

Data-Intensive Cloud Control for GENI

GEC 8 demo
Orca control framework
July 20, 2010

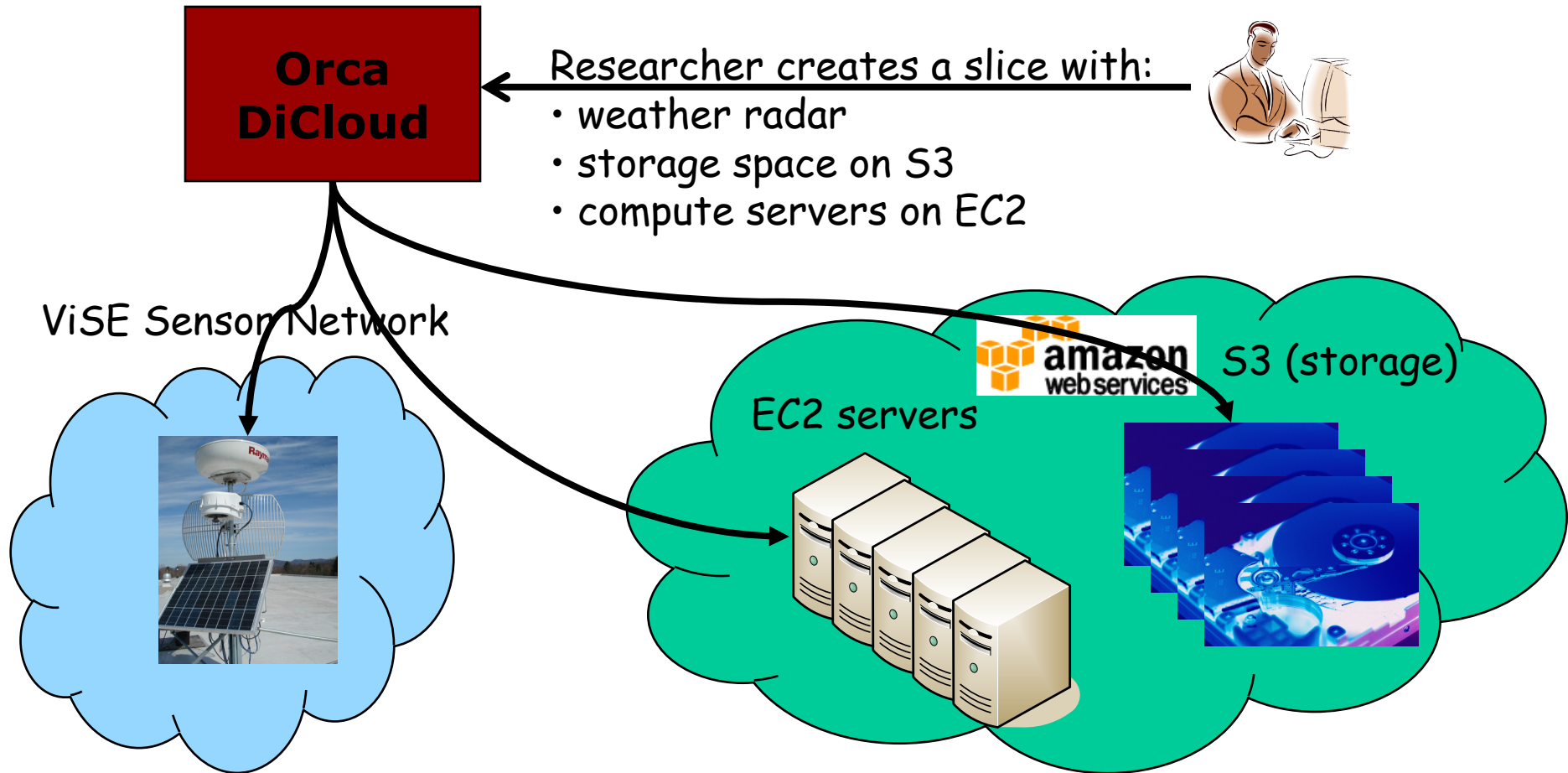
Michael Zink, Prashant Shenoy, Jim Kurose,
David Irwin and Emmanuel Cecchet
cecchet@cs.umass.edu

