Service Oriented BI

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

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

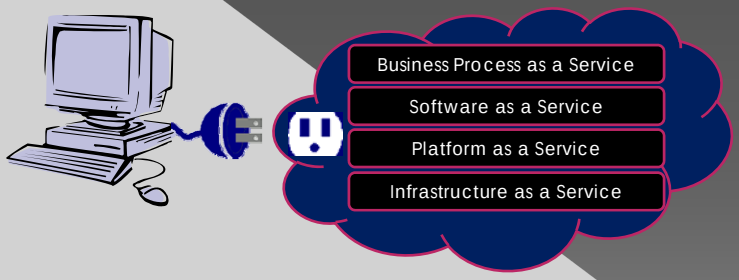
"All managerial themes unique to services are founded in customers providing significant inputs into the production process."

Sampson

Past seminars (Sep'09-Jul'12)

- Antoni Olivé: Què és un Servei?
- Ruth Raventós: Service Management
- Enric Mayol, Tomàs Aluja: Service Is Front Stage
- Ernest Teniente: Reaching the Goal
- Joan Antoni Pastor: CMMI for Services
- Pere Botella: Plataformes i iniciatives en serveis
- Albert Abelló: Understanding Service Businesses. Applying Principles of the Unified Services Theory (UST)
- Antoni Olivé: Introducció a la tecnologia RFID i a la seva aplicació en hospital
- Enric Martínez: SOA: Service Oriented Architecture
- Carles Farré: Web Usability
- Antonio Villegas: Everything you always wanted to know about Web services but were afraid to ask
- Ferran Martí: Breu Introducció a ITIL
- Jordi Pradel: SOA des de les trinxeres
- Antonio Valle: Modelo de Costes. Caso práctico de aplicación
- Jordi Torres: The changing role of the computer scientist in the cloud
- Carme Quer: SERVQUAL. Un instrument de mesura de la qualitat dels serveis
- Frederic Marimon: Instruments de mesura de la qualitat dels serveis online
- Pablo Casado: Browser + Device App Stores. Noves perspectives per al desenvolupament de la web
- Jesús Bisbal: Towards negotiable SLA-based QoS Support for Data Services
- Michael D. Myers: Conducting Critical Research in IS, and its relationship with IS Design Research
- Kazuyosi Hidaka: Service Science Research and Education in Japan
- Miquel Barceló: Per on va la recerca europea en TIC

Service layers



Outline

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

Business Process as a Service

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

Software as a Service

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

Platform as a Service

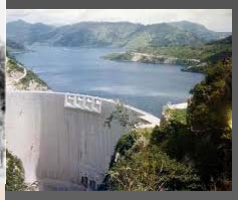
BigTable
SimpleDB
SDS
DB
ytics

Infrastructure as a Service

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

Infrastructure as a Service IaaS

Electricity as a utility

Computation as a utility



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)

Characteristics

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

What's new

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

A cow or bottled milk?

Buy a cow

- High upfront investm
- High maintenance c
- Produces a fixed am
- Step-wise scaling

Security

Buy bottled milk

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

Low cost

Daniel Abadi analogy

Providers' Challenges

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

Obstacles/Opportunities

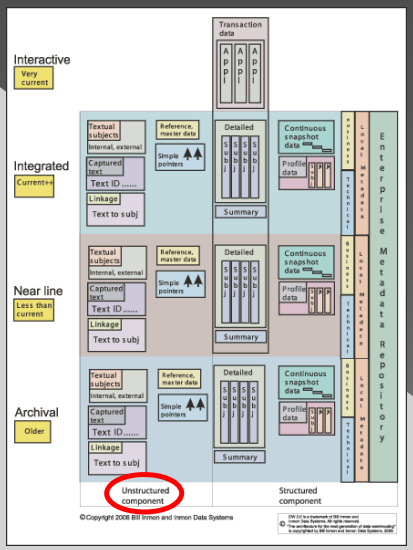
- Availability of service
- Data lock-in
- Data confidentiality
- Data transfer bottlenecks
- Performance unpredictability
- Scalable storage
- Debugging
- Scaling quickly
- Reputation fate sharing
- Software licensing

