Open access for ALICE analysis based on virtualisation technology

Predrag Buncic[1], Mihaela Gheata[2], Yves Schutz[3] on behalf of the ALICE Collaboration

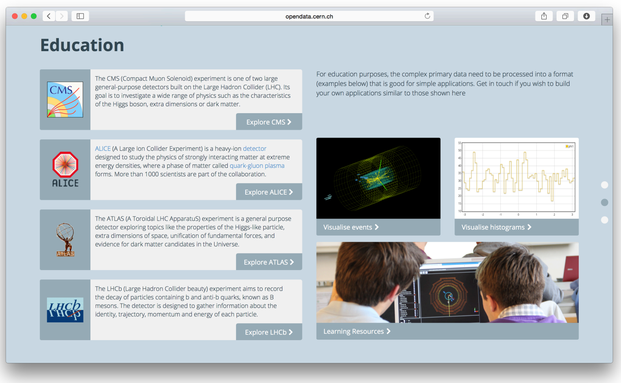
[1]CERN, Geneva, Switzerland - [2]Institute of Space Sciences, Bucharest, Romania - [3]Laboratoire de Physique Subatomique et des Technologies Associees

Open access objectives
- Provide open access to scientific LHC data, including software and documentation
- Allow reproducibility of analysis in the future by third party

Open access web portal hosting data and applications
- Coherent style and organisation for all LHC experiments
- Possibility to host and make public data releases according each experiment policy

Target audience: mainly education in the current approach
- Interested citizen, students of physics

opendata.cern.ch



CERN portal as user entry point to the ALICE environment, applications and data.
- Outreach information regarding ALICE experiment
- Detailed instructions on the virtual machine installation procedure and usage
- Downloadable ALICE datasets
- Analysis tools used with the open access ALICE application

cernvm-online.cern.ch

ALICE data accessible via *xrootd* from EOS CERN storage
- 8 TBytes of *Pb-Pb* and *p-p* 2010 reconstructed event summary data at $\sqrt{s}$ = *3.5* TeV
- Filtered data collections needed for the ALICE *masterclasses*
- Tools for running ALICE masterclasses and analysis software

The ALICE open access application is bootstrapped from a dedicated context publicly available on the ***CernVM Marketplace***.
- Creating a default user account and customizing its profile, such as environment, home page location and automatic download of required software.

The context is applied on a standard *microCernVM* image, downloadable from the *CernVM* site, while the required applications are downloaded on demand
- This approach minimizes the amount of data to be downloaded by the user compared to storing separate VM images when changing the application

Datasets metadata stored in JSON format

```
Dataset": {
    "name": "LHC2010h_PbPb_ESD_138275",
    "description": "Pb-Pb ESD data sample at 3.5 TeV",
    "path": "/eos/opendata/alice/2010/…",
    "files": 100,
    "file_details": [
      {
        "file_path": "/eos/opendata/alice/…",
        "file_size": 200762886,
        "file_checksum": "41b466d6eb9b…",
        "file_timestamp": "2014-09-24 12:32:31"
      }, …
```

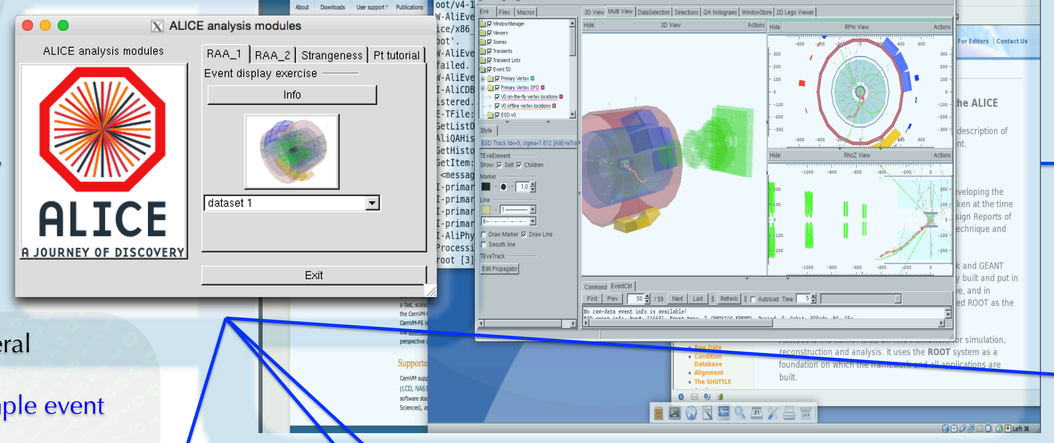
CernVM File system

After bootstrapping the virtual machine, the ALICE software environment is automatically installed using *CernVM-FS*
- Currently taking one of the builds of *AliRoot* Git tags stored in *CernVM-FS*

Starting a graphics user interface based on root
- Entry point for downloading and running ALICE masterclasses and analysis examples

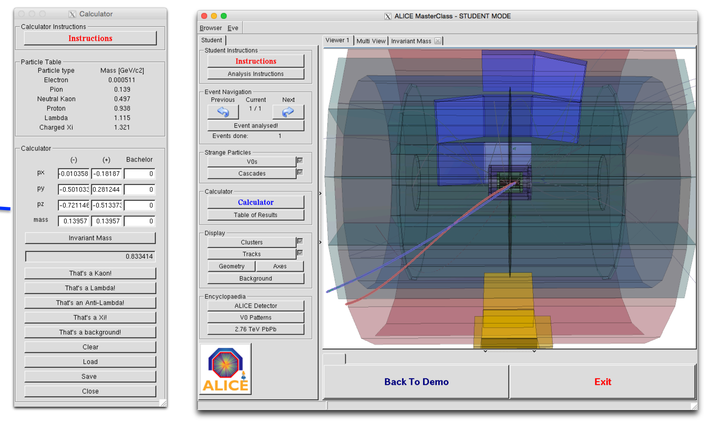
No need to download manually datasets, these can be selected from the interface and copied automatically on the VM client

