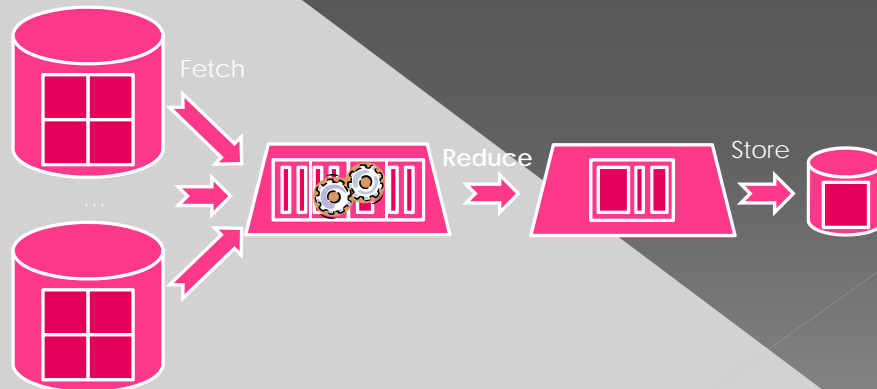


Alberto Abelló & Oscar Romero 05/07/2011

36

## Algorithm: Shuffle and Reduce

12. Reducers fetch data from mappers
13. Mappers output is merged and sorted
14. Reduce function is executed per key
15. Store the result into disk



Alberto Abelló & Oscar Romero 05/07/2011

37

## Local-global aggregation

- Combine is executed locally
  - > Assumes uniform random distribution of input
  - > Reduces the number of tuples sent to reducers
- Only possible when the reducer function is:
  - > Commutative
  - > Associative
- Only makes sense if
 
$$|I|/|O| \gg \#CPU$$

Alberto Abelló & Oscar Romero 05/07/2011

38

## MapReduce performance problems

- ◉ Does not benefit from compression
- ◉ Writes intermediate results to disk
  - › Reduce tasks pull intermediate data
- ◉ Defines the execution plan on the fly
  - › Schedules one block at a time

## MapReduce improvements

- ◉ Bypass the storage system or even OS
- ◉ Add/Use indexing structures
- ◉ Follow some programming rules
- ◉ Provide sorting alternatives to merge-sort
- ◉ Offer alternatives to block granularity for scheduling

Jiang et al.  
VLDB 2010

## MapReduce extensions

- Ashish Thusoo, et. al. "*Hive - A Warehousing Solution Over a Map-Reduce Framework*". VLDB' 2009
- Sergey Melnik, et. al. "*Dremel: Interactive Analysis of Web-Scale Datasets*". VLDB' 2010
- Jens Dittrich, et. al. "*Hadoop++: Making a Yellow Elephant Run Like a Cheetah (Without It Even Noticing)*". VLDB' 2010
- Alexander Alexandrov, et al. "*MapReduce and PACT - Comparing Data Parallel Programming Models*". BTW' 2011

