中山大學
SUN YAT-SEN UNIVERSITY

Lecture 27
Cloud Computing & Semantic Web

SE-805 Web 2.0 Programming (supported by Google)

http://my.ss.sysu.edu.cn/courses/web2.0/

School of Software, Sun Yat-sen University

Outline

- **Challenges of the Current Web**
- **Cloud Computing**
  - **What is cloud computing?**
  - **Who is in this game?**
  - **The supporting technologies**
- **Semantic Web**
  - **A short history of Web**
  - **Introduction to semantic web**
  - **The supporting technologies**

Challenges of the Current Web

Size of the current Web → **24.32 billion** pages

-- WorldWideWebSize.com, 2010-06-16

- Challenge 1: The size, calling for a new generation of powerful web infrastructure.

  Cloud Computing

- Challenge 2: The content, calling for a new way of organizing the web.

  Semantic Web

What is Cloud Computing?

It's a fuzzy concept ! ☹

Widely distributed,
network based,
storage,
computation,
SaaS models.

What is Cloud Computing?

- *Bussiness Week:* any situation in which computing is done in a remote location (out in the clouds), rather than on your desktop or portable device.

- *Wikipedia:* **Cloud computing** is **Internet** ('Cloud') based development and use of computer technology ('Computing'). It is a style of **computing** where IT-related capabilities are provided "**as a service**", allowing users to access technology-enabled services from the Internet **without knowledge of**, expertise with, or control over the technology **infrastructure** that supports them.

What is Cloud Computing?

- Key concepts
  - Changes the economics of Computing from being **a Capital investment to Utilities** (You buy electricity you don't buy generators)
  - Changes the way software is developed – **Hardware provisioning , Deployment** and **Scaling** now part of **developer lifecycle as a Program / script** as compared to a Purchase order
  - Automates a whole bunch of **infrastructure** related tasks and activities leading **efficiencies** and **cost savings**

Why We Want It?

Eventually users can focus on what the service delivers rather than how they are implemented or hosted

- **Cloud = Less Investment**
  (not own data center, hardware; use outside provider of servers, storage, and bandwidth)
- **Cloud = Scale**
  (tens of thousands of server computers)
- **Cloud = Flexible and Efficiency**

The Time is Ready...

Key factors for the popularity of cloud computing

- Commoditization and standardization of technologies
- Virtualization
- Service-oriented computing architects
- The growth of the Internet and bandwidth

How Big is the Market?

- "There's a whole industry emerging,“
  says Marc Benioff, Salesforce.com's CEO

- Merrill Lynch: 2012 = the annual global market for cloud computing will surge to $95 billion

Who is in this Game?

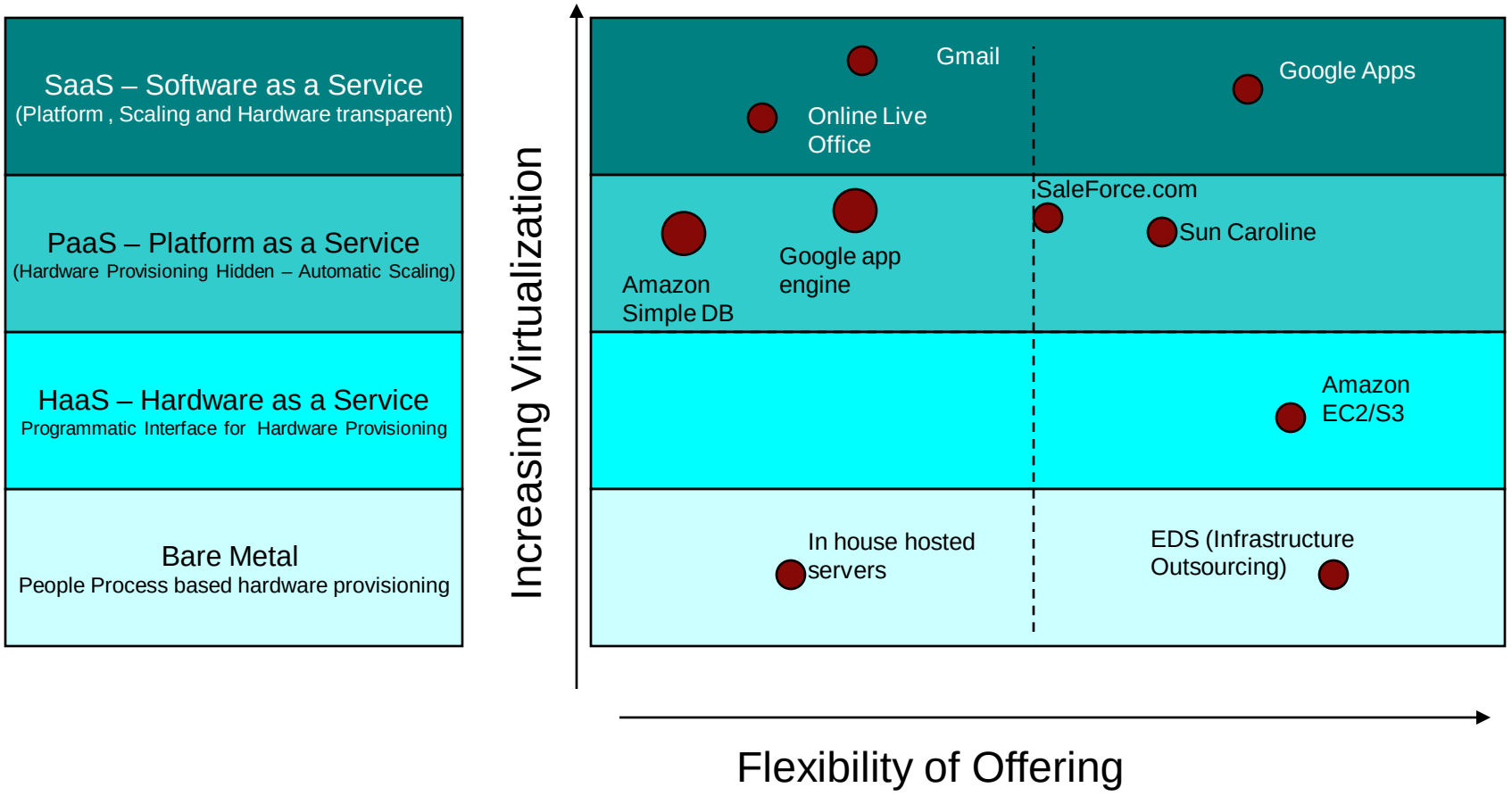
- Saleforce.com
  - Startup since 1999, CRM Software as a service
  - Now: 40K customers, 2,500+ full-time employees, 247 Million quarter revenue
  - Just push out new “platform as service”
- Amazon: Offer EC2/S3 since 2002, most flexible and popular service so far
- Google
  - One of earliest that explores cloud computing architects internally
  - Offer Google App Engine since april 2008.
- A long list: Microsoft, IBM and Sun……

