

A hand is shown from the bottom left, reaching upwards towards the top of the frame. The background is a clear blue sky with scattered white clouds. The hand is positioned as if reaching for something or gesturing upwards.

# Alternative Deployment Models for Cloud Computing in HPC Applications

Society of HPC Professionals  
November 9, 2011  
Steve Hebert, Nimbix

# The case for Cloud in HPC

Build it in house



Assemble in the cloud?



## Trend 1: Computation Challenge

### In Bioinformatics:

Human genome takes many days of CPU time on large-memory, multiprocessor computers

In 2011, the worldwide capacity for whole human genome sequencing will easily reach 50,000

Beijing Genomics Institute alone has purchased 128 of these instruments

Illumina's HiSeq 2000 gene sequencer can sequence two whole human genomes per week

## Trend 2: Compute Capacity Constraints

- ▶ Building in house means capital equipment investment, DC operating costs, and fixed capacity for growing workloads
- ▶ Building in the cloud offers elastic hourly capacity expansion, but brings challenge around management, ease of use, and data movement
- ▶ How best to leverage cloud resources in HPC business process?

# “as-a-Service” Cloud Models

## Infrastructure

- Virtual Machine server instances
- On-demand, spot-pricing, dynamically scalable



## Platform

- Middleware on top of VM server instances
- On-demand, dynamically scalable, one-click cluster creation



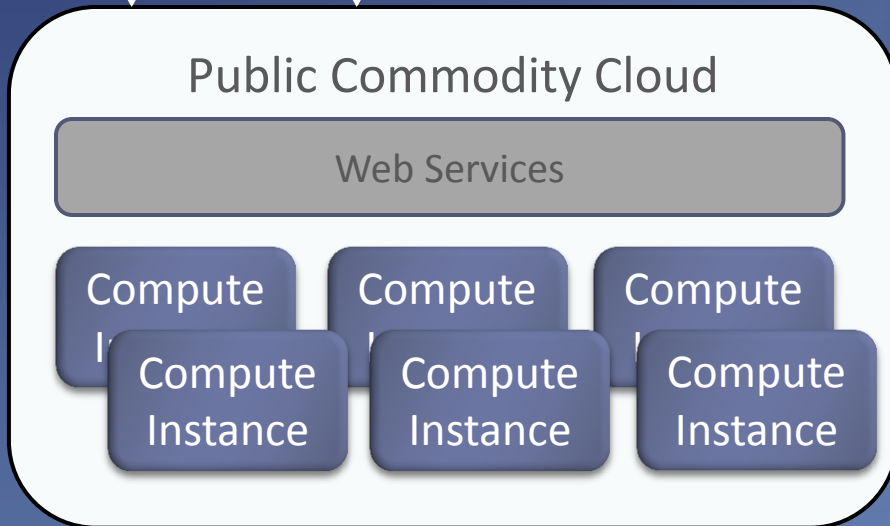
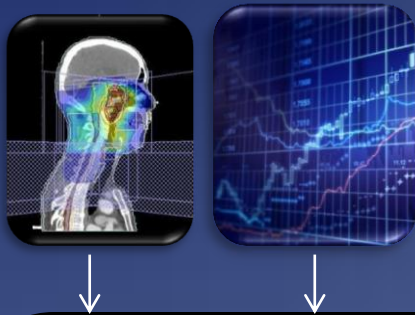
## HPC / Application

- Bare-metal, purpose built supercomputers
- Schedule jobs, no infrastructure mgmt



# Elastic Virtual Infrastructure: IaaS

User constructs HPC cluster with VM instances

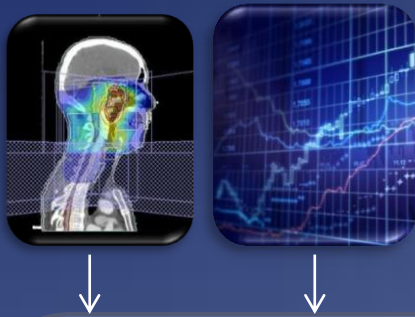


- ▶ Cost model: hourly rate for compute instances + BW ingress/egress
- ▶ Dynamically scalable (but with real-world limits)
- ▶ Machines must be imaged, applications installed along with any tools used to deploy jobs

