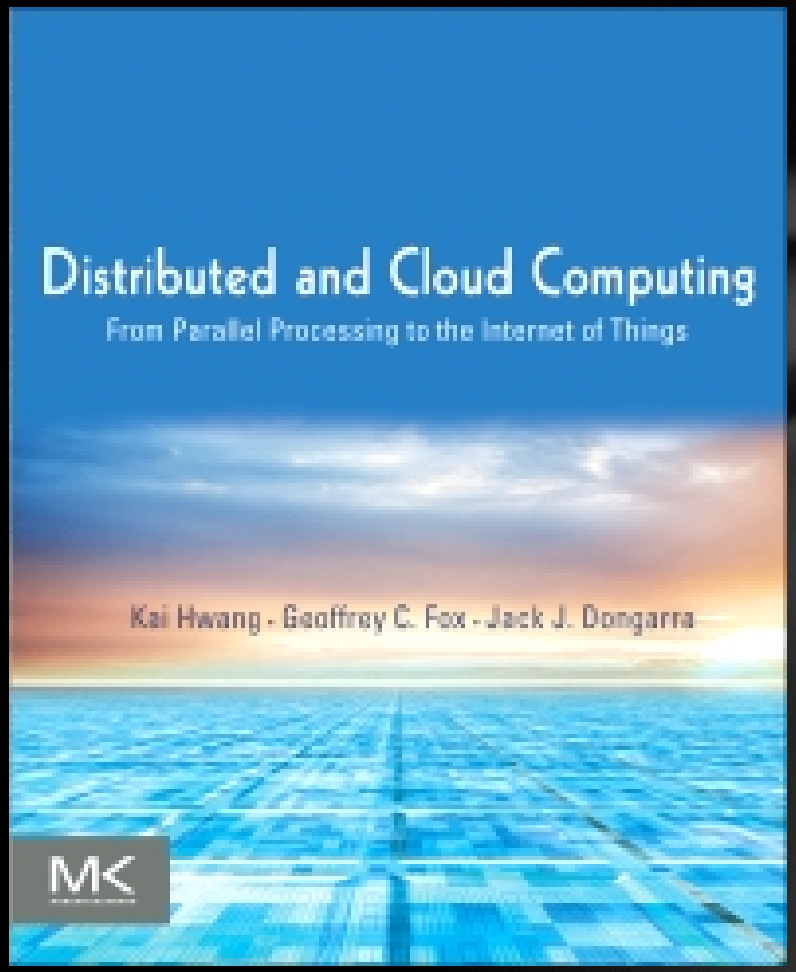
Psi5120 Tópicos de computação em nuvem

6a. Aula Parte1

20. Período de 2013

Livro texto


Distributed and Cloud Computing
From Parallel Processing to the Internet of Things
Kai Hwang · Geoffrey C. Fox · Jack J. Dongarra
MK

Cloud Applications

- **Scientific / Tech Applications**
- **Business Applications**
- **Consumer / Social Applications**

Science and Technical Applications

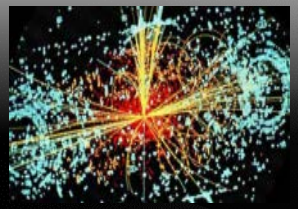
Business Applications

Consumer / Social Applications

The Data Explosion Era

Experiments

Simulations



Archives

Literature

Cornell University Library

arXiv.org

Consumer

Petabytes
Doubling &
Doubling

The Challenge
Enable Discovery

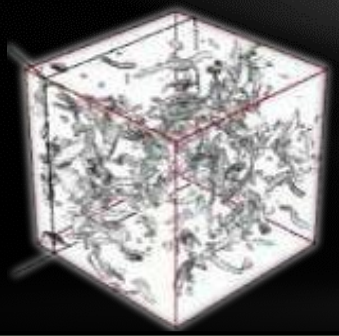
Deliver the capability to mine, search and analyze this data in near real time

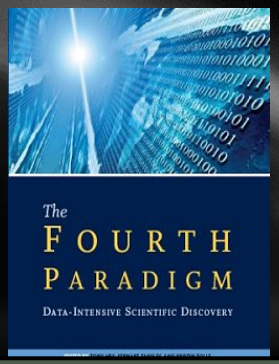
The Response

Science itself is evolving

The Changing Nature of Research

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - \mathrm{K}\frac{c^2}{a^2}$$



