Usage of JINR SSO authentication and authorization system with distributed data processing services

A.Iachmenev, I.Matveev, D.Oleynik, A.Petrosyan

Grid 2021, JINR, Dubna
July 8, 2021

The role of the authentication and authorization system in distributed systems

Common authentication and authorization are unifying parts of a distributed data processing system.

Historically, GRID uses certificates X.509. This approach is time-tested, but it is often inconvenient for both users and administrators.

To solve this problem, it’s proposed to use single sign on service (SSO).

URMS purposes

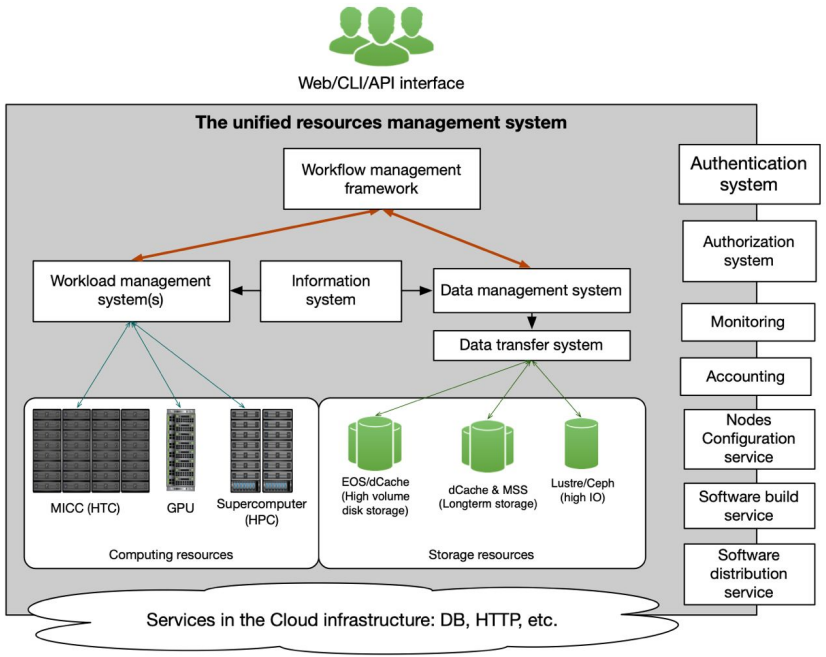
Unify of access to the data and compute resources in a heterogeneous distributed environment

Automate most of the operations related to massive data processing

Avoid duplication of basic functionality, through sharing of systems across different users (if it possible)

As a result – reduce operational cost, increase the efficiency of usage of resources

