GSAW 2010
Working Group Session 11D

Data Center Migration for Ground Systems: Geospatial Clouds

Dr. Craig A. Lee, lee@aero.org
Senior Scientist, CSD/CSTS/CSRD, The Aerospace Corp.
President, Open Grid Forum

March 3, 2010

Motivations for this Workshop

- Cloud computing offers the potential for significant economies of scale, improved utilization of servers, more flexible allocation of resources, and workload management
  - *Cloud computing entails the dynamic provisioning of processing, storage, and networks in a data center to essentially become a generic hosting environment, prompting the concept of "Data Center Migration" for ground system operators*
- How do we apply cloud computing in support of satellite ground systems?
  - *Serious challenges concerning security, performance management, portability, interoperability, costing models, lack of standards, etc.*
- How do we integrate geospatial standards and tooling with dynamically provisioned resources?
  - *Geospatially referenced data are central to many ground systems*

Agenda

- 13:00-13:15 Welcome and Introductions
  - *Craig Lee, OGF & The Aerospace Corp.*
- 13:15-13:50 Cloud Computing in Ground Segments: Earth Observation Processing Campaigns
  - *Fabrice Brito, Terradue, s.r.l.*
- 13:50-14:25 Geoprocessing in the Cloud
  - *Brian Levy, Open Solutions Group & DIA*
- 13:25-15:00 <Title TBD>
  - *Dan Mandl, NASA Goddard & Open Geospatial Consortium*

- 15:00-15:15 Break

- 15:15-15:50 Eucalyptus-based Event Correlation
  - *Nehal Desai, The Aerospace Corp.*
- 15:50-16:25 Developing Cloud Standards
  - *Craig Lee, OGF & The Aerospace Corp.*
- 16:25-17:00 Open Floor Discussion

GSAW 2010
Working Group Session 11D

Developing Cloud Standards

Dr. Craig A. Lee, lee@aero.org
Senior Scientist, CSD/CSTS/CSRD, The Aerospace Corp.
President, Open Grid Forum

