ESGF Compute Node Certification and ESGF Compute Service Certification

White Paper

Charles Doutriaux
On behalf of
ESGF Compute Working Team

Table of content

Abstract

With CMIP6 rapidly approaching and requiring the accessibility of petabytes of data, it is imperative that the ESGF compute team's work be available in order to help scientists process volumes of data that they cannot possibly download to their institutions in a timely manner.

The CWT already has some services in place and a few servers (here and here) are ready to be deployed.

However during the 2017 ESGF F2F in San Francisco, concerns were raised about which services will be available and how the quality of the services and operational management (or servers) offered could reflect poorly upon ESGF. In this document "server" refers to a "Compute" node, and all aspect discussed here relate to the "Compute" aspect of the server/node. It is quite possible that the requirements expected by ESGF for other nodes (such as data nodes) would be different.

In order to continue fostering creativity and innovation w/o compromising ESGF's reputation the ESGF CWT proposes to establish a set of rules for both the servers and services to obtain an official ESGF certification. We envision two separate certifications, one to ensure the server itself is robust enough, and another one to ensure the services offered are reliable and compatible with ESGF. This approach is similar to data nodes having Tier1 and Tier2 groups, a "certified" server would be the equivalent of a data node "Tier 1". Certification will not be required to participate in ESGF, but will reflect a higher degree of confidence from ESGF.

These certifications will be issued for a year, and reviewed every year. Also on an annual basis the ESGF CWT will review proposals for new services certification (submitted 3 to 4 months ahead of the F2F) and will submit its findings and recommendations to the ESGF executive committee at F2F for final approval shortly thereafter.

Centers will be free to choose which services they will host (even if their compute node is not ESGF certified) but it should be made obvious to the user which of the services offered are ESGF certified and that, while based on ESGF, results processed by non-ESGF certified operators should be used as-is without any endorsement by ESGF.

Similarly centers can choose to operate a non-certified server (not even an ESGF node) but still offer ESGF-certified services, and propose new services for ESGF certification.

ESGF Certifications

Server Certification

This section describes the requirement we envisage in order for a “compute” node (server) to be certified. Several aspects of the server/node are being considered in order to get certification. In the first stage, the required? components are: “Official ESGF Operators”, “Access to Official Datasets”, “Security”, “Metrics”, “Stress Test”.

The CWT is currently working on the implementation of a “server score” service that will produce detailed results for each of the section bellow. These results will be the basis for the CWT recommendations to the ESGF executive committee. Results not available via an automated procedure should be provided by the entity requesting certification (such as for example cache capacity and minimum length of storage cache).

A list of official ESGF Compute node will be maintained on the ESGF website.

A presentation on the certification process can be found here.

Official ESGF Operators

Detailed description of official operators can be found here.

The ESGF Executive Committee proposed the following set of services required for ESGF certification (with more to be added later), the following list should be considered as a “core” (minimal) set of ESGF-certified services that the server needs to provide, however additional service can be hosted as well (both ESGF certified and non ESGF-certified):

- Aggregate files across time
- Spatial and temporal subset
- Regridding (with specified regridder)
- Minimum across one or many dimensions
- Maximum across one or many dimensions
- Average across one or many dimensions, area weighted for latitude, longitude. Need to describe what is done for other dimension (use of bounds or not).
- Server Metrics
    - Health Metrics
        - Number of jobs running

    - Number of jobs in queue
    - Number of active users
    - Number of users including queue
    - Current CPU load
    - Number of nodes
  - Usage Metrics
    - Files accessed (per user).
    - Services used (per user), timestamps and calls urls.
    - Time for each request.
    - Volume of data accessed locally.
    - Volume of data downloaded and from which data node/center.
    - Volume of data uploaded.
- Test suite service for all certified services

The “core” set of services will be reviewed on an annual basis by the ESGF executive committee, based on CWT recommendations and users feedbacks during the F2F.

Once the compute node is integrated into the ESGF installation process, the compute node will come with one or multiple implementation of these core services.

Official Dataset

The ESGF Executive Committee will approve a set of datasets to be used for testing services. These datasets should:

- Cover multiple years
- Include high and low resolutions
- Include different grid types, including curvilinear and generic.
- Be distributed AND replicated at various data nodes

CWT recommends to use the datasets available there:

Security

Servers should be subjected to the same scrutiny as any other component of ESGF. Services should be protected via valid user credentials, and ensure software integrity, i.e., all APIs should pass software security scans (static and dynamic analysis).

Metrics

Servers should be able to capture certain metrics, to be expanded by the Executive Committee. These metrics should be stored and made available (possibly to a restricted audience) for a certain amount of time (at least 6 months). At the minimum daily average should be stored. We propose to start with the set of easily capturable metrics described in the section above.

Stress

- Server should have internal “priority” queues, to process some jobs faster than others (e.g., based on input size as EDAS does).
- Node are expected to have a reasonable amount of storage assigned for user output and cache. This amount of storage should be proportional to the expected number of users, as a start point, the CWT recommends about 1Gb of long term (72 hours at a minimum) storage per expected user. Additionally the server itself should have dedicated cache storage for intermediate results.