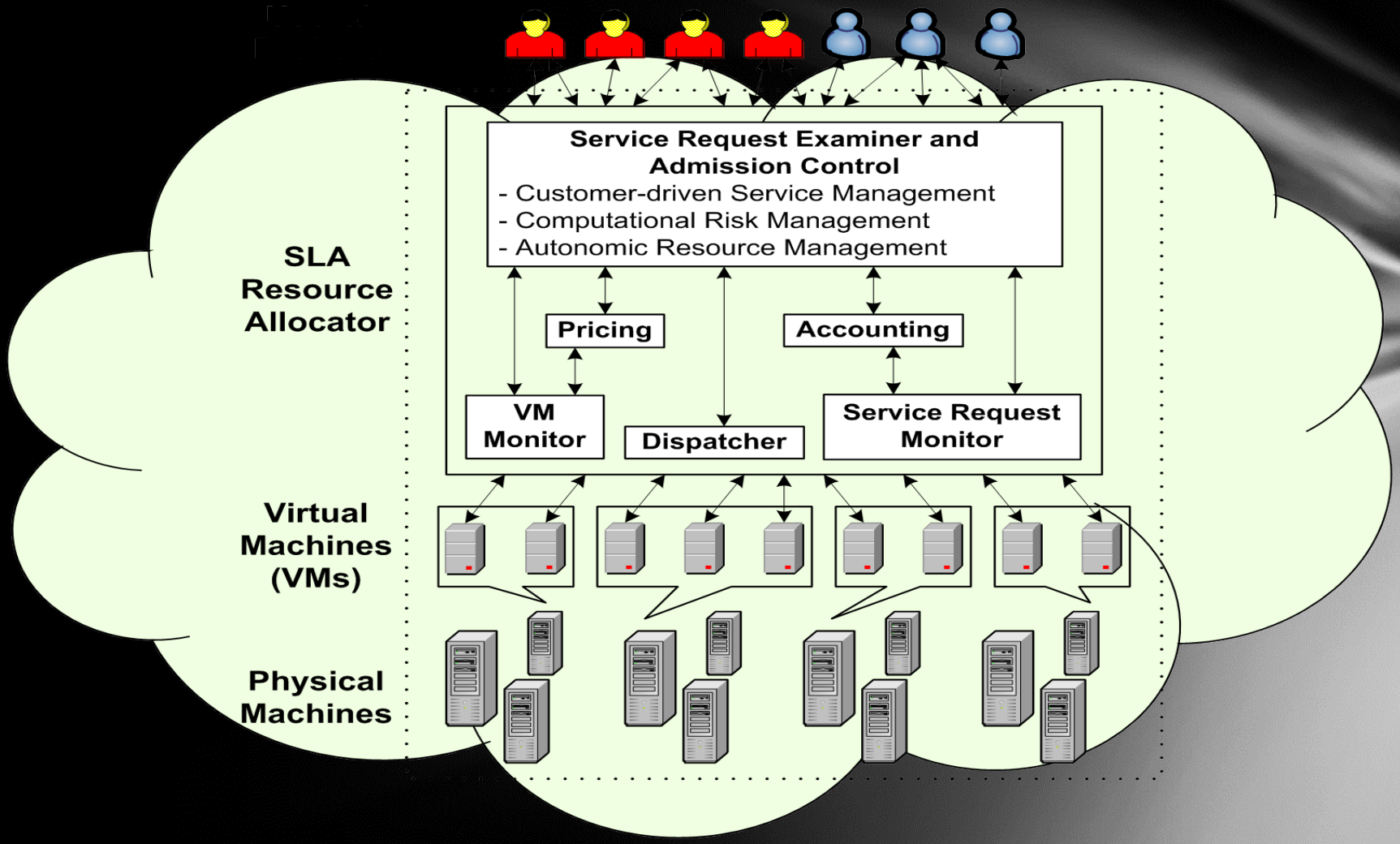


Experimental	Theoretical	Computational	The Fourth Paradigm
Thousand years ago	**Last few hundred years**	**Last few decades**	**Today and the Future**
Description of natural phenomena	***Newton's laws, Maxwell's equations…***	***Simulation of complex phenomena***	***Unify theory, experiment and simulation with large multidisciplinary Data***
			Using data exploration and data mining (from instruments, sensors, humans…)
			Distributed Communities

Cloud Ecosystem Requirements:

- At the system level, the cloud ecosystem include the cloud platform and infrastructure, resource management policies, etc.
- At the service level, the SLAs, globalized standards, reputation system, billing and accounting system, cloud business models, etc.
- At the user (client) level, Application programming interfaces (APIs), cloud programming environment, Quality of Service control, etc.

