Enabling Volunteered Cloud Computing for Disaster Management

Earth Science Technology Forum 2014
October 28, 2014

John D. Evans, Wei Hao, Samir Chettri
Global Science & Technology, Inc. (GST)

NASA ESTO Earth Science Technology Office

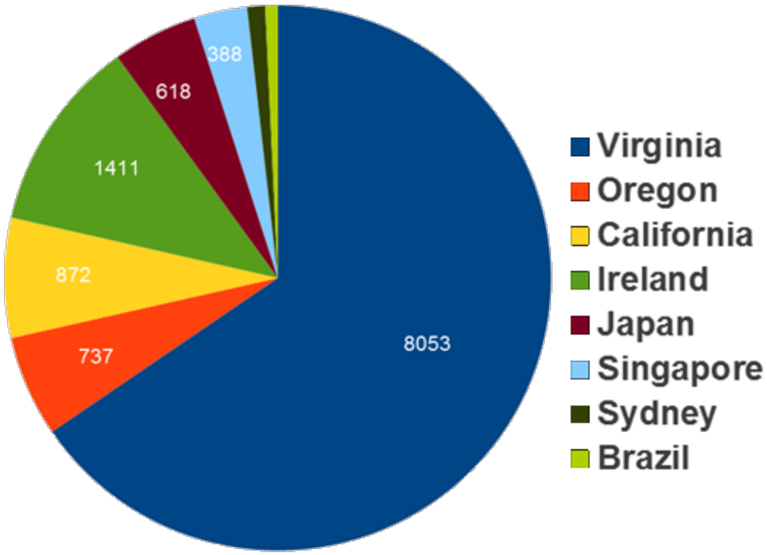
Global Science & Technology, Inc.

Key concepts

- *Volunteered Computing*
  - Broad public participation in large computations, using slack resources on desktop computers
    - SETI@Home, Folding@Home, PrimeGrid, …
  - *However:* Slack capacity may diminish in coming years, as mobile devices gain market share
- Might *Volunteered Cloud Computing* (VCC) fill the gap?
  - Participants contribute cloud computing resources
  - Esp. resources paid for but unused or underused
    - Virtual machine time, virtual storage volumes, *etc.*
- VCC may also offer improvements in…
  - Bandwidth
  - Scalability
  - Flexibility for complex code or data
  - Optimal use of cloud computing resources

Feasibility

- Amazon Web Services computing resources:
  - *(March 2014 estimate)*
  - 12,300 racks worldwide (incl. ~8,000 in us-east-1)
  - 12M CPU cores worldwide (8M in us-east-1)