Platform as a Service
PaaS

Kinds of platforms

- Storage as a Service
- Database as a Service
- Processes (Workflow) as a Service
- Service as a Service
- Security as a Service
- Management/Governance as a Service

DW 2.0



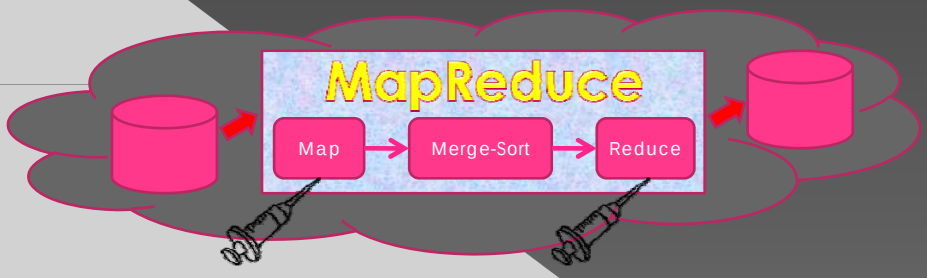
Kinds of software

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

Open source platform

- Hadoop
  - Based on Google MapReduce (2004)
- Hbase (also Dynamo and Cassandra)
  - Based on Google BigTable (2006)

MapReduce



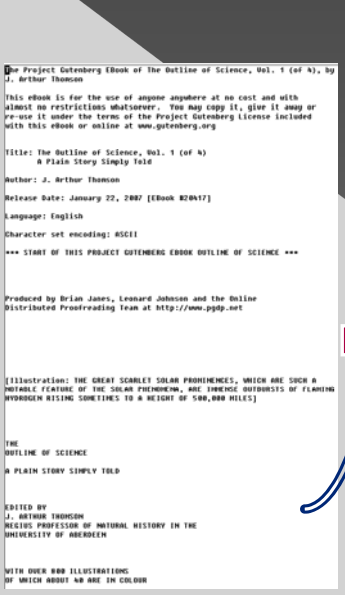
- Processes pairs [key, value]
- Hides parallelization, fault-tolerance, data distribution and load balancing