Transparent accounting of usage of resources



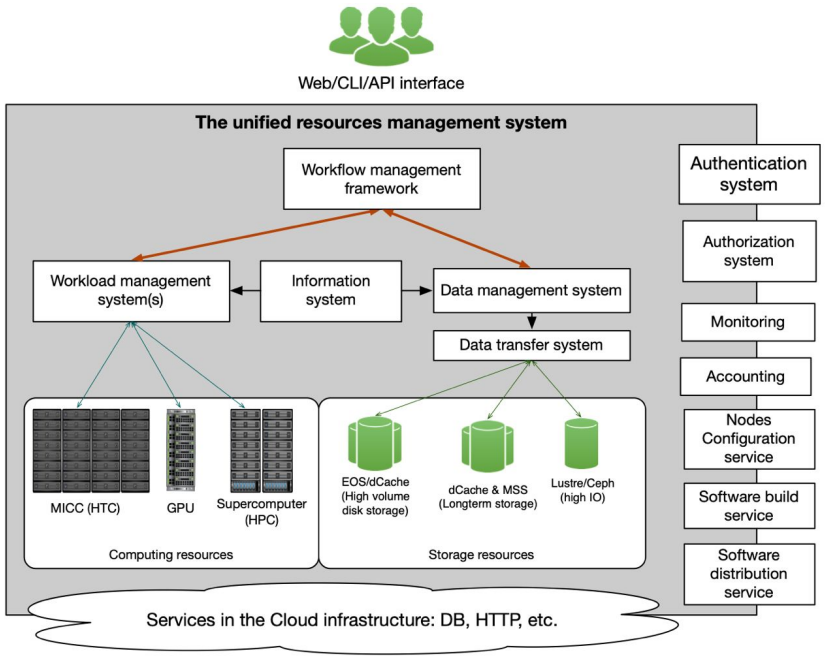
URMS

Workflow management system – controls the process of processing of data on each step of processing. Produce tasks, which required fo processing of certain amount of data, manages of tasks execution.

Workload management system – processes tasks execution by the splitting of the task to the small jobs, where each job process a small amount of data. Manage the distribution of jobs across the set of computing resources. Takes care about generation of a proper number of jobs till task will not be completed (or failed).

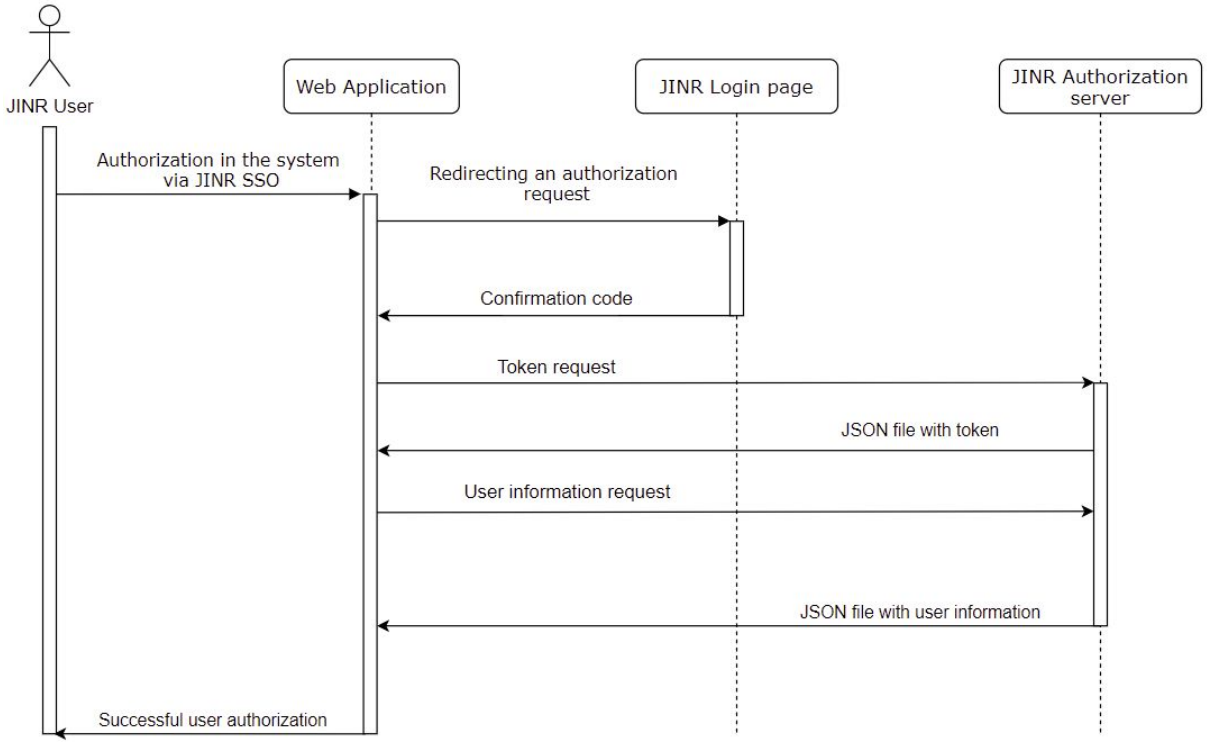
Data management system – responsible for distribution of all data across computing facilities, managing of data (storing, replicating, deleting etc.).