Alberto Abelló & Oscar Romero 05/07/2011

41

## NoSQL drawbacks

- No ACID
- No standard
- Low-level query

Michael Stonebraker

Alberto Abelló & Oscar Romero 05/07/2011

42

## Brewery or bottled beer?

### Do It Yourself

- Expensive
- Ad hoc development

### Off the Shelf

- Economies of scale
- Concrete functionalities



Specificity

Low Cost

Florian Waas analogy

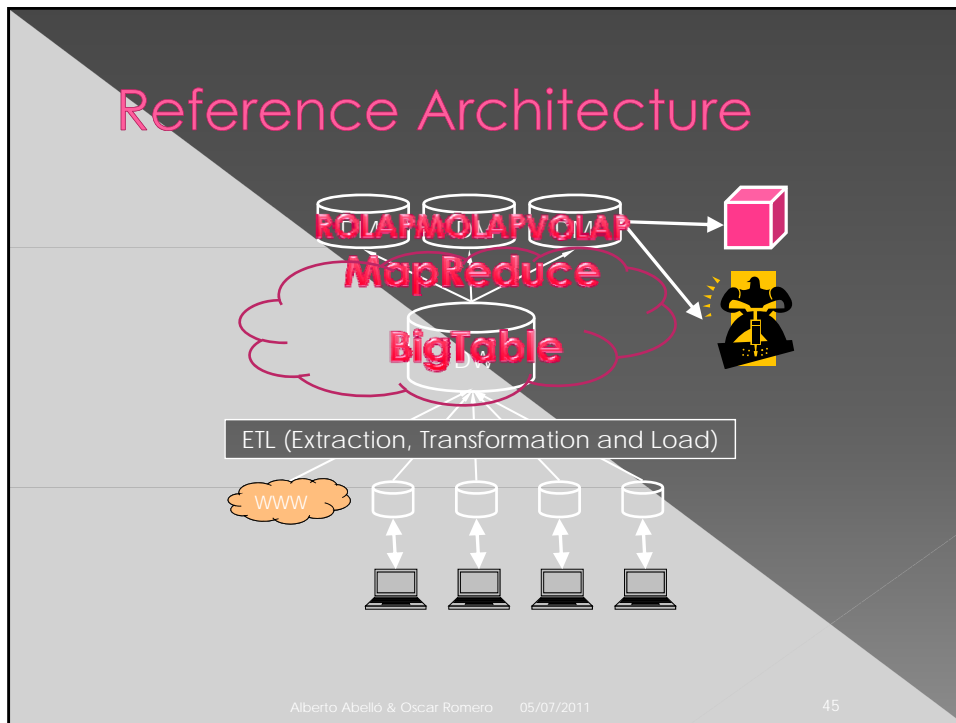
Alberto Abelló & Oscar Romero 05/07/2011 43

## Market tools

“ ... But to really unlock the power of Hadoop, you must be able to efficiently extract data stored across multiple (often tens or hundreds) of nodes with a user-friendly ETL (extract, transform and load) tool that will then allow you to move your Hadoop data into a relational data mart or warehouse where you can use BI tools for analysis. ”

**Ian Fyfe**  
*Pentaho*

Alberto Abelló & Oscar Romero 05/07/2011 44



## Big elephant or little elephant?

**Huge**

**Trademarks**

- Expensive
- Many functionalities
- ...

**Adaptative**

**Open source**

- Free
- Simple functionalities
- ...

The image shows a golden scale of justice. The left pan is significantly lower than the right pan, indicating it is heavier. On the left pan, the word 'ORACLE' is written in red. On the right pan, the 'hadoop' logo is written in blue. The scale is positioned in the center of the slide, between the 'Huge' and 'Adaptative' text blocks.

Alberto Abelló & Oscar Romero 05/07/2011 46

# Friends or foes

The diagram illustrates the relationships between several technologies. At the top, three pink rounded rectangular boxes labeled 'MapReduce', 'Java', and 'Oracle' are arranged horizontally. Below them is a single, wider red rounded rectangular box labeled 'Oracle Grid Engine'. At the bottom is a grey rounded rectangular box labeled 'Hadoop Distributed File System (HDFS)'. The background features a diagonal split between light grey and dark grey.

Alberto Abelló & Oscar Romero 05/07/2011 47

# Software as a Service SaaS

The slide features the text 'Software as a Service' in a blue, 3D-style font, with 'SaaS' below it in a similar style. The background is a diagonal split between light grey and dark grey.