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));
}
```


WordCount execution example



Map

The	1
Project	1
Gutemberg	1
Ebook	1
of	1
The	1
Outline	1
Of	1
Science,	1
Vol.	1
1	1
(of	1
4),	1
by	1

Reduce

The	57631

["The",[1,1,1,1,...]]

Benefits of MapReduce

- Programming model simple yet expressive

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

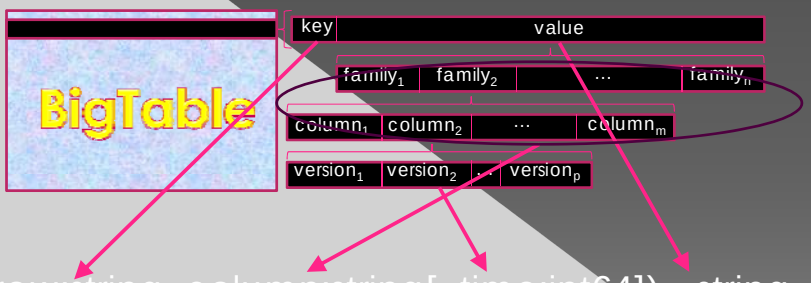
$$\texttt{reduce}(\text{key } ik, \text{vset } ivs) \mapsto [ov_1, .., ov_{r(ik,ivs)}]$$

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

MapReduce performance problems

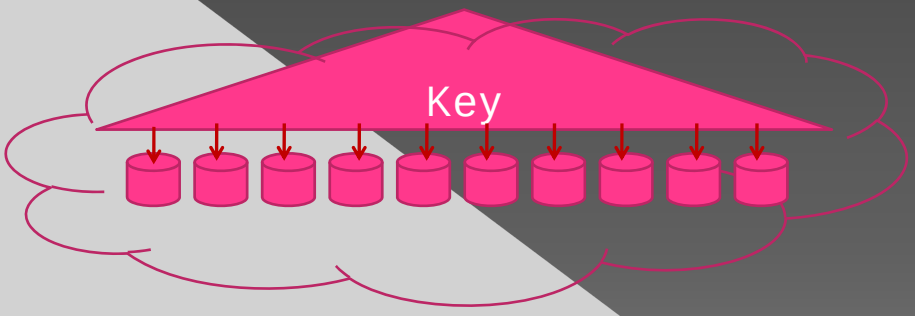
- Does not benefit from compression
- Writes intermediate results to disk
  - Reduce tasks pull intermediate data
- Defines the execution plan on the fly
  - Schedules one block at a time

BigTable