DiCloud Project Overview

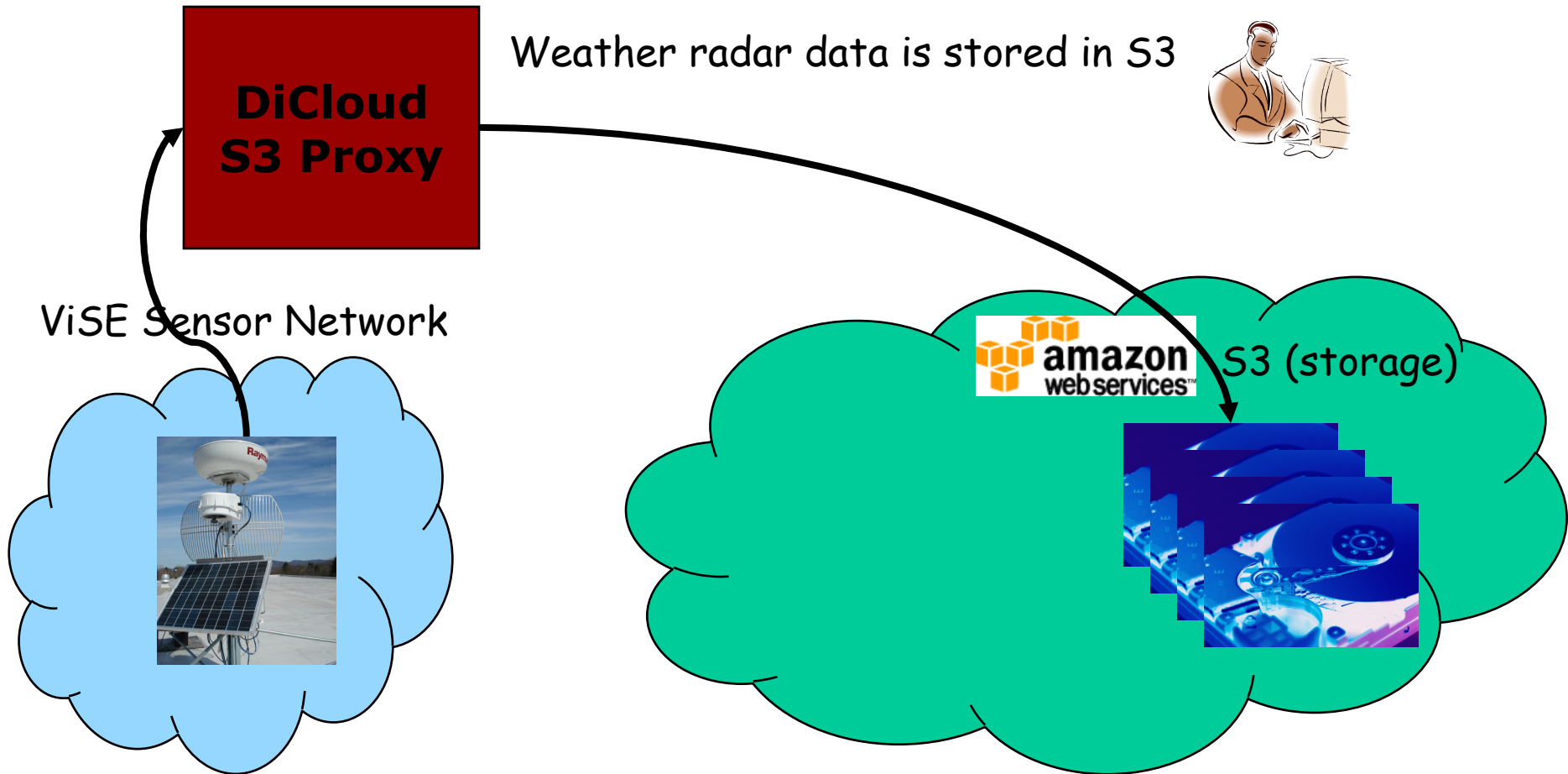
- Support researchers conducting *data-intensive* experiments
 - Sensors → network → storage volumes → processing

- Extend Orca with
 - *Data-centric Slices*: storage as a first-class resource with Amazon Simple Storage Service (S3) and Elastic Block Storage (EBS)
 - *Cloud Computing*: connect Amazon Elastic Compute Cloud (EC2) resources to GENI

Demo - Step 1 (Allocate resources)