March 3, 2010

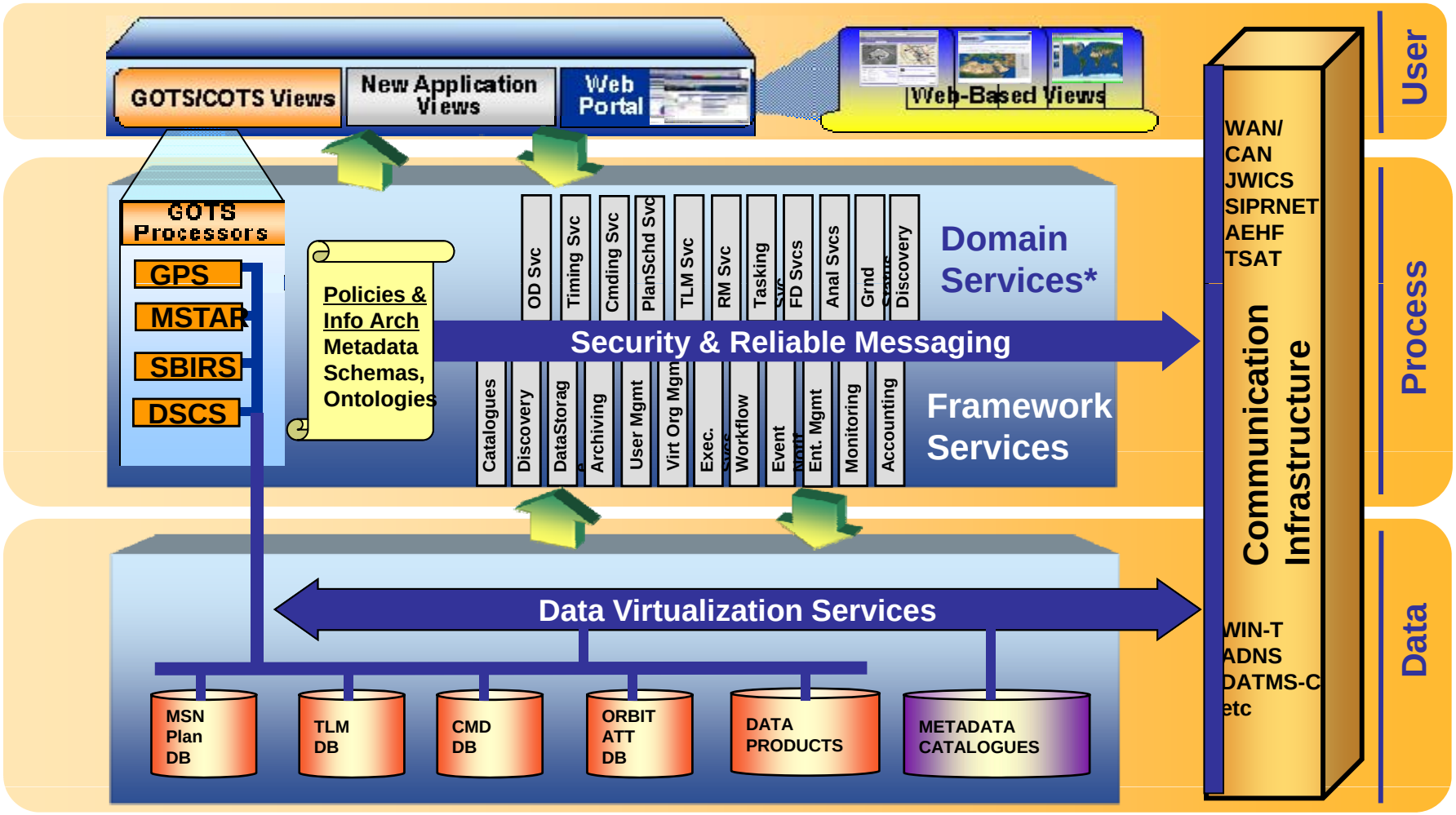
Standards Will Be Critical to Clouds

- Most of the touted benefits will only be realized if there are commonly accepted interfaces, protocols, etc.
- At what levels in the System Stack are standards needed?
  - *Infrastructure, Platform, Services?*
- In which priority order?
  - *What market forces will drive standardization?*

Application Level	• *Software as a Service (SaaS)* • Build an application from pre-defined services • Example: Salesforce.com
Platform Level	• *Platform as a Service (PaaS)* • Acquire a set of hosting environments • Example: Google App Engine (Python)
Infrastructure Level	• *Infrastructure as a Service (IaaS)* • Acquire a set of machines you can login to • Example: Amazon EC2

What Ground System Areas Are Candidates?

A Proposed Reference Model for a Netcentric Ground System Service Architecture

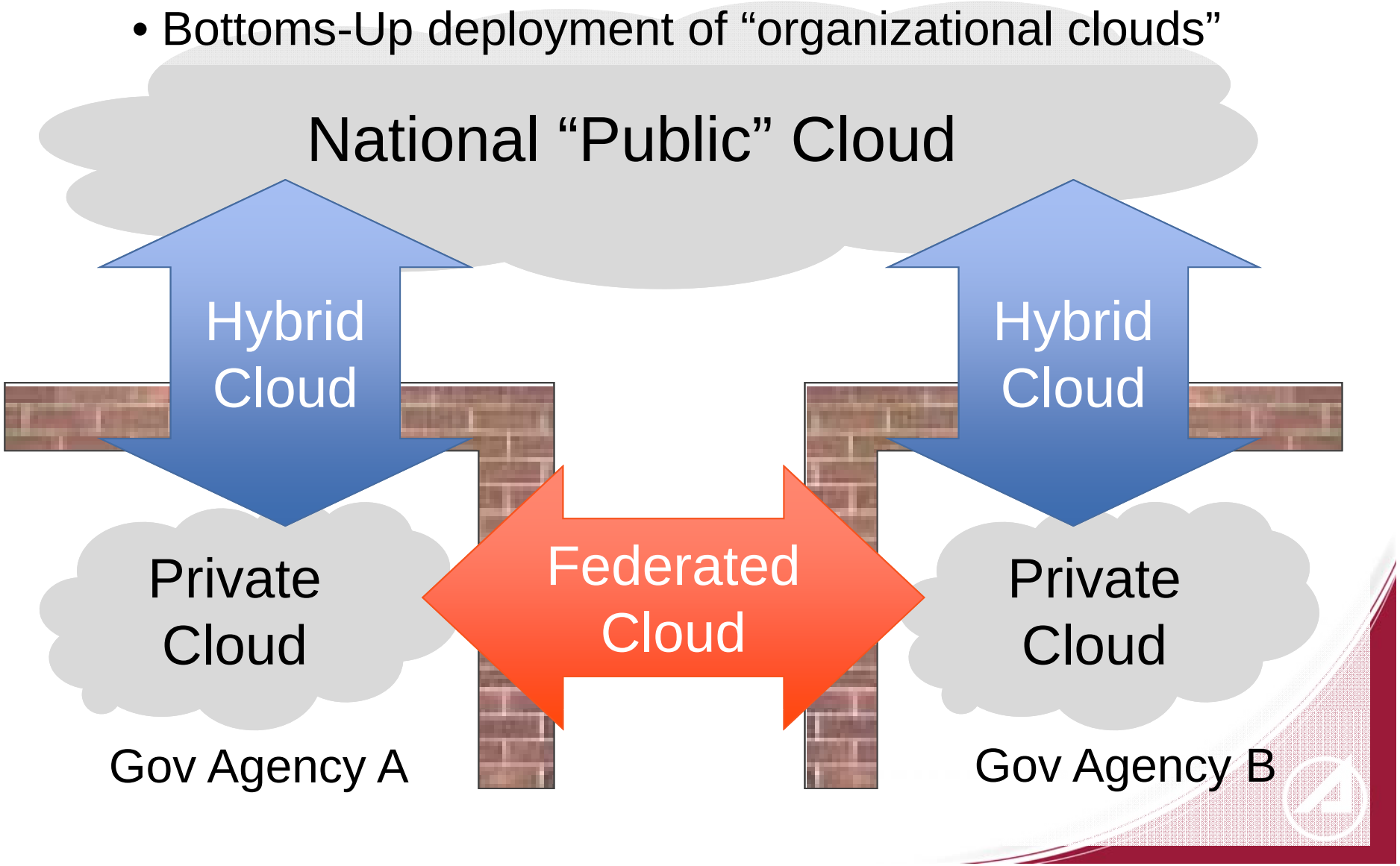


Clouds compliment SOAs and the concept of Netcentricity!

