A 501(c)(3) not-for-profit operating clouds for science.

The Open Science Data Cloud

April 4, 2012

Robert L. Grossman

University of Chicago
and Open Cloud Consortium

Open Cloud Consortium

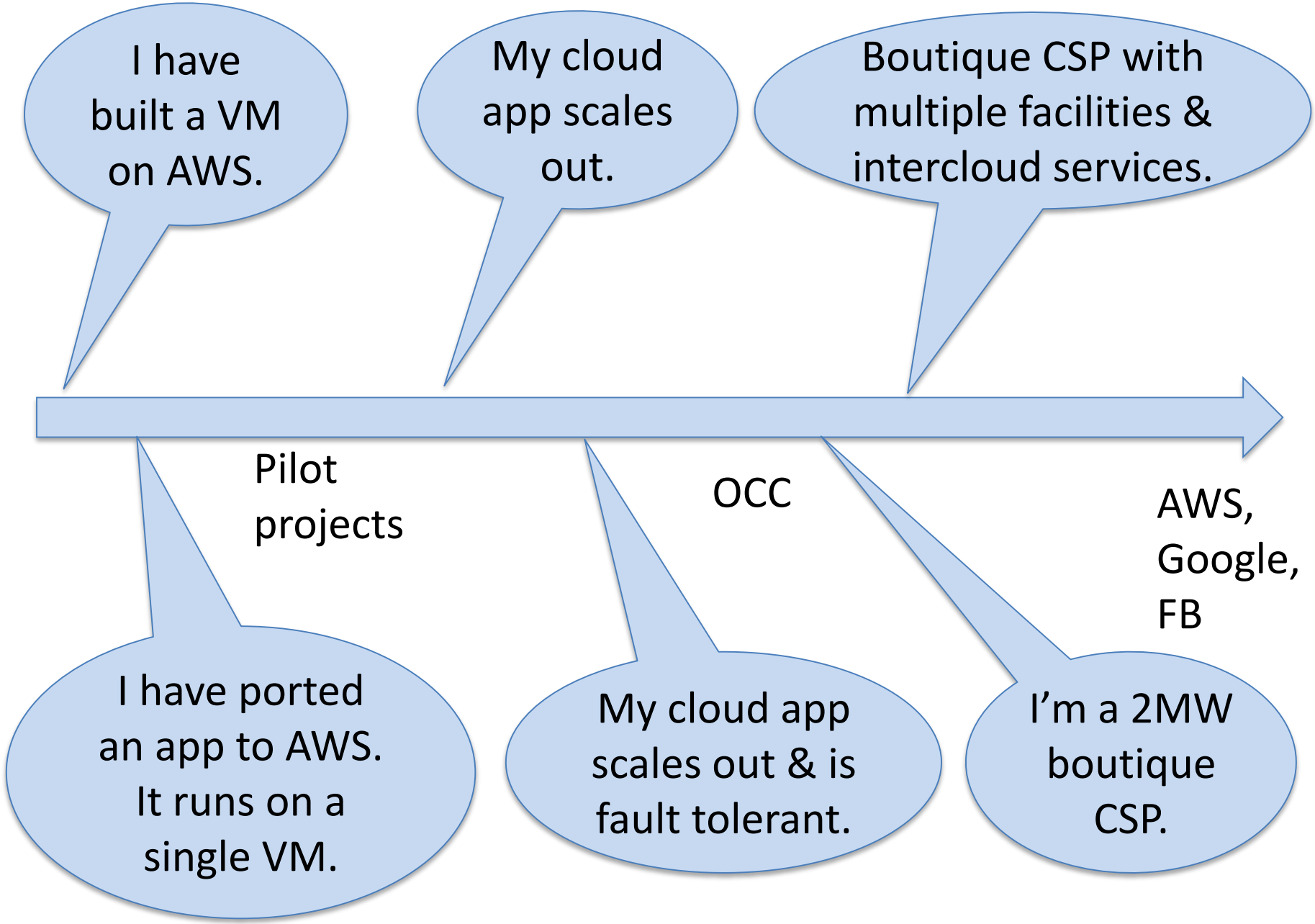
- U.S based not-for-profit corporation.
- Manages cloud computing infrastructure to support scientific research: Open Science Data Cloud.
- Manages cloud computing testbeds: Open Cloud Testbed.
- Develops reference implementations of standards based software for clouds

