OSG News

Frank Würthwein
OSG Executive Director
Professor of Physics
UCSD/SDSC

Two Slides of Standard PR

The Scope of Open Science

- All of open science irrespective of discipline
- Advance the maximum possible dynamic range of science, groups, and institutions
  - From individual undergraduates to international collaborations with thousands of members.
  - From small colleges, museums, zoos, to national scale centers of open science.
- Advancing this entire spectrum requires us to have a diversified portfolio of services

OSG serves 4 distinct groups

- The individual researchers and small groups on OSG-Connect
- The campus Research Support Organizations
  - Teach IT organizations & support services so they can integrate with OSG
  - Train the Trainers (to support their researchers)
- Multi-institutional Science Teams
  - XENON, GlueX, SPT, Simons, … many more
  - Collaborations between multiple campuses
- The 4 "big science" projects:
  - US-ATLAS, US-CMS, LIGO, IceCube

NSF CC* Program

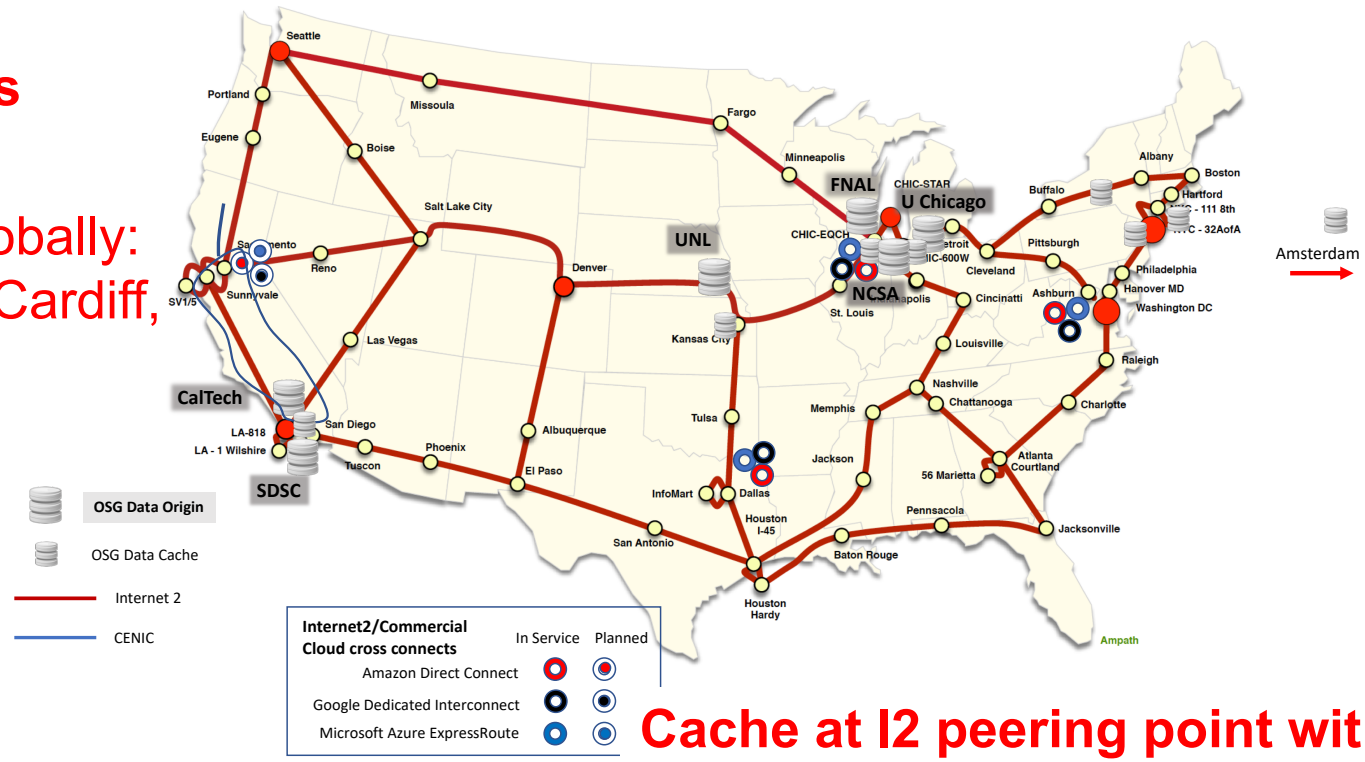
- NSF funded 12 clusters at various institutions at ~$400k each.
- Each of these pledged in their proposals to make 20% of their capacity available to the general community via OSG.
- We had an initial workshop to engage with these institutions, and are bringing them up via hosted CEs.
- More on this in Lauren's presentation tomorrow.

