HELMHOLTZ

RESEARCH FOR
GRAND CHALLENGES

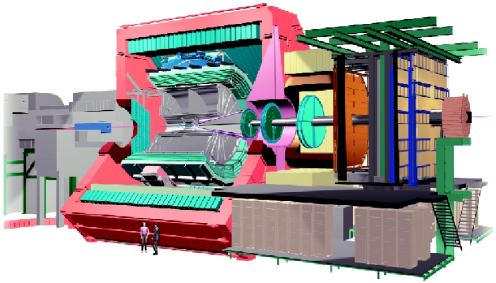
Generic tools for a high throughput online processing

Andrey Lebedev on behalf of SDE group*
GSI Helmholtzzentrum für Schwerionenforschung GmbH

** Mohammad Al-Turany, Radoslaw Karabowicz, Dennis Klein,*
Dmytro Kresan, Matthias Kretz, Anar Manafov,
Alexey Rybalchenko, Christian Tacke, Florian Uhlig

Introduction

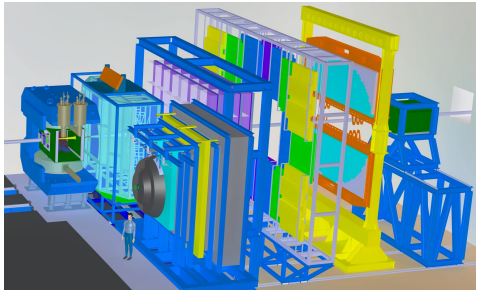
High demands of modern HEP experiments



ALICE@CERN
in 2021

3,4 TB/s

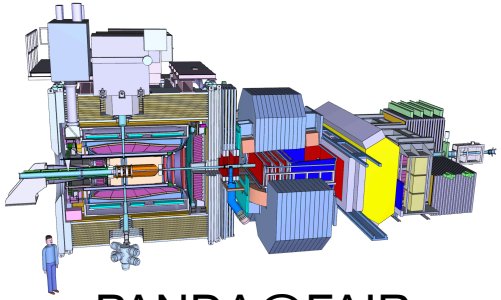
50 PB/year



CBM@FAIR
in 2025

1,0 TB/s

22 PB/year



PANDA@FAIR
in 2025

0,2 TB/s

24 PB/year

High throughput requirements for FAIR experiments as well as for the upcoming upgrade of the ALICE experiment.

Motivation

High throughput online processing

- **ALICE** and **FAIR** experiments are moving from traditional single to **multiprocessing tools** for simulation and reconstruction.
- This requires a **highly parallelised** and **data flow driven processing pipelines** with a **distributed architecture**, i.e. a collection of highly maintainable, testable, loosely coupled, independently deployable processes.
- Managing such an environment requires a system, which is able to **spawn and control** hundreds of thousands of different processes which are tied together by a topology. It can run on computing clusters using different resource management systems (RMS) or even on a laptop and can be controlled by external tools.