Ecosystem for Market-Oriented Clouds



...ket-Oriented Cloud Computing:
...r Delivery IT Services as
... HPCC, Sept. 25-27, 2008 [4])

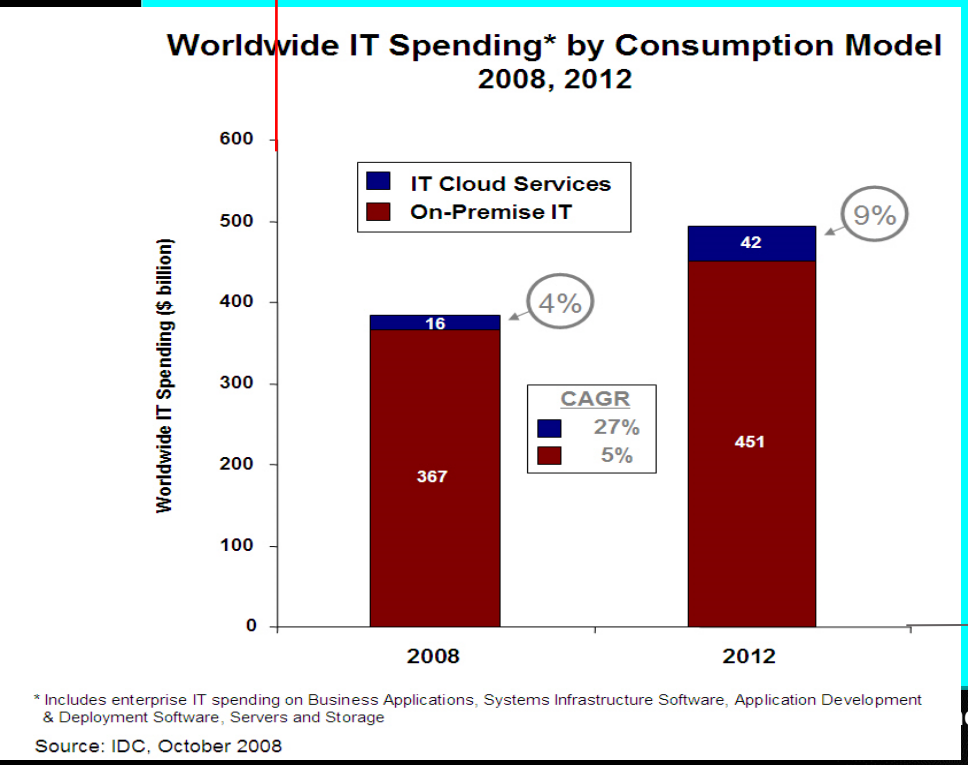
Cloud Software Packages and Features

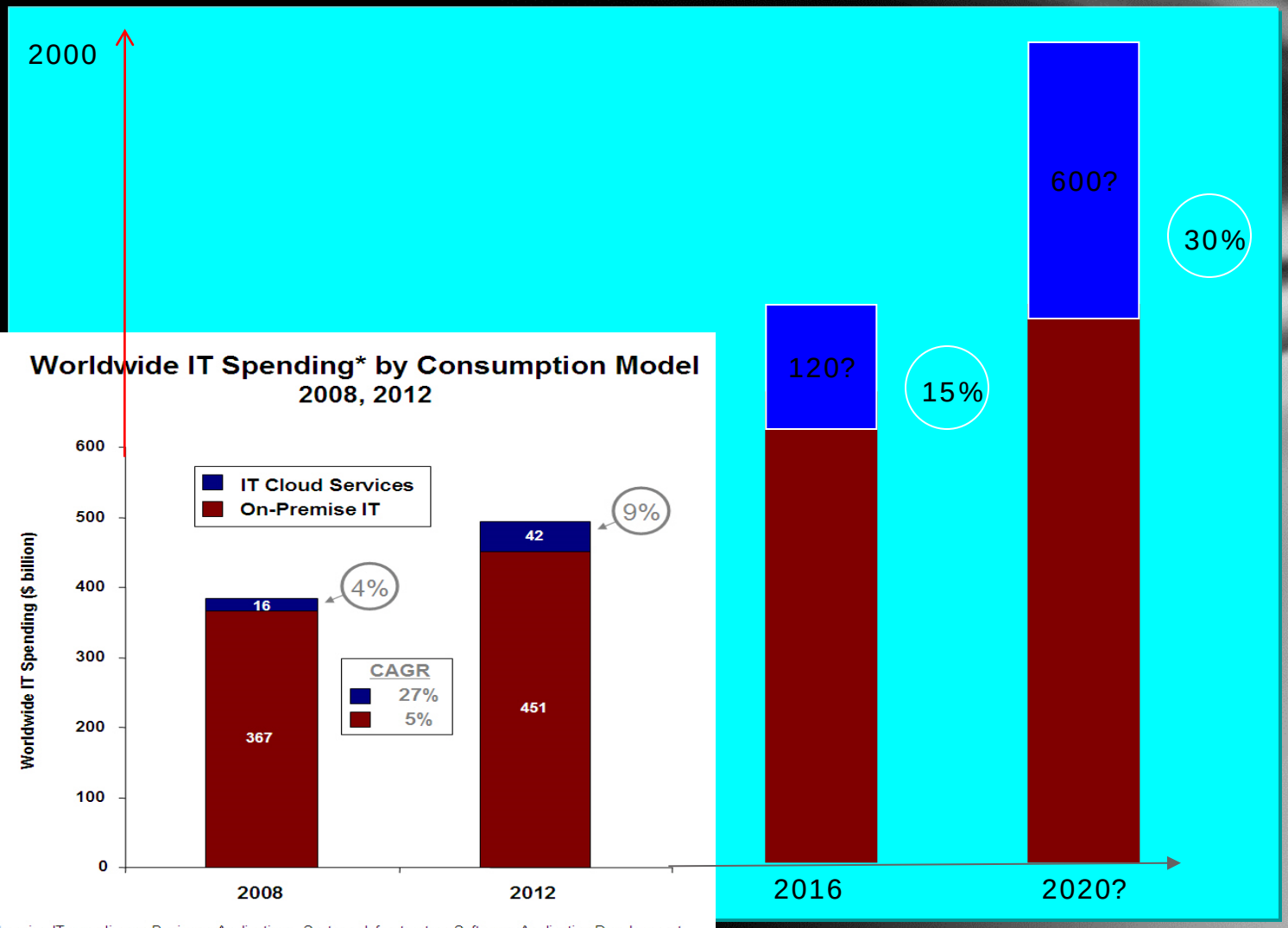
Software	Cloud Type	License(s)	Language	Linux/ Windows	EC2/S3	Xen/KVM/ VMWare	VirtualBox	OCCI/ vCloud
Fluid Operations	Iaas, Paas, LaaS, SaaS, TaaS, DaaS, BaaS	Proprietary	Java, C	Yes/Yes	Yes/No	Yes/Yes/ yes	?	No /Yes
ApplScale	Paas	BSD	Python, Ruby, Go	? / ?	Yes/ Yes	Yes/Yes/ yes	Yes	? / ?
Cloud Foundry	PaaS	Apache	Ruby, C	Yes/No	Yes/No	Yes/Yes/ yes	Yes	No /Yes
Cloud.com	IaaS	Proprietary, GPLv3	Java, C	? / ?	? / ?	Yes/Yes/ yes	?	? / ?
Eucalyptus	IaaS	Proprietary, GPLv3	Java, C	? / ?	Yes/ Yes	Yes/Yes/ yes	?	? / ?
Nimbus	IaaS	Apache	Java, Python	? / ?	Yes/No	Yes/ Yes/?	?	? / ?
OpenNebua	IaaS	Apache	C++,C,Ruby, Java, lex, yacc, Shellscript	Yes/ ?	Yes/ ?	Yes/ Yes/?	?	Yes/Yes
OpenStack	IaaS	Apache	Python	Yes/ ?	Yes/ Yes	Yes/ Yes/?	?	? / ?