The pairing procedure requires the connection to the *cernvm-online* site while having the *microCernVM*-based machine up and running



ALICE open access GUI allows running several applications from a single interface
- Nuclear modification factor (R_{AA}) and simple event display for strange particle "hunting"
- Demonstrator for ALICE analysis framework plotting simple observables
- Standard ALICE event display software

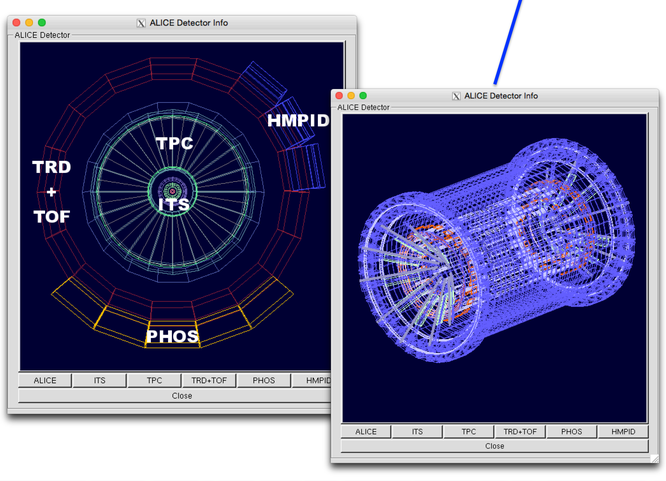
The graphics user interface is extensible to more analysis modules
- A plug-in approach foreseen in future, allowing to add modules dynamically

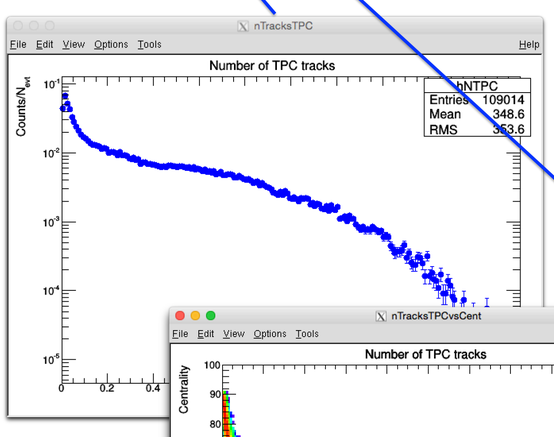


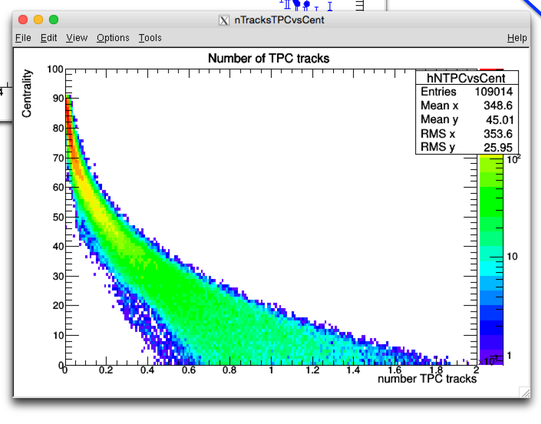
ROOT Data Analysis Framework

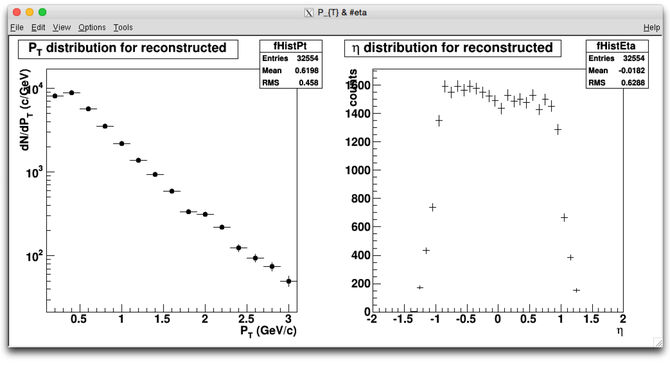
In case of reconstructed data analysis, the GUI allows for selecting one or more datasets
- Currently data is copied to the VM client using *xrdcp*, then processed locally

The analysis examples are compiled dynamically using ROOT
- The analysis is not "frozen", but allows for changing input parameters, analysis cuts or output histograms









The future extension of this tool will integrate with the *Data Analysis Preservation Framework* (DAPF). This framework is currently under development in collaboration with the IT and GS groups at CERN and is designed to provide a common infrastructure for long term preservation of analysis procedures.
- Full description of the analysis steps, from primary datasets to the final publication
- Preservation of the analysis software and full documentation

The tool will offer the practical possibility to reproduce published ALICE results, using the VM technology

ALICE experiment works in close collaboration with DPHEP, the other CERN experiments and the CERN IT/GS for implementing a common approach and generic solutions for data preservation and open access
- Same data preservation principles and experiment policy guidelines
- Open access portal, common analysis preservation framework, use of virtualisation technology

This open access initiative is only the beginning of a large scale effort to be pursued even beyond the experiment's lifetime

21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 - 17, 2015