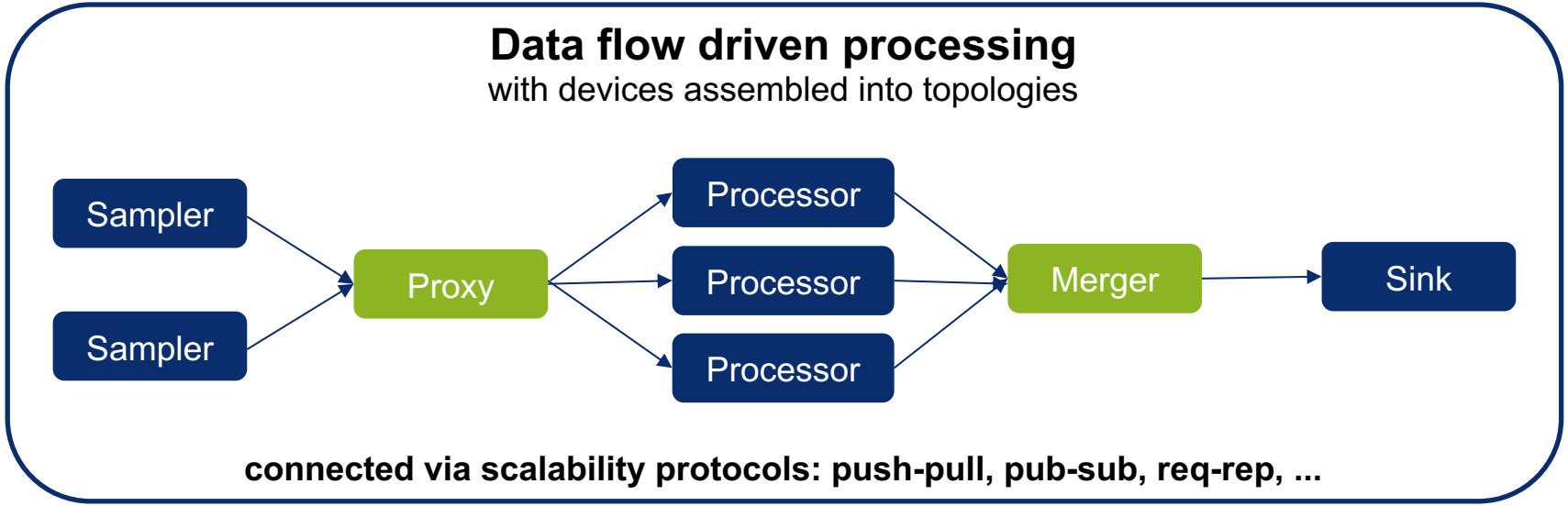
Highlights

Generic tools to solve the experiment challenges

- **Transport layer**
  - FairMQ - framework for high throughput distributed data processing
- **Deployment and control of distributed workflows**
  - DDS – The Dynamic Deployment System
  - ODC – Online Device Control
- **Other projects of our group (not covered in this talk)**
  - FairRoot – a simulation, reconstruction and analysis framework
    - https://github.com/FairRootGroup/FairRoot
  - VC – portable, zero-overhead C++ types for explicitly data-parallel programming
    - https://github.com/VcDevel/Vc
  - FairSoft – installation of common software packages
    - https://github.com/FairRootGroup/FairSoft

FairMQ

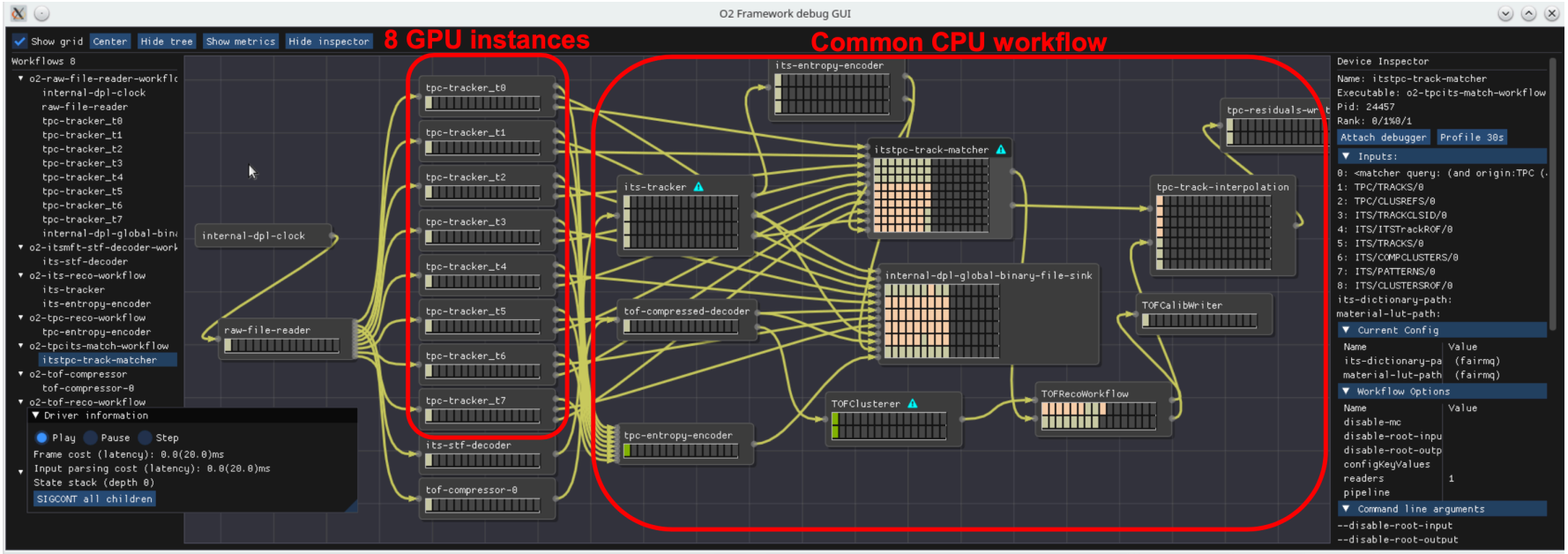
FairMQ is the core of data transport layer



- A data flow driven computing model by providing building blocks to construct distributed processing pipelines.
- Each processing stage in the pipeline transmits data through message queues via an abstract interface.

