ESGF: Compute Working Team Progress update & future roadmap

ESGF F2F Workshop, Washington, DC, December 2016

Charles Doutriaux
Tom Maxwell, Jason Boutte, Dan Duffy, Dean N. Williams,
ESGF Compute Working Team

December 6th, 2016

Lawrence Livermore National Laboratory

Major Achievements

- General API for CWT Servers
  - WPS based with CWT extensions
    - Domain, Variable,
  - No Restriction on Server Side Implementation
  - Language Independent

- Toy Server (http:/aims2.llnl.gov)
  - Open Source/Python-based: https://github.com/ESGF/esgf-compute-wps

- Several groups already started to independently to develop API compliant Servers
  - NASA (scala)
  - Ouranos (Birdhouse)
  - CMCC (Ophidia)

- Python End User API

ESGF
Earth System Grid Federation

Relevant Survey Responses

Question	# Responses	% Responded	Weight	Score
Access to large volumes of data with computational resources for server-side (i.e., remote) analysis and visualization	119	34.69%	4.22	1.46
Access to enough computational and storage resources	99	28.86%	3.89	1.12
Direct data delivery into ESGF computing systems from distributed data resources	95	27.70%	3.99	1.10
Reproducibility	106	30.90%	3.57	1.10
Web documentation	128	37.61%	3.89	1.46

Alignment with User Requirements

- **ESGF CWT Charge:** Develop general APIs for exposing ESGF distributed compute resources (such as computer clusters, cloud servers and HPCs) to multiple analysis tools
  - *API* - Well-defined interface to large-scale, distributed, data-proximal analytics and visualization capabilities for data accessible through ESGF.
  - *Analytics Operations (canonical operations)* - A set of analytical operations that can be accessible through the API for server-side and distributed analytics, such as sub-setting, averaging, variation and anomaly calculations, etc.
  - *Compute Platform* - Backend high-performance computational platforms to allow for both server-side processing capabilities and the ability to perform distributed analytics.

- The CWT is in direct alignment with the needs of the user community and their responses in the survey.

Roadmap to 2020 - APIs

- End-user API documentation (2017)

- Workflows (2017-2018)

- Services naming (2017-2018)

- Authentication (2017-2018) (if necessary)

- Caching (2017-2018)

Roadmap to 2020 - Features

- Implement Services Needed by Community (Ongoing)

- Authentication fully integrated (2017)

- Caching (2017-2018)

- Resource Management (2017-2018)
  - Is a user authorized to access data?
  - How much CPU/download is a user allowed?

- Pick from various implementation available on server (user can override this) (2017: via user-api, 2018: automated)

- Based on resources, auto pick best server from all available ones providing the requested service (2019: via api, 2020: automated)

ESGF
Earth System Grid Federation

Collaborations Needed

- Authentication (resources allocation, data access, etc...)

- Node Manager (to know status and holdings of all nodes)

- Install Team (for compute node install)

ESGF
Earth System Grid Federation

Resources

- Email: esgf-cwt@llnl.gov

- Webex meetings:
  - First Monday of the month: General Meeting
  - Third Monday of the month: Implementation Meeting

- Documentation
  - Web (requires login)
  - API

- Code (github)
  - Server: https://github.com/ESGF/esgf-compute-wps
  - End-user: https://github.com/ESGF/esgf-compute-api