Discovery 2015: Cloud Computing Workshop
June 20-24, 2011
Berkeley, CA

Introduction to Cloud Computing

Keith R. Jackson
Lawrence Berkeley National Lab

What is it?

NIST 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. This cloud model promotes availability and is composed of five essential characteristics, three service models, and four deployment models.

NIST Essential Characteristics

- On-demand self-service.
  - A consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service's provider.
- Broad network access.
  - Capabilities are available over the network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs).

NIST Essential Characteristics

- Resource pooling.
  - The provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to consumer demand. There is a sense of location independence in that the customer generally has no control or knowledge over the exact location of the provided resources but may be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter). Examples of resources include storage, processing, memory, network bandwidth, and virtual machines.

NIST Essential Characteristics

- Rapid elasticity.
  - Capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.
- Measured Service.
  - Cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

NIST Service Models

- Cloud Software as a Service (SaaS).
  - The capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based email). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

NIST Service Models

- Cloud Platform as a Service (PaaS).
  - The capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

NIST Service Models

- Cloud Infrastructure as a Service (IaaS).
  - The capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

NIST Deployment Models

- Private cloud.
  - The cloud infrastructure is operated solely for an organization. It may be managed by the organization or a third party and may exist on premise or off premise.
- Community cloud.
  - The cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be managed by the organizations or a third party and may exist on premise or off premise.

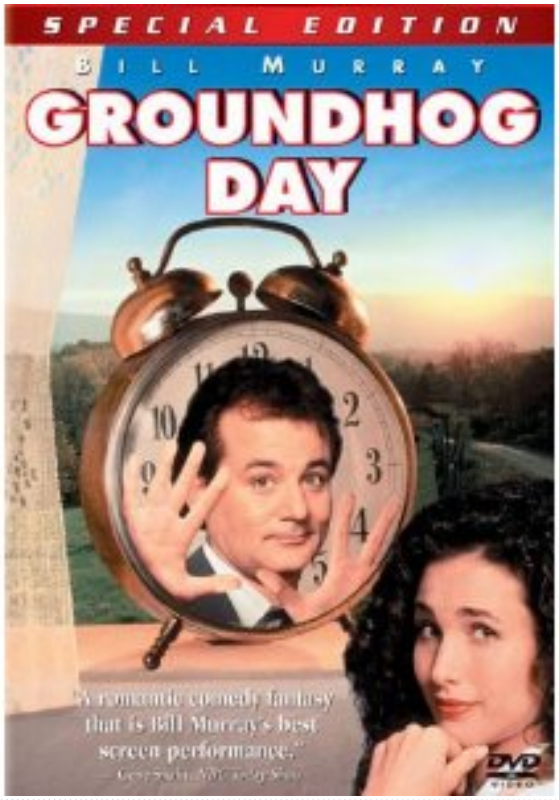
NIST Deployment Models

- Public cloud.
  - The cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.
- Hybrid cloud.
  - The cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).
- Note: Cloud software takes full advantage of the cloud paradigm by being service oriented with a focus on statelessness, low coupling, modularity, and semantic interoperability.

Above the Clouds: A Berkeley View of Cloud Computing

- The illusion of infinite computing resources available on demand, thereby eliminating the need for Cloud Computing users to plan far ahead for provisioning.
- The elimination of an up-front commitment by Cloud users, thereby allowing companies to start small and increase hardware resources only when there is an increase in their needs.
- The ability to pay for use of computing resources on a short-term basis as needed (e.g., processors by the hour and storage by the day) and release them as needed, thereby rewarding conservation by letting machines and storage go when they are no longer useful.

Haven't We Been Here Before?