FairMQ

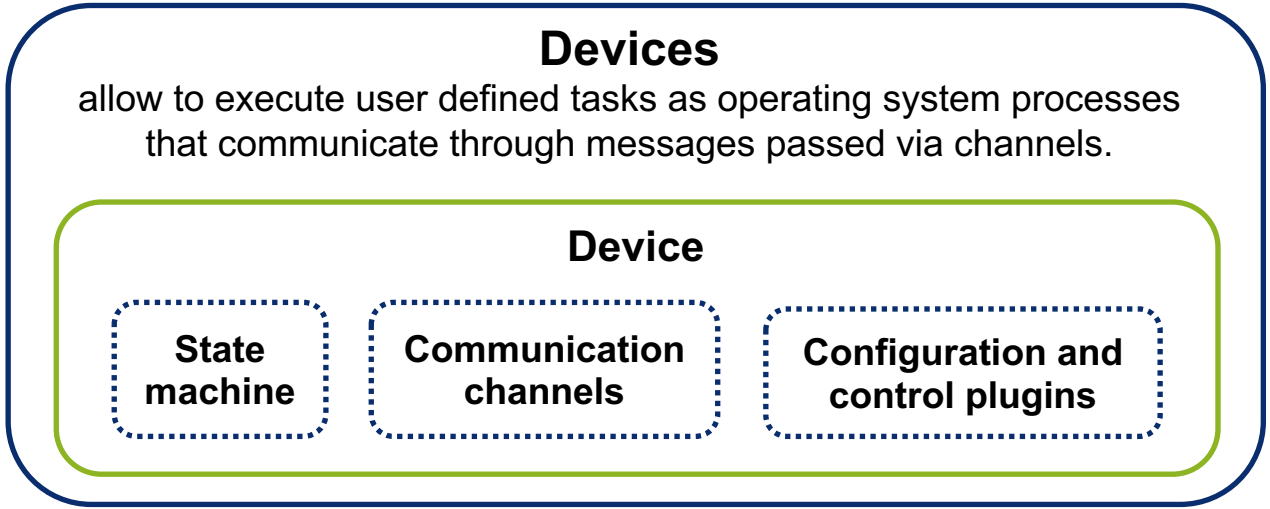
EPN workflow from ALICE



David Rohr and Giulio Eulisse

FairMQ

Device and SDK



- FairMQ based processes can be **controlled and orchestrated** via different systems by implementing the corresponding plugin.
  - **DDS plugin** is provided out of the box.
- **FairMQ.SDK** provides API to control the state machine of a **single *device*** as well **as a topology of *devices***.

FairMQ

Multiple implementations of data transport interface

- FairMQ supports most of the **existing data transport technologies**:
  - ZeroMQ,
  - Shared memory,
  - RDMA transport based on libfabric.
- A single process can use either single or **multiple different transports** at the same time.
- The transport implementation can be selected **at runtime** per input/output channel.
- Messages can be passed from one channel to another **regardless of the transport**.

All this helps the user to define the most suitable transport technology for a given task.