Hardware as a Service: Virtualization

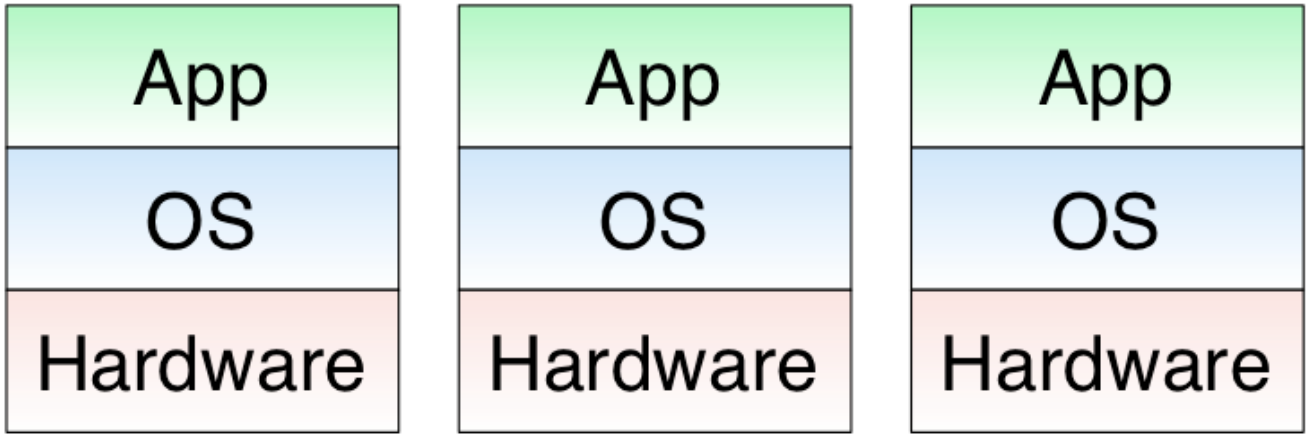
Run multiple virtual computers on one physical box

- Lots of way to do it
  - Xen
  - VMWare
  - Parallels
  - Amazon AMI
  - Microsoft Hype V

