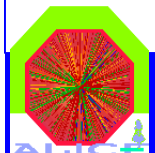
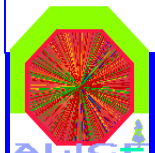
DISTRIBUTED ANALYSIS PRACTICAL EXAMPLES

Panos Christakoglou
University of Athens – CERN



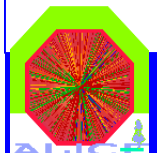
OUTLINE

- News on analysis – Analysis Task Force:
  - Web pages for all analysis topics.
- Scheme of the analysis framework.
- Interaction with the file catalog.
- Interactive analysis sessions.
- Batch analysis sessions.
- Next steps - Summary.

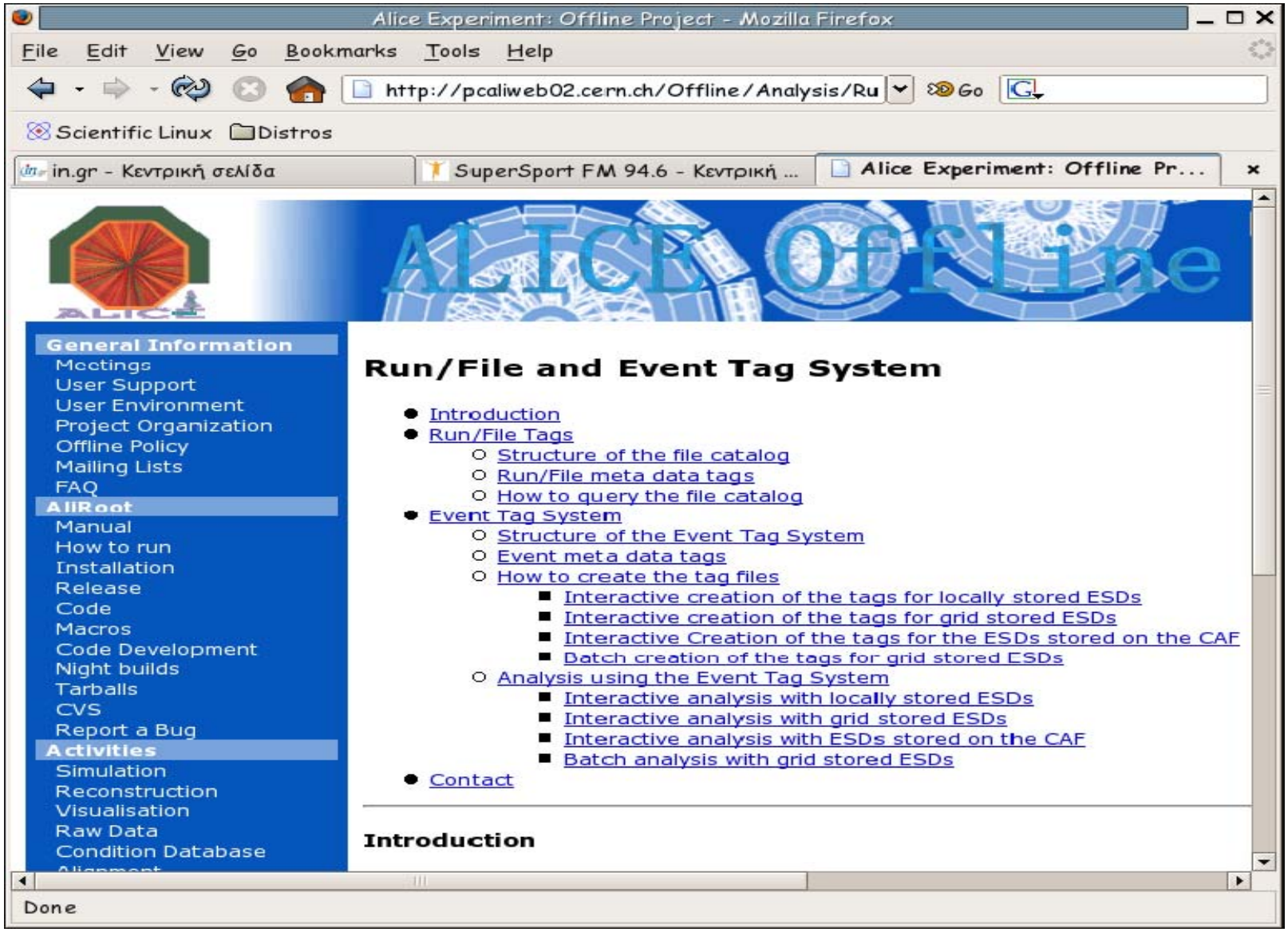


NEWS ON ANALYSIS

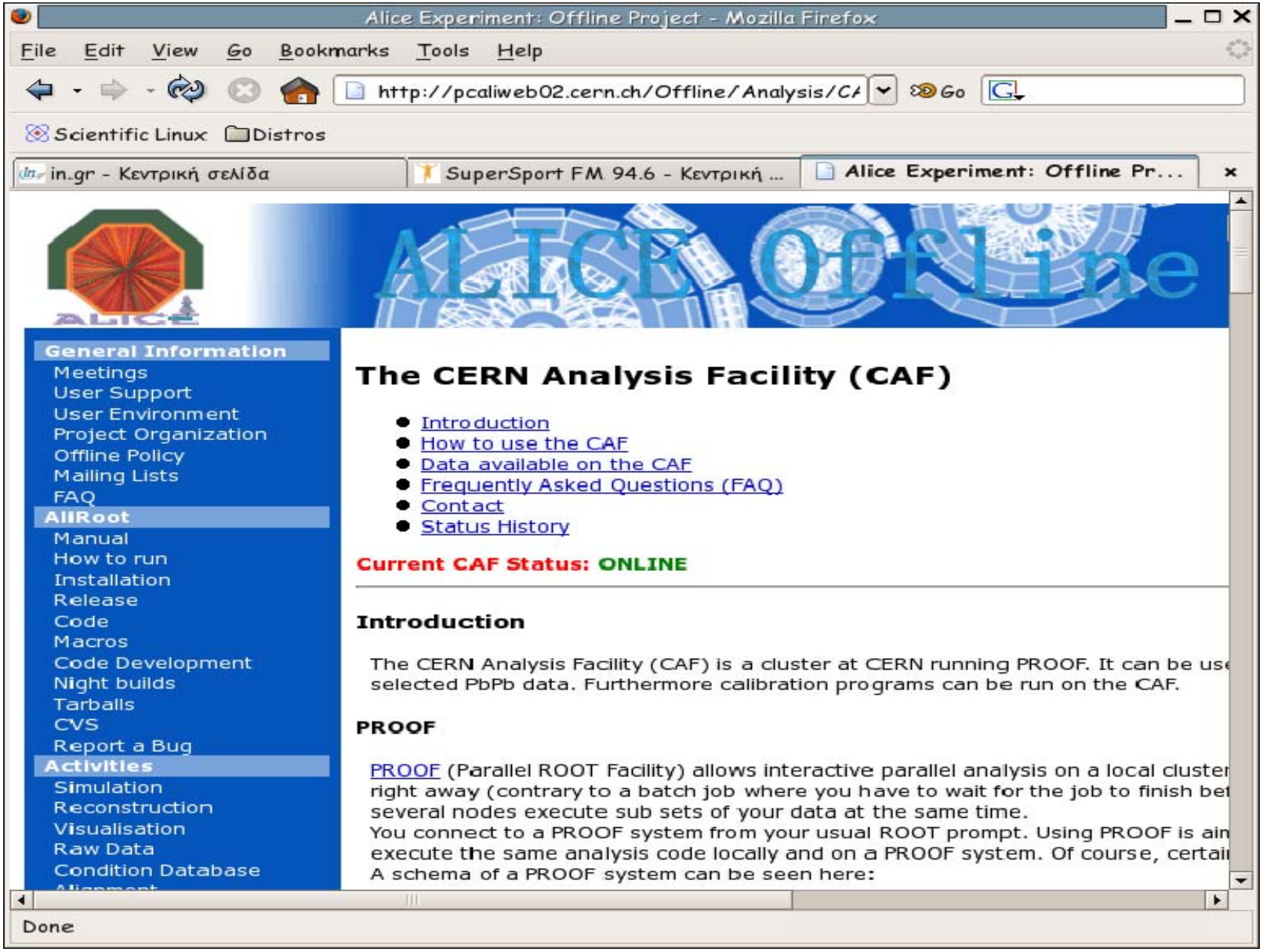
- A dedicated task force has been formed that deal with this task.
  - Many problems have been identified the majority of which have already been solved.
- There are dedicated links for every analysis topic under the official offline web page:
  - http://pcaliweb02.cern.ch/Offline/Analysis/RunEventTagSystem/
  - http://pcaliweb02.cern.ch/Offline/Analysis/CAF/
  - http://savannah.cern.ch/bugs/?group=aliroot
  - http://pcaliweb02.cern.ch/Offline/Analysis/FAQ/
- An internal note on the "Event Tag System" has been sent to the editorial committee. I received comments that will be included in the updated version.
- A new note on the analysis framework along with some practical examples is on the way. Both the interactive and the batch sessions will be addressed.