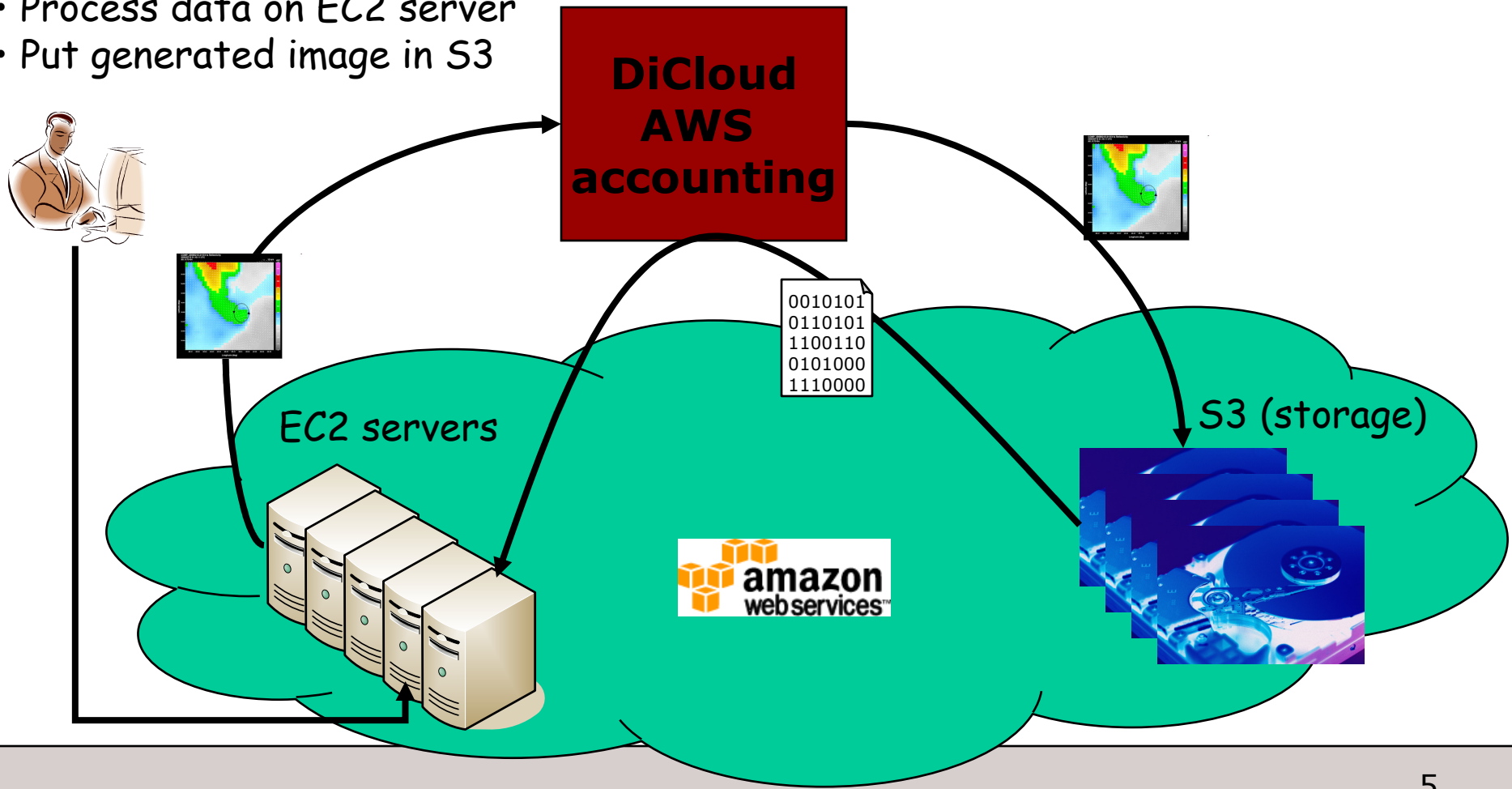


Demo - Step 2 (Storing data)