Classification of Cloud Computing Services

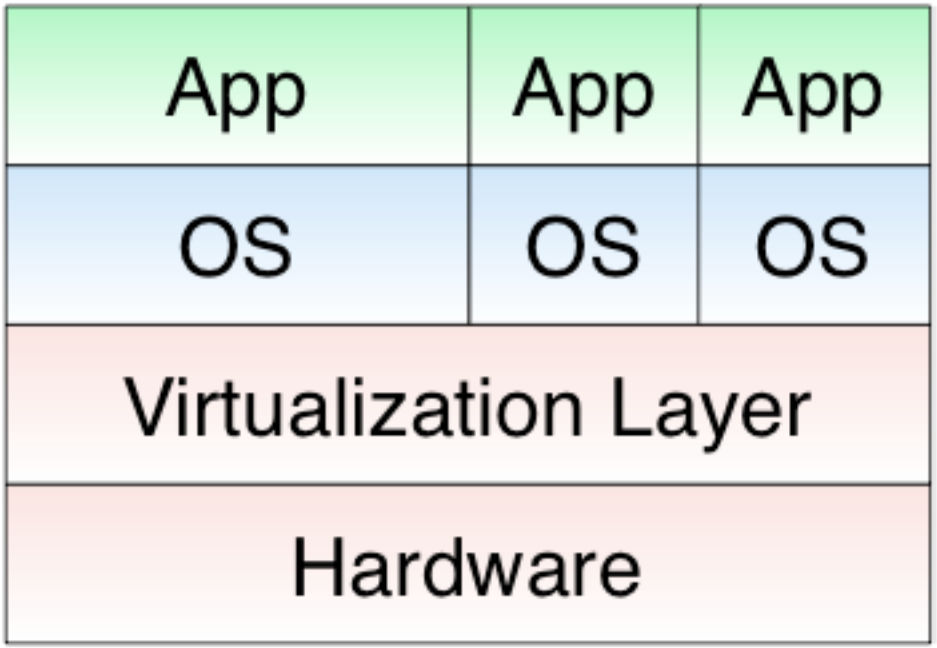


Virtualization – Benefit



- 5 to 15 % utilization only

Virtualization – Benefit



High utilization and standardization

Platform as a Service

- Question: Given 100 computers, how do you use them to compute the frequency of words in 1T text files?

  To utilize the underlying computing power, you need

  a framework for storing and processing large scale of data

  *Storage: Distributed File/Database system

  * Processing: Map-reduce

A Brief History in this Area

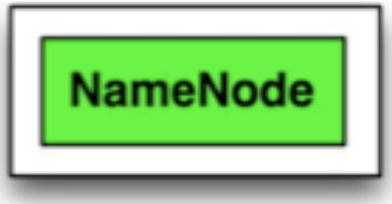
- 2003, First MapReduce Lib developed in Google
- 2003, 2004, and 2006, Google published papers on GFS/MapReduce/BigTable.
- 2005- Now, Hadoop project (An open source implementation of GFS/BigTable/MapReduce)
  - Yahoo use Hadoop for their underlying search service

HDFS (Hadoop Distributed File Systems)

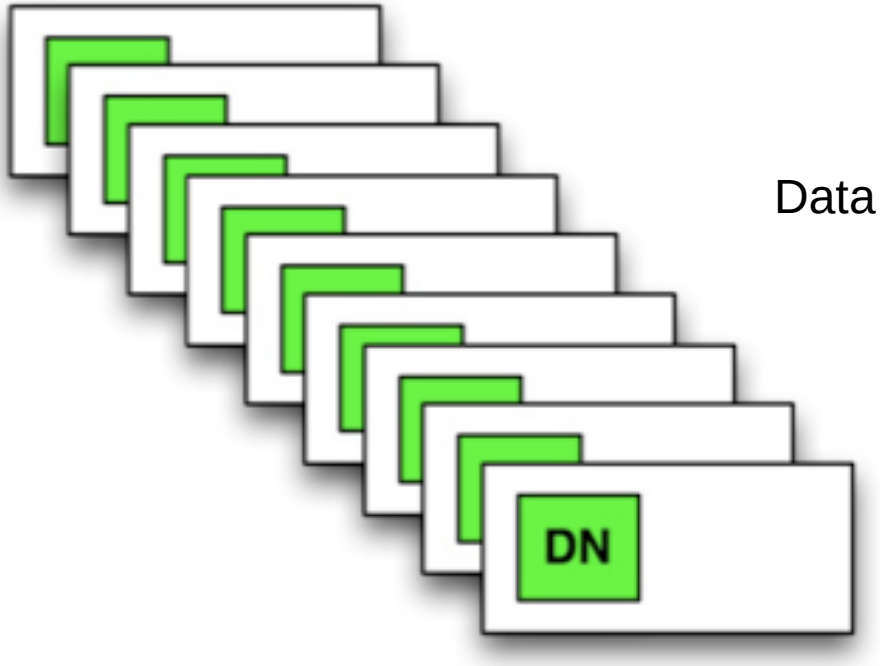
Hbase (Hadoop Distributed Database)

MapReduce Framework

HDFS - Design



Name node manages meta data
and block placement



Data node handles block storage

HDFS - Features

- Fault-tolerant
  - Default is 3x replication
  - Dynamic control of replication factor
- Load balancing
  - Balancer application to rebalance cluster in the background

HBase

- Modelled on Google's Bigtable
- Row/Column store
- Billions of rows * millions of columns
- Column-oriented – nulls are free
- Untyped – stores bytes.

Hbase – Data Model

Row	Timestamp	Column family animal:		Column family repairs:
		animal:type	animal:size	repairs:cost
enclosure1	t2	zebra		1000 EUR
	t1	lion	big	
enclosure2	...	...	...	...

