



AliECS

A New Experiment Control System for the ALICE Experiment

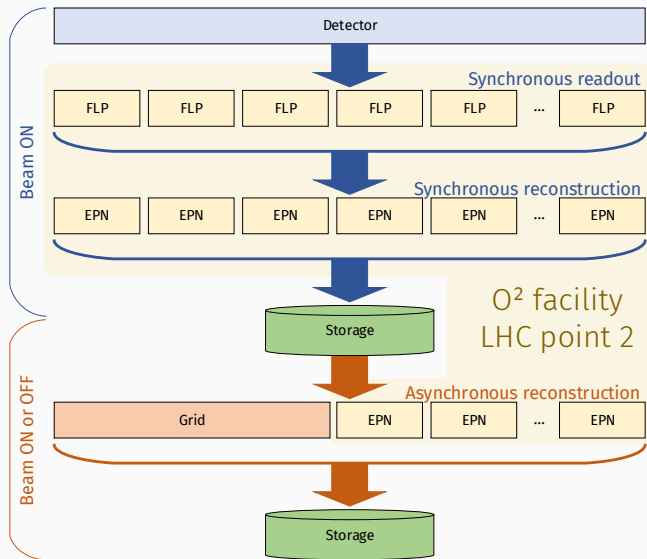
Teo Mrnjavac

CERN EP-AID-DA

on behalf of the ALICE O²/FLP project

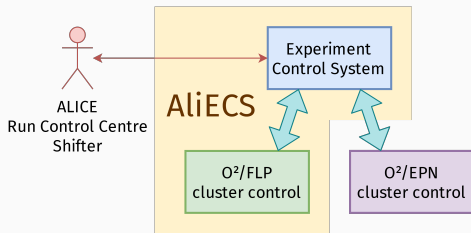
9 October 2019

The ALICE Online-Offline computing system



- Multiprocess **data flow and processing** framework
- **100,000s of processes**, ~1000 machines
- **Synchronous and asynchronous** (grid-like) workflows
- One computing system, 2 types of node arranged in 2 clusters:
FLP - First Level Processors
EPN - Event Processing Nodes
- Operations start in **2021**

ECS and O² cluster control



- Manage the lifetime of thousands of **stateful processes** in the O²/FLP cluster (control of O²/EPN delegated to a specialized O²/EPN cluster control)
- Minimize the waste of beam time by reusing processes and avoiding time-consuming process restart operations
- Interface with the LHC, the trigger system, the Detector Control System and other systems through common APIs

Managing a cluster with Apache Mesos

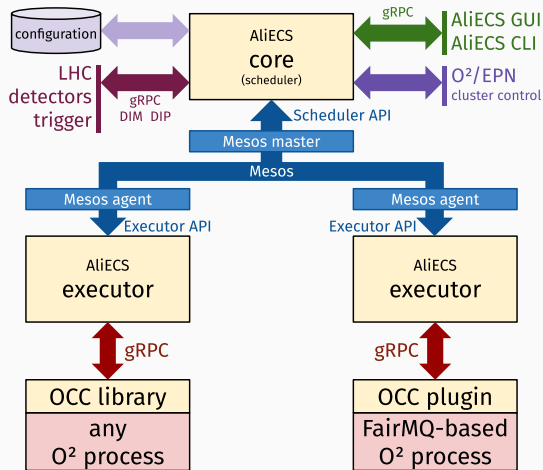
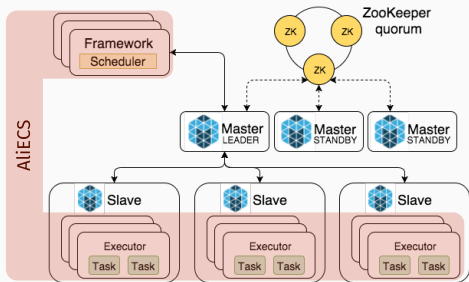
“Program against your datacenter like it’s a single pool of resources.”

Managing a cluster with Apache Mesos

“Program against your datacenter like it’s a single pool of resources.”

- Mesos acts as a unified **distributed execution environment**
which streamlines how AliECS manages its components, resources and tasks inside the O²/FLP farm.
- Benefits:
 - **knowledge** of what runs where,
 - **resource management** (ports, CPU, RAM, ...),
 - **transport** for control messages,
 - task event **notification** (dead, failed to launch, ...),
 - node attributes, high availability, checkpointing, ...