Source: http://en.wikipedia.org/wiki/Cloud_computing_comparison (read 02/02/2012)

Cloud Bisiness Potential:

A trillion $ business/year by 2020?

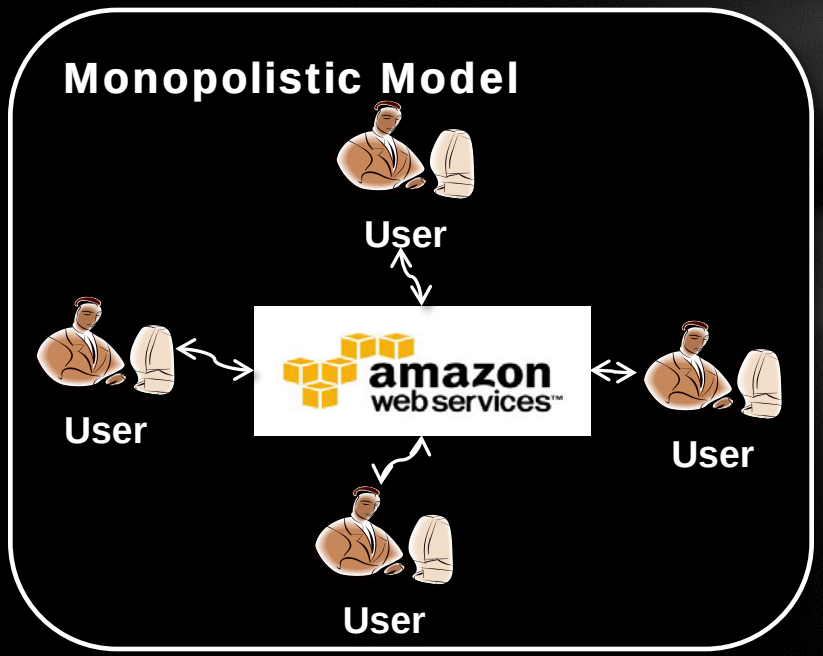


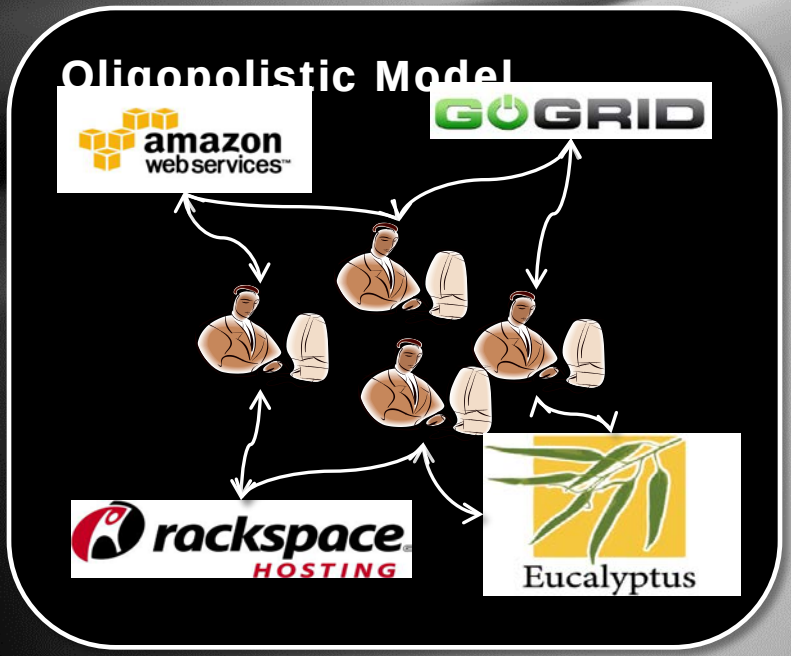


Inter-Cloud Business Models

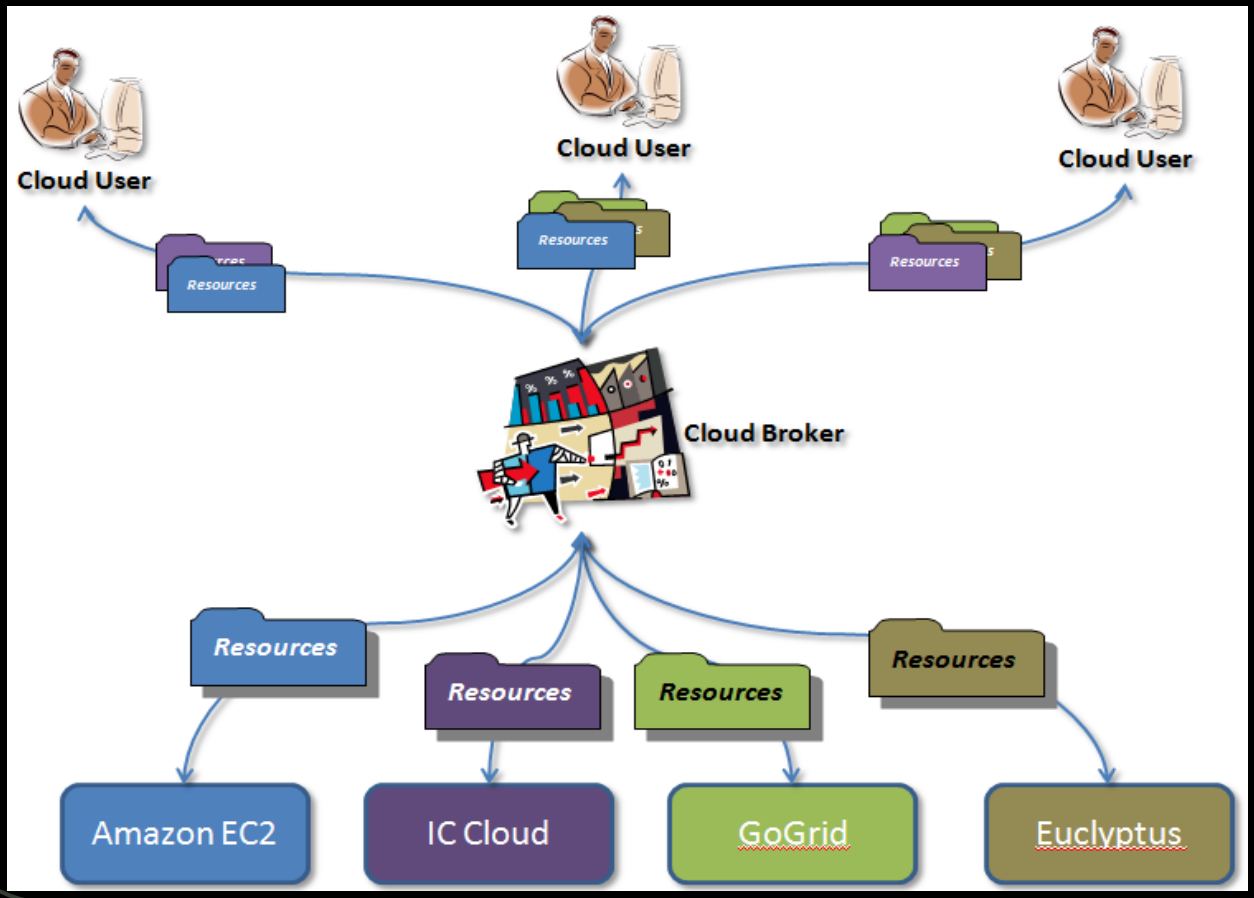
1. Single Provider Model (Amazon, GoGrid, Rackspace etc)
2. Inter-Cloud Brokering Model (Rightscale)
3. Inter-Cloud Federation Model (A Missing Opportunity)