# Platform Middleware: PaaS

User constructs HPC cluster via platform middleware

- ▶ Cost model: hourly rate for compute instances + BW ingress/egress
- ▶ Generally run on commodity public cloud I/F



Cloud Process Automation Layer

- Cluster Mgmt
- Auto scaling
- Dashboarding

Public Commodity Cloud

API Integration ↔ Web Services

Compute

Compute

Compute

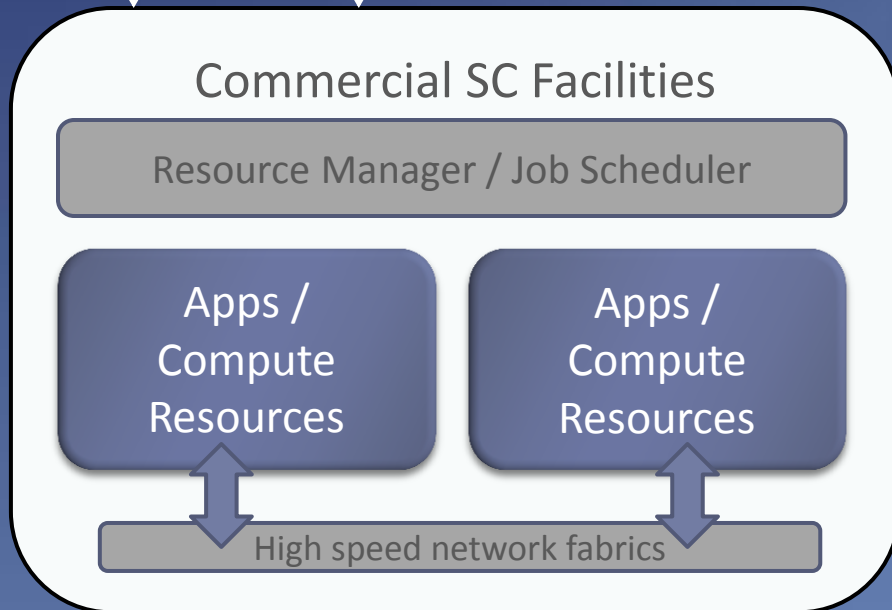
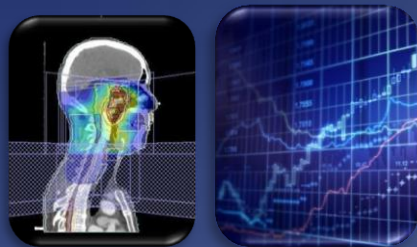
Compute Instance

Compute Instance

Compute Instance

# Bare-Metal Supercomputing: HPCaaS

User submits job via ssh or web-based job scheduler



- ▶ Cost model: Pre-paid blocks of time
- ▶ Low-latency network interconnect
- ▶ High performance file systems
- ▶ Not dynamically scalable
  - ▶ Future: move jobs between providers?