Data schema

Column family animal:

(enclosure1, t2, animal:type)	zebra
(enclosure1, t1, animal:size)	big
(enclosure1, t1, animal:type)	lion

Disk storage

Hbase – Design & Features

- Design similar to HDFS
  - Name node → Master server
  - Data node →Region server, organized in columns and cells
- Features
  - Fault tolerant and auto load balancing
  - Fast access to cells, and fast scan over the ranges of rows.
  - More flexible schema than traditional database.

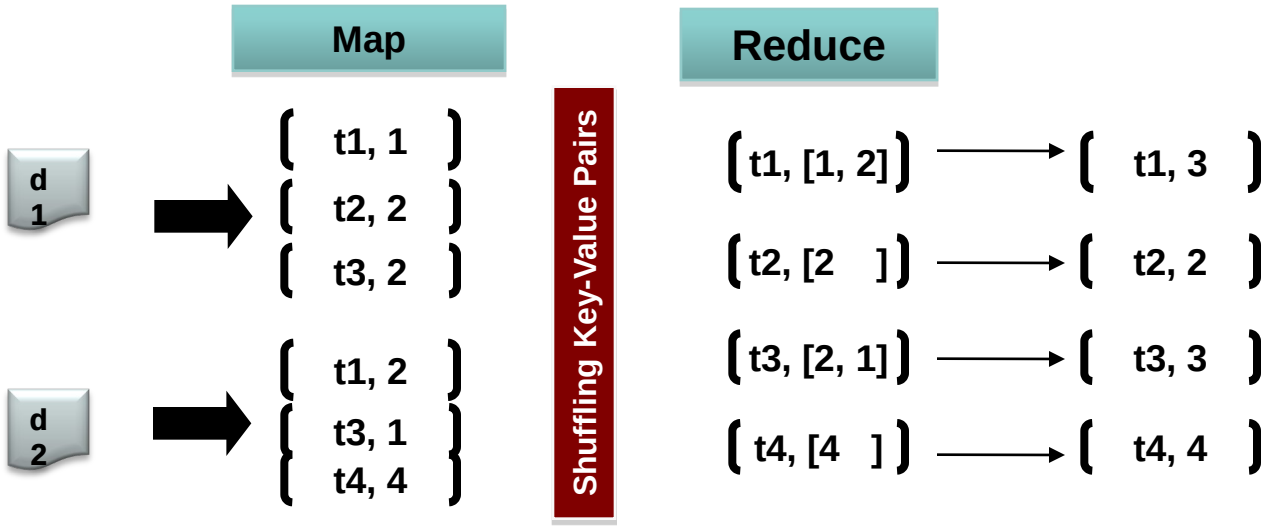
MapReduce

Break up a problem,
allocate to many machines,
reassemble for use.

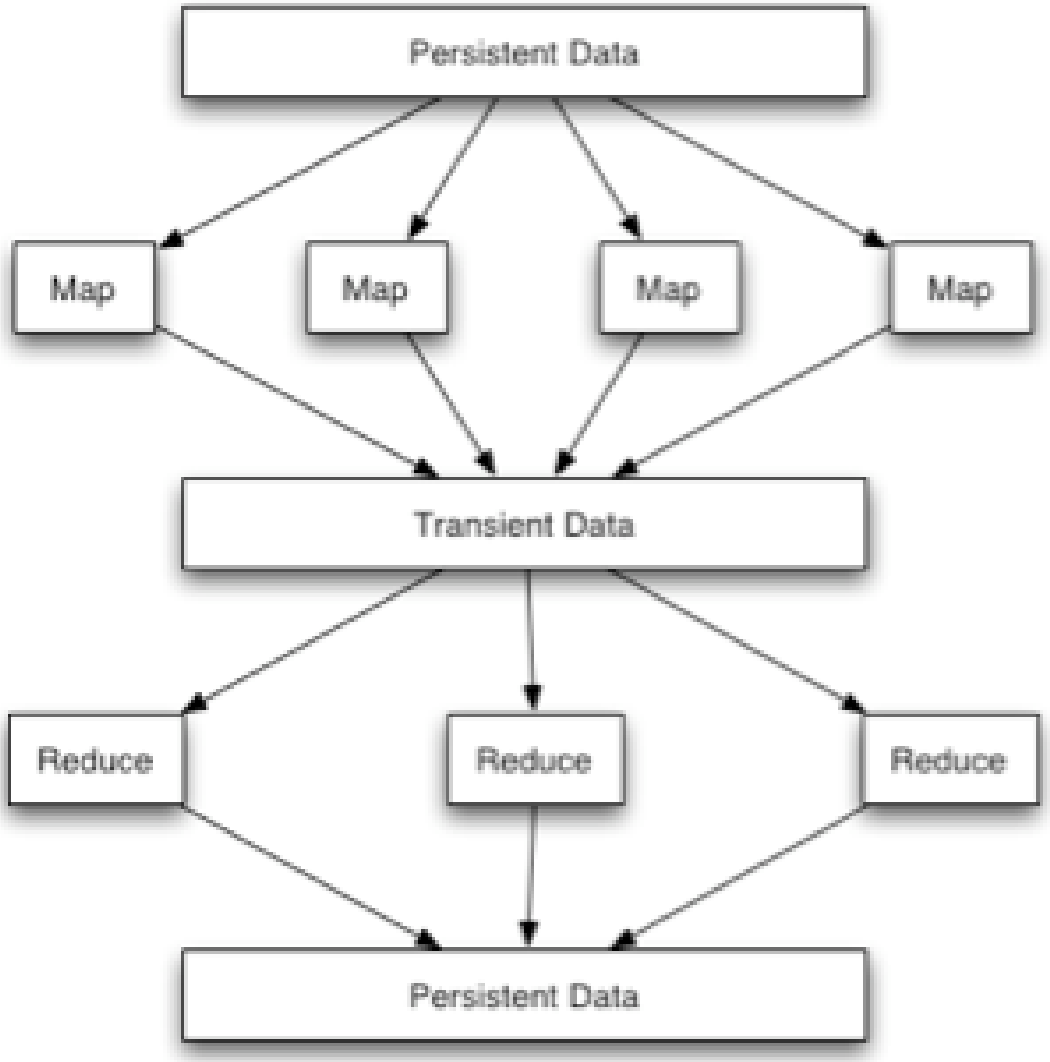
Simple programming model: key-value pairs

