ESGF + DOCKER

ESGF F2F Workshop,
Washington, DC, December 2016

Luca Cinquini
NASA/Jet Propulsion Laboratory + California Institute of Technology
Docker in a Nutshell

- Docker is the leading “containerization” technology: run an application as a “black box” on any Docker-enabled server
  - Build: images are built as bundles that include the application itself, all of its dependencies, and “just-enough-operating-system”
  - Ship: images are hosted on online repositories such as DockerHub
  - Run: images are run as containers on any host that includes a Docker daemon

- Why using Docker for ESGF?
  - Promises to greatly improve installation and maintenance of an ESGF node
  - Part of DREAM strategy for modularizing the ESGF architecture and porting it to other domains
Demo: Install and Run an ESGF Node

- Pre-requisite: install Docker Engine on host (Linux, MacOSX, Windows)
- Instructions:
  - `git clone https://github.com/ESGF/esgf-docker.git`
  - `cd esgf-docker`
  - `export ESGF_HOSTNAME=<host FQDN>`
  - `export ESGF_CONFIG=<some directory>`
  - `./esgf_node_init.sh`
  - `docker-compose up`

https://my-node.esgf.org/

https://hub.docker.com/esgfhub/

CoG

TDS

Solr
• All ESGF services are run as independent, interacting Docker containers
• Specific Node configuration (certificates, passwords, XML files) saved in $ESGF_CONFIG
• Specific Node data (Postgres db, Solr indexes, TDS catalogs, NetCDF files, CoG site media) stored on Docker volumes
• Custom networks isolate applications for additional security (for example, Postgres db)
Docker Details

- **Dockerfile:** "recipe" for building a Docker image
- **docker-compose.yml:** configuration file for bundling several images
ESGF/Docker Data Node Architecture

docker-compose -f docker-compose-data-node.yml up
ESGF/Docker Index Node Architecture

docker-compose -f docker-compose-index-node.yml up

Using standard Solr replication + distributed search
ESGF/Docker Index Node with Solr Cloud

docker-compose -f docker-compose-solr-cloud.yml up

- Solr-Cloud advantages:
  - Automatic distributed indexing and searching (no custom configuration)
  - Load balancing and high availability
  - Fault tolerance
  - Scalability
Advantages of using Docker for ESGF

• Installation:
  ‣ Installation scripts are much more modular and much smaller
  ‣ Installation process is much easier: simply download the images, no compilation involved
  ‣ Easier upgrades, possibly one module at a time, and reversible
  ‣ Everybody runs exactly the same software

• Architecture:
  ‣ Can define and deploy new architectures by simply writing new configuration files
  ‣ Can introduce new modules by simply writing & wiring new images (solr-cloud, nginx, …)

• Portability: ESGF node will run on any platform (Linux, MacOSx, Windows) including Cloud

• Scalability: modules can be scaled arbitrarily by running more containers (e.g. TDS, WPS…)
  ‣ Caveat: application must be written to enable distributed access to data

• Other Docker/Swarm advanced features:
  ‣ Scalability onto multiple hosts clusters
  ‣ High availability, fault tolerance, rolling updates, …
Disadvantages of using Docker for ESGF

• Docker is a **new paradigm** for building and running applications:
  ‣ New knowledge for application developers
  ‣ New training for node administrators
  ‣ Excellent and up-to-date documentation is a must

• Must port the remaining ESGF modules to Docker:
  ‣ ESGF Publisher Client
  ‣ Globus
  ‣ MyProxy or new OAuth server
  ‣ Desktop + Dashboard
  ‣ LAS

• Must develop a process to migrate all application data:
  ‣ ESGF Postgres database (users and data)
  ‣ TDS catalogs
  ‣ Solr indexes
  ‣ CoG Postgres database and site data
Possible Future Roadmap

• Establish a “whale team” of experts to work on:
  ▸ Test the current infrastructure
  ▸ Port the remaining modules
  ▸ Work on other outstanding tickets
  ▸ Develop a testing suite
  ▸ Develop data migration tools
  ▸ Revise and expand the documentation
  ▸ Re-use current esgf-iwt biweekly meetings?

• Migrating to Docker would take 6-12 months
  ▸ Must support current or upgraded installed in the meantime
  ▸ Might want to switch after CMIP6 - looking into the long-term longevity of ESGF…