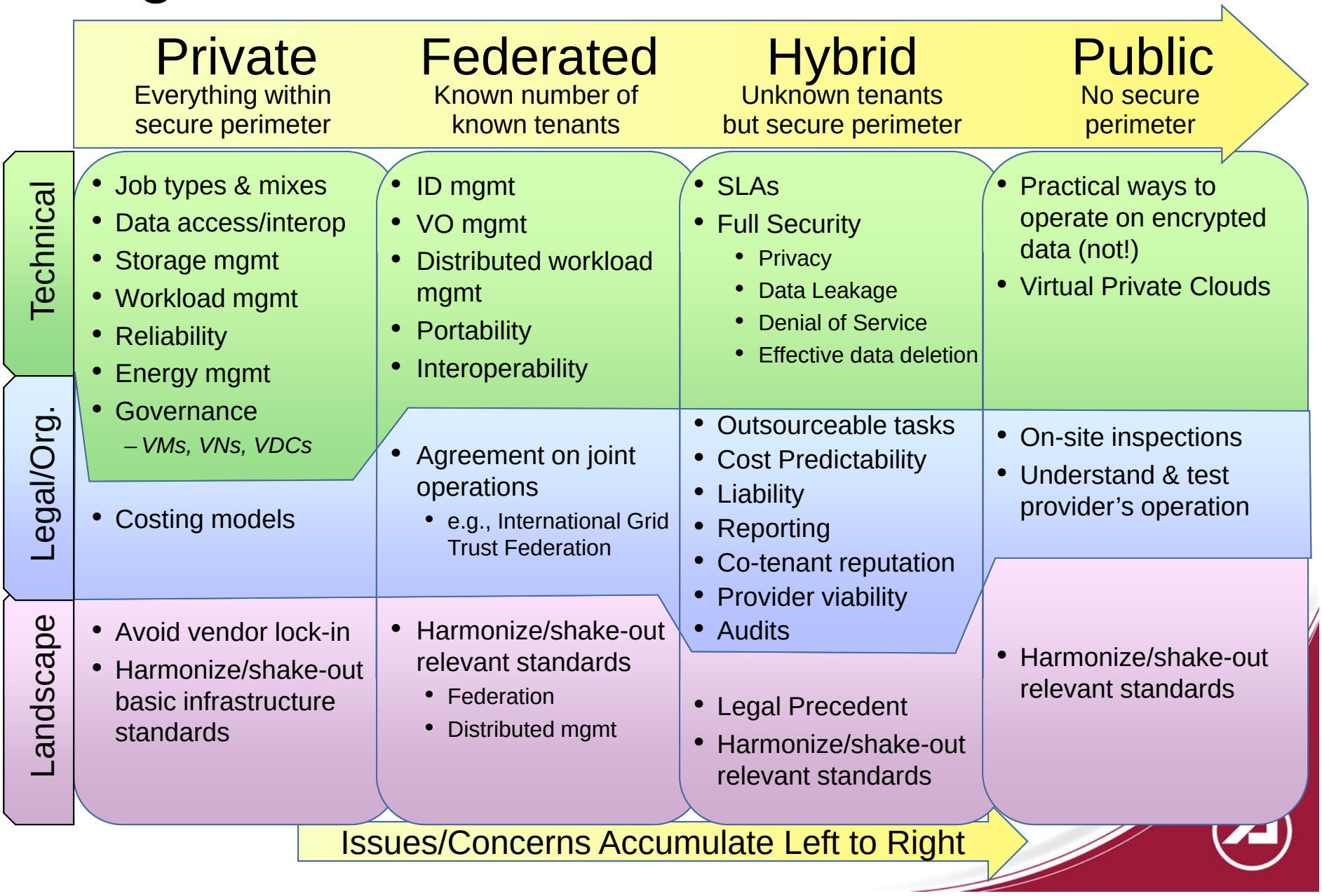
*Domain Services listed are notional may be augmented in any concrete architecture.
Heavily modified from USAF, Distributed Common Ground Architecture (DCGS-A).

Cloud Deployment Trajectory?

- Top-Down deployment of “national public cloud”
- Bottoms-Up deployment of “organizational clouds”



Progression of Issues & Concerns



These Concerns Driving Lots of Activity

- Series of meetings and workshops (all dates in 2009)
  - *Enterprise Cloud Conference, Open Group, Feb. 3, San Diego*
  - *SATCCI, OMG, March 23, Arlington/Crystal City*
  - *Cloud Standards Summit, OMG, July 13, Arlington/Ballston*
  - *Federal Cloud Symposium, July 15, Washington, DC*
  - *Cloud Interoperability Roadmaps, OMG, December 10, Long Beach*
- Wide interest from federal agencies
  - *GeoCloud NSF Workshop, Indianapolis, Sept. 17-18*
  - *Standards for the US Cloud Storefront, NCOIC, Sept. 21, Fairfax*
  - *GeoINT Technical Exch. Meeting hosted at MITRE, McLean VA, Sept. 24*
- Multiple national cloud initiatives
  - *US Cloud Storefront, Japanese Kasumigaseki, UK G-Cloud*
- Development of potential standards
  - *OCCI, vCloud, delta-Cloud, Fujitsu API, Simple-cloud, …*
- Led to creation of *Cloud-Standards.org*