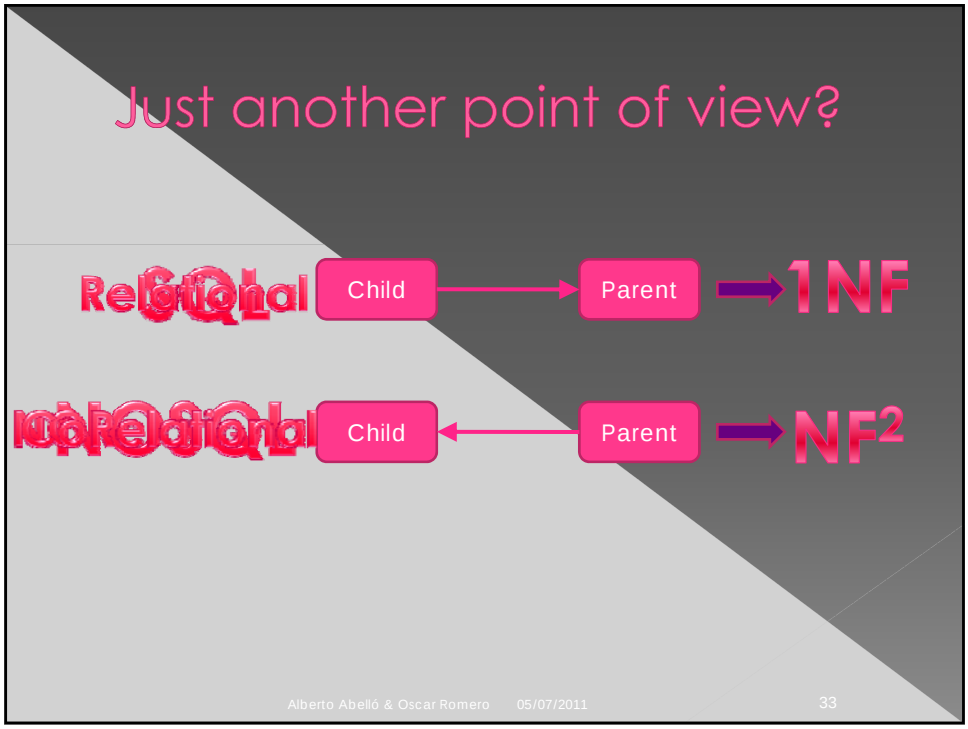


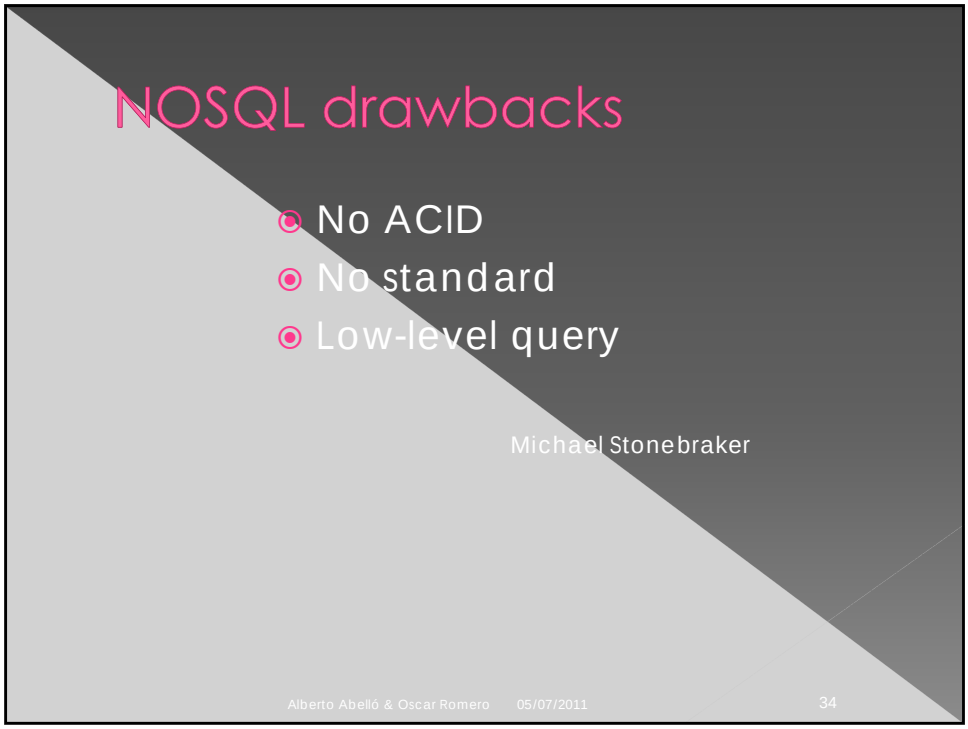
(row:string, column:string[, time:int64])→string

Physical implementation



- Supports single row transactions
- Compression per block can be enabled
- The schema determines the locality of data


Just another point of view?
Relational
Child
Parent
1NF
NoRelational
Child
Parent
NF²
Alberto Abelló & Oscar Romero
05/07/2011
33


NOSQL drawbacks
No ACID
No standard
Low-level query
Michael Stonebraker
Alberto Abelló & Oscar Romero
05/07/2011
34

Brewery or bottled beer?

Specificity

Do It Yourself
- Expensive
- Ad hoc development

Off the Shelf
- Economies of scale
- Concrete functionalities

Low Cost

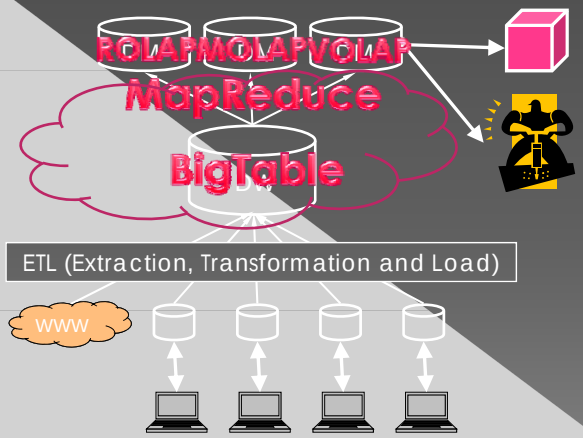
Florian Waas analogy

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

Reference Architecture