DDS – The Dynamic Deployment System

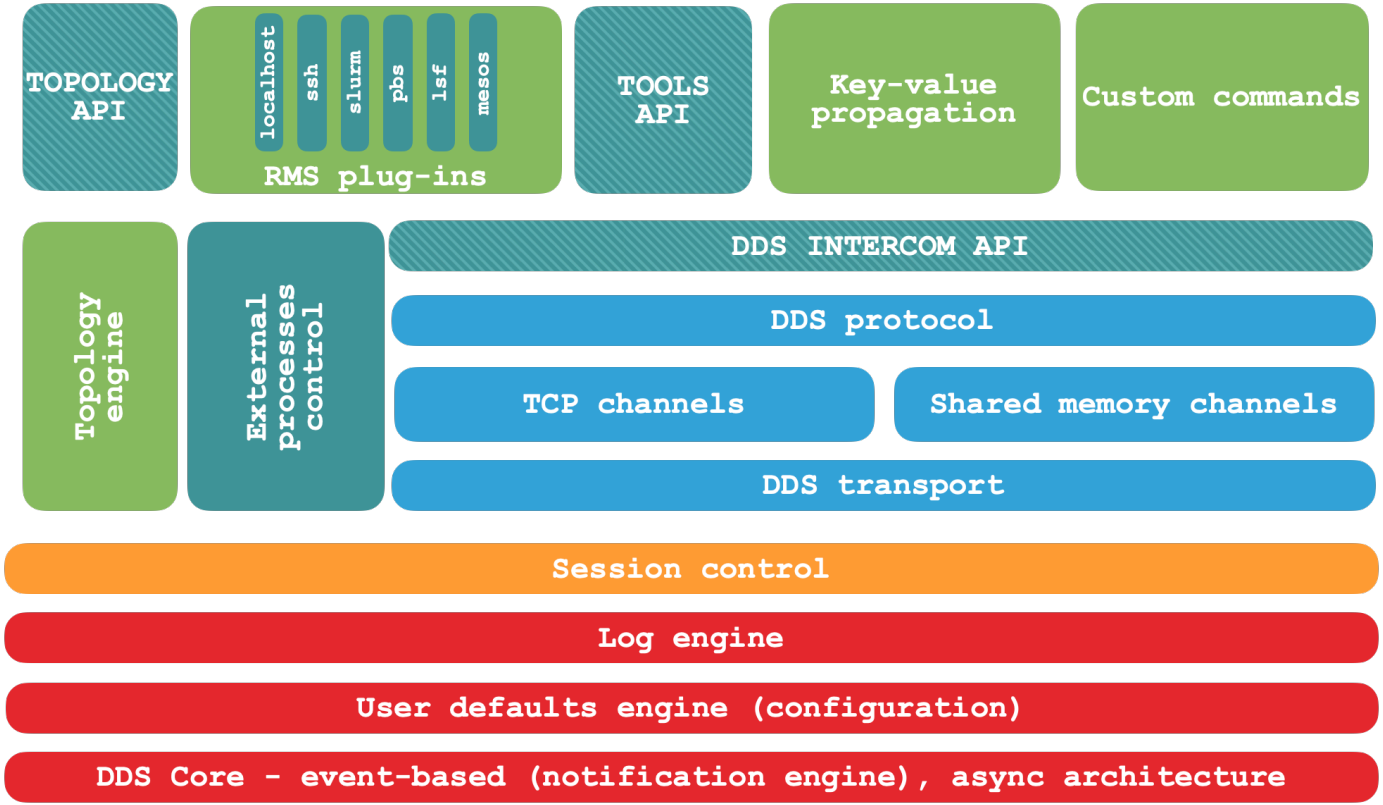
Basic concepts

DDS is a tool-set that automates and simplifies the deployment of user-defined processes and their dependencies using a predefined processing graph (topology).

- A single responsibility principle command line **tool-set and API**;
- users' task is a black box – it can be an **executable or a script**;
- **watchdogging**;
- rule-based execution of tasks;
- **plug-in system** to abstract from RMS including **SSH** and a **localhost** plug-ins;
- **doesn't require pre-installation and pre-configuration** on the worker nodes;
- private facilities on demand with **isolated sandboxes**;
- key-value propagation and custom commands.