(1). Single Provider Model





(2). Inter-Cloud Brokering Model



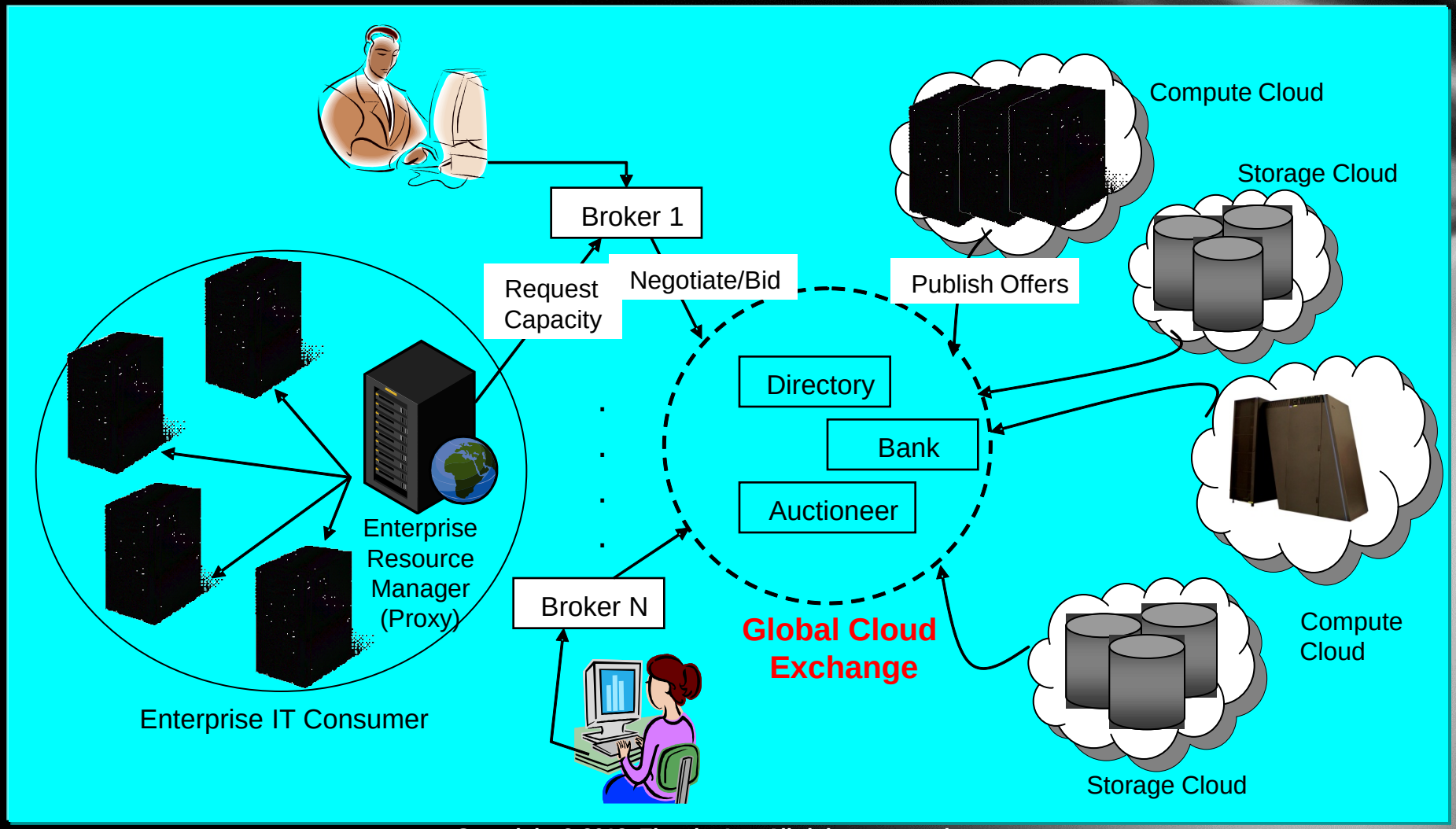
(3). Inter-Cloud Federation Model

Cloud providers are grouped together as a community which aims providing better services and attracts larger user population.