Big elephant or little elephant?

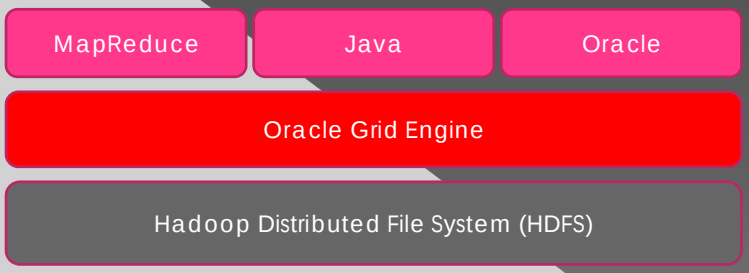
Trademark
- Expensive
- Many functionalities
- Mature

Huge

Open source
- Free
- Simple functionalities
- Young

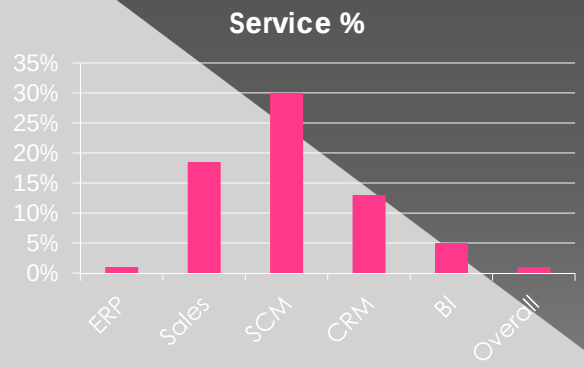
Adaptative

Friends or foes



Software as a Service
SaaS

Gartner's market analysis



Gartner's considerations

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

Business processes as a Service
BaaS

SaaS vs BaaS

- Business Process as a Services are focused on providing existing **business processes through a cloud**. If there is an existing process with steps that are known it can be provided as a service within the catalog. This allows the Cloud Service Provider to automate any steps within the process while leaving the changes transparent to the Cloud Service Consumer.