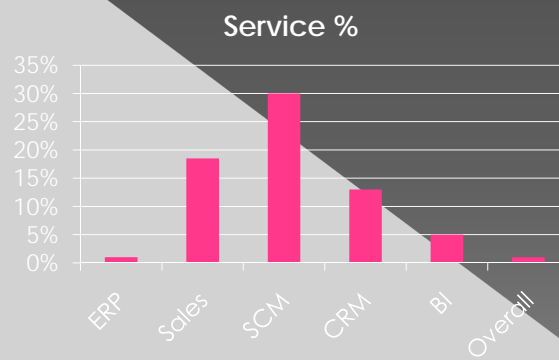
Alberto Abelló & Oscar Romero 05/07/2011 48



## Gartner's considerations

	On-premises	Service-based
Customization	+	+/-
Implementation time	+/-	+/-
Application shut-off	+	-
Hidden fees	-	-
Security of data	+	-
Process integrity	+	-
Guarantee of quality	+	-

## Gartner's market analysis



## Supply Chain Management

- ◉ Increase the efficiency of transactions
- ◉ Reduce the costs of maintenance
  - > Reduce obsolescence costs
- ◉ Reduce the loss of sales
- ◉ Reduce the average time to deliver a product
  - > Improve the cash flow
    - Shorter delivery and charging time
  - > Reduce the costs of dealing with urgent orders
  - > Improve the service to the client
    - Smaller delivery time and price

Alberto Abelló & Oscar Romero 05/07/2011

51

## Customer Relationship Management

- ◉ 360° view of clients
- ◉ Automate and administering sales processing
- ◉ Reduce service costs
- ◉ Improve collaboration
- ◉ Speed-up in sales cycle
- ◉ Improve efficiency in the management of contacts
- ◉ Increase sales in the clients base

Alberto Abelló & Oscar Romero 05/07/2011

52

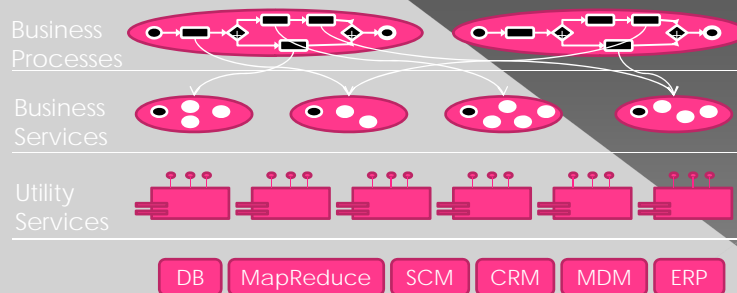
## Enterprise Resource Planning

- Only one repository
  - > Finances and accounting
  - > Material management
  - > Human resources
  - > Manufacturing
  - > Sales and marketing
- Modules include
  - > SCM
  - > CRM

## Business processes as a Service BaaS

## Technological Challenges

- Business Process Management
- Service Composition
- Service Infrastructure and Management



Alberto Abelló & Oscar Romero 05/07/2011

55

## SOA principles

- Reusability
- Loose coupling
- Contract
- Abstraction
- Composability
- Autonomy
- Statelessness
- Discoverability

Alberto Abelló & Oscar Romero 05/07/2011

56

## Paradigm

	Distributed components	SOA
Design	Functionality	Process
Designed to ...	Last	Change
Development cycle	Long	Interactive and iterative
Centered on ...	Cost	Business
Coordination	Blocks	Orchestration
Coupling	Tight	Loose (agile and adaptive)
Technologies	Homogeneous	Heterogeneous
Programming	Objects	Messages
Encapsulation	Partial	Full (contracts)