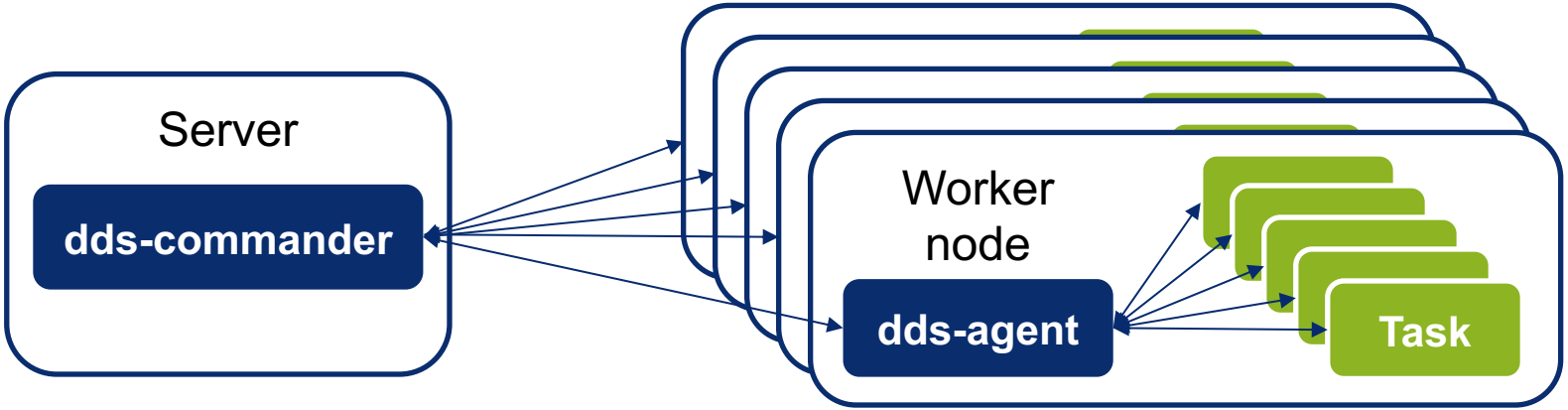
DDS – The Dynamic Deployment System

10000 feet view



DDS – The Dynamic Deployment System

Workflow



- **dds-session** *start*
- **dds-submit** *-r ssh -c ssh_hosts.cfg*
- **dds-topology** *–activate topology.xml*
- **dds-topology** *–update new_topology.xml*

Single DDS agent manages multiple user tasks

DDS – The Dynamic Deployment System

From user's perspective

Topology
Topology API

```
<topology id="myTopology">

  [... Definition of tasks,
properties, and collections ...]

  <main name="main">

    [… Definition of the topology
itself, including groups...]

  </main>

</topology>
```

CLI tools
Tools API

dds-session
dds-agent-cmd
dds-custom-cmd
dds-info
dds-prep-worker
dds-server
dds-stat
dds-submit
dds-test
dds-topology
dds-user-defaults

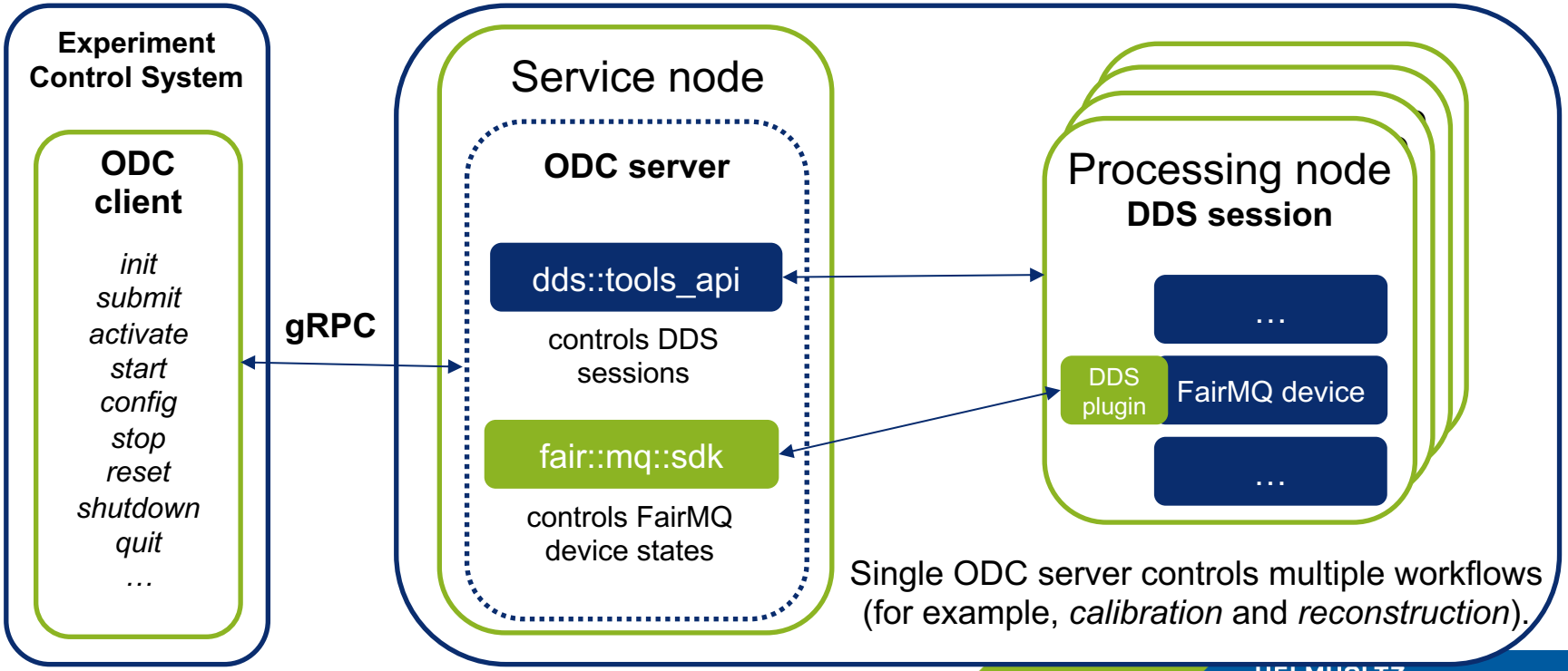
KV and CC
Intercom API

```
CIntercomService service;
CKeyValue keyValue(service);

// Subscribe on key update events
keyValue.subscribe([](
      const string& _propertyID,
      const string& _key,
      const string& _value)
{…});

// Start listening to events we
// have subscribed on
service.start();
```