Each cloud provider can register its own resources and products to a Cloud Community Hub (CCH) . The CCH will provide billing, payment, SLA report, credit report etc. for all participating cloud resource providers and customers. (C2C)

The CCH also provides models and tools that enable the cloud providers to form communities so they can share resources within the scope of the community (fusion etc.) (B2B)

InterCloud: "Federation of Clouds" for Scaling Application Services

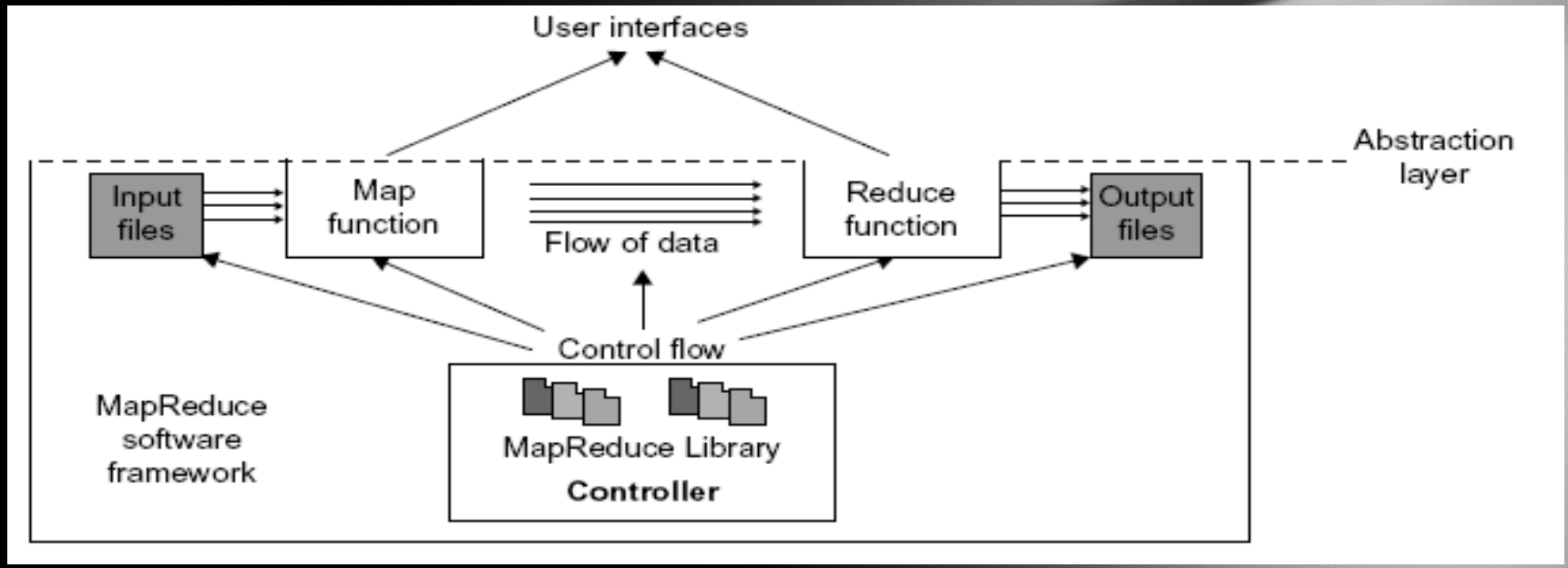


Cloud Computing Software Tool Packages from Google

- The Google file system (GFS) – already covered in Lecture 8
- The MapReduce package – Read Ref. paper [5]
- The Bigtable package – Read Ref. paper [4]
- Google 101 Seminars in Cloud Computing (http://videovoo.com/2007/12/14/ibm-teams-up-with-goole-google-101-cloud-computing-drowning-with-data/)

MapReduce : Scalable Data Processing on Large Clusters

- A web programming model for fast processing large datasets
- Applied in web-scale search and cloud computing applications
- Users specify a *map function* to generate intermediate key/value pairs
- Users use a *reduce function* to merge all intermediate values with the same key.



Batch Processing framework: MapReduce

Map: applies a programmer-supplied function to each logical input record

- Runs on thousands of computers
- Provides new set of key-value pairs as intermediate values

Reduce: collapses values using another programmer-supplied function

Programming Models and Workloads

MapReduce runtime environment schedules map and reduce task to WSC nodes

Availability:

- Use replicas of data across different servers
- Use relaxed consistency:
  - No need for all replicas to always agree

Workload demands

- Often vary considerably

Copyright © 2012, Elsevier Inc. All rights reserved.