Alberto Abelló &amp; Oscar Romero 05/07/2011

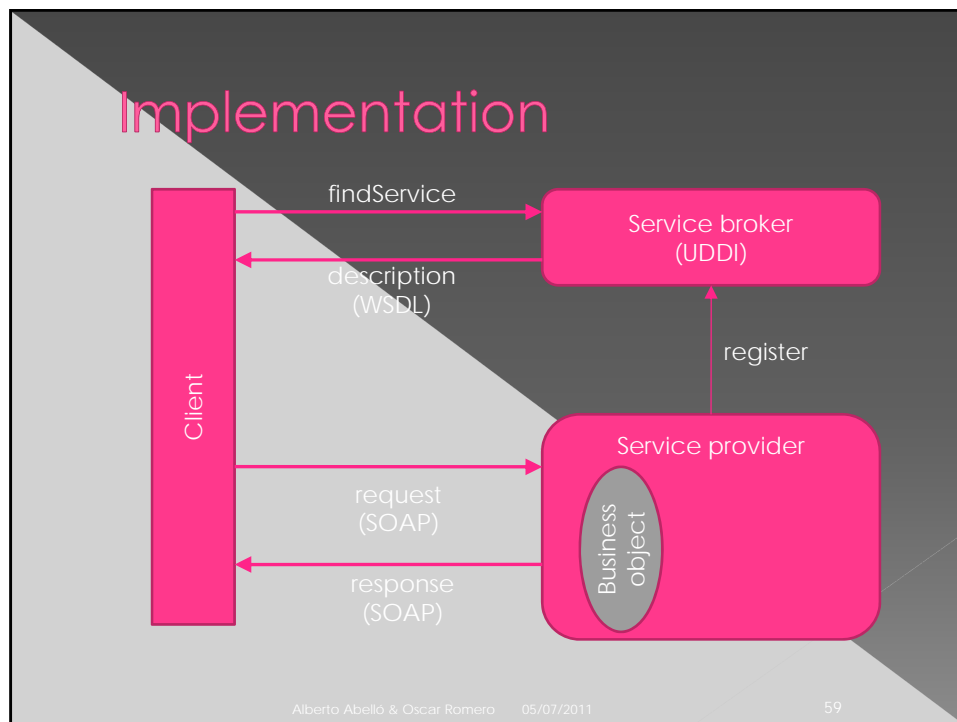
57

## Service composition

- ◉ Primitive activity
- ◉ Complex activity
  - > Atomic transaction
  - > Business activity
    - Orchestration
    - Choreography

Alberto Abelló &amp; Oscar Romero 05/07/2011

58



## Quality of Service

- ◉ Definition
  - > Difference between perceived and expected
- ◉ Negotiation
  - > Service Level Agreement
    - Service Level Objectives
- ◉ Assurance

Alberto Abelló & Oscar Romero 05/07/2011

61

## QoS measures

- ◉ Input
  - > Supply
  - > Cost
- ◉ Process
  - > Performance
  - > Security
- ◉ Outcome
  - > Customization
  - > Satisfaction
- ◉ Systemic
  - > Reproducibility
  - > Sustainability

Alberto Abelló & Oscar Romero 05/07/2011

62

# Business Intelligence on Services

Alberto Abelló & Oscar Romero 05/07/2011

63

# Customer Relationship Management

- ◉ Turns manufacturers into services
- ◉ Becomes a barrier for competitors
- ◉ By coding customers allows to:
  - > Instruct staff
  - > Manage queues in call centers
  - > Target offers
  - > Share data with other firms

