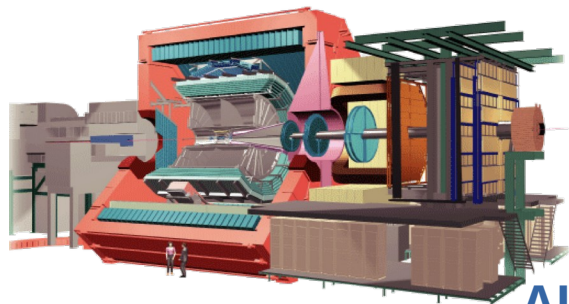


Innovative Workflows in Astro and Particle Physics
March 8 – 12, 2021



Tools for distributed data processing at FAIR and in ALICE experiment

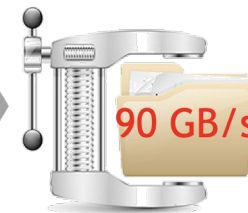
D. Kresan on behalf of SDE Group (GSI, Darmstadt)



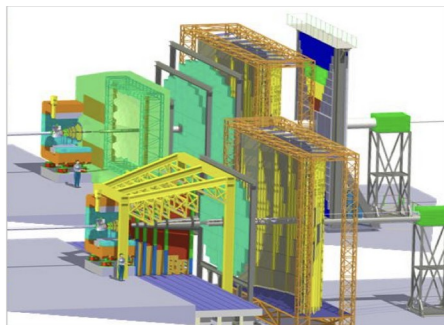
3.4 TB/s

ALICE

- + 463 FPGAs
- + 100'000 CPU cores
- + 3000 GPUs



2021: 50
PB/year



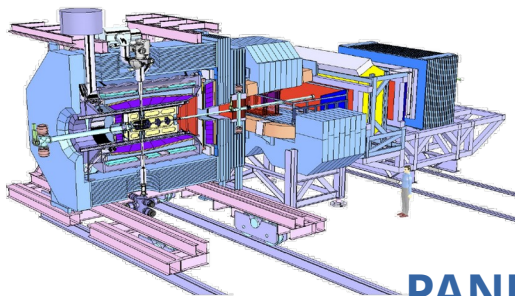
1.0 TB/s

CBM

- + ??? FPGAs
- + 60'000 CPU cores
- + ??? GPUs



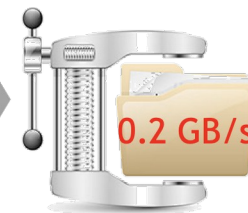
2025: 22
PB/year



0.2 TB/s

PANDA

- + ??? FPGAs
- + 66'000 CPU cores
- + ??? GPUs



2025: 24
PB/year

Software for distributed data processing

- FairRoot: simulation, reconstruction and data analysis
- FairMQ: framework for high throughput distributed data analysis
- Dynamic Deployment System (DDS): management and controlling of distributed workflows

10.5281/zenodo.3896282

<https://fairroot.gsi.de>

10.5281/zenodo.1689985

<https://github.com/FairRootGroup/FairMQ>

10.5281/zenodo.3854975

<http://dds.gsi.de>



Use case

- Provides building blocks
 - for experiment-specific algorithms
 - distributed data analysis on small to very large data scale
 - in heterogeneous hardware
 - multiple languages environment
- DDS is used on computing cluster to start, initialize and manage large amount of user processes
- Each process is a separated reconstruction/analysis step
 - communicate over network or shared memory via message passing

Data flow driven processing

with devices assembled into topologies

```
graph LR; S1[SAMPLER] --> P[PROXY]; S2[SAMPLER] --> P; P --> PR1[PROCESSOR]; P --> PR2[PROCESSOR]; P --> PR3[PROCESSOR]; PR1 --> M[MERGER]; PR2 --> M; PR3 --> M; M --> S[SINK];
```

The diagram illustrates a data flow driven processing topology. It starts with two green boxes labeled 'SAMPLER' on the left. Arrows from both samplers point to a central blue box labeled 'PROXY'. From the 'PROXY' box, three arrows point to three separate green boxes labeled 'PROCESSOR'. Arrows from all three 'PROCESSOR' boxes point to a blue box labeled 'MERGER'. Finally, an arrow from the 'MERGER' box points to a green box labeled 'SINK' on the right.

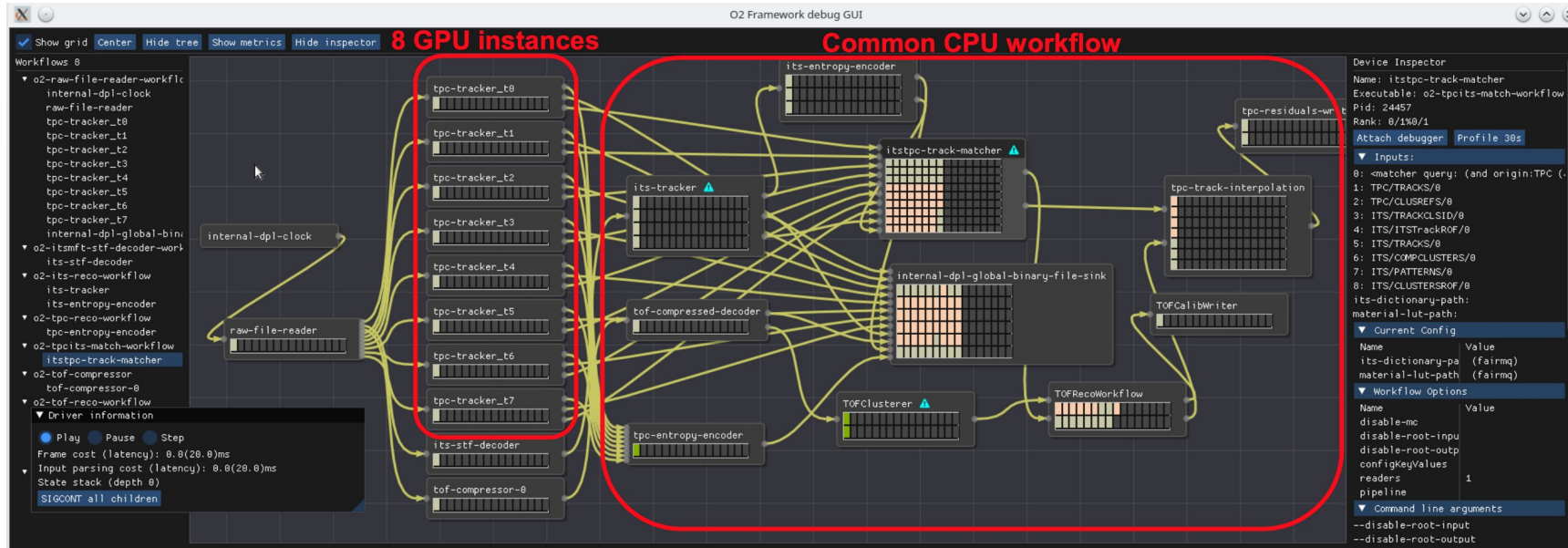
connected via scalability protocols: push-pull, pub-sub, req-rep, ...

A. Rybalchenko

connected via scalability protocols: push-pull, pub-sub, req-rep, ...

ALICE EPN workflow

D. Rohr, G. Eulisse



D. Kresan

Tools for Distributed Data Processing