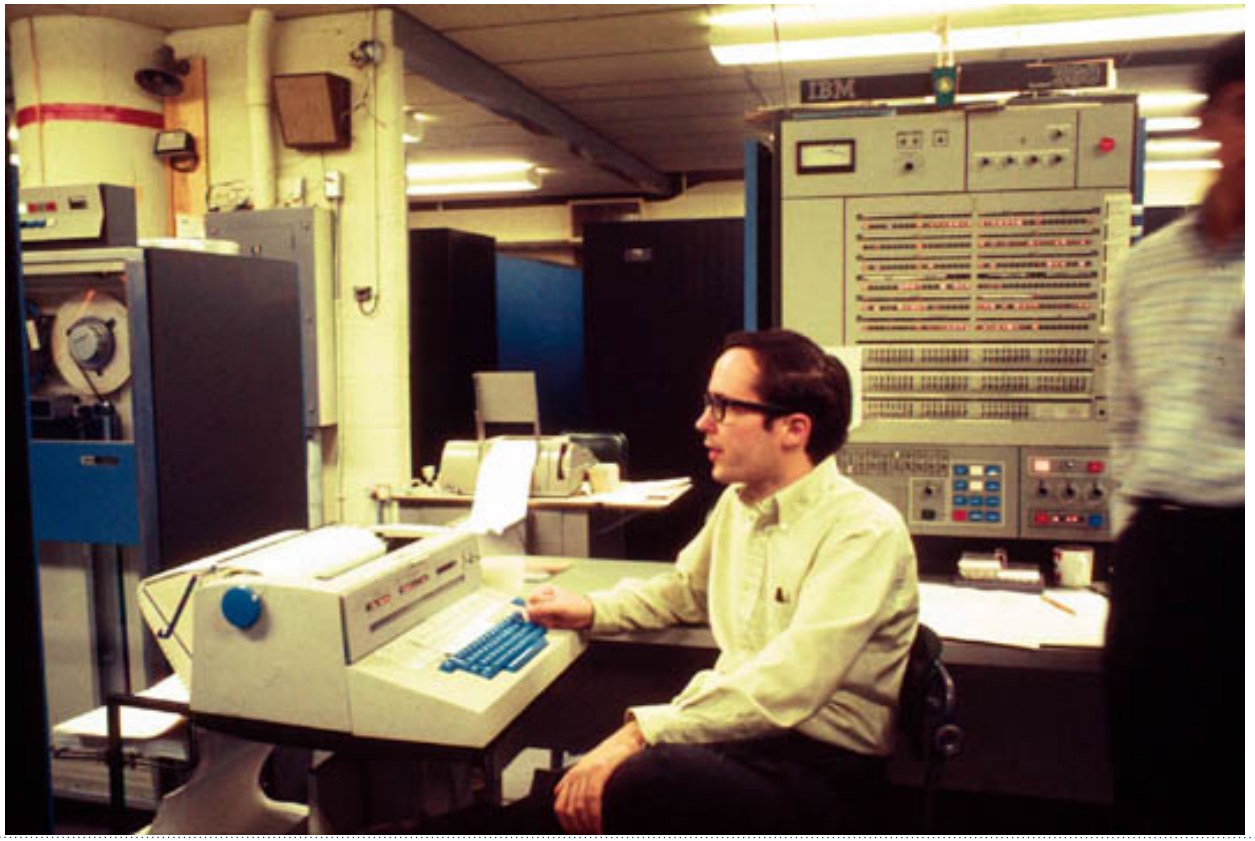


Utility Computing

- John McCarthy predicted in 1961 that:
  - “computation may someday be organized as a public utility”