Service/Operator Certification

This section describe the minimum set of requirements that need to be met in order for a service (or operator) to be “Certified”. At the moment the CWT is planning to look into: “API compliance”, “Performance Compliance”, “Stress”, “Public Test Results” and (eventually) “Provenance”. While these are necessary to obtain certification, they are not a guarantee either. The “value” of the operator to the ESGF community will also be heavily weighted in the final decision by the XC.

API compliance

- Service should understand requests as formulated here
- In addition to satisfying API conformance the services should return (at least) all parameters identify by ESGF Certification Specifications
- When a new API version is available, services will be expected to move their implementation to the newer version. With an overlap period for both implementations.
- Service naming should obey the “library.service”/”library.operator” convention, e.g. (EDAS.aggregate or LLNL.aggregate)
- Documentation should describe algorithm assumptions or hypotheses that deviate from standard expectations and outline differences with other implementations. E.g is standard deviation center or not centered, etc… Or at the minimum link to a document describing these.
- Once the CWT reaches consensus on federated computing, services will be expected to fully comply in order to maintain certification..

Performance (Timing/Rate) Compliance

The CWT is still developing a set of tools to obtain these metrics, at this time nothing is set in stone but the CWT envisage metrics based on amount of data processed, possibly weighted by server load. Operators will be tested against standard datasets. Operator failing performance compliance but not hosted on and official ESGF node might be temporarily installed on a certified dev machine to ensure the operator performance is not hinged by the original server's performance.

- If a similar service implementation already exists, subsequent implementations should be within reasonable range of other "certified" services response time on the same hardware/server.
- If new service implementations appear, existing services could lose their certification (at their next re-certification evaluation) if they end up performing more poorly than other certified services in a way that affects server performances.

Stress Test

- The CWT team will devise stress strategies to ensure services still perform in an acceptable fashion when stressed. For example, the server can implement limits on service with adequate documentation on why service returned an error. Stress can be defined as:
    - High number of datasets requested concurrently.
    - High volume of data to process (per dataset)

Public test suite results

- For certified services, a public facing test page should display the results from the test suite in order to build user confidence in the implementation. The results should provide enough information for users to compare their results with those of the provided services (input files, output result file, request parameters, links to test suite source code, etc).
- For new implementations, before obtaining certification the CWT will ensure that the service provides a test suite and the results are correct, provider should do best effort to demonstrate this.
- New implementation of an existing service should yield same results as the existing implementation or provide sufficient documentation on why this implementation differs.
- The CWT will develop a "generic" test suite that calls various implementations of same services and reports on failures and successes of the different implementations.
- The test suite should be able to account for difference in results based on implementation algorithm. Maybe a threshold or in worst case suggesting renaming the service to reflect that the results are significantly different (yet valid).

Provenance

Once CWT establishes a provenance standard, services will be expected to implement it. At the very least result should contain metadata about the original URL, and the some information on the server, service(s) and input files used (version, origin, etc…).

Interactive Services

- For version 1.0 of the certification, there will be no official certification for these interactive services, but we would like to introduce some validations in future versions (2.0 and up). These might include comparisons between on-the-fly operations and their offline counterparts in order to validate results for some common operations.
- New services are being introduced that can be executed on-the-fly by users in a variety of contexts. Such services include the ViSUS Server or VCDAT.
- These services perform regridding and subsetting interactively based on the user's navigation into the data, similar to the capabilities of Google Maps.
- Interactive scripting is available through these services which allows users to provide their own (e.g., Python) scripts to perform custom analyses, and with some of these languages it's possible to show the results incrementally.
- While not required, having a test/playground area for these is highly recommended.

Certification Process

The CWT is in the process of implementing the component of the certification process, but the general idea is that the tools used to certify either a server or a service will be made available to the community so they can test and prepare for certification.

It is hoped to automate as much as possible of the certification process in order to not burden too much the CWT.

Both new and previously certified servers and operators will be subjected to these tools on a yearly basis. The CWT will notify developers/maintainers of the findings in order to let them address potential issues, and a final report will be presented to the executive committee at the F2F for “certification”.

Servers

The CWT will run its “server certification” tool on a server. The result will be made available first to the entity requesting certification for review. The CWT will fix a deadline after which no report can be ran. The final “report” will be reviewed by CWT and “passing” servers and their associated reports will be recommended for certification to the executive committee at the F2F.

Services

In order to certify a service the CWT plans on “hosting” the service on a dedicated server and run the certification tool on it. Originally LLNL will be the “host” but if enough support comes from the community it is possible and desirable that multi server act as “hosts”. Report(s) will then be sent to the provider for review and changes. Here again an hard deadline will be fixed for submitting new services.

“Passing” services will then be collected and submitted to the executive committee at the F2F for a final review. It is expected that scientific value of the service will weight heavily in the final decision of the executive committee.