AliECS in a nutshell



AliECS in a nutshell

- Components:

- AliECS core (incl. Apache Mesos scheduler)
- AliECS executor
- O² control and configuration FairMQ plugin (FairMQPlugin_OCC)
- O² control and configuration library (libocc)
- AliECS control and configuration utility (coconut)
- Single-node OCC debug utility (peanut)

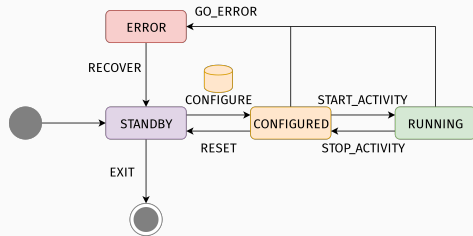


- Also available:

- The web-based AliECS GUI
- AliECS deployment mechanism



AliECS concepts



- AliECS schedules, configures and controls **tasks**
- A task represents a stateful process, which implements a **role**

- A collection of AliECS roles (arranged in a workflow) along with their configuration is an **environment**
- Tasks, roles and environments have their own **state machines**
- An environment represents the collective state of its constituent roles
- When an environment reaches the **RUNNING** state, it is granted a unique **run number** which remains valid until the **RUNNING** state exits

AliECS workflow and task configuration

- Based on **Git**, multiple repositories per AliECS instance
- Task descriptors and workflow templates are **YAML** (plus template system)
- Once loaded, every task type and workflow is **uniquely identified** by
git repository + task/workflow file name + git revision

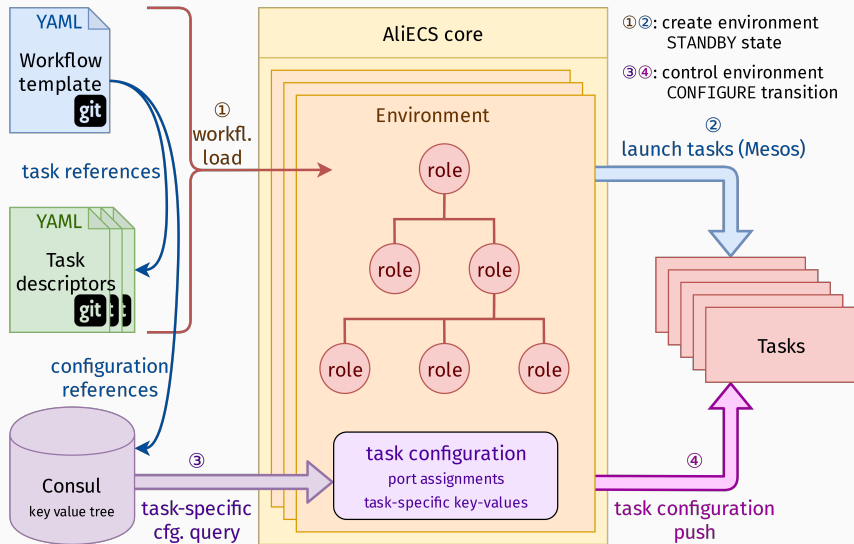
```
[* ](+ 172.17.0.1) tee@alidock-1000 ~/workspace/go/src/github.com/AliceO2Group/Control (master +%$) $ coconut env show ebd281b-c7fc-11e9-b194-0242ac110002 -tw
environment id: ebd281b-c7fc-11e9-b194-0242ac110002
created: 2019-08-26 14:27:51 CEST
state: RUNNING
run number: 15
```

