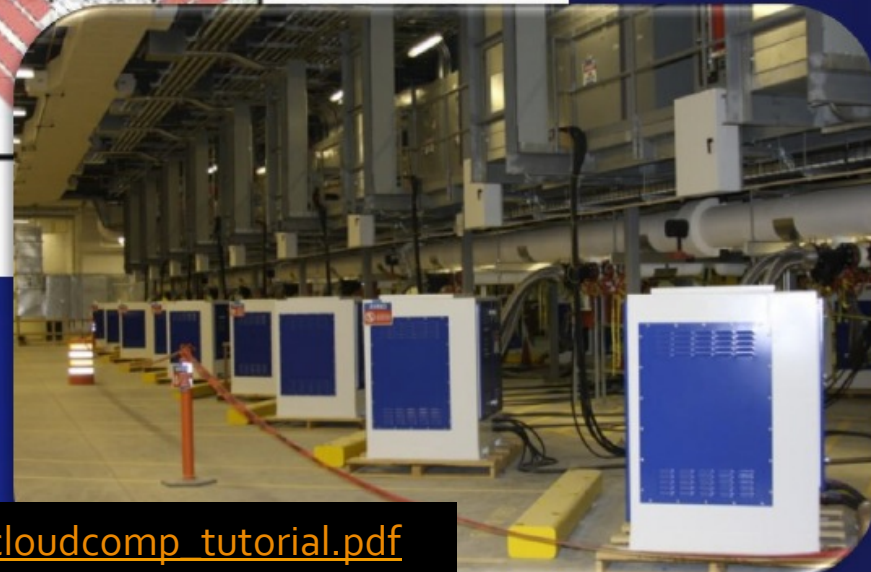# Advances in DC deployment

- Conquering complexity.
  - Building racks of servers & complex cooling systems all separately is not efficient.
  - Package and deploy into bigger units:



**Introducing ICE Cube™**
**Rackable** systems   The world's most efficient modular data center

**Microsoft**

# Containers: Separating Concers

http://research.microsoft.com/en-us/um/redmond/events/cloudfutures2012/tuesday/
Keynote_OpportunitiesAndChallenges_Yousef_Khalidi.pdf

# Green Clouds

- Cloud Centers optimize life cycle costs and power use

$$PUE = \frac{Total\ facility\ power}{IT\ equipment\ power}$$

- http://www.datacenterknowledge.com/archives/2011/05/10/uptime-institute-the-average-pue-is-1-8/

- Average PUE = 1.8 (was nearer 3) ; Good Clouds are 1.1-1.2

- 4th generation data centers (from Microsoft) make everything modular so data centers can be built incrementally as in modern manufacturing

- http://loosebolts.wordpress.com/2008/12/02/our-vision-for-generation-4-modular-data-centers-one-way-of-getting-it-just-right/

- Extends container based third generation

# Some Sizes in 2010

- [http://www.mediafire.com/file/zzqna34282frr2f/koomeydatacenterelectuse2011finalversion.pdf](http://www.mediafire.com/file/zzqna34282frr2f/koomeydatacenterelectuse2011finalversion.pdf)
- 30 million servers worldwide
- Google had 900,000 servers (3% total world wide)
- Google total power ~200 Megawatts
  - < 1% of total power used in data centers (Google more efficient than average – **Clouds are Green**!)
  - ~ 0.01% of total power used on anything world wide
- Maybe total clouds are 20% total world server count (a growing fraction)

# Some Sizes Cloud v HPC

- **Top Supercomputer** Sequoia Blue Gene Q at LLNL
  - 16.32 Petaflop/s on the Linpack benchmark using 98,304 CPU compute chips with 1.6 million processor cores and 1.6 Petabyte of memory in 96 racks covering an area of about 3,000 square feet
  - 7.9 Megawatts power
- **Largest (cloud)** computing data centers
  - 100,000 servers at ~200 watts per CPU chip
  - Up to 30 Megawatts power
  - Microsoft says upto million servers
- So **largest supercomputer** is around **1-2% performance of total cloud computing systems** with Google ~20% total