Data Federation Update

OSG Data Federation

6 Data Origins
12 Data Caches

Caches deployed globally: Amsterdam, Korea, Cardiff, … more coming.



Cache at I2 peering point with Cloud providers in Chicago

Reads from Data Federation 9/1/2018-2019

Dune	~ 2.6PB
LIGO public	~ 1.5PB
LIGO private	~ 0.5PB
DES	~ 1.1PB
Minerva	~ 1.0PB

Depending on community, files were read 10-30,000 times during typical 60 day period.

Data Federation Goals

- People come with their data on their storage systems.
- OSG offers to operate a Data Origin Service to export your data into the OSG Data Federation.
  - We give you a globally unique prefix for your filesystem namespace, and then export your namespace behind it.
  - We allow you to decide who can access what.
- OSG then strives to guarantee ”uniform” performance across the nation by operating caches to:
  - Hide Access Latencies
  - Reduce unnecessary network traffic from data reuse
  - **Protect the data origins from overloads**

OSG operates overlay system(s) as services to all of science

New Deployment and Operations Paradigm

OSG has started offering services as containers that can be deployed via a container orchestration system.

We are presently using Kubernetes for that.

We are presently planning to adopt SLATE for that.

Long Term Vision

- Capacity Providers
  - Commercial cloud “competing” with on-premise
  - Different regions in the world will invest differently, and yet, capacity needs to be integrated globally.
- Service Providers
  - Software based services
  - Human based services (“consulting, training, …”)
  - “Content” providers
- Scientists organized at all scales
  - Individuals to 1000’s of collaborators

Increased Engagement with Cloud

We are engaged in two mutually supporting cloud projects.

A cloud GPU burst for IceCube

An IO bandwidth and latency measurement campaign

Cloud Bursting Proposal

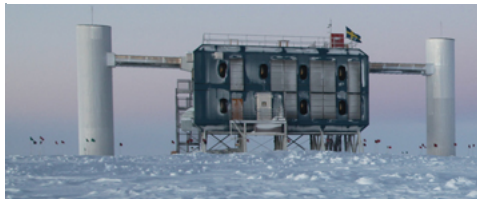
- NSF award: Use 80,000 V100s for 1h to process 250TB of input data that generates 500TB of output data. => Exaflop hour in commercial Cloud . $270,000 for the burst hour + $25k for testing & R&D ahead of time plus storage.
- Reality: 80,000 V100 equivalent GPU capacity on demand does not exist today. We understood capacity limit only after proposal submission.
- Our response: **Want to buy the entire global GPU capacity that is for sale across AWS, Azure, and Google.**
  - Prepared to run on any GPU accelerator (K80 and up) in an x86 host system anywhere in the world. Stage input/output data as necessary. Working with Internet2 to achieve necessary connectivity to providers.

Tentative Date: Saturday November 16th 2019

Depending on how much capacity we can buy that day, we may try again on a more obvious vacation date, e.g. thanksgiving or Xmass.

Cloud Bursting Team

Frank Wuerthwein, Executive Director of OSG



Benedikt Riedel, Computing Manager IceCube

Miron Livny, Lead of HTCondor & Director of CHTC

Larry Smarr, PI of PRP

Howard Pfeffer, President and CEO of Internet2

Mike Norman, Director SDSC & Vince Kellen, CIO UCSD

Questions ?