Map:	(K1, V1)	→ list(K2, V2)
Reduce:	(K2, list(V2))	→ list(K3, V3)

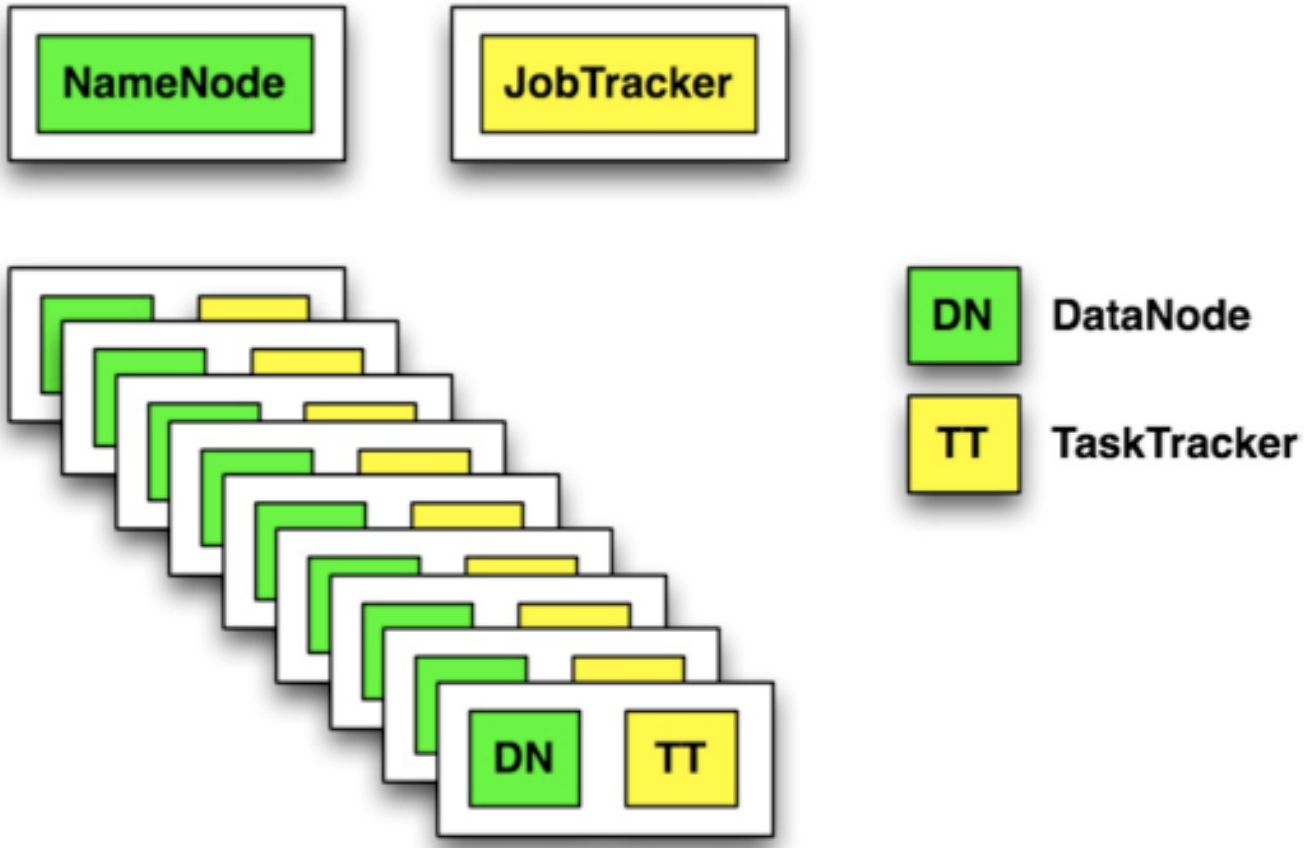
MapReduce – a "hello world" example



MapReduce – Logical Data Flow



MapReduce - Design



Moving computation is cheaper than moving data

Outline

- Challenges of the Current Web
- Cloud Computing
  - What is cloud computing?
  - Who is in this game?
  - The supporting technologies
- **Semantic Web**
  - **A short history of Web**
  - **Introduction to semantic web**
  - **The supporting technologies**