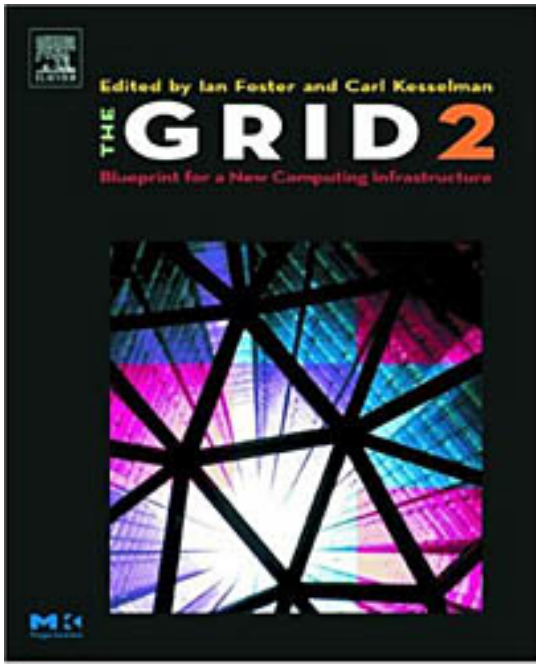
Time Sharing & Virtualization

- IBM System 360



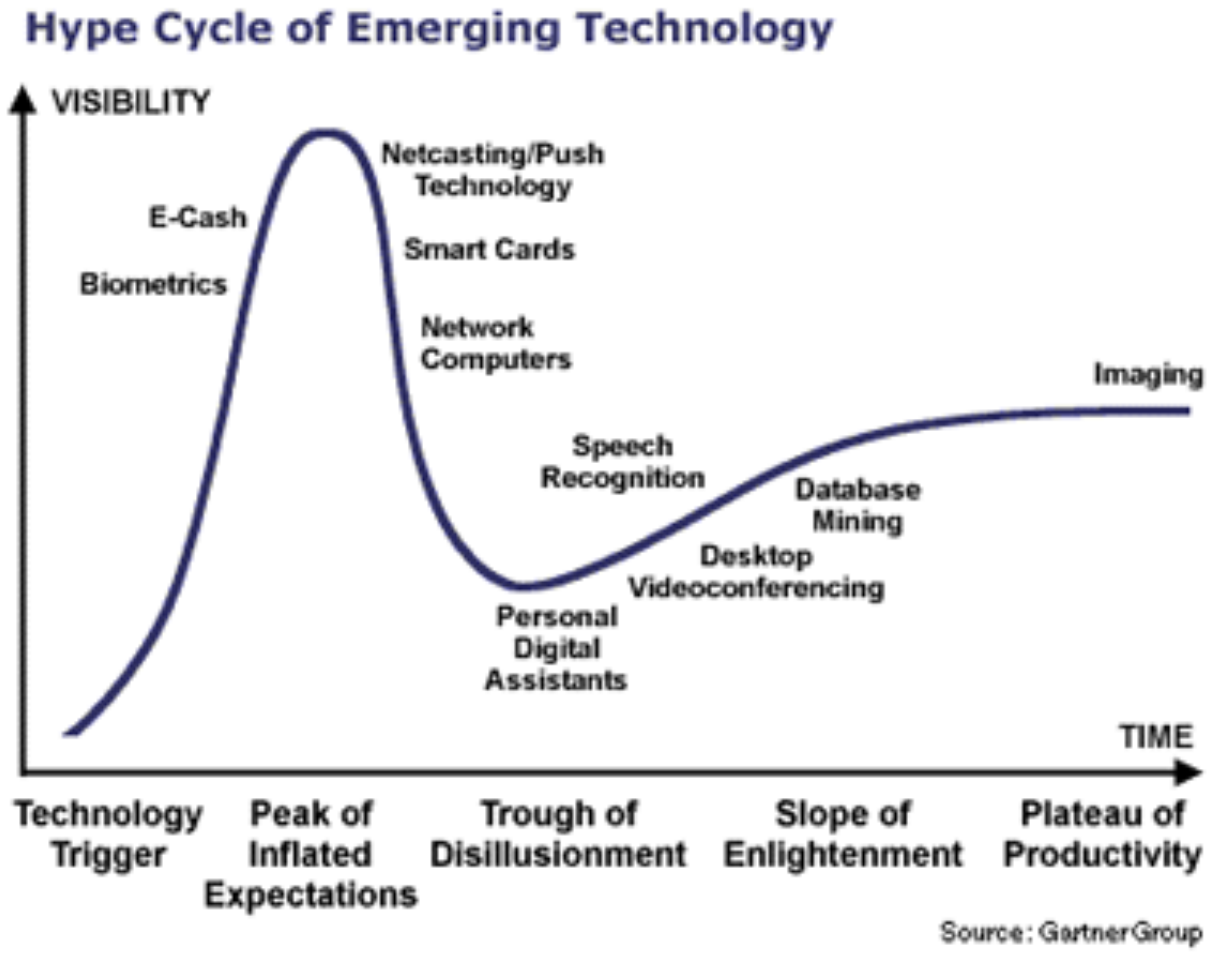
Grid Computing

- The 1990's saw the rise of Grid Computing

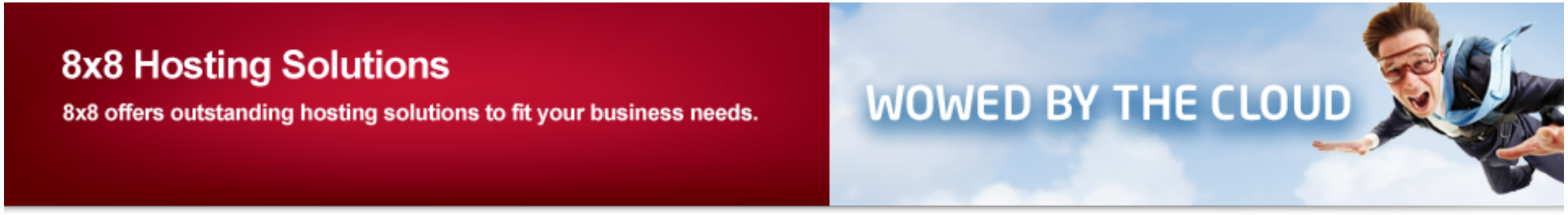


Hype or Reality?