- Software Services allow a Cloud Service Consumer to select a specific **software instance** that they want created without the need to be aware of where and how it will be hosted. … This allows the Cloud Service Consumer to focus on the characteristics of the application and gives the Cloud Service Provider the freedom to fulfill the request with any resources that will meet the need.

NIST (National Institute of Standards and Technology)

SOA principles

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

Paradigm

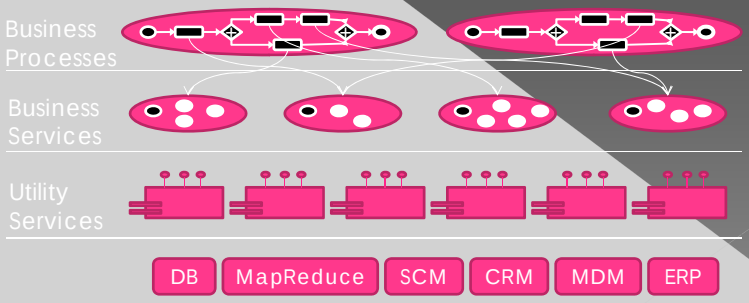
	Distributed components	SOA
Design	Functionality	Process
Designed to ...	Last	Change
Development cycle	Long	Interactive and iterative
Centered on ...	Cost	Business
Coordination	Blocks	Orchestration
Coupling	Tight	Loose (agile and adaptive)