A Short Web History

Mostly flat information

Some databases but content is very functional

Little engagement or interactivity

A Short Web History

Greater interactivity

Growth of social media / social networking

Online communities created / social capital

A Short Web History

Joining up of information

Data portability

Browsers and search engines become more ‘intelligent’

What is Semantic Web?

"The **Semantic Web** is an extension of the current web in which information is given well-defined **meaning**, better enabling computers and people to **work in cooperation**."

[Berners-Lee et al., 2001]

A side note: the same guy invented the Web in 1989.

An Example

Semantic Web allows to recognize *people, places, events, products movies*, etc, and it can understand relationships between things

The Solution

- Top-down approach
  - Information analysis, web scraping, natural language processing
  - Expensive, and need human intervention, hard to maintain!
- Bottom-up approach
  - Embedding *semantical annotations* into the data
  - Available options:
    - RDFa
    - Microformats

RDF

- RDF = Resource Description Framework
  - A W3C standard for describing resources in the Web
  - RDF identifies things using URI (Uniform Resource Identifiers)
  - RDF uses simple statements (Triples) to describe things

    Things – Property – Value, or

    Subject – Predicate – Object

RDF Graph Representation

RDF/XML

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:contact="http://www.w3.org/2000/10/swap/pim/contact#">
  <contact:Person rdf:about="http://www.w3.org/People/EM/contact#me">
    <contact:fullName>Eric Miller</contact:fullName>
    <contact:mailbox rdf:resource="mailto:em@w3.org"/>
    <contact:personalTitle>Dr.</contact:personalTitle>
  </contact:Person>
</rdf:RDF>
```

RDFa & SPARQL

- RDFa = RDF in attributes
  - It provides a set of XHTML attributes (Dublin Core vocabulary) that express RDF data
- SPARQL = a query language for RDF data

  sample query: friends of Alice who created items who title contains ‘bob’.

```
<div about="/alice/posts/trouble_with_bob">
    <h2 property="dc:title">The trouble with Bob</h2>
    Bob takes much better photos than I do:
    <div about="http://example.com/bob/photos/sunset.jpg">
        <img src="http://example.com/bob/photos/sunset.jpg" />
        <span property="dc:title">Beautiful Sunset</span>
        by <span property="dc:creator">Bob</span>.
    </div>
</div>
```

Microformats

- Simple conventions for embedding semantics in HTML
- Designed for human first, and machine second
- No new tags, built upon existing standards: vCard, iCalendar, Atom, etcs

vCard Example

- Represent people, companies, places, and organizations

```
<div id="hcard-Hatem-Mahmoud" class="vcard">
    <span class="fn">Hatem Mahmoud</span>
    <a class="email" href="mailto:hatem@expressionlab.com">
        hatem@expressionlab.com
    </a>
    <div class="adr">
        <span class="locality">Alexandria</span>,
        <span class="postal-code">21523</span>
    </div>
    <div class="tel">0123456789</div>
</div>
```

hCalendar Example

- Represent calendar events

```
<div class="vcalendar">
    <p class="vevent">
        My event:
        <span class="summary">Web 3.0</span>
        (<span class="location">Alexandria</span>), Sunday 12th Jul,
        <abbr class="dtstart" title="2009-07-12T15:45">2:45</abbr> -
        <abbr class="dtend" title="2009-07-12T16:15">4:15pm</abbr>
    </p>
</div>
```

hReview Example

- Represent reviews of products, services, businesses, events etc