Gartner Hype Cycle



Hype


8x8 Hosting Solutions
8x8 offers outstanding hosting solutions to fit your business needs.
WOWED BY THE CLOUD

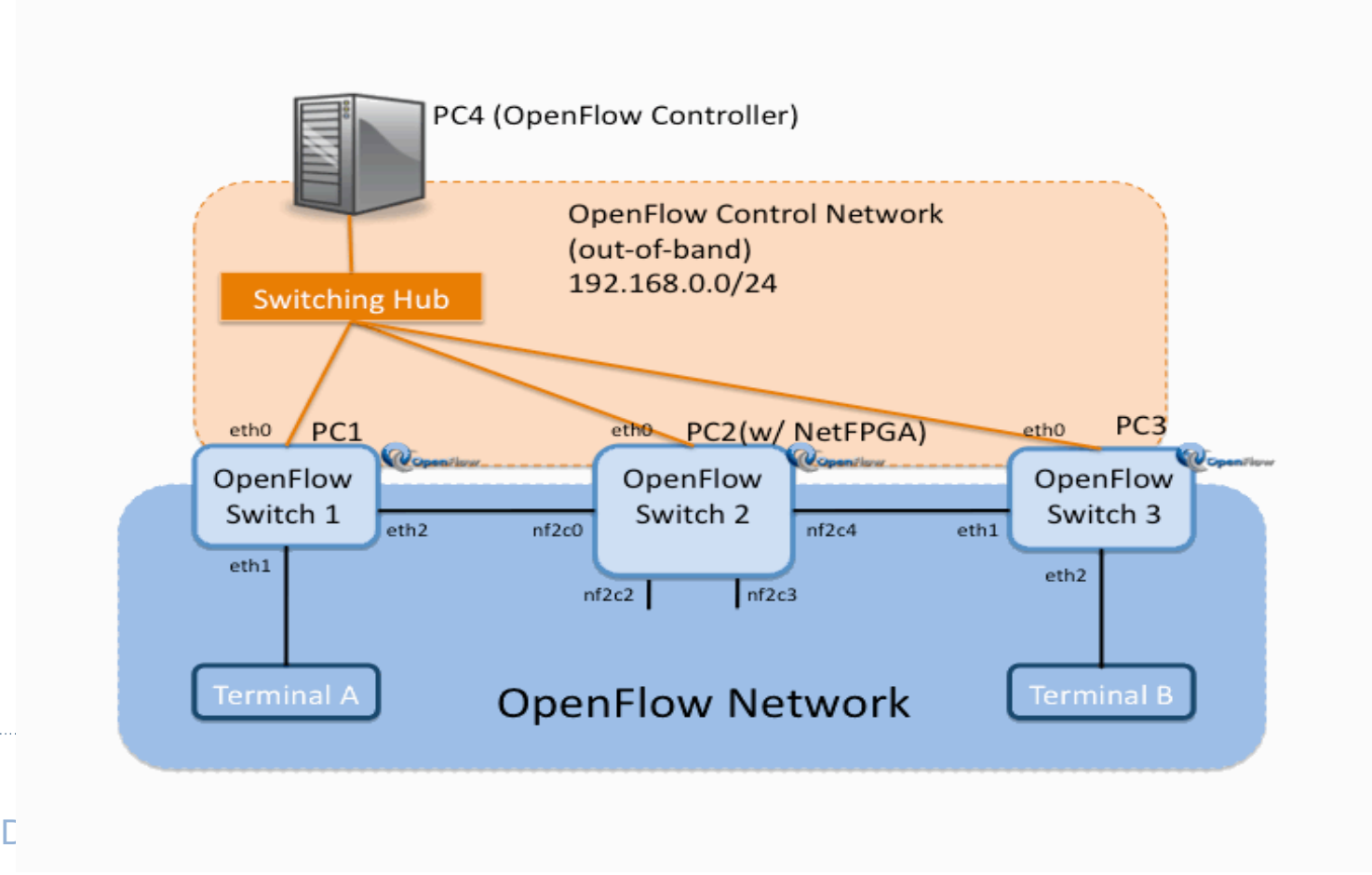
Enabling Technologies and Programming Models