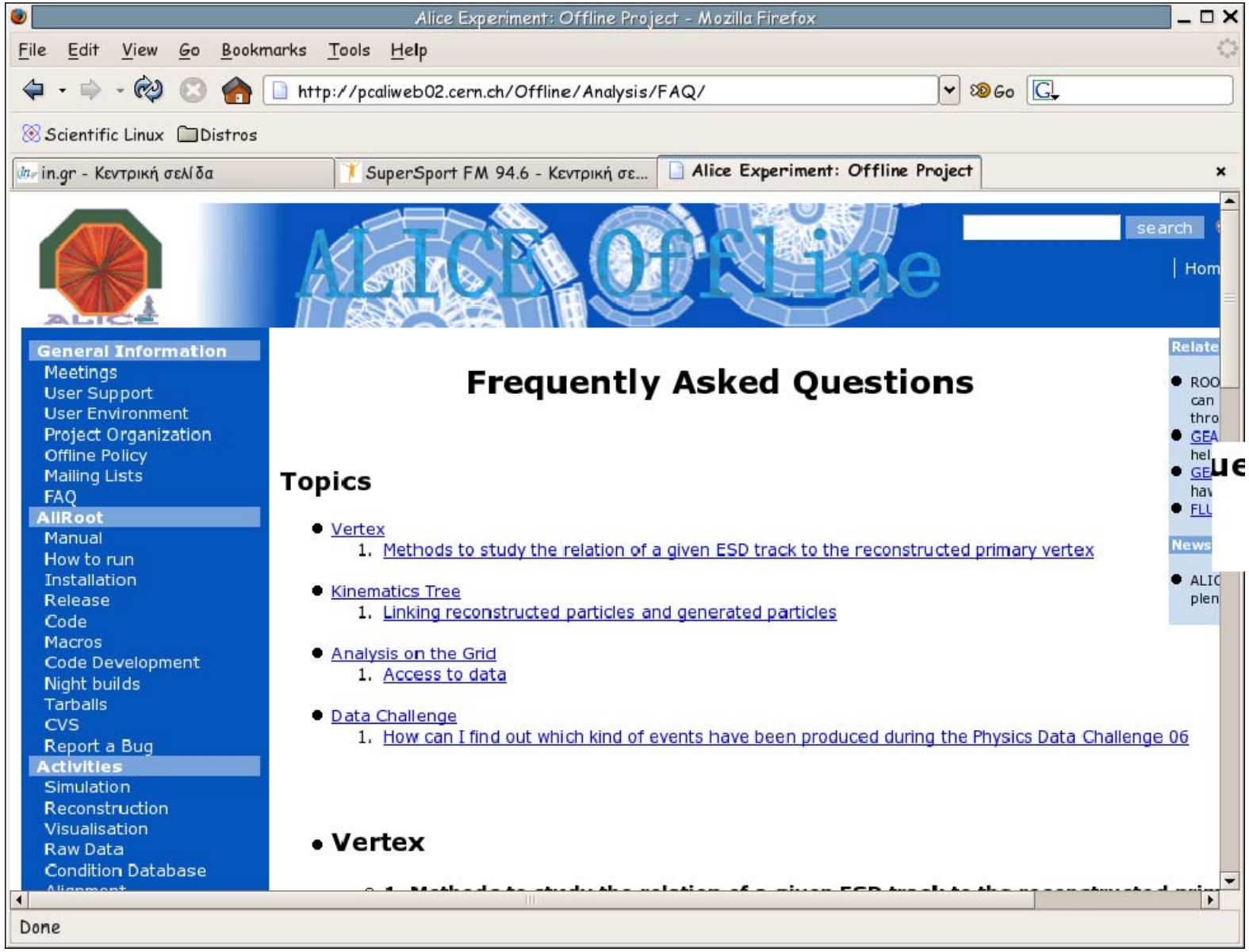
RUN-EVENT TAGS WEB PAGE



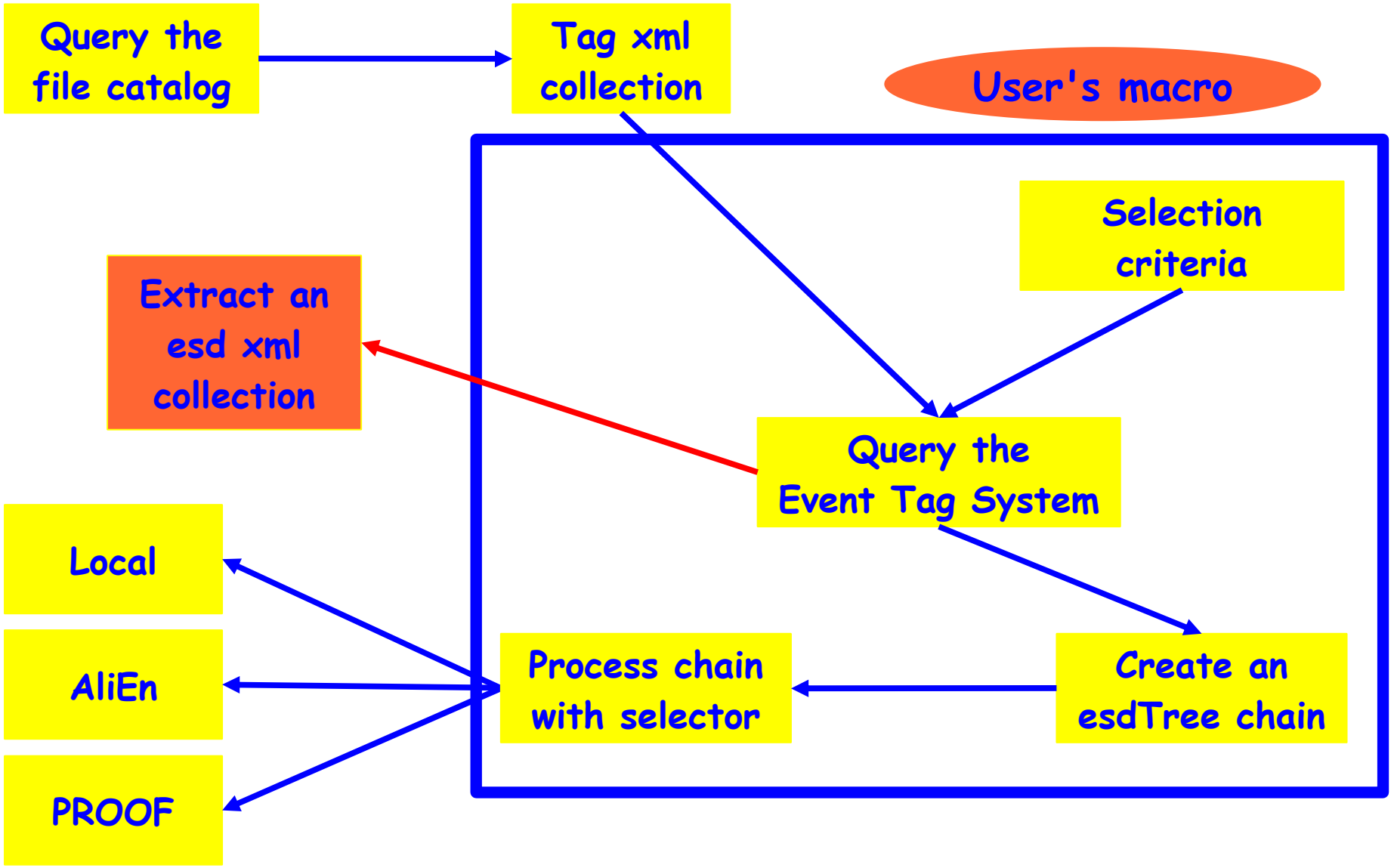
CAF WEB PAGE


Alice Experiment: Offline Project - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://pcaliweb02.cern.ch/Offline/Analysis/CA
Scientific Linux Distros
in.gr - Κεντρική σελίδα
SuperSport FM 94.6 - Κεντρική ...
Alice Experiment: Offline Pr...
ALICE OffLine
General Information
Meetings
User Support
User Environment
Project Organization
Offline Policy
Mailing Lists
FAQ
AliRoot
Manual
How to run
Installation
Release
Code
Macros
Code Development
Night builds
Tarballs
CVS
Report a Bug
Activities
Simulation
Reconstruction
Visualisation
Raw Data
Condition Database
The CERN Analysis Facility (CAF)
Introduction
How to use the CAF
Data available on the CAF
Frequently Asked Questions (FAQ)
Contact
Status History
Current CAF Status: ONLINE
Introduction
The CERN Analysis Facility (CAF) is a cluster at CERN running PROOF. It can be use
selected PbPb data. Furthermore calibration programs can be run on the CAF.
PROOF