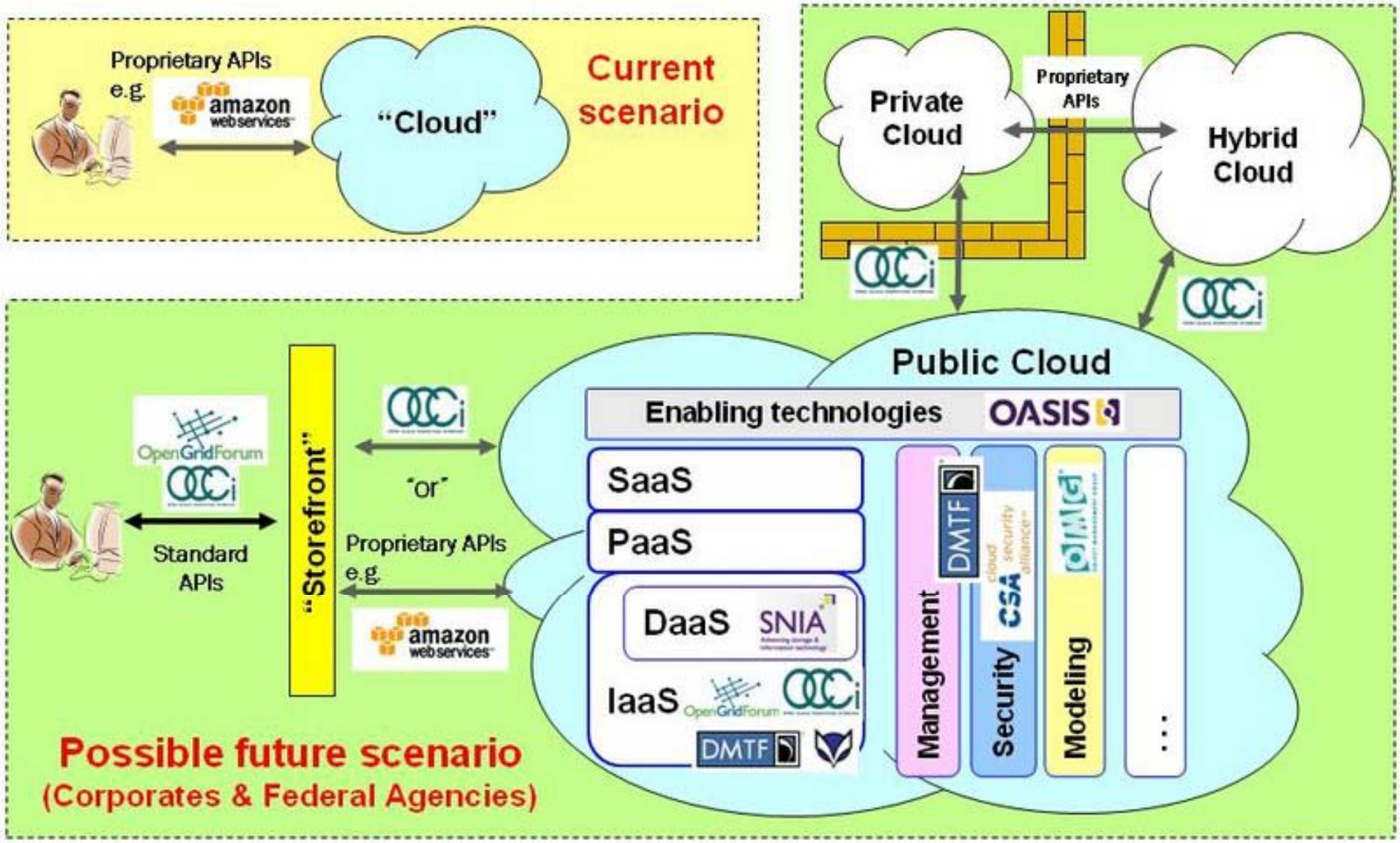
Cloud-Standards.org

- An informal group of Standards Development Organizations (SDOs) collaborating to coordinate and communicate standards for cloud computing, networks and storage
  - *Wiki: cloud-standards.org*
  - *Mailing List: groups.google.com/group/CloudStandards*

Open Cloud Consortium


- Different SDOs bring different but complementary technologies & capabilities
  - *Storage, execution models, deployment models, service level agreements, security, authentication, privacy*
- All interested, committed persons and organizations with relevant technical skills can participate

