

Clouds and Other Computational Frameworks

Everett Toews, Cybera Inc.



Todd King, UCLA



Presentation Overview

- The cloud can be a great fit for your computational and storage needs
 - Projects overview
 - The cloud
 - Simple Storage Service (S3)
 - SPASE
 - Advantages/Disadvantages
 - Conclusion

Space Science IT Projects in Alberta

- CANARIE Network Enabled Platforms for Space Science
- **Cloud-Enabled Space Weather modeling and data assimilation Platform (CESWP)**
 - www.ceswp.ca
 - Geographically distributed cloud infrastructure
 - Use cases
 - Virtualize CSSDP
- **Canadian Space Science Data Portal (CSSDP)**
 - www.cssdp.ca
 - Data federation
 - Collaboration
 - Scientific workflows
 - SPASE utilization

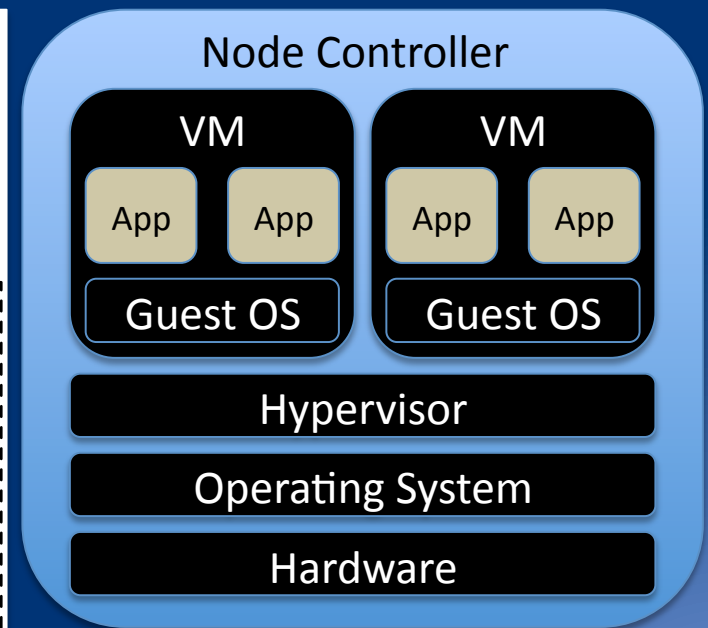
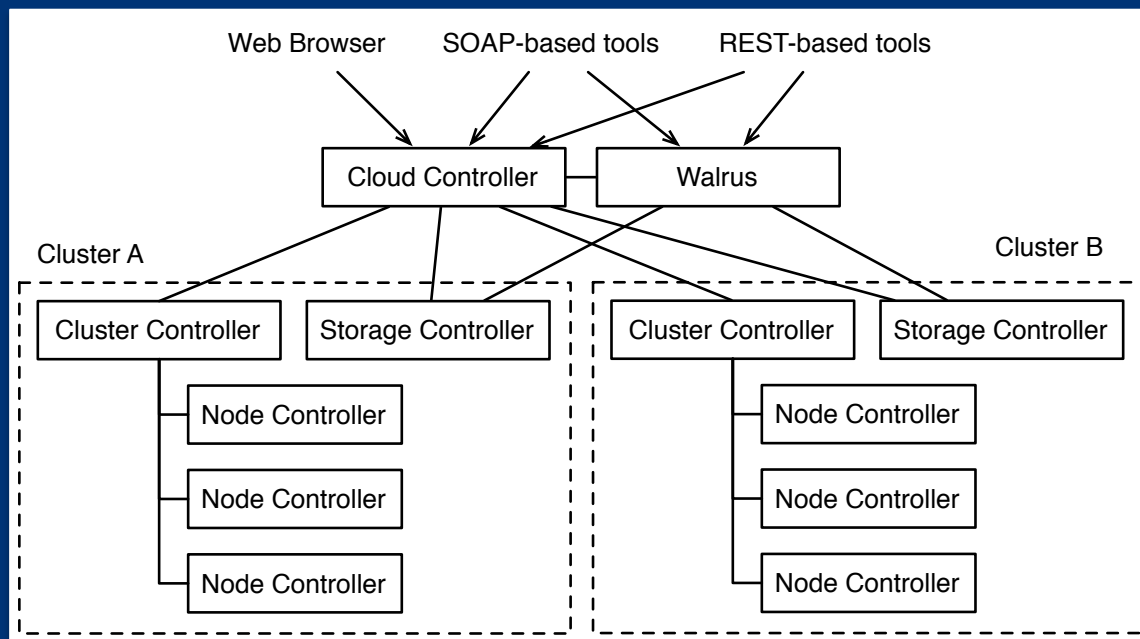
The Cloud - A Working Definition

NIST: <http://csrc.nist.gov/groups/SNS/cloud-computing/cloud-def-v15.doc>

- **Essential Characteristics:**
 - **On-demand self-service**
 - **Broad network access**
 - **Resource pooling**
 - **Rapid elasticity**
 - **Measured service**
- **Deployment Models:**
 - **Private cloud**
 - **Community**
 - **Public cloud**
 - **Hybrid cloud**
- **Service Models:**
 - **Software as a Service (SaaS):** web-based email
 - **Platform as a Service (PaaS):** Google AppEngine
 - **Infrastructure as a Service (IaaS):** Amazon Web Services