PROOF (Parallel ROOT Facility) allows interactive parallel analysis on a local cluste
right away (contrary to a batch job where you have to wait for the job to finish bet
several nodes execute sub sets of your data at the same time.
You connect to a PROOF system from your usual ROOT prompt. Using PROOF is ain
execute the same analysis code locally and on a PROOF system. Of course, certai
A schema of a PROOF system can be seen here:
Done

FAQ WEB PAGE


Alice Experiment: Offline Project - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://pcaliweb02.cern.ch/Offline/Analysis/FAQ/
Scientific Linux Distros
in.gr - Κεντρική σελίδα
SuperSport FM 94.6 - Κεντρική σε...
Alice Experiment: Offline Project
ALICE Offline
search
General Information
Meetings
User Support
User Environment
Project Organization
Offline Policy
Mailing Lists
FAQ
AliRoot
Manual
How to run
Installation
Release
Code
Macros
Code Development
Night builds
Tarballs
CVS
Report a Bug
Activities
Simulation
Reconstruction
Visualisation
Raw Data
Condition Database
Frequently Asked Questions
Topics
Vertex
1. Methods to study the relation of a given ESD track to the reconstructed primary vertex
Kinematics Tree
1. Linking reconstructed particles and generated particles
Analysis on the Grid
1. Access to data
Data Challenge
1. How can I find out which kind of events have been produced during the Physics Data Challenge 06
Vertex
Done

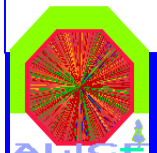
SCHEME OF THE ANALYSIS FRAMEWORK



FILE CATALOGUE STRUCTURE

- The path name is decided to be the following:
  - for 'real' data: **/data/<Year>/<AcceleratorPeriod>/<RunNumber>/**
  - for simulated data: **/sim/<Year>/<ProductionType>/<RunNumber>/**
- Subdirectories will be called:
  - **raw/**
  - **cond/**
  - **reco/<PassX>/ESD/**
  - **reco/<PassX>/AOD/**
  - **…**
- File names will look like this: **<xxxx>.AliESDs.root**
- **For further information see**
  http://indico.cern.ch/conferenceDisplay.py?confId=3280
  http://cern.ch/Oldenburg/MetaData/MetaData.doc

MARKUS OLDENBURG AN INTERNAL NOTE IS ON THE WAY

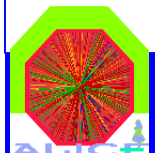


HOW DO WE QUERY THE FILE CATALOGUE (1)

- Get a valid alien session.