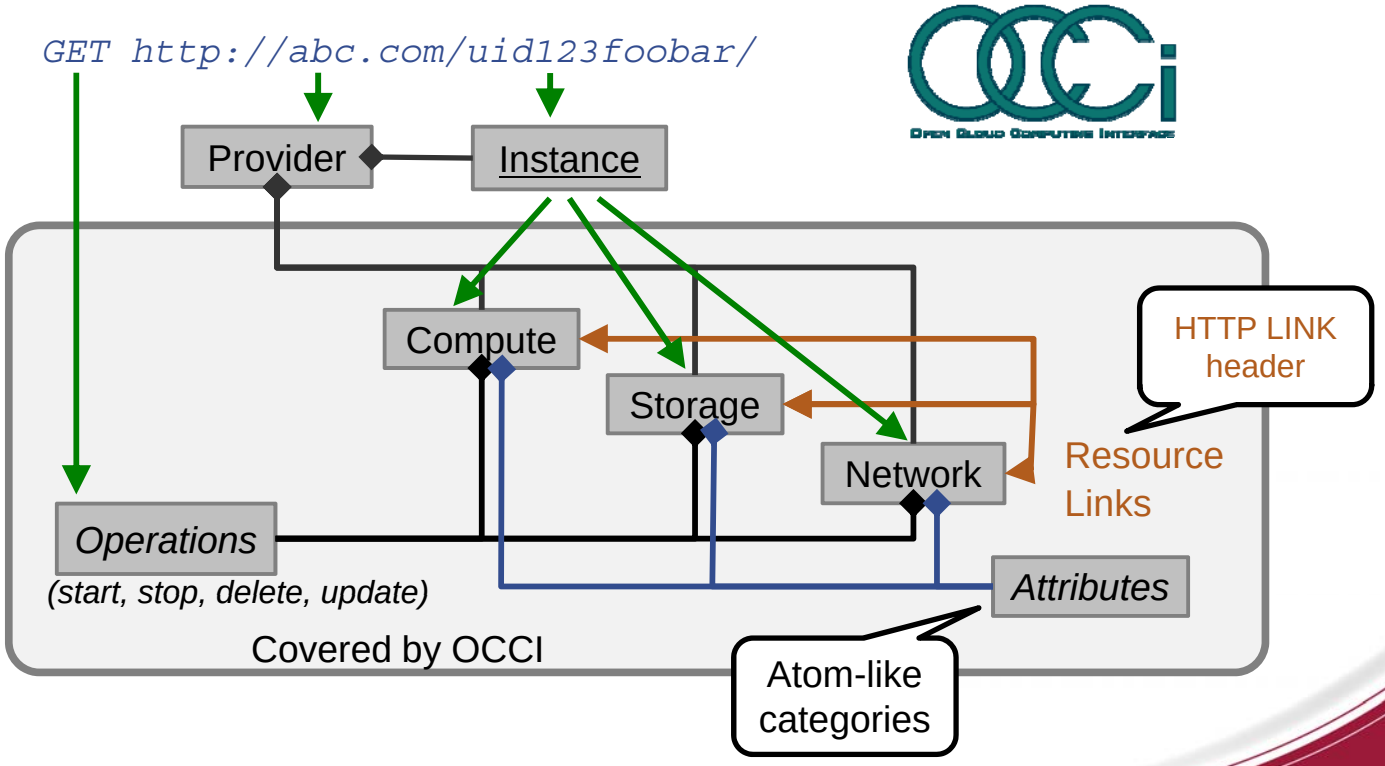
A Positioning of Cloud Standards



Courtesy of Enrico Ronco, Telecom Italia

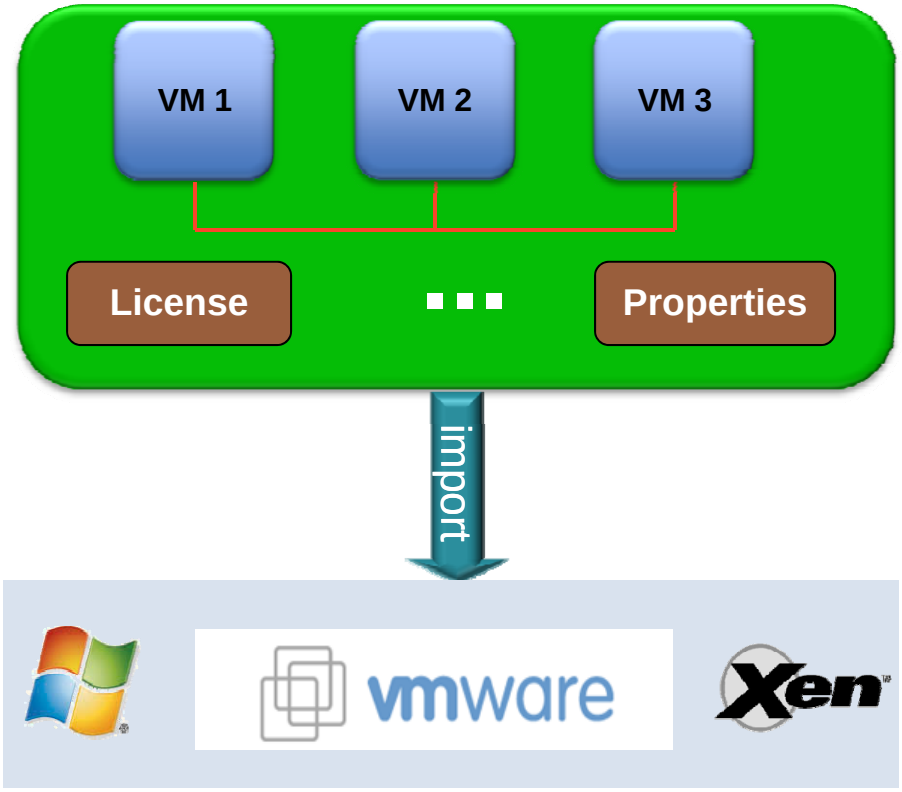
OGF Open Cloud Computing Interface

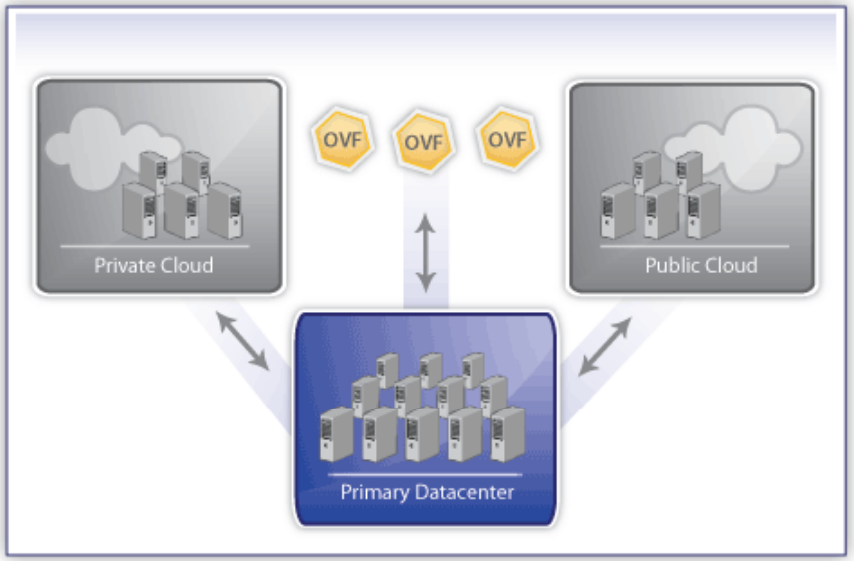
- Interoperable IaaS Cloud API Standard
- Simple, RESTful API
  - *~15 commands – very extensible*
- Solid community interest: 160 members on mailing list
- *www.occi-wg.org*



DMTF OVF -- Open Virtualization Format

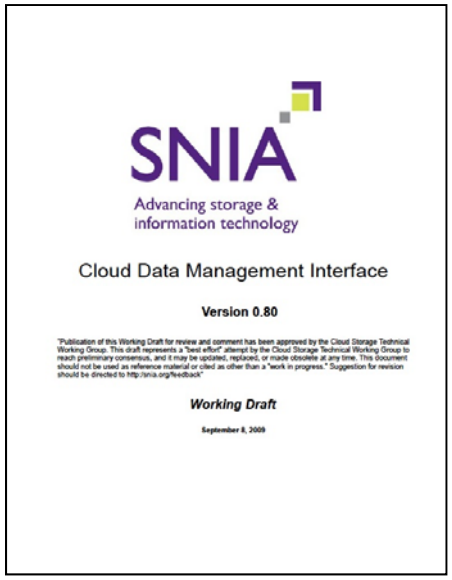
- A multi-vendor format enabling interoperability