CESWP Focus: IaaS

- Amazon Web Services (AWS)
- Eucalyptus (AWS API Compatible)



CESWP Logical View (circa Sept. 2011)



CESWP Use Cases

- Collaboration
- Model development
- Run simulations
- SPASE utilization



Provenance

Amazon's Simple Storage Service (S3)

- Amazon S3 provides a simple web services interface that can be used to store and retrieve any amount of data, at any time, from anywhere on the web
 - Eucalyptus Walrus is S3 compliant storage (subset)
- The storage model is a bucket
 - A bucket can contain unlimited objects up to 5 GB each of any type (text file, photo, video, binary, etc.)
- Supports both a REST and SOAP API over HTTP
 - Retrieve using HTTP GET
 - <http://bucketname.s3.amazonaws.com/ObjectName>
 - Retrieve using HTTP GET with a Torrent
 - <http://bucketname.s3.amazonaws.com/ObjectName?torrent>

S3 Security

- Query String Authentication
 - Create a URL that is valid for a limited amount of time
- Access Control List
 - Grant permissions (read, write) to the owner, AWS account holders, anyone with an AWS account or to the public
- Bucket Policy
 - Access control for buckets and the objects in them
 - Regular expressions, IP addresses, dates, user agents, the HTTP referrer, and transports (http/https)

Describing S3 sources with SPASE

- Data in S3 storage can be described with SPASE metadata.
- If the data resource is all stored in one S3 bucket then in the NumericalData resource it can be referenced as:

```
<AccessInformation>  
<RepositoryID>spase://amazon</RepositoryID>  
<AccessURL>  
  <URL>http://example.s3.amazon.com</URL>  
</AccessURL>  
<Format>CDF</Format>  
<Encoding>S3_Bucket</Encoding>  
</AccessInformation>
```

← Encoding value "S3_bucket" not currently in SPASE

- Each object in the bucket is a Granule:

```
<Granule>  
  <ResourceID>spase://vxo/Granule/Example/granule1</ResourceID>  
  <ReleaseDate>2010-08-02T00:00:00</ReleaseDate>  
  <ParentID>spase://vxo/NumericalData/Example</ParentID>  
  <StartDate>2001-02-03T00:00:00</StartDate>  
  <StopDate>2001-02-04T00:00:00</StopDate>  
  <Source>  
    <SourceType>Data</SourceType>  
    <URL>http://example.s3.amazon.com/granule1.dat</URL>  
  </Source>  
</Granule>
```

Advantages

- Cost
- Don't need to acquire/maintain hardware
- Handles large volumes of data
- Redundant storage
- Standards compliant
- Security
- Fast connection
- Many simultaneous connections

Disadvantages

- Cost?
- Data no longer stored locally?
- Dependent on service outside your control

References

- CESWP: www.ceswp.ca
- CSSDP: www.cssdp.ca
- Eucalyptus: open.eucalyptus.com
- AWS: aws.amazon.com
- S3: aws.amazon.com/s3
- S3Fox Organizer: www.s3fox.net

Conclusion

- Questions
- Demo
 - SPASE on S3 with GUI
 - SPASE on S3 with web browser
 - SPASE on S3 with CLI
 - CESWP

Everett Toews: everett.toews@cybera.ca

Todd King: tking@igpp.ucla.edu

Tech Detail: S3 Bucket Names

To comply with Amazon S3 requirements, bucket names:

- Can contain lowercase letters, numbers, periods (.), underscores (_), and dashes (-)
- Must start with a number or letter
- Must be between 3 and 255 characters long
- Must not be formatted as an IP address (e.g., 192.168.5.4)

To conform with DNS requirements, we recommend following these additional guidelines when creating buckets:

- Bucket names should not contain underscores (_)
- Bucket names should be between 3 and 63 characters long
- Bucket names should not end with a dash
- Bucket names cannot contain two, adjacent periods
- Bucket names cannot contain dashes next to periods (e.g., "my-.bucket.com" and "my.-bucket" are invalid)

To use the REST API bucket names must conform to DNS requirements. In addition bucket names can only contain lowercase letters.

Tech Detail: S3 Object Names

- An Object Name is also called a "key"
- A key is a sequence of Unicode characters whose UTF-8 encoding is at most 1024 bytes long
- Keys often have a suffix that describes the type of data in the object. For example, ".jpg" indicates that an object is an image

Tech Detail: S3 API

- Supports both a REST and SOAP API

REST API

- Uses HTTP Request and Response headers to transfer metadata.
Example: When putting an object Content-MD5 is used to verify integrity of transfer.

Retrieving an object in S3

- Using HTTP GET
`http://bucketname.s3.amazonaws.com/ObjectName`
- Using HTTP GET with a Torrent
`http://bucketname.s3.amazonaws.com/ObjectName?torrent`