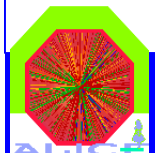
```
$ find -x pp
/alice/cern.ch/user/p/pchrista/production/pp/PDC06/*
AliESDs.root > pp.xml
```

Redirect the output to the xml collection.



HOW DO WE QUERY THE FILE CATALOGUE (3)

```
$ find -x pp
/alice/data/2008/LHC08a/*/reco/Pass3/*
*tag.root
Run:collision_system="pp" and
Run:stop<"2008-03-20 10:20:33" and
Run:start>"2008-03-19" > pp.xml
```



INTERACTIVE ANALYSIS SESSION (1)

Setup par archive
Load the needed libraries

```
TGrid::Connect("alien://pcapiserv01.cern.ch:10000","pchrist")
```

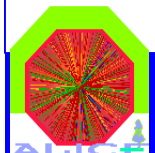
```
AliEventTagCuts *EvCuts = new AliEventTagCuts();
EvCuts->SetMultiplicityRange(0,1500);
```

```
TAlienCollection* coll = TalienCollection::Open("tag100.xml");
TGridResult* TagResult = coll->GetGridResult("");
```

```
AliTagAnalysis *TagAna = new AliTagAnalysis();
TagAna->ChainGridTags(TagResult);
```

```
analysischain = TagAna->QueryTags(EvCuts);
```

```
const char *selectorfile = "esdPt.C";
analysischain->Process(selectorfile);
```



INTERACTIVE ANALYSIS SESSION (2)

Setup par archive
Load the needed libraries

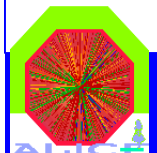
```
TGrid::Connect("alien://pcapiserv01.cern.ch:10000","pchrist")
```

```
AliEventTagCuts *EvCuts = new AliEventTagCuts();
EvCuts->SetMultiplicityRange(0,1500);
```