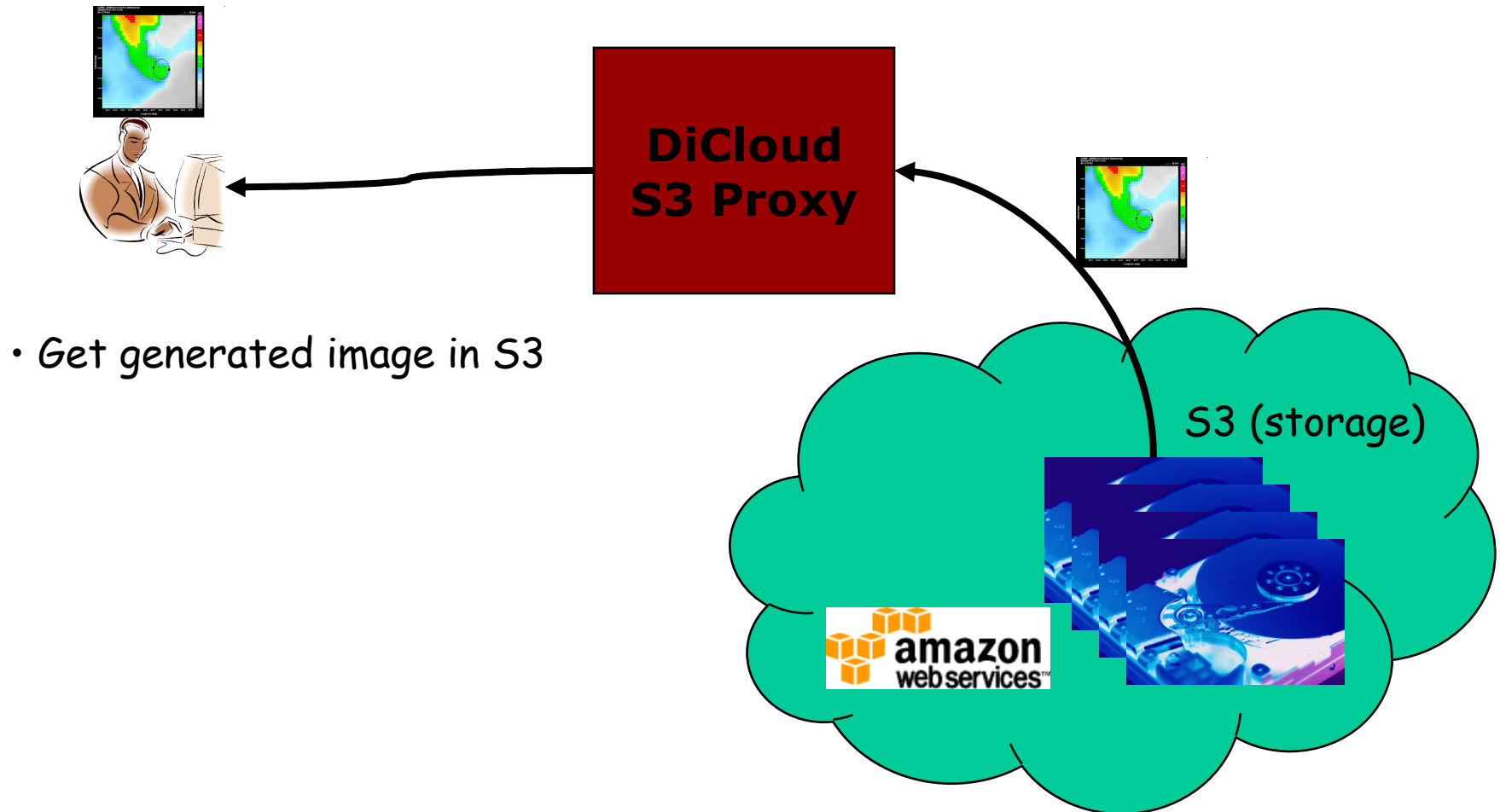


Demo - Step 3 (Processing data)

- Get radar data from S3
- Process data on EC2 server
- Put generated image in S3



Demo - Step 4 (Visualizing data)



Technical details

- Implemented on top of Amazon Web Service API
- DiCloud monitors
 - EC2 Server hourly usage
 - Network usage (in and out traffic using CloudWatch statistics)
 - S3 storage space and put/get operations using a dedicated proxy
 - EBS disk usage (storage space and number of IOs using CloudWatch statistics)
- All accounting operations are logged in a database
- Resources are automatically revoked when budget has expired

Challenges

- Network connectivity with AWS
- Monitoring
 - CloudWatch does not differentiate between free and paying network traffic or disk IO
 - S3 Proxy in the cloud would save network traffic and cost
- What security model for storage
 - How to share data with others?
 - Storage leases can last years
- What budget to allocate to make this useful?