Data transfer service: takes care about major data transfers. Allow asynchronous bulk data transfers.



JINR SSO

The JINR SSO service provides an opportunity to authorize users who are registered in JINR.



Integrating SSO with CRIC

CRIC is a central information system of middleware, which is designed to describe the topology of the ATLAS experiment at the Large Hadron Collider (LHC).

The information system stores and provides a description of resources in a consistent form, since resources have different architectures, and they must be presented in an abstract, but in general, form.

The CRIC implements an authentication and authorization mechanism for the specifics of CERN.

The CRIC authentication system has been redesigned for JINR SSO.

The CRIC authorization system provides a wide set of rights, which allows it to be adjusted to JINR SSO.

Integrating SSO with Airflow

Airflow is a platform to programmatically author, schedule and monitor workflows. In URMS, it is used as workflow management system.

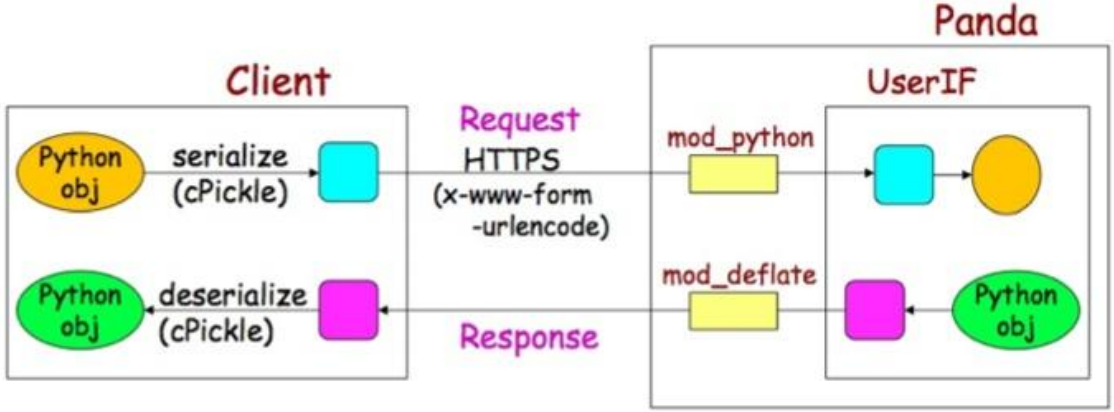
Airflow supports different backends for integration with different SSO. The backend was written that interacts with JINR SSO by exchanging information using http requests.

```
def login(self, request):
    log.debug('Redirecting user to JINR login')
    return self.JINR_oauth.authorize(callback=url_for(
        'jinr_oauth_callback',
        _external=True),
        state=request.args.get('next') or request.referrer or None)

def get_JINROAuth_user_profile_info(self, JINROAuth_token):
    url = self.JINROAuth_user_info_json + JINROAuth_token
    resp = self.JINR_oauth.get(url, token=JINROAuth_token)


    if not resp or resp.status != 200:
        raise AuthenticationError(
            'Failed to fetch user profile, status ({0})'.format(
                resp.status if resp else 'None'))

    return resp.data['login'], resp.data['mail']
```

PanDA

PanDA is a task execution system, by distributing tasks across the computing resources and further monitoring of the execution of these tasks. In URMS, it is used as a workload management system. Tasks created in the workflow management system are submitting to the workload management system for execution. Resources configuration comes from the CRIC information system. Thus, a person without a certificate can start task processing in a distributed computing environment using a workload management system which is integrated with Airflow.

To interact with Panda, the Airflow uses the PanDA Client. This is a set of classes that implements interaction with the PanDA API.



Conclusion

We are building a system that consists of different subsystems, each of which requires authentication and authorization. We managed to create a single entry point using JINR SSO and leave certificates only at the service level, which is much more convenient for users.

At the moment, we have two applications integrated with the JINR SSO. This confirms that the methodology is working and we are going to connect the rest of our services in the same way.