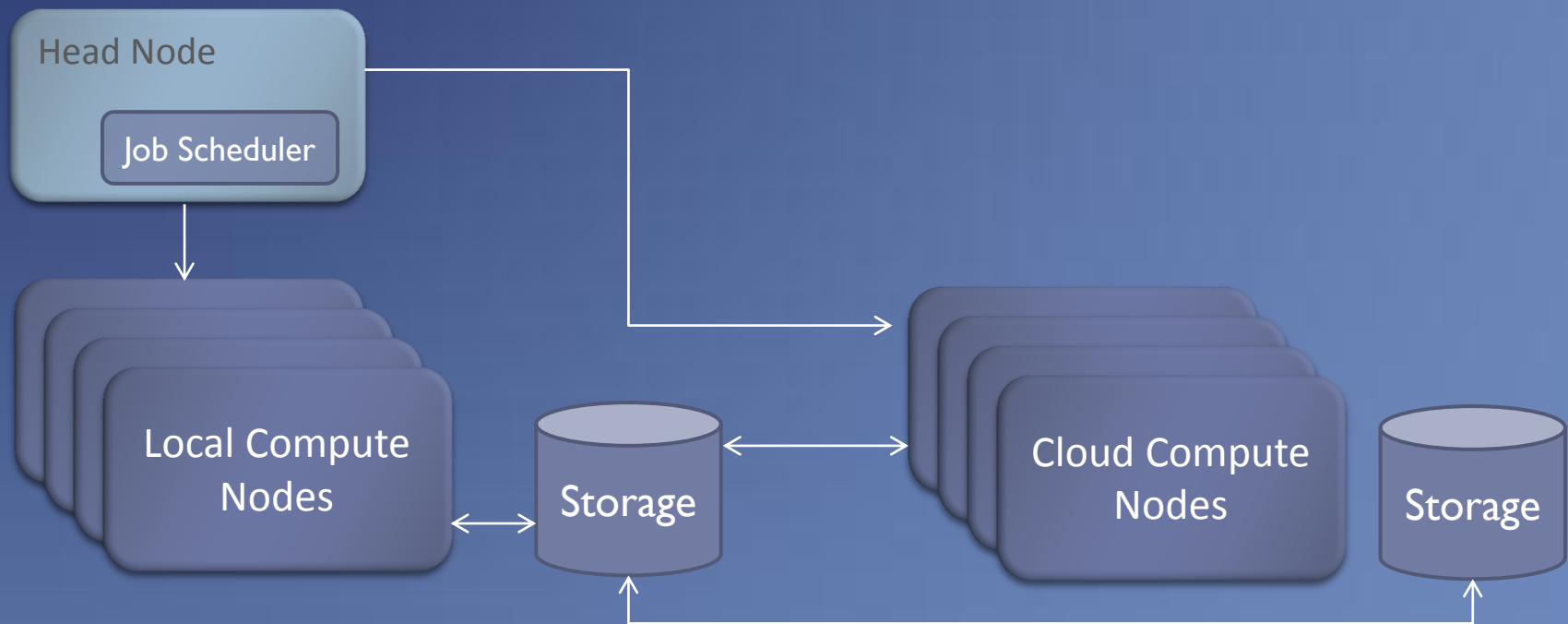
# Managing Data



- ▶ **Data size & movement**
  - ▶ Transfers with SCP, FTP, sFTP, gridFTP
    - ▶ BW limitations can be impediment for large datasets
  - ▶ Ship on disk
  - ▶ Pipelined data sets
- ▶ **Data persistence and storage**
  - ▶ Cloud instance local data is volatile
  - ▶ Block storage and attached shared volumes
  - ▶ Data security and encryption

# Integrating Workflow

- ▶ Console or API?
- ▶ Integrated cloud node extensions or autonomous clusters?
- ▶ Job automation or manual queuing?



# Applications

- ▶ What are application characteristics?
  - ▶ Parallel
  - ▶ Streaming data
  - ▶ Operate on large datasets or many smaller data sets?
- ▶ In house, open source, or licensed?



# Strategies for Cloud Adoption in HPC

- ▶ Review internal capacity, workload attributes and workload volume growth
- ▶ Evaluate candidate applications for cloud trial
- ▶ Select deployment model that aligns best with application and minimum performance requirements
- ▶ Launch controlled trial with identifiable goals
  - ▶ Cost / benchmarks
- ▶ Anticipate evolution of software applications and track computing and datacenter operations economics

# Conclusions

- ▶ Cloud deployment models address HPC and application needs differently
  - ▶ Is desired performance achievable?
- ▶ Understand specific deployment challenges and associated costs
- ▶ Define and implement specific strategies around cloud adoption that address real business or technical goals

# NIMBIX

[www.nimbix.net](http://www.nimbix.net)