Virtualization

- Separate the running OS from the underlying hardware
- Full hardware virtualization
  - Present a full machine abstraction so an unmodified OS can run
  - May or may not be hardware assisted
    - Intel VT-x and AMD-V extensions
- Paravirtualization
  - Guest OS is modified to use special calls to the hypervisor

Network Virtualization

- Abstract network administration and management from the underlying hardware



Why Virtualization for Clouds?

- Allow more then one VM to run on the underlying hardware
  - Allows for maximizing hardware utilization
- Provides a convenient unit for management
  - Allows the customer to have complete control over their environment
  - Can be used for scaling by instantiating more or less VM's
  - Cloud provider can migrate VM's to improve utilization
- Provides a mechanism for isolating different applications running on the same hardware
- Provides the ability to manage the network in complicated ways

Who uses it?

- Amazon Web Services
- Eucalyptus
- Windows Azure

Map/Reduce

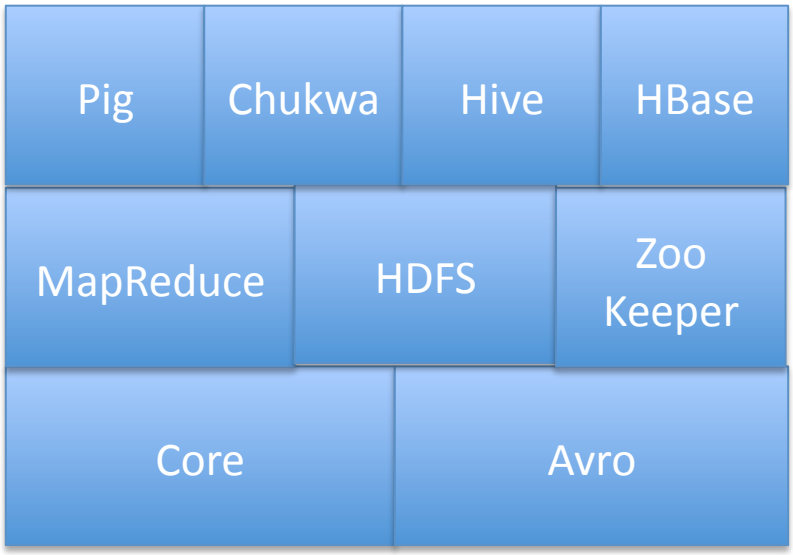
- Distributed computing framework for parallel processing of large data sets
  - Inspired by the map and reduce functions in functional programming languages
  - Made popular by Google
    - Driven by needs of search
- Formally defined as:
  - Map(k1, v1) -> list(k2, v2)
  - Reduce(k2, list(v2)) -> list(v3)

Apache Hadoop

- Open source implementation of Map/Reduce
  - Fault tolerant
  - Data aware scheduling
- Written in Java
  - Bindings in C++ (Pipes), support for other languages through "streaming"
- Comes with a high-performance distributed file system (HDFS)
  - Based on the Google File System (GFS)
  - Configurable data replication
    - For data locality and redundancy