Technologies	Homogeneous	Heterogeneous
Programming	Objects	Messages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

Service composition

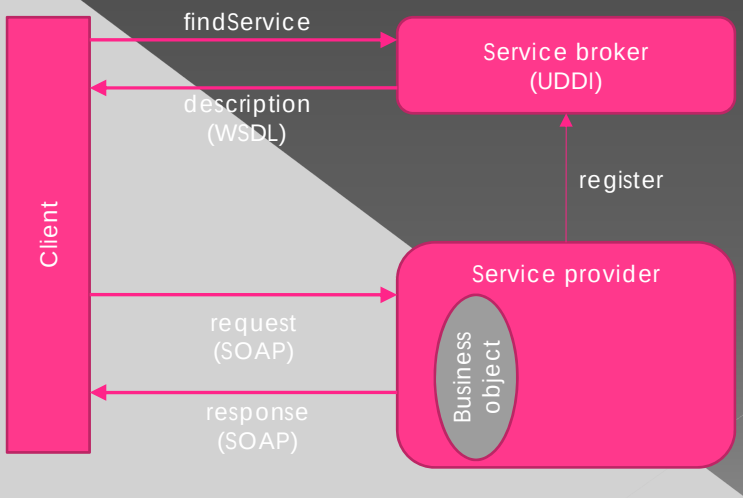
- Primitive activity
- Complex activity
  - Atomic transaction
  - Business activity
    - Orchestration
    - Choreography

Technological Challenges

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



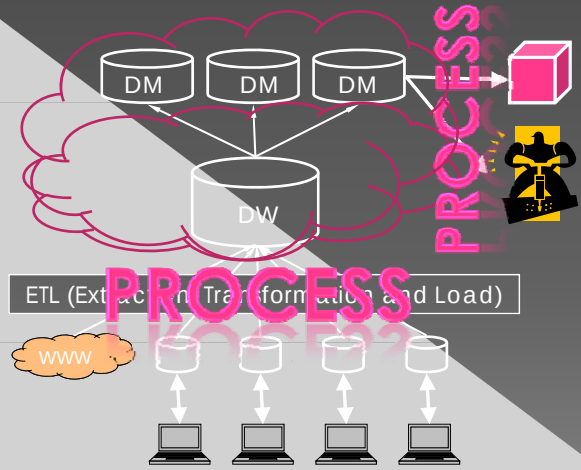
Implementation



Engineering challenges

- Service Engineering and Design
- Service Adaptation and Monitoring
- Service Quality

Reference Architecture

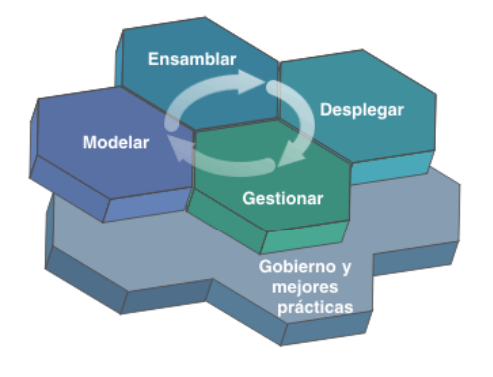


Business Intelligence on Services

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
  - Sell data to other firms

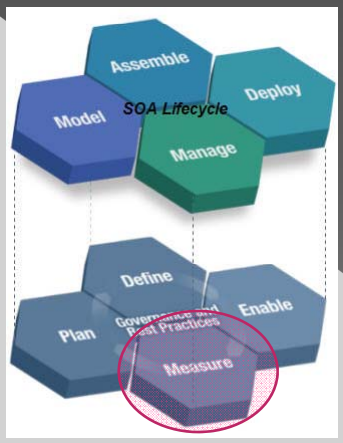
SOA lifecycle



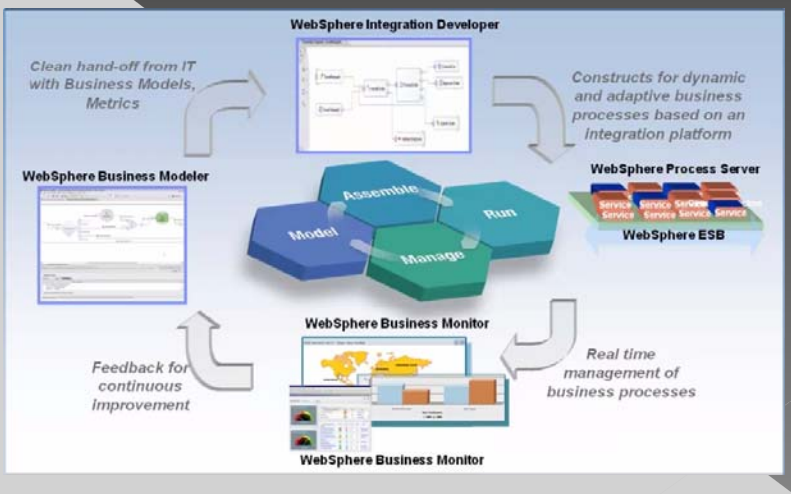