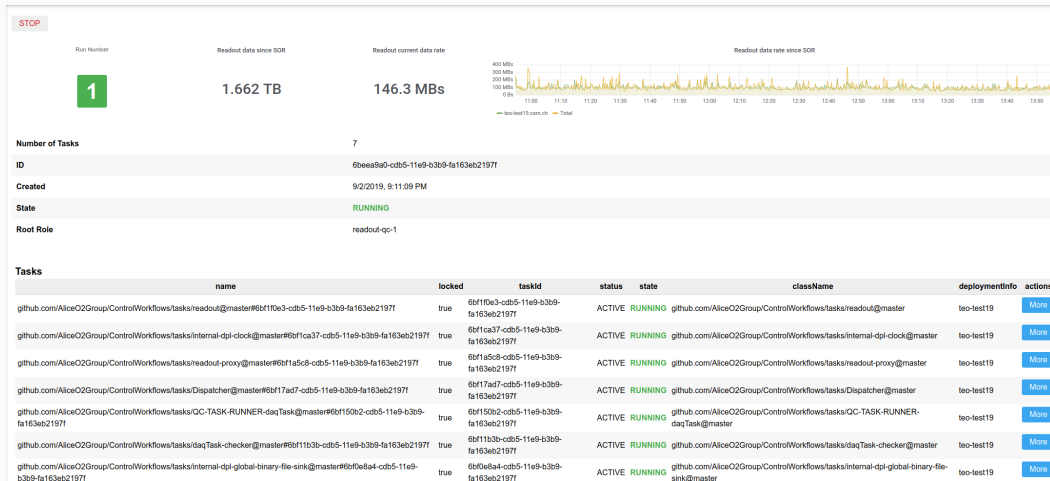
TASK ID (7 TASKS)	CLASS NAME	HOSTNAME	STATUS	STATE
ebe2c9d0-c7fc-11e9-b194-0242ac110002	github.com/AliceO2Group/ControlWorkflows/tasks/readout@readout-qc	aido2-bld4-lab107.cern.ch	ACTIVE	RUNNING
ebe1e122-c7fc-11e9-b194-0242ac110002	github.com/AliceO2Group/ControlWorkflows/tasks/internal-dpl-clock@readout-qc	aido2-bld4-lab107.cern.ch	ACTIVE	RUNNING
ebe182c1-c7fc-11e9-b194-0242ac110002	github.com/AliceO2Group/ControlWorkflows/tasks/readout-proxy@readout-qc	aido2-bld4-lab107.cern.ch	ACTIVE	RUNNING
ebe12486-c7fc-11e9-b194-0242ac110002	github.com/AliceO2Group/ControlWorkflows/tasks/Dispatcher@readout-qc	aido2-bld4-lab107.cern.ch	ACTIVE	RUNNING
ebe0acdc-c7fc-11e9-b194-0242ac110002	github.com/AliceO2Group/ControlWorkflows/tasks/QC-TASK-RUNNER-daqTask@readout-qc	aido2-bld4-lab107.cern.ch	ACTIVE	RUNNING
ebe04951-c7fc-11e9-b194-0242ac110002	github.com/AliceO2Group/ControlWorkflows/tasks/daqTask-checker@readout-qc	aido2-bld4-lab107.cern.ch	ACTIVE	RUNNING
ebdf838-c7fc-11e9-b194-0242ac110002	github.com/AliceO2Group/ControlWorkflows/tasks/internal-dpl-global-binary-file-sink@readout-qc	aido2-bld4-lab107.cern.ch	ACTIVE	RUNNING

```
workflow:
[ RUNNING ] readout-qc-1
├── [ RUNNING ] readout-role                → task ebe2c9d0-c7fc-11e9-b194-0242ac110002
├── [ RUNNING ] internal-dpl-clock          → task ebe1e122-c7fc-11e9-b194-0242ac110002
├── [ RUNNING ] readout-proxy              → task ebe182c1-c7fc-11e9-b194-0242ac110002
├── [ RUNNING ] Dispatcher                 → task ebe12486-c7fc-11e9-b194-0242ac110002
├── [ RUNNING ] QC-TASK-RUNNER-daqTask     → task ebe0acdc-c7fc-11e9-b194-0242ac110002
├── [ RUNNING ] daqTask-checker            → task ebe04951-c7fc-11e9-b194-0242ac110002
└── [ RUNNING ] internal-dpl-global-binary-file-sink → task ebdf838-c7fc-11e9-b194-0242ac110002
```

Documentation: https://github.com/AliceO2Group/Control/blob/master/coconut/doc/coconut_repository.md

AliECS workflow and task configuration





Conclusions

- The new ALICE O² computing system requires a **new control system**
- AliECS carries both **ECS** and **O²/FLP cluster control** duties
- Opportunity to leverage technologies such as **Mesos** and **Go** for a high performance, low latency ECS
 - Mesos gives us resource management, transport and much more
 - Minimize waste of beam time
 - Improved operational flexibility

AliECS on GitHub: github.com/AliceO2Group/Control

Configuration examples: github.com/AliceO2Group/ControlWorkflows