Example : Counting the number of occurrences of each word in a large collection of documents

```
map(String key, String value):
  // key: document name
  // value: document contents
  for each word w in value:
    EmitIntermediate(w, "1");
```

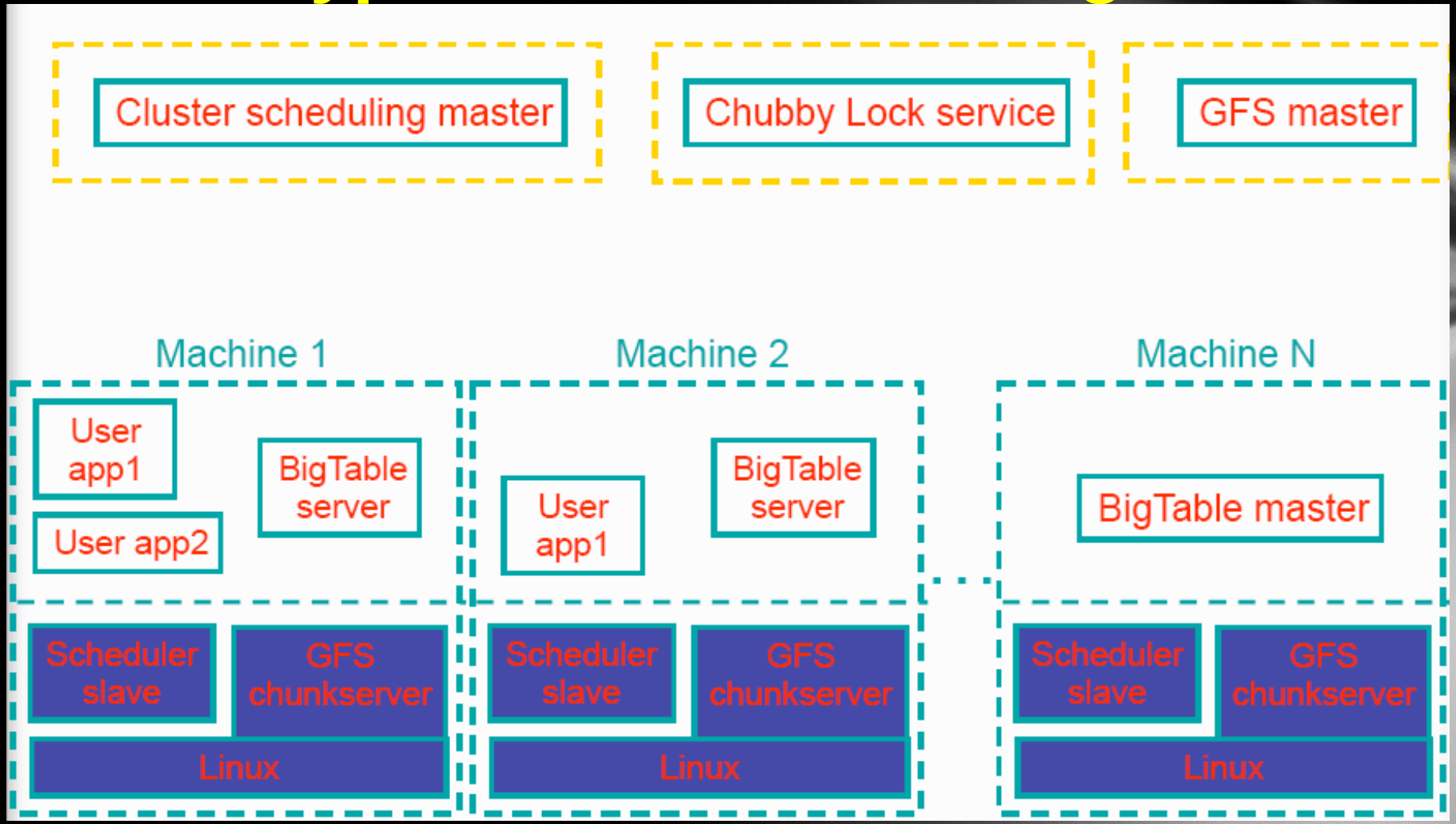
The map function emits each word *w* plus an associated count of occurrences (just a "1" is recorded in this pseudo-code)

Example : Counting the number of occurrences of each word in a large collection of documents

```
reduce(String key, Iterator values):
  // key: a word
  // values: a list of counts
  int result = 0;
  for each v in values:
    result += ParseInt(v);
  Emit(AsString(result));
```

The reduce function sums together all counts emitted for a particular word

Typical Cluster at Google



(Courtesy of Jeffrey Dean, Google, 2008)

Data Processing: MapReduce

- Google's batch processing tool of choice
- Users write two functions:
  - Map: Produces (key, value) pairs from input
  - Reduce: Merges (key, value) pairs from Map
- Library handles data transfer and failures
- Used everywhere: Earth, News, Analytics, Search Quality, Indexing, …

(Courtesy of Jeffrey Dean, Google, 2008)