Alberto Abelló & Oscar Romero 05/07/2011

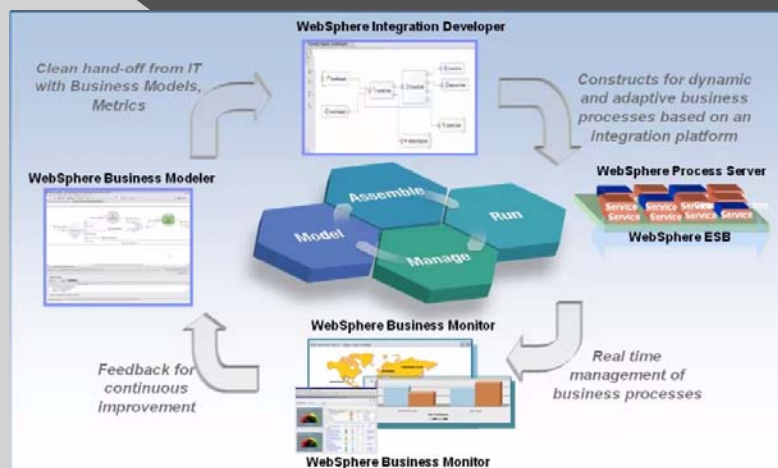
64



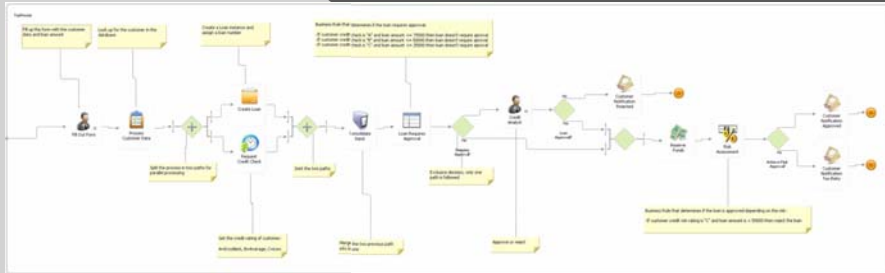
# Key Performance Indicators

"You cannot control what you cannot measure"