Key Performance Indicators

"You cannot control what you cannot measure"

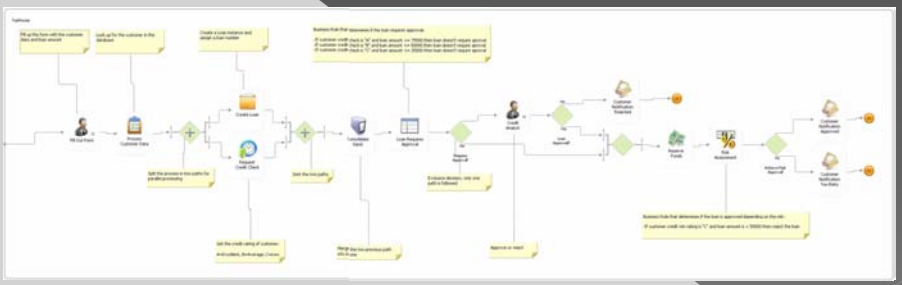
SOA Governance



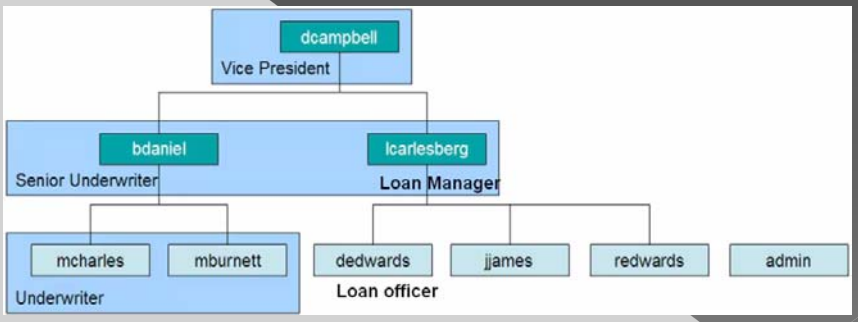
IBM WebSphere architecture



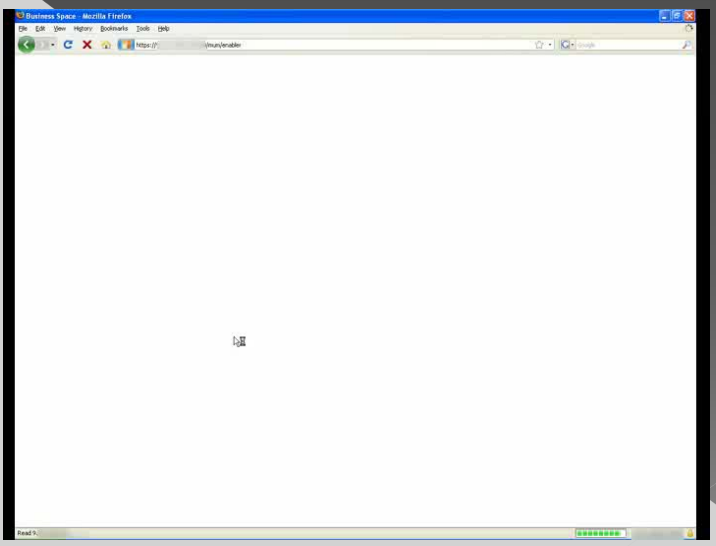
IBM e-Xtend World Bank demo



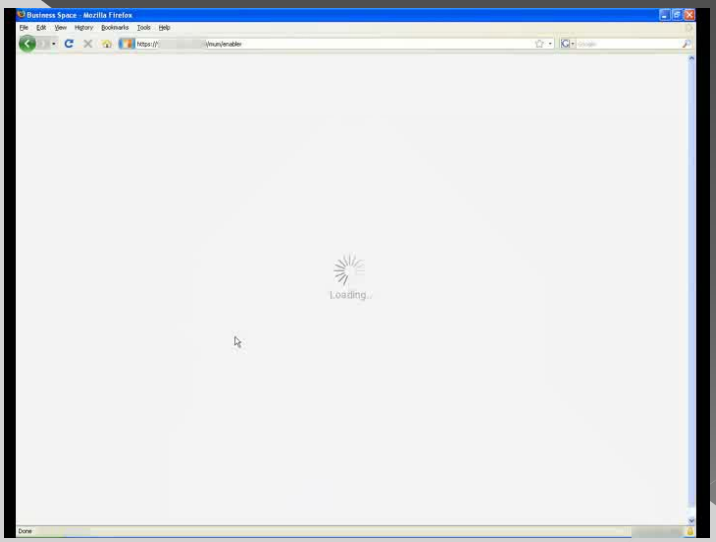
Users in e-Xtend World Bank



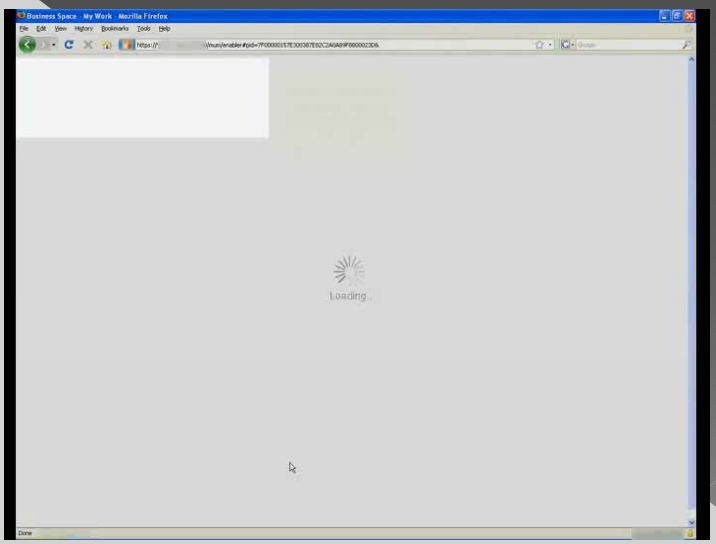
Loan manager reviews loan status



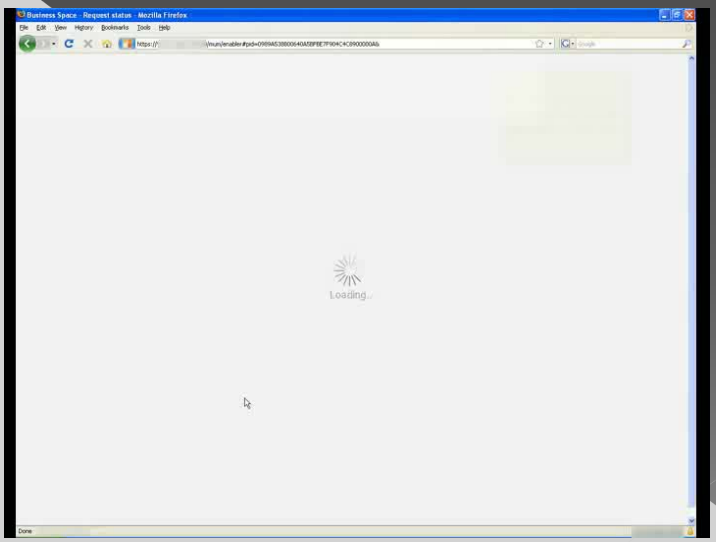
Loan officer creates a new loan request



Senior underwriter works in loan request



Loan manager checks updates



Conclusions

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

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
- Josep Carmona, Jordi Cortadella: Process Mining Meets Abstract Interpretation. ECML/PKDD (1) 2010: 184-199