www.opencloudconsortium.org


Number
1000's
100's
10's
amazon web services
Individual scientists &
small projects
Open Cloud Consortium
Community based
science via Science as a
Service
LSST
OOI
OCEAN OBSERVATORIES INITIATIVE
CERN
very large projects
Data Size
Small
Medium to Large
Very Large
Public
infrastructure
Shared community
infrastructure
Dedicated
infrastructure

Varieties of Clouds

- Types of clouds
  - Utility Clouds (e.g. Amazon)
  - Data Clouds (e.g. Hadoop)
  - Storage Clouds (e.g. Dropbox)
  - HPC Clouds
- Types of cloud service providers (CSP)
  - Commercial (e.g. Amazon)
  - Science CSP
  - Health related (CSP) (sign BAA)
  - Defense related (CSP)


I have built a VM on AWS.
My cloud app scales out.
Boutique CSP with multiple facilities & intercloud services.
Pilot projects
OCC
AWS, Google, FB
I have ported an app to AWS. It runs on a single VM.
My cloud app scales out & is fault tolerant.
I'm a 2MW boutique CSP.

Science Clouds

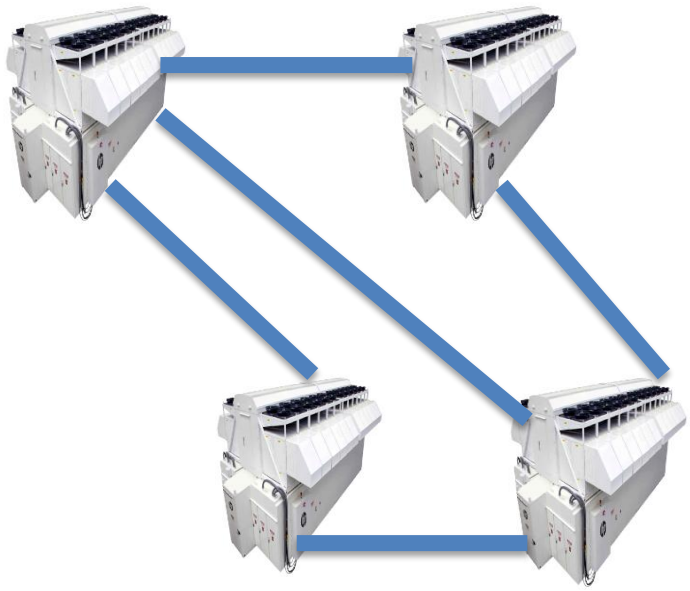
	Science Clouds	**Commercial CSP**
Perspective	Democratize access to data. Integrate data to make discoveries. Long term archive.	As long as you pay the bill; as long as we keep our business model.
Data	Data intensive computing	Internet style scale out
Flows	Large data flows in and out	Lots of small web flows
Lock in	Data liberation	Lock in is good

New Opportunities

1. Simplify the management, analysis and sharing of data, especially big data for researchers.
2. New opportunities for discovery by integrating data at scale from within a discipline and across disciplines.
3. By thinking of a data center as an instrument, you have a new instrument for creating a new science of data.

Key Challenges

- Building the required infrastructure at the required scale.
- Developing new software tools for exploring and integrating data at scale.
- Brining together (either physically or virtually) enough data to change the way we make discoveries.
- Security, compliance & payments.



Next Steps

- Start some pilot projects.
- Create some data hubs.
- Create some large scale services between the data hubs.
- Create some open source blue prints for container based facilities.

Questions?

robert.grossman@uchicago.edu

info@opencloudconsortium.org