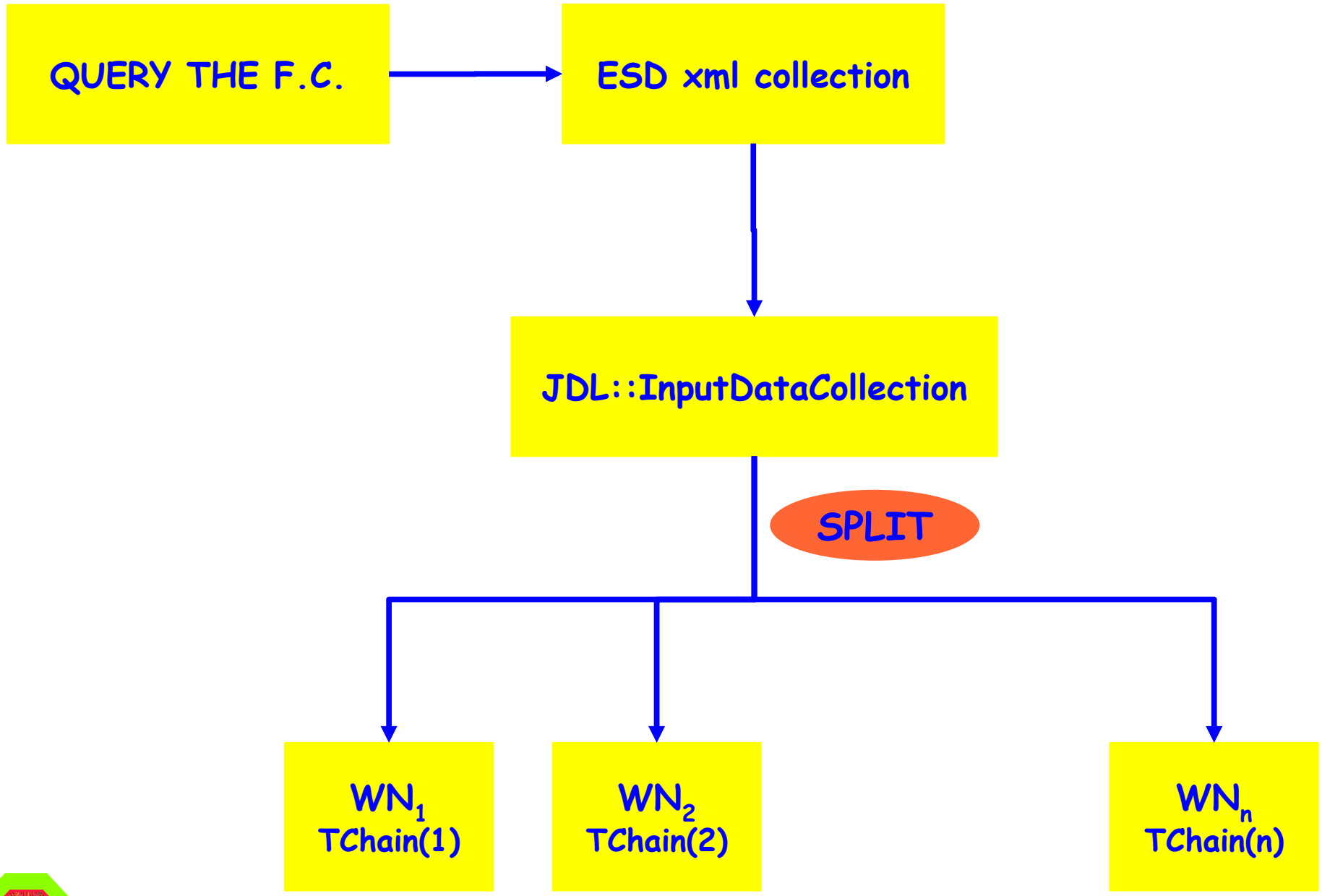
```
AliTagAnalysis *TagAna = new AliTagAnalysis();
TagAna->ChainLocalTags("home/pchrist/ALICE/PDC06/Tags");
```

```
analysischain = TagAna->QueryTags(EvCuts);
```

```
const char *selectorfile = "esdPt.C";
analysischain->Process(selectorfile);
```

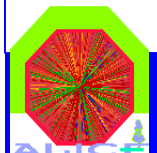


BATCH ANALYSIS SESSION (1)



BATCH ANALYSIS SESSION (2)

- The Event Tag System was created to provide a global TChain along with the associated TEventList.
- The iteration over the entries of this list works if we are performing interactive sessions.
- In the case where the batch job needs to be splitted we needed to find a way to have this functionality on every worker node where the job will run.
- A few modifications and developments were needed by the ROOT team:
  - At the moment an overall TEventList is assigned to a TChain.
  - This will change and a TEntryList will be assigned to every entry of the TChain (TChainElement).
  - I will do extensive tests within the next 2 weeks, which is the time needed by the ROOT team to provide me with a working version of the latest developments.



BATCH ANALYSIS SESSION (3)