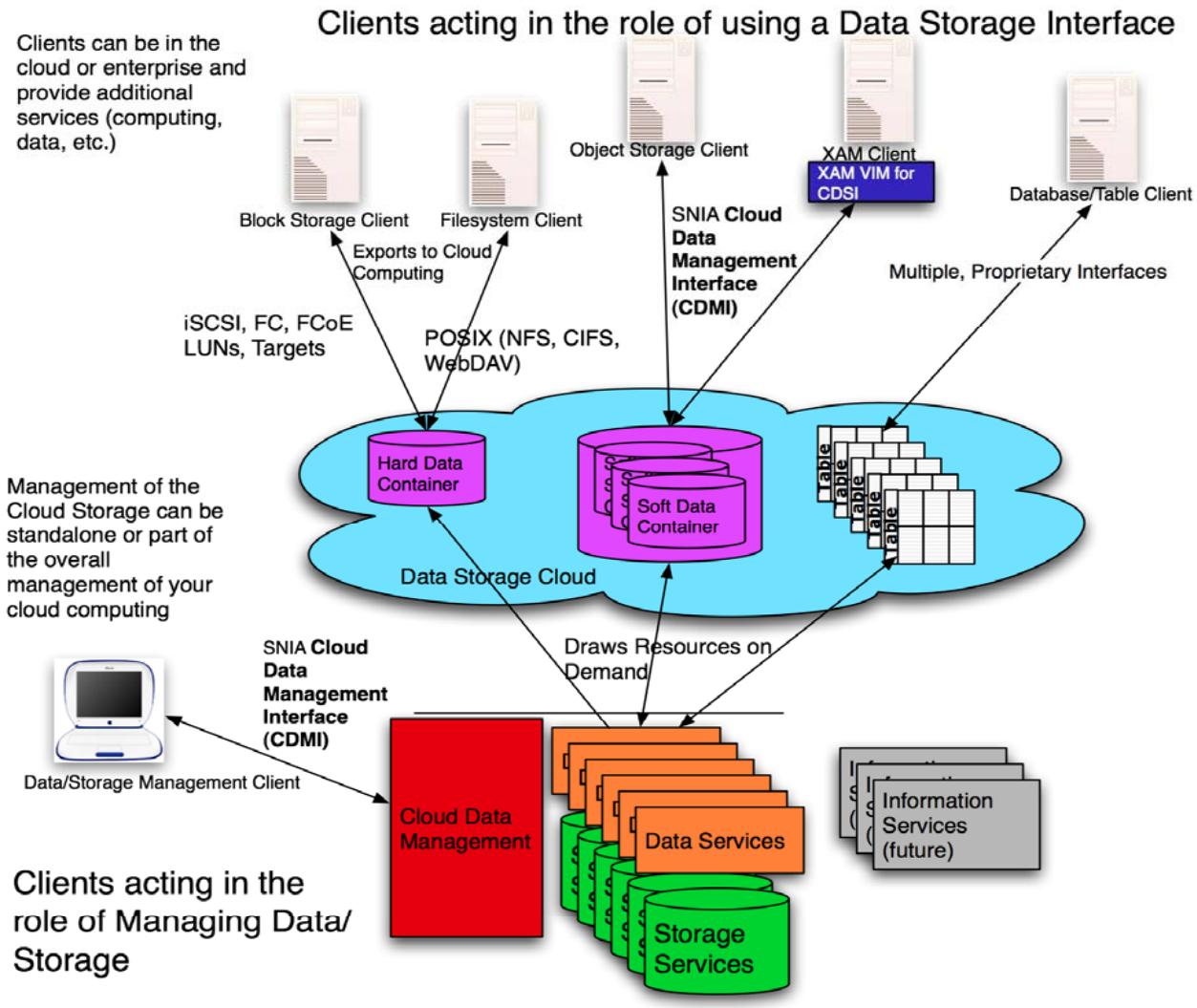




SNIA CDMI – Cloud Data Mgmt Interface

Manages the provisioning of block-oriented, file-oriented & object-oriented storage





Standardization Areas Briefed at Federal Cloud Symposium (July 15, 2009)

- Security (e.g. authentication, authorization)
- Interfaces to IaaS (e.g., compute, storage)
- PaaS & deployment formats for Cloud applications
  - *Resource descriptions (required, available)*
  - *Service & SLA models*
- Management Frameworks
  - *Governance and Policy Enforcement*
  - *Regulatory agreements (e.g. data location and security)*
  - *SLA formats (e.g. performance, availability)*
- Portable component descriptions (e.g. VM's)
- Data exchange formats (to and from Clouds)
- Cloud Taxonomies and Reference Models

Courtesy Richard Soley, OMG

Next Steps?

- Driving Cloud Adoption
  - *Which specific satellite programs/ground systems?*
  - *Which specific functions?*
- Driving Cloud Standards
  - *"Developing Cloud Standards" -- turning the adjective into a verb*
  - *Getting major stakeholders to demand standards from vendors*
- Technology Roadmapping
- Technology Demonstrations
- Leverage National Cloud Initiatives
  - *US Cloud Storefront, UK G-Cloud, Japanese Kasumigaseki*
- Leverage Existing Groups & Resources
  - *Open Cloud Consortium Testbed*
  - *Aerospace deploying corporate cloud resource & small classified cloud resource*

Open Discussion …

- Hot Button Issues
  - …
  - …
- Roadmap Requirements
  - …
  - …
- Programs Considering Cloud Computing
  - …
  - …
- Possible Demonstrations
  - …
  - …
- Available Resources (Time, Money & People!)
  - …
  - …

Contact me:
Craig A. Lee
lee@aero.org

BACK-UPS

Motivations?

- Commodification of compute infrastructure
  - *General infrastructure that can support many programs, functions*
- Improving server utilization
  - *Flexibility in mapping work to servers*
- Managing surge requirement with a pool of common resources
  - *Sizing for the average case, rather than the worst case*
- Improving reliability
  - *Easier fail-over between servers*
- Greener IT
  - *Reduce energy costs through consolidation, improved utilization & moving work to where the energy is cheaper*
- ***Many benefits will only be realized when operating "at scale"***

No Shortage of Cloud Challenges

- Data access and interoperability
  - *Must be done at the application domain level, by the domain users*
- Security
  - *Different models will expose different security threats*
- Reliability
  - *Managing redundancy, live migration, etc., across the infrastructure*
- Frameworks
  - *How to manage sets of resources, e.g., VMs and VOs?*
- Performance management
  - *What job mix needs to be supported, e.g., e-commerce, HPC, transactional, database, data streaming?*
- Costing models
  - *How to compare your own infrastructure costs with a cloud providers?*

Cybersecurity Issues

- Private cloud avoid many security issues
- Governance and Policy
- Compliance and Audit
- Tamper-proof logging, integrity monitoring tools
- Information Lifecycle Management
  - *Backups and recovery tests*
  - *Logical separation of information and protective controls*
  - *Compartmentalization of job duties*
- Incident Response, Notification, Remediation
- Users cannot instantiate arbitrary machine images
  - *Must select from an approved set of images conforming to internal policy*
- Identity Management, SAML, WS-Federation
- Virtual Organization Management, SAML and XACML
- Delegation of Trust
- Virtualization has security advantages by creating isolated environments
- VM-specific security mechanisms may need to be embedded in hypervisor APIs
- Secure communication among sites