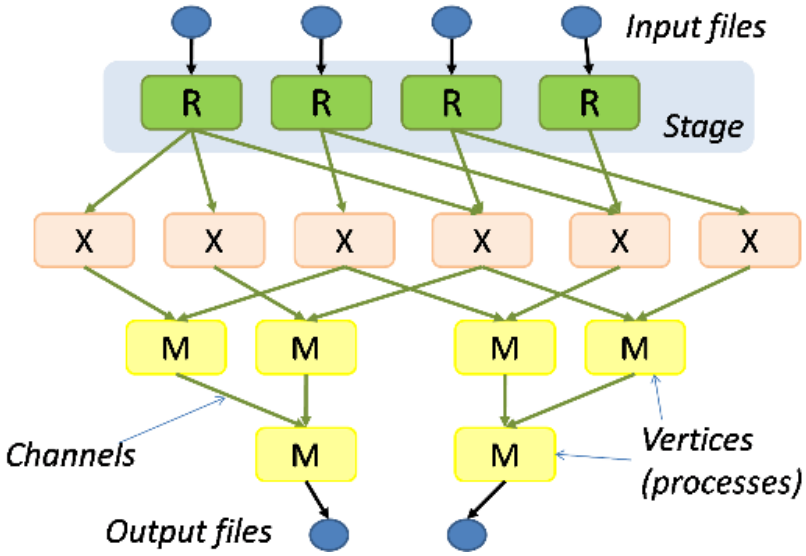
Apache Hadoop Ecosystem

- Pig
  - A high-level data flow language and execution framework
- Hive
  - A data warehouse framework with SQL like query language
- HBase
  - Scalable, distributed NoSQL database
  - Similar to Google's BigTable
- Zookeeper
  - Coordination service for distributed applications



Microsoft Dryad

- Infrastructure for writing scalable, distributed, data parallel applications
- Separate serial programs can be connected together
- Computations structured as DAG's
  - DAG's are a superset of Map/ Reduce



Building Internet Scale Distributed Apps

- Google, Microsoft, and Amazon have all evolved towards similar best-practices for building loosely-coupled distributed applications
- Common set of basic building blocks from all three
  - Google's however is typically not public
    - Apache Hadoop project re-implements from Google papers

Basic Building Blocks

- Queues
  - Used for communication between components
  - Decouples the components from each other in case of failure
- Tabular storage
  - Non-ACID, high-performance and distributed
  - Used for key/attribute storage for structured data
- Blob storage
  - High-performance, distributed object store for non-structured data

Cloud Computing Companies

Amazon Elastic Compute Cloud (EC2)

- Amazon's IaaS offering
- Offers support for running Linux and Windows in VM's in the Amazon datacenters
- Provides a simple web service based API for managing the complete VM lifecycle
- Uses paravirtualization
  - Xen based hypervisor
  - Requires modified Xen kernels
  - The new cluster compute instance type uses hardware assisted virtualization

AWS Storage Options

- Local disk attached to VM instance
  - Not persistent
- Simple Storage Service (S3)
  - Scalable persistent object store
  - HTTP based with a simple REST API
- Elastic Block Storage (EBS)
  - Persistent, block level storage
  - Exposed as raw block level device to VM

Additional AWS Components

- Simple DB
  - Scalable, high availability tabular storage
  - Non-ACID
- Simple Queue Service (SQS)
  - Reliable, scalable message based queue system
- Relational Database Service (RDS)
  - Hosted, replicated MySQL database service
- Elastic MapReduce
  - Hosted Hadoop framework

Eucalyptus

- Open source IaaS implementation
  - API compatible with Amazon AWS
- Manage virtual machines
  - Automated translation from Amazon EC2 images
- Walrus
  - Object store
  - Interface compatible to S3
- Block Storage
  - Interface compatible to EBS

Windows Azure

- Microsoft's cloud platform
- Based on virtual machines
  - VM's are managed for you
  - Two types of VM's
    - Web role
    - Worker role
- Provides queue's, object storage, tabular storage, and hosted SQL Server instances

Google Cloud Offerings

- Google Storage
  - S3 interface compatible, high availability, object store
- Prediction API
  - Access to Google's machine learning algorithms
  - Simple REST based API
- Big Query
  - SQL like query language for analyzing tabular data
- Google AppEngine
  - Platform for developing scalable web applications that run in the Google cloud

Cloudera

- Bundles Apache Hadoop into an integrated top-to-bottom data platform
  - Integrates the various components
  - Provides support
  - Builds value added services

Conclusions

- Cloud Computing builds upon 50 years of CS research
- Cloud Computing is reaching the top of the hype curve
  - Avoid the “Trough of Disillusionment” by:
    - Taking a data-driven approach to understanding what Clouds are good at and what they are not.

References

- M. Armbrust, A. Fox, R. Griffith, A. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, M. Zaharia. Above the Clouds: A Berkeley View of Cloud computing. Technical Report No. UCB/EECS-2009-28, University of California at Berkley, USA, Feb. 10, 2009.
- I. Foster, Y. Zhao, I. Raicu, and S. Lu, “Cloud Computing and Grid Computing 360-Degree Compared,” Proc. IEEE Grid Computing Environments Workshop, pp. 1-10, 2008.
- Mell, P. and Grance, T. 2009. The NIST Definition of Cloud Computing. National Institute of Standards and Technology.

Questions?