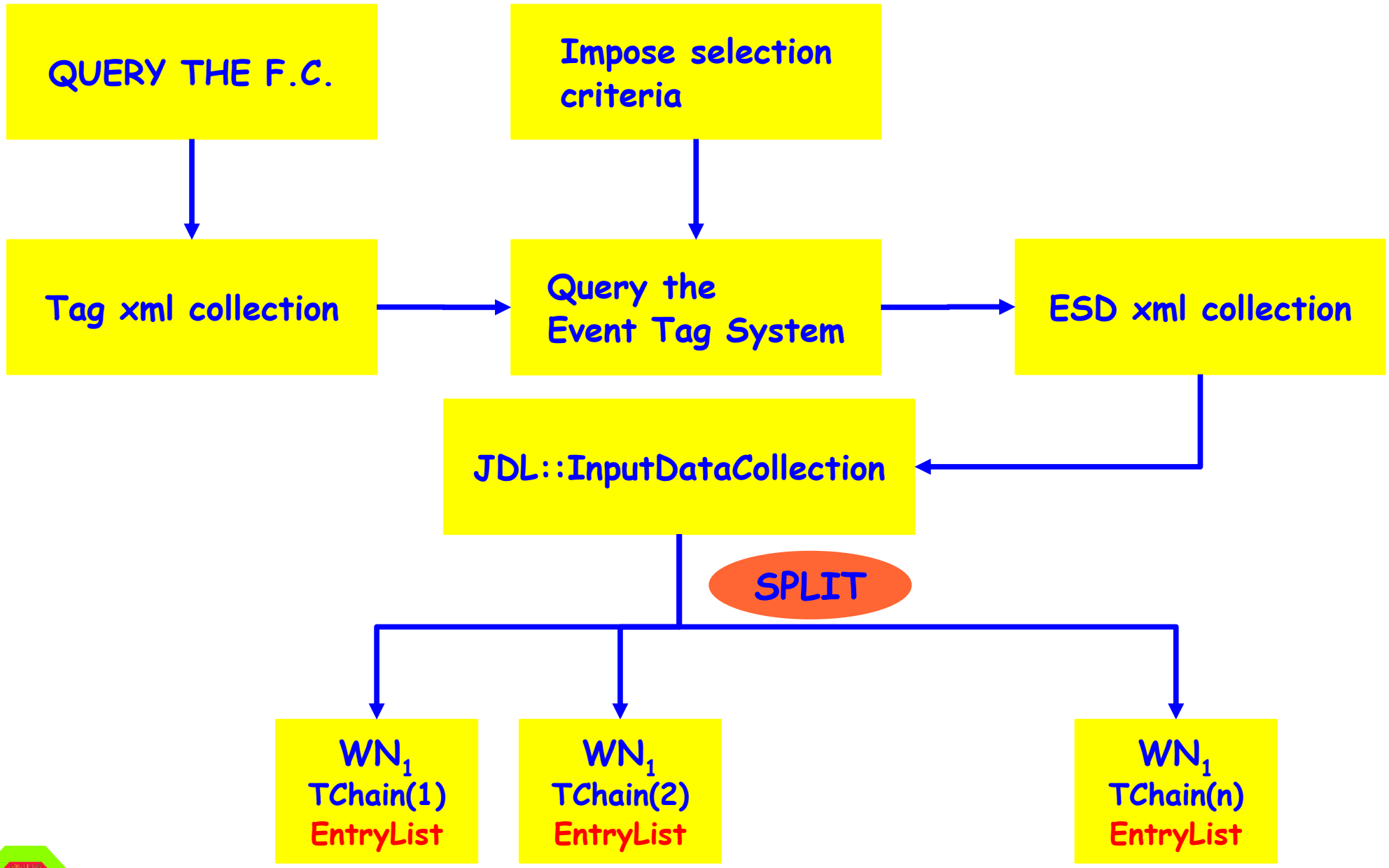
QUERY THE F.C.

Impose selection criteria

Tag

```
<?xml version="1.0"?>
<tags>
  <collection name="test">
    <event name="1">
      <file name="AliESDs.root"
       guid="0096dce6-f62b-14de-97bf-5519e793beef"
       eventlist="4,14,44,61,66,67,74,77,88,89,90" />
  </collection>
</tags>
```

BATCH ANALYSIS SESSION (3)



NEXT STEPS - SUMMARY

- A note on the Event Tag System has been sent to the editorial committee. The note will be submitted next week.
- A new note that will address the issues of the distributed analysis (interactive and batch) is on the way.
- Users are encouraged to join the ATF and try the analysis framework. In case of problems you should report to savannah.
- Following the developments of ROOT, there will be a few modifications on the existing framework for the batch sessions.
- I will do extensive tests within the next 2 weeks, which is the time needed by the ROOT team to provide me with a working version of the latest developments.
- A GUI that will be used for job submission is being built. A preliminary prototype has been developed by Jose Lo which will be the base of the new attempt.