Dynamic integration of distributed, Cloud-based HPC and HTC resources using JSON Web Tokens and the INDIGO IAM Service

Stefano Dal Pra (stefano.dalpra@cnaf.infn.it), INFN-CNAF

Roberto Alfieri\(^1\), Andrea Ceccanti\(^2\), Davide Salomoni\(^2\), Daniele Spiga\(^4\),

1. University of Parma and INFN (IT), 2. INFN - CNAF (IT), 3. INFN -Perugia (IT)
Outline

- Introduction and Motivations
- Use case overview
- Description of the chosen technologies
  - Connecting the dots
- Summary and future plans
Introduction & Motivations

- **On the demand side**, several user communities would like to exploit an integrated use of HPC and HTC resources. I.e., they would like to run mixed workloads through a common entry point or method for resource access.
- **On the supply side**, not being able to efficiently and effectively federate HPC and HTC resources often leads to under- or over-utilization of resources, which could be at least mitigated through an integrated approach.
- **Any solution must not be disruptive** in terms of the computing models used by the communities, and should be generally applicable to multiple infrastructures (i.e. proven, open, dynamic, provider-independent).
- **The concrete use case that originated this work** is the integration of HPC and HTC resources funded and provisioned in the Emilia Romagna region (Italy).
  - However, the solutions worked out here are immediately applicable in any similar HPC and HTC integration scenario.
Requirement analysis

We seek an (integrated) solution:

- To dynamically instantiate or join batch systems on Cloud
  - Need to allow both remote and local access, possibly with distinct priorities

- To adopt a federated, VO-aware system to handle AuthN/Z

- To enable overflow mechanisms between batch systems (opportunistic model)
  - Overflow could be based on job requirements, and batches are supposed to be geographically distributed

- To mix resources types
  - HPC Slurm-based batch system must be transparently accessible to HTC workflows (opportunistically)
The Bologna-Parma Testbed

- **INFN CNAF, Bologna**
  - Cloud@CNAF, based on Openstack, part of the larger INFN computing infrastructure, with about 80,000 CPU cores, 60 PB of disk space and 90 PB of tape space

- **University and INFN, Parma**
  - HTC cluster based on HTCondor 8.8.5
  - HPC cluster based on Slurm 17.11.5, including 2200 cores (800 BDW, 1100 KNL, 300 SKL) and 16 NVIDIA GPUs (14 P-100, 2 V-100).
    - All nodes are interconnected with Intel OmniPath.
DODAS: Infrastructure as code

- Open source, on-demand creation and configuration of analysis clusters on any cloud infrastructure with almost zero effort

- Platform as a Service for hybrid clouds
  - High level of automation
  - Oriented to the ZeroOps model
  - Supports user-tailored applications and software for data analysis
  - Flexible Authentication and Authorization model

See: The DODAS experience on EGI Federated Cloud
Dynamic, on-demand HTCondor pool on Cloud

- DODAS generated HTCondor batch system on Openstack
  - Remote submit enabled, fully integrated with monitoring

@ CNAF

stefano.dalpra@cnaf.infn.it

CHEP 2019, 4-8 November, Adelaide Australia
INDIGO IAM & HTCondor Authentication

IAM as a flexible solution for authN/Z

- User registration & VO management
- User/group provisioning via SCIM
  - enables VO-aware condor-mapping on the instantiated condor cluster (see next slide)
- WLCG JWT profile & SciTokens compatibility
  - Integrates with HTCondor v. >= 8.9.3 via the SCITOKENS authentication method

See talk on Thursday: Beyond X.509: token-based authN/Z in practice
IAM & GSI AuthN method

- Synchronize authorized condor X509 DNs from IAM groups:
  - Use a IAM client with SCIM read capabilities to retrieve user info for a certain IAM group
  - Generate DN from IAM userID and username
  - Create a condor mapfile
  - Use it to verify user and condor daemon X509 proxy authZ

GSI "^\V\C\=IT\V\O\=\=CLOUD@CNAF\V\CN\=0931b26e-89f6-4118-a5c4-dd7f9e9ec85a@dodas-iam$" spiga
Distributed Pools Federation

- We identified the Flocking mechanisms as a suitable solution for our use case
- Configured **flock from CNAF HTCondor pool** and **flock to Parma HTCondor pool**
  - We started flocking ALL jobs and now evaluating selective Flocking
  - This allow to flock based on user selection (through requirements)

```
FLOCK FROM=131.154.192.239
FLOCK TO=192.135.11.37
```
The Parma site has both HPC and HTC clusters. We use them as a testbed to fine tune the proper solution to the resource mixing.

Work in progress: set up Job Routing to dispatch user payloads to the specified cluster.

```plaintext
JOB_ROUTER_ENTRIES @=jre
[
    TargetUniverse = 5;
    name = "HTC-batch";
]
[
    GridResource = "batch slurm";
    TargetUniverse = 9;
    name = "Slurm-batch";
]
@jre
```
Connecting the dots

Based on the lesson learned, we currently envision the best solution for our use case to:

- Rely on HTCondor to overlay distributed resources as well as to:
  - Support mixing HPC and HTC resources through a Job Routing daemon, implementing custom policies for job transformation
  - to match static user requirements
  - to dynamically (and opportunistically) exploit idle HPC cycles
  - Implement a Pool Federation via Flocking mechanisms
  - A suitable solution as we expect a limited number of “trusted” schedds, with good networking connectivity

- Use DODAS as deployer manager to dynamically exploit hybrid Cloud providers
  - Including automation of CVMFS and data ingestion solution
  - Will be also tested as solution for NON cloud resources

- Rely on IAM Authorization Services for managing groups and roles
  - And act as SciTokens issuer for HTCondor
Summary and Future plans

- We described the successful activity done in order to create the building blocks needed to implement a regional federation of resources.
- We described the technological implementation strategy, based on three pillars:
  - HTCondor, DODAS and INDIGO-IAM
- For the future, we plan to evaluate additional solutions such as a centrally supported HTCondor-CE
  - E.g. similar to the OSG Hosted-CE model
- We also plan to extend these mixing models to cover additional, trans-regional or trans-national HPC and HTC resources.