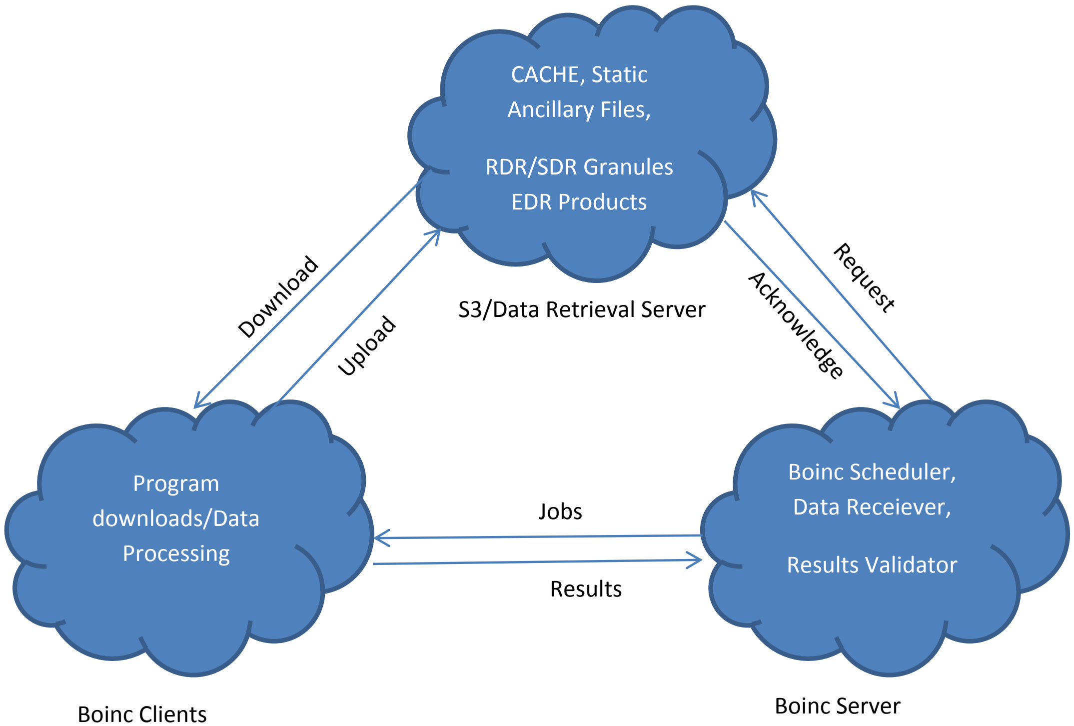
Amazon EC2 server racks by region

- AWS Volunteered cloud computing resources:
  - *(@ 0.1% participation)* 1,600 core-hours/hr worldwide (900 core-hours/hr in us-east-1)
- Disaster mgmt. may provide a compelling cause

Platforms and Applications

- Candidate distributed computing platforms:
  - Hadoop / MapReduce
  - Berkeley Open Infrastructure for Networked Computing (BOINC)
  - Parabon / Frontier© Grid Platform
- Candidate computing applications:
  - Hurricane / storm coastal surge model
  - Landslide modeling
  - CREST hydrology model
  - Near-Real-Time processing of S-NPP VIIRS observations