```
<div class="hreview">
    <div class="item">
        <h2 class="fn">Aliens</h2>
        <p class="summary">
            My favorite movie.
        </p>
        <p>
            <span class="rating">9</span>/
            <span class="best">10</span>
        </p>
    </div>
</div>
```

Applications

- YahooTech uses hReview for product reviews
- Linkedin uses hResumes for resumes
- YahooUpcoming uses hCal for events

Egyptian Art at The Metropolitan Museum of Art (Friday July 10, 2009) - Upcoming - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://upcoming.yahoo.com/event/2572014/

Contacts (1) Events (1) Locations (1) Tagspaces (1) Bookmarks Resources (5) Options

Egyptian Art

Egyptian Art
Export All

Export Event
Add to 30 Boxes
Bookmark with Firefox
Add to Google Calendar
Add to Yahoo! Calendar

Welcome, Guest

Web Search

upcoming
EVENTS & THINGS TO DO

search for jazz, Lakers, etc...

New York, New York 10028

SEARCH EVENTS

HOME MY EVENTS FRIENDS MORE + ADD AN EVENT this is YAHOO!

DETAILS & PHOTOS MAPS & WHAT'S NEARBY

Egyptian Art

Buy Tickets

Friday July 10, 2009 from 9:30am - 9:00pm

The Metropolitan Museum of Art
1000 Fifth Avenue
New York, New York 10028 Get Directions

Have a photo? Add it here

This exhibit features statuary, reliefs, stelae, funerary objects, jewelry, daily implements and architecture from prehistoric Egypt through the Old, Middle and New Kingdoms to the Roman period (4th century C.E.).

Ticket Info: Included in admission

Buy Tickets

Category: Performing/Visual Arts

Add to calendar Invite Friends Print

Discovered by Upcoming Robot on May 7, 2009 Report a problem

Additional Local Dates

Date & Time	Venue
Sat, Jul 11 at 9:30am	The Metropolitan Museum of Art
Sun, Jul 12 at 9:30am	The Metropolitan Museum of Art

Let the community know:

I'm Going or I'm Interested

Also at The Metropolitan Museum of Art

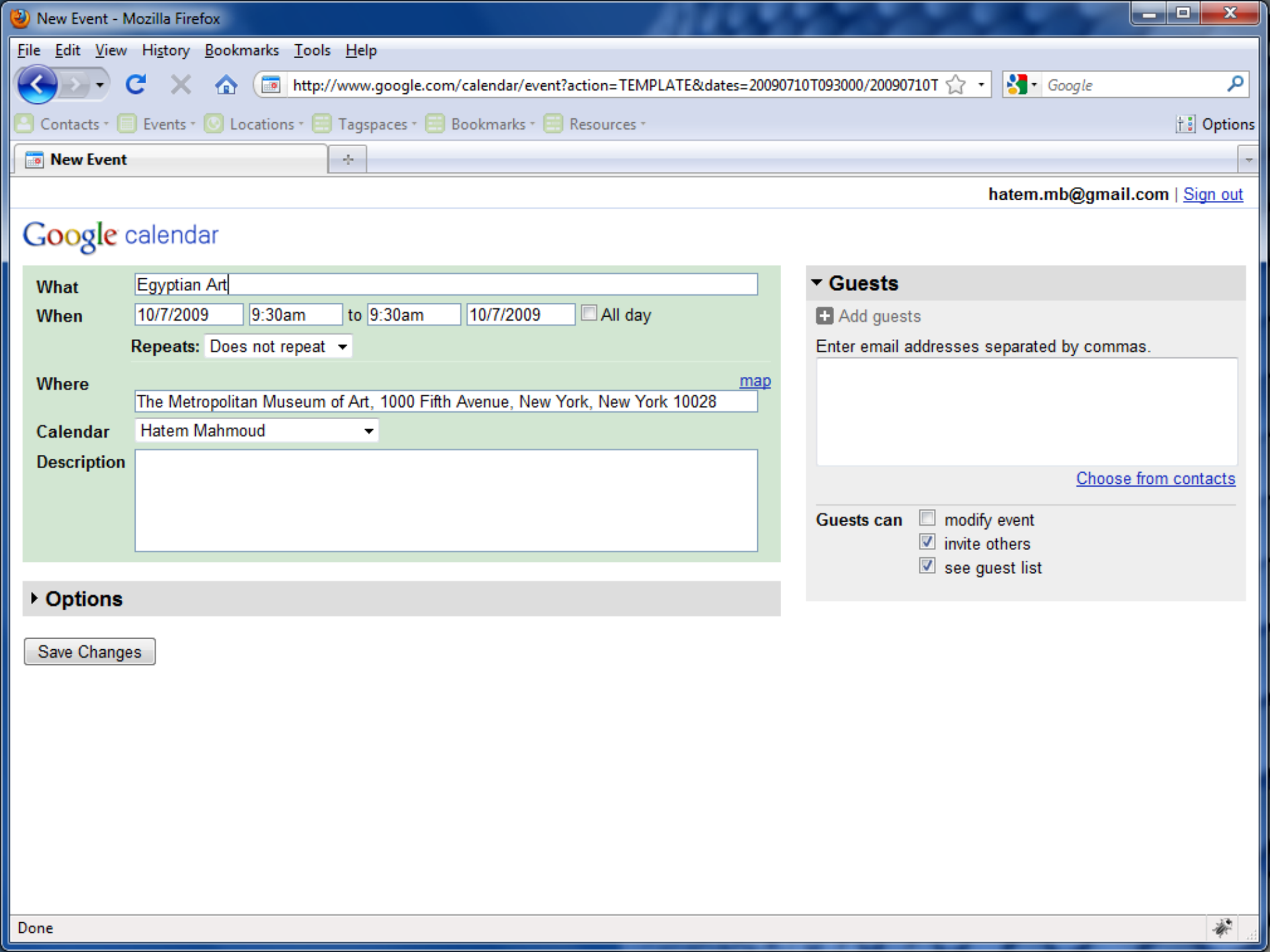
Fri, Jul 10	Gallery for the Art of Native North America
Fri, Jul 10	Photographs
Fri, Jul 10	Living Line: Selected Indian Drawings From the Sub...
Fri, Jul 10	Roxy Paine on the Roof: Maelstrom
Fri, Jul 10	Modern Art

See all 1938 events

Tags

Theater

Done


New Event - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.google.com/calendar/event?action=TEMPLATE&dates=20090710T093000/20090710T
Google
Contacts Events Locations Tagspaces Bookmarks Resources Options
New Event
hatem.mb@gmail.com | Sign out
Google calendar
What Egyptian Art
When 10/7/2009 9:30am to 9:30am 10/7/2009 All day
Repeats: Does not repeat
Where map
The Metropolitan Museum of Art, 1000 Fifth Avenue, New York, New York 10028
Calendar Hatem Mahmoud
Description
Guests
Add guests
Enter email addresses separated by commas.
Choose from contacts
Guests can modify event
invite others
see guest list
Options
Save Changes
Done

Egyptian Art at The Metropolitan Museum of Art (Friday July 10, 2009) - Upcoming - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://upcoming.yahoo.com/event/2572014/

Google