ODC – Online Device Control

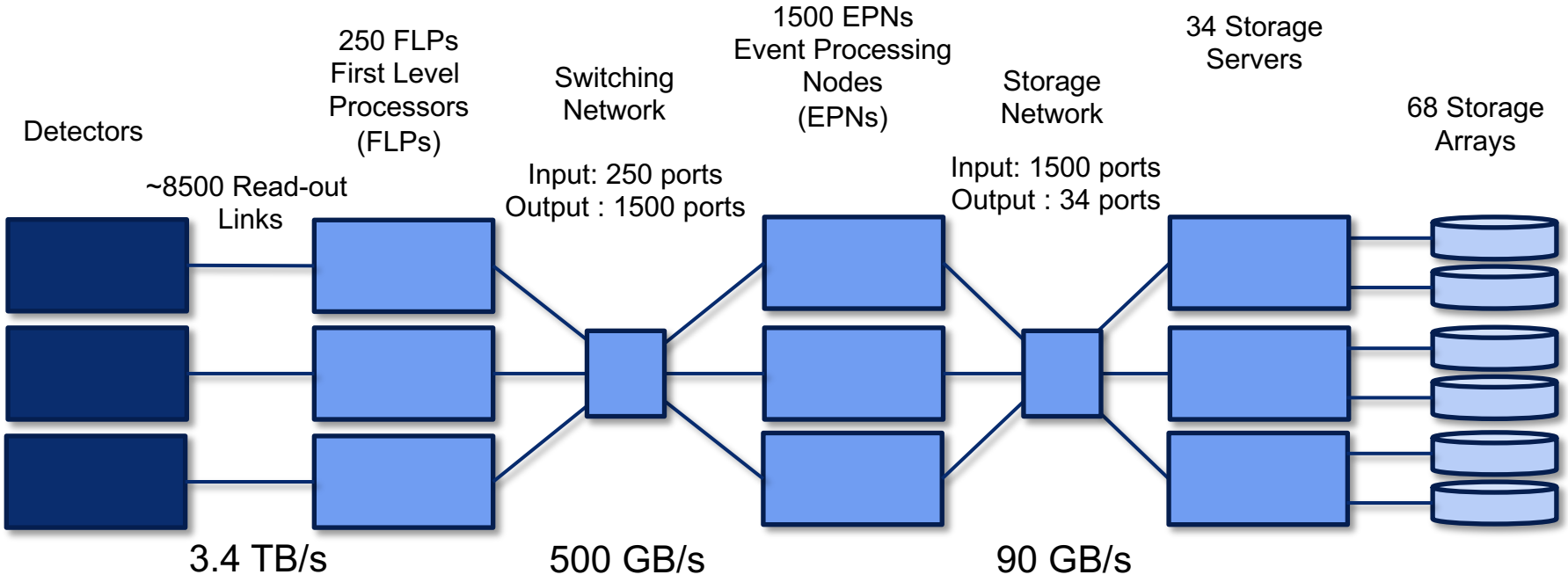
Deploys and controls a topology of FairMQ devices