Public vs. Private Cloud Issues

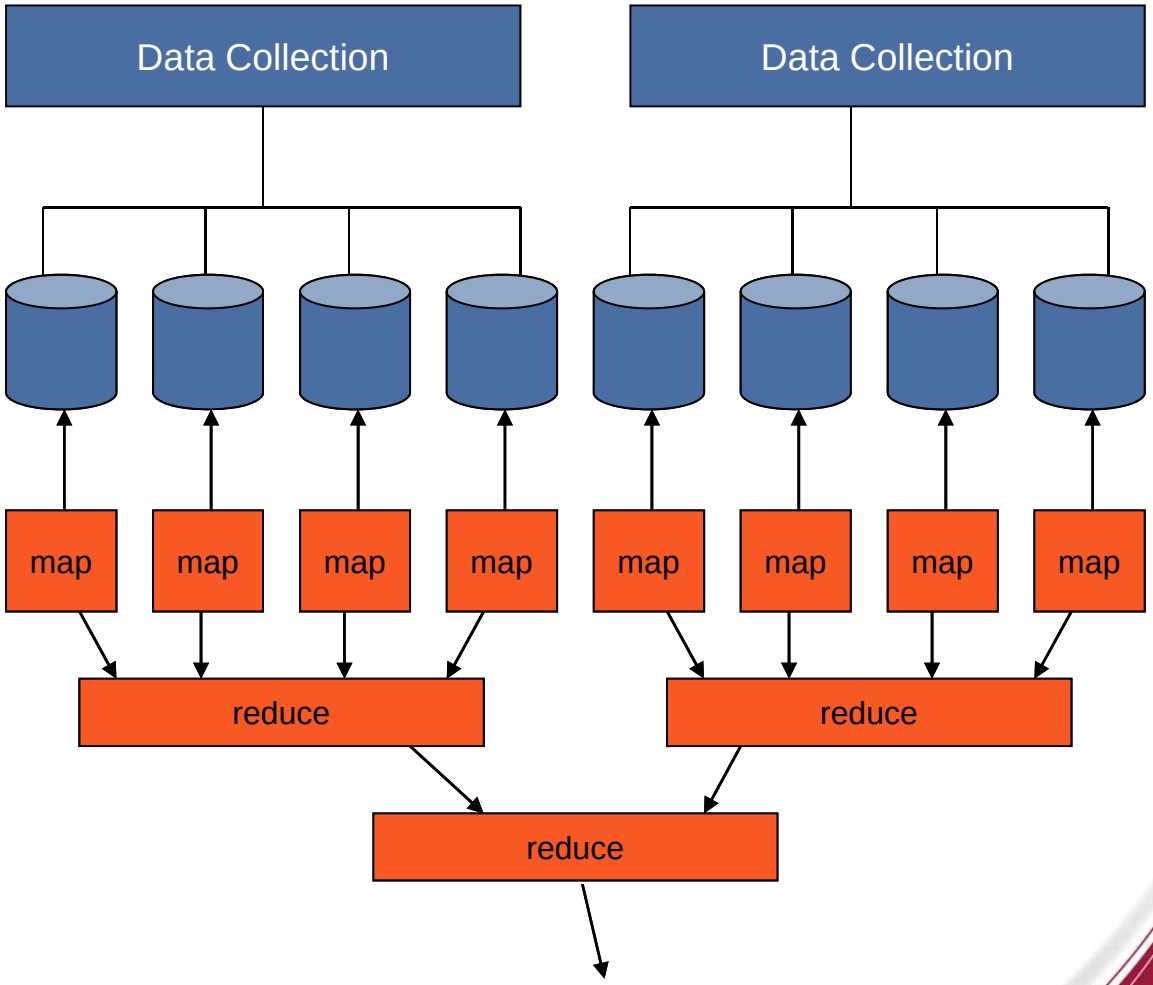
- Cost & Cost Predictability
- Users expect to monitor & manage "their" infrastructure
  - *Will a public cloud provider expose enough information for a client to troubleshoot when something goes wrong?*
- Security & Privacy
  - *You can store encrypted data in a clouds, but can you compute on it?*
- Regulation
  - *Physical location of data*
  - *Long-term audit trails (15-20 years)*
- Individual vs. Corporate Requirements
  - *Corporate use of public clouds may entail legal & contracting overheads*
  - *Ease of use and quick provisioning may tempt individuals to ignore corporate procedures*
  - *Trade-off between quick results and risk exposure*
- Internal IT departments may want to offer their own "seed cloud”
- Interoperability & portability between private and public clouds

Map-Reduce

Example: counting the number of occurrences of each word in a large set of documents

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

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));
```



Hadoop

- Implements MapReduce using the Hadoop Distributed File System (HDFS)
  - *MapReduce divides applications into many small blocks of work*
  - *HDFS creates multiple replicas of data blocks for reliability, placing them on compute nodes around the cluster*
  - *MapReduce can then process the data where it is located*
- Efficient: by distributing the data, can process it in parallel on the nodes where the data is located
- Reliable: automatically maintains multiple copies of data and automatically redeploys computing tasks based on failures
- Scalable: can reliably store and process petabytes
- Demonstrated on clusters with 2000 nodes
  - *Current design target is 10,000 node clusters*
- An open source volunteer project under the Apache Software Foundation
  - *hadoop.apache.org/core*

Other APIs / Interfaces