Proof of concept

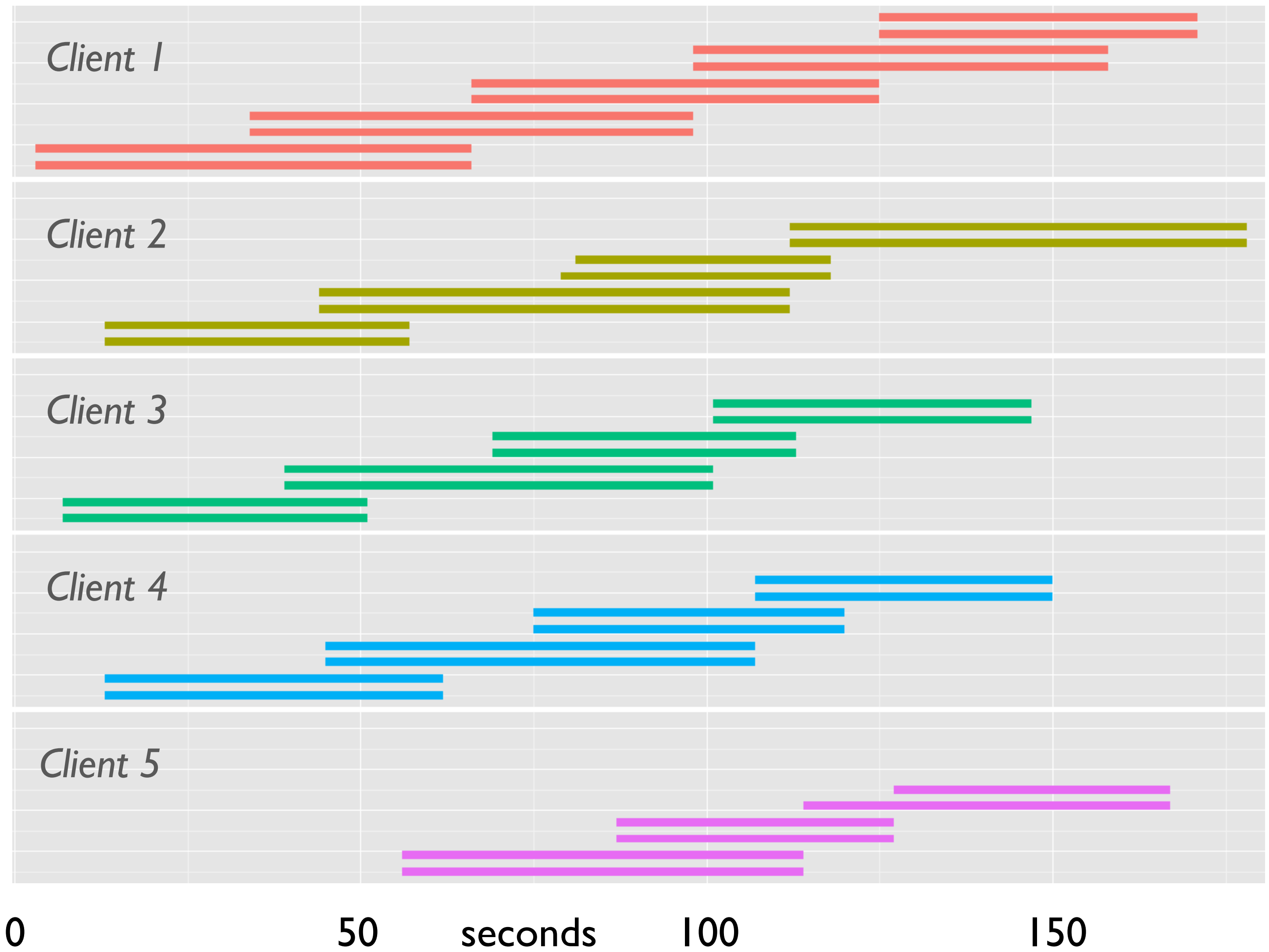


- Using BOINC protocol applied to
  - CREST hydrology model runs
  - S-NPP VIIRS fire detection algorithm

Preliminary experiments

- CREST model ensembles:
  - Deploy ("stage") 20 input datasets for download
    - Random perturbation of rainfall estimates
  - Create 20 work units for execution
    - 1 per scenario
  - Use the BOINC dashboard to monitor progress on work units
  - Once results are complete & validated, review the (20) overland flow rates for a particular day.

Work unit execution



Preliminary findings

- Simple proof of concept
  - Automated data management processes
  - Transferring many files
    - Bundled into .zip files, with pre/post scripts
- Scalable to larger computing tasks
  - S-NPP VIIRS near real time processing
  - Manage data files on AWS Simple Storage Service (S3)
    - Clients receive S3 URLs for data input files
    - Clients upload outputs to S3 via signed URLs
  - Many server & client configuration options

Ongoing work & future plans

- Extend basic capabilities:
  - Larger data volumes (VIIRS SDR + ancillary)
  - "Interstitial" computing
    - Use the remainder of a billing hour for free
  - "Repurposing" clilent machines for more complex code / data dependencies
    - Virtual storage volumes
    - Virtual machine images (e.g., Docker containers)
- Capture insights & recommendations
  - Disaster Management use cases
  - Broader applicability
- Engage volunteers & scale up

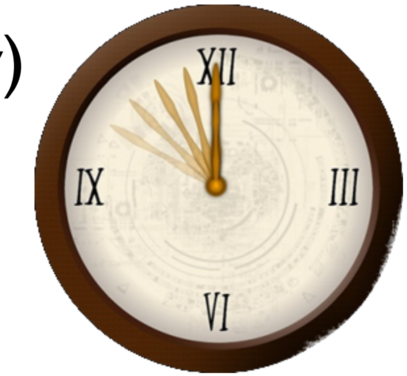


Image © deux-chi via flickr.com

Discussion

Or contact: john.evans@gst.com

Global Science & Technology, Inc.