ODC uses DDS and FairMQ.SDK as well as gRPC based interface for client



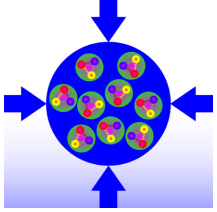
O2 Facility for ALICE@CERN

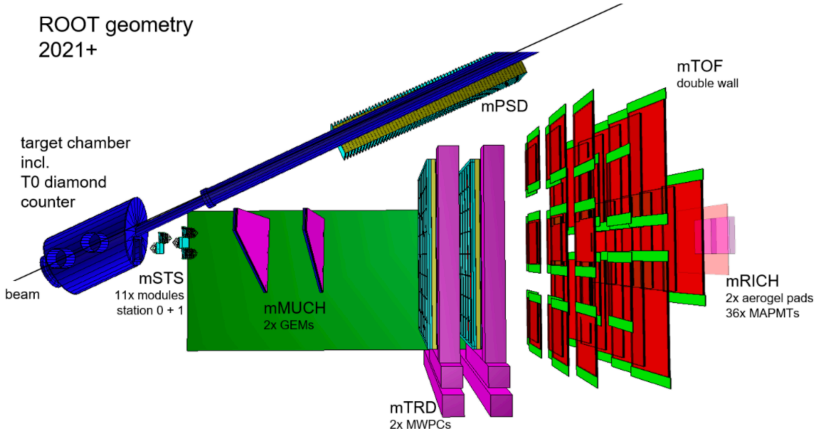
In the process of commissioning



mCBM@GSI

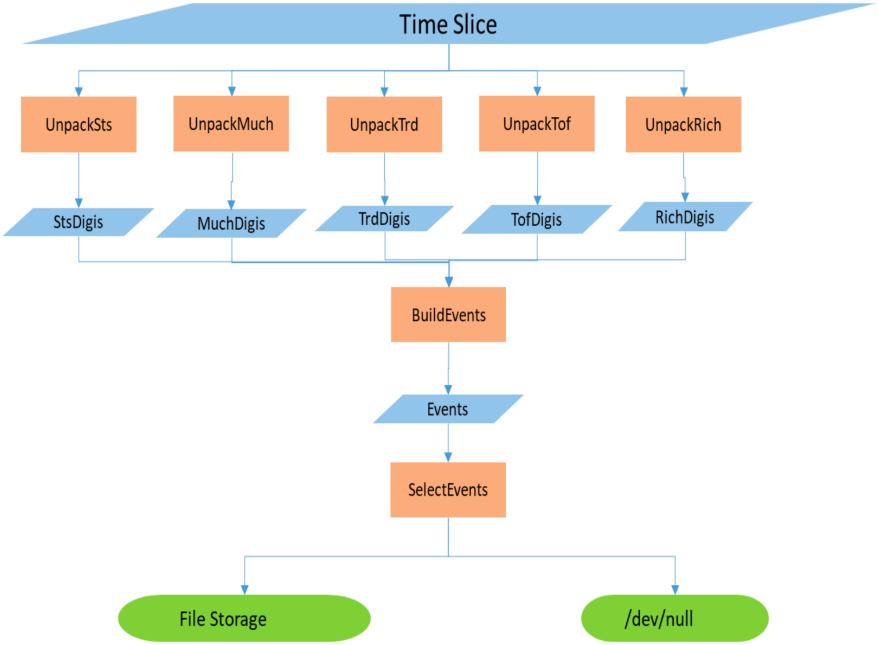
Various tests planned for this year, preparation in progress





Minimal online approach for 2021

(Volker Friese & Florian Uhlig)



Summary

- The developed generic tools provide building blocks for highly parallelised and data flow driven processing pipelines required by the next generation of experiments.
- It allows:
  - to write a message based code without going into details of the transport;
  - to deploy and control topologies on computing and online clusters as well as on a single laptop.

- GitHub projects
  - FairMQ - https://github.com/FairRootGroup/FairMQ
  - DDS - https://github.com/FairRootGroup/DDS
  - ODC - https://github.com/FairRootGroup/ODC