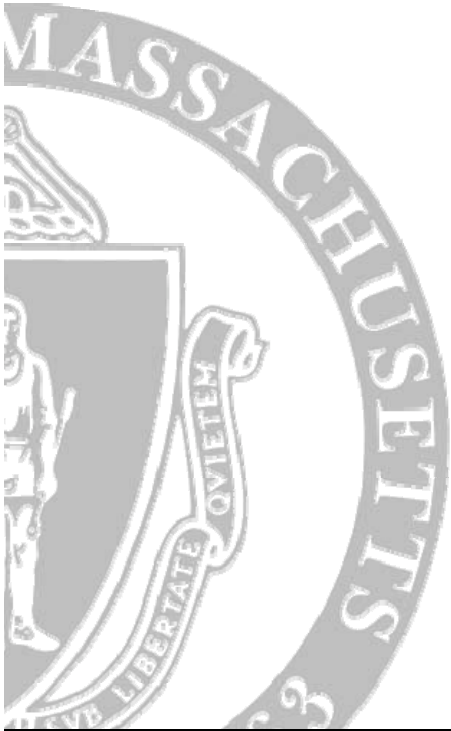


Data-Intensive Cloud Control for GENI

Cluster D Session

Michael Zink, Prashant Shenoy, Jim Kurose,
Emmanuel Cecchet, and David Irwin

November 16th, 2009

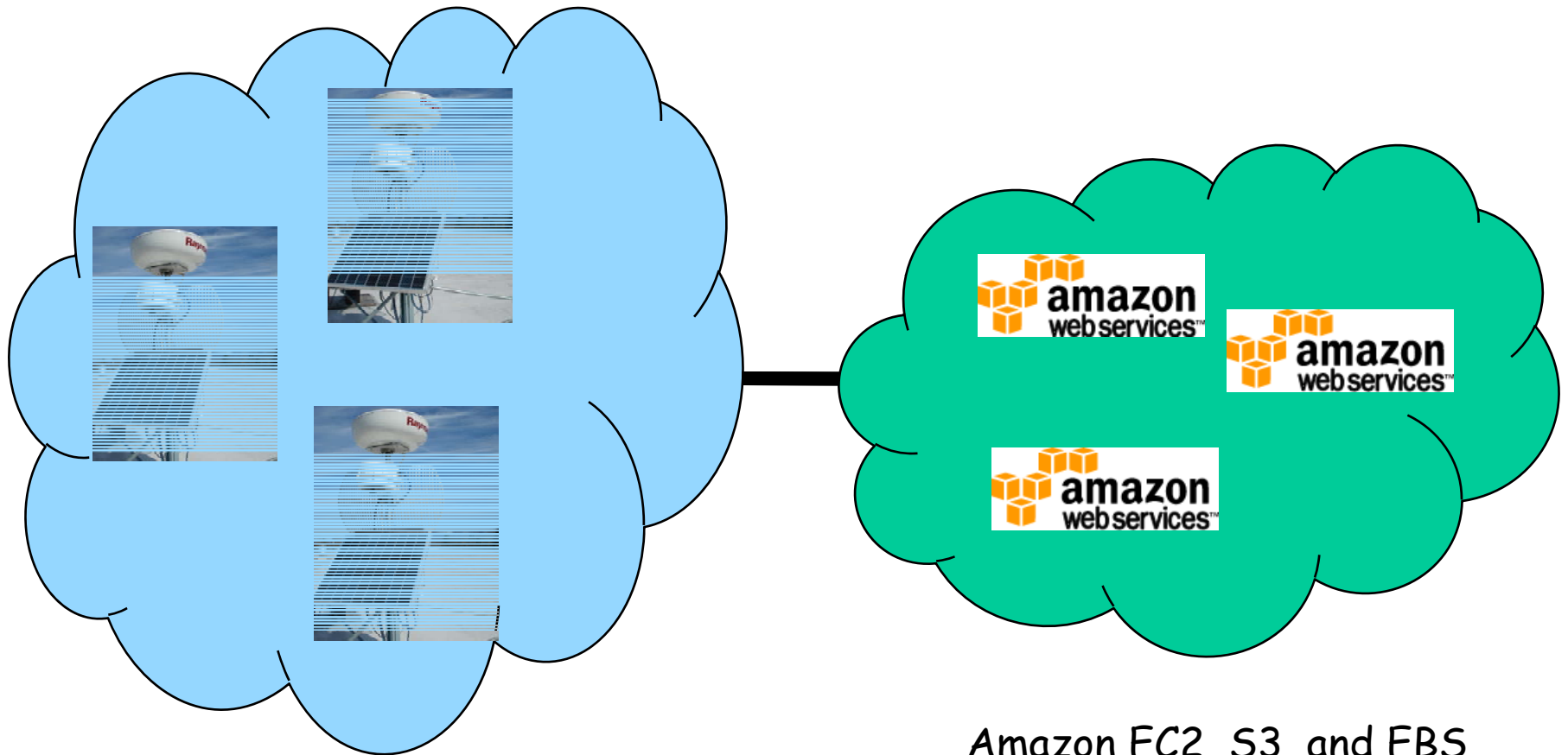


Project Overview

- Support researchers conducting *data-intensive* exp.
 - Sensors → network → storage volumes

- Extend ViSE/Orca with
 - *Data-centric Slices*: storage as a first-class resource
 - *Cloud Computing*: connect Amazon Services to GENI
 - *Experiment Workflow*: interface for executing experiments

Example Slice



ViSE Sensor Network

Amazon EC2, S3, and EBS

Simple End-to-End Experiment Example

- 1 Collect high-bandwidth radar data
 - 10s to 100s of Mbits/second
- 2 Transmit data over network to backend storage volumes
 - Storage volumes leased from "the cloud"
- 3 Process/Merge data from overlapping sensors
 - Feature detection: tornadoes, hail, etc.
 - Weather forecasting

Year 1 Goals

- Enable GENI to allocate resources from Amazon's cloud
 - Lease EC2 servers, EBS volumes, and S3 objects "for free"
 - GPO has allocated a per-year cloud budget
- Enable leasing of storage-centric resources
 - EBS volumes and S3 objects
 - Critical for data-intensive sensing experiments
 - "Stitch" together resources: storage + processing

Year 1 Goals Cont'd

- Develop budget-based resource allocation policy
 - Ensure researchers don't go over our allotted budget
 - Experiment with virtual currency policies
- Allow researchers to store/archive/process data from sensors and other components in the cloud

Year 2 Goal

- Incorporate higher-level programming paradigms
 - MPI for tightly-coupled parallel algorithms
 - Hadoop for loosely-coupled data processing
- Make it easier for researchers to “do something” with their data!
 - Example: generating radar reflectivity data from raw moment voltage data is simple MapReduce job

Year 3 Goal

- Incorporate Experiment Workflow tool
 - ViSE portal is an interface to request resources
 - Does not have mechanisms to program resources for a specific experiment
 - "Orca-enable" Gush for this purpose
- Make it easier for researchers to program diverse "end-to-end" slices!

Questions?