# Data Center vs Supercomputers

- Scale
  - Blue Waters = 40K 8-core "servers"
  - Road Runner = 13K cell + 6K AMD servers
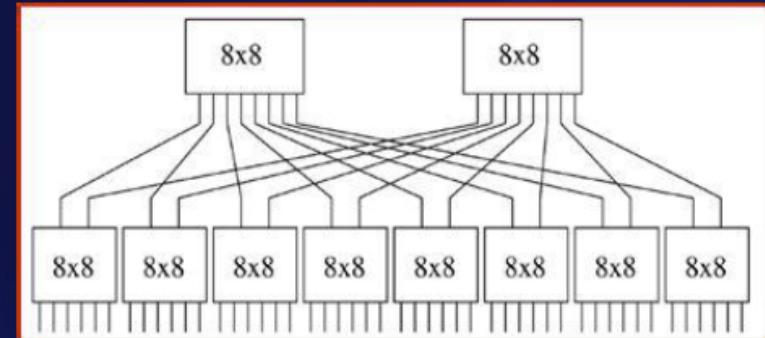  - MS Chicago Data Center = 50 containers = 100K 8-core servers.
- Network Architecture
  - Supercomputers: CLOS "Fat Tree" infiniband
    - Low latency – high bandwidth
    - protocols
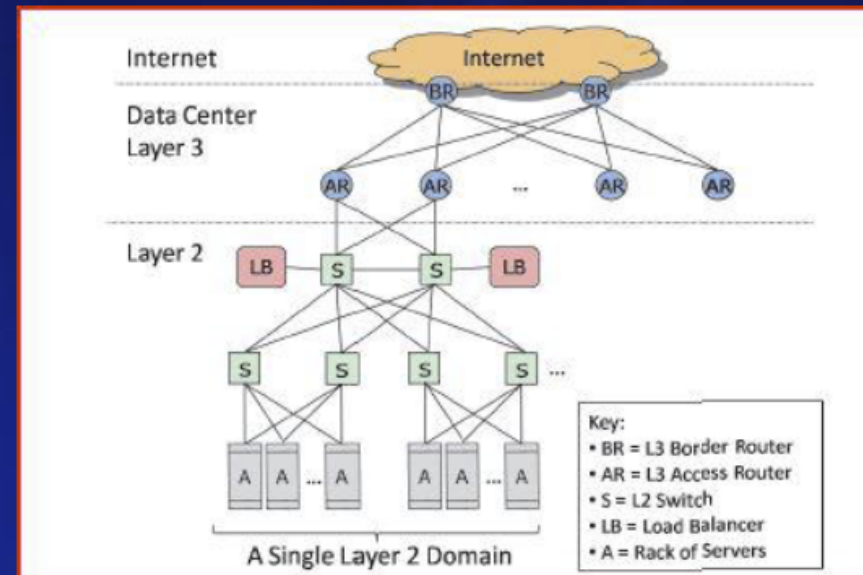  - Data Center: IP based
    - Optimized for Internet Access
- Data Storage
  - Supers: separate data farm
    - GPFS or other parallel file system
  - DCs: use disk on node + memcache

Fat tree network

Standard Data Center Network

23

# Cloud Industry Players

## Players: Providers

Programmatic access via Web Services and/or Web APIs

"Pure" virtualized resources

   CPU, memory, storage, and bandwidth

   Data store

versus

Virtualized resources plus application framework
(e.g., RoR, Python, .NET)

   Imposes an application and data architecture

   Constrains how application is built

[MM]

22

# Players: Cloud Intermediaires

Resells (aspects of) raw cloud resources, with added value
    propositions

- Packaging resources as bundles
- Facilitating cloud resource management,
      e.g., setup, updates, backup, load balancing, etc.
- Providing tools and dashboards

Enabler of the cloud ecosystem

[MM]

## Players: Application Providers

Software as a Service (SaaS):
  Applications provided and consumed over the Web
Infrastructure usage (mostly) hidden



[MM]

# Cloud Talk

▲ *The Public Cloud "Big Four"*

▲ *Amazon*
▲ *Google*
▲ *Microsoft*
▲ *Salesforce.com*

johnmwillis.com

http://www.slideshare.net/botchagalupe/introduction-to-clouds-cloud-camp-columbus
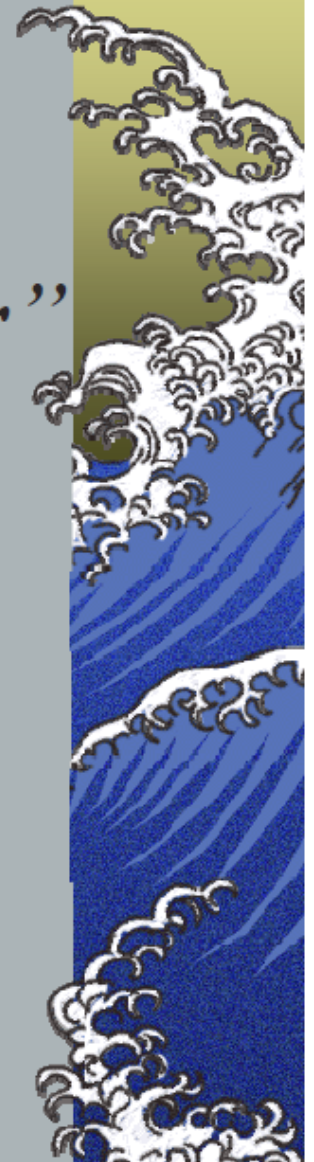
# Cloud Talk

⤴ *The Private Cloud "Big Four"*

⤴ ***IBM***

⤴ ***VMware***

⤴ ***Sun/Oracle***

⤴ ***3Tera***

johnmwillis.com

http://www.slideshare.net/botchagalupe/introduction-to-clouds-cloud-camp-columbus