Contacts (1) Events (1) Locations (1) Tagspaces (1) Bookmarks Resources (5) Options

Egyptian Art at The Metrop

Egyptian Art

Export All KML

Export KML

Find with Google Maps

Find with MapQuest

Find with Yahoo! Maps

Welcome, Guest Sign In | Sign Up

Web Search

upcoming
EVENTS & THINGS TO DO

SEARCH EVENTS search for jazz, Lakers, etc... New York, New York 10028 SEARCH EVENTS

HOME MY EVENTS FRIENDS MORE + ADD AN EVENT this is YAHOO!

DETAILS & PHOTOS MAPS & WHAT'S NEARBY

Egyptian Art

Buy Tickets

Friday July 10, 2009 from 9:30am - 9:00pm

The Metropolitan Museum of Art
1000 Fifth Avenue
New York, New York 10028 Get Directions

Have a photo? Add it here

This exhibit features statuary, reliefs, stelae, funerary objects, jewelry, daily implements and architecture from prehistoric Egypt through the Old, Middle and New Kingdoms to the Roman period (4th century C.E.).

Ticket Info: Included in admission

Buy Tickets

Category: Performing/Visual Arts

Add to calendar Invite Friends Print

Discovered by Upcoming Robot on May 7, 2009 Report a problem

Additional Local Dates

Date & Time	Venue
Sat, Jul 11 at 9:30am	The Metropolitan Museum of Art
Sun, Jul 12 at 9:30am	The Metropolitan Museum of Art

Let the community know:

I'm Going or I'm Interested

Also at The Metropolitan Museum of Art

Fri, Jul 10	Gallery for the Art of Native North America
Fri, Jul 10	Photographs
Fri, Jul 10	Living Line: Selected Indian Drawings From the Sub...
Fri, Jul 10	Roxy Paine on the Roof: Maelstrom
Fri, Jul 10	Modern Art

See all 1938 events

Tags

Theater

Done

Summary

- **Challenges of the Current Web**
  - The size → a new generation of powerful web infrastructure.
  - The content → a new way of organizing the web.
- **Cloud Computing**
  - What is cloud computing?
  - Who is in this game?
  - The supporting technologies
    - Virtualization/MapReduce/Bigtable
- **Semantic Web**
  - A short history of Web
  - Introduction to semantic web
  - The supporting technologies
    - RDF
    - Microformat

Further Readings

- Academic Papers
  - "Map-Reduce-Merge: Simplified Relational Data Processing on Large Clusters" — paper by Hung-Chih Yang, Ali Dasdan, Ruey-Lung Hsiao, and D. Stott Parker; from Yahoo and UCLA; published in Proc. of ACM SIGMOD, pp. 1029--1040, 2007
  - MapReduce: Simplified Data Processing on Large Clusters, Jeffrey Dean, Sanjay Ghemawat, *OSDI'04: Sixth Symposium on Operating System Design and Implementation*, 2004, pp. 137-150.
  - Above the Clouds: A Berkeley View of Cloud Computing, Micheal Armbrust etc. from U.C. Berkley.
  - Google's Tutorial on Cluster Computing and MapReduce
- Reference Books
  - Hadoop: The Definitive Guide, by Tom White, published by O'Reilly, June, 2009. ISBN: 978-0596521974.
  - Pro Hadoop, by Jason Venner, published by APress, June, 2009. ISBN：978-1430219422.
- Online Materials
  - Cloud computing A-Z: a complete list
  - Hadoop Summit and Data-Intensive Computing Symposium 2008

Thank you &
Welcome to Cloud Era!