# IBM WebSphere architecture



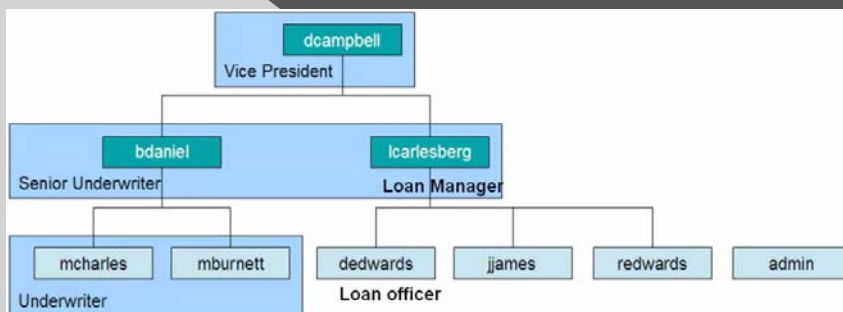
## IBM e-Xtend World Bank demo



Alberto Abelló & Oscar Romero 05/07/2011

67

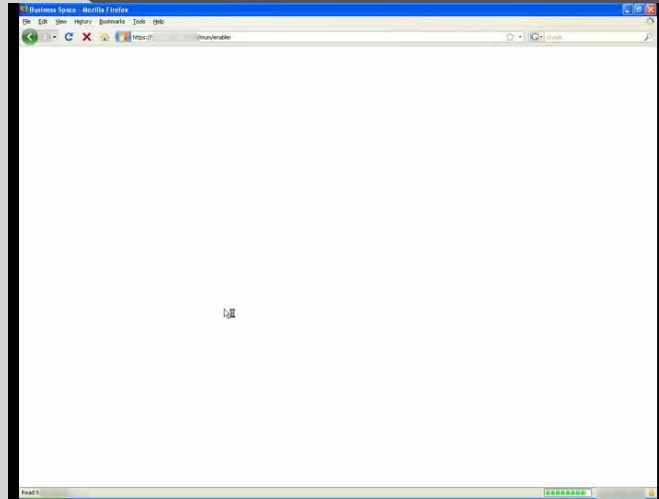
## Users in e-Xtend World Bank



Alberto Abelló & Oscar Romero 05/07/2011

68

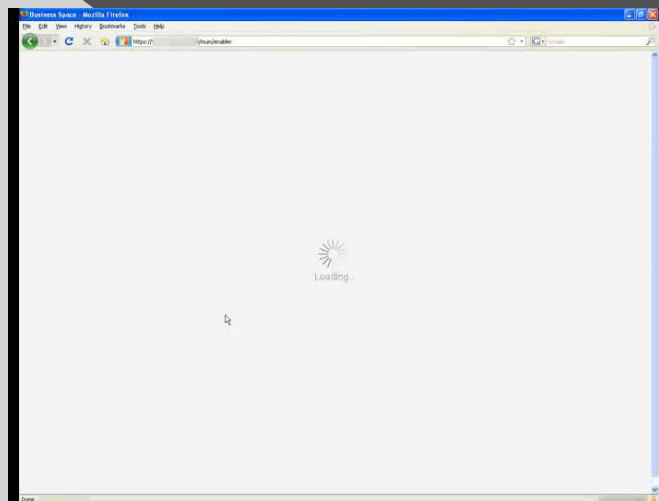
## Loan manager reviews loan status



Alberto Abelló & Oscar Romero 05/07/2011

69

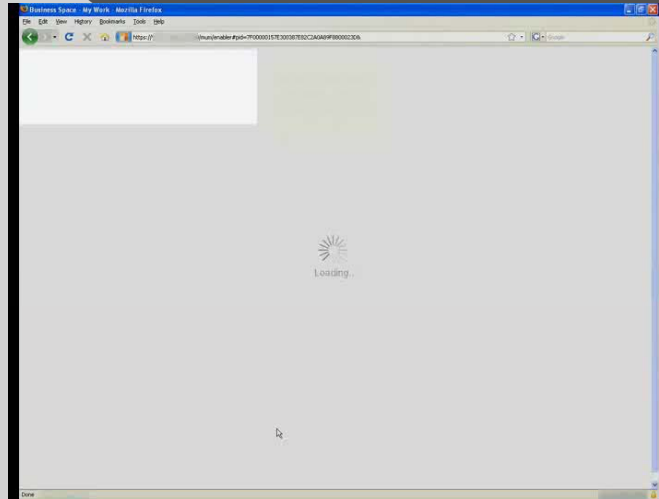
## Loan officer creates a new loan request



Alberto Abelló & Oscar Romero 05/07/2011

70

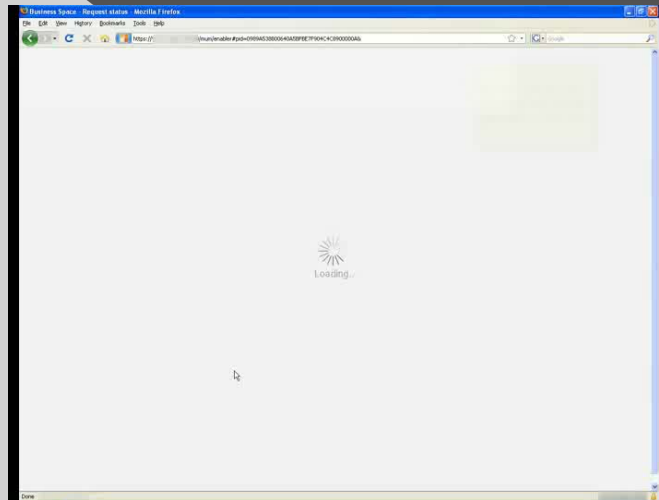
## Senior underwriter works in loan request



Alberto Abelló & Oscar Romero 05/07/2011

71

## Loan manager checks updates



Alberto Abelló & Oscar Romero 05/07/2011

72

## Conclusions

- BI can benefit from services at four levels
  - > IaaS
  - > PaaS
  - > SaaS
  - > BaaS
- Services benefit from BI
  - > KPI and Balance Scorecards

## Bibliography

- Mell, P., Grance, T.: The NIST Denition of Cloud Computing. Special Publication 800-145, National Institute of Standards and Technology (January 2011), draft
- Abadi, D.J.: Data management in the cloud: Limitations and opportunities. IEEE Data Engineering Bulletin 32(1), 3-12 (2009)
- Stonebraker, M., et al.: MapReduce and parallel DBMSs: friends or foes? Communication of ACM 53(1), 64-71 (2010)
- Hostmann, B.: Business Intelligence as a Service: Findings and Recommendations. Research G00164653, Gartner (January 2009)
- Erl, T.: Service Oriented Architecture. Prentice Hall (2006)
- Castellanos, M., et al.: Automating the loading of business process data warehouses. In EDBT'2009. pp. 612-623. ACM