



A Large Ion Collider Experiment

European Organisation for Nuclear Research



LHC End-User Analysis

Andreas Morsch

CERN / ALICE

Grid Fest

CERN, October 3, 2008

[Data Processing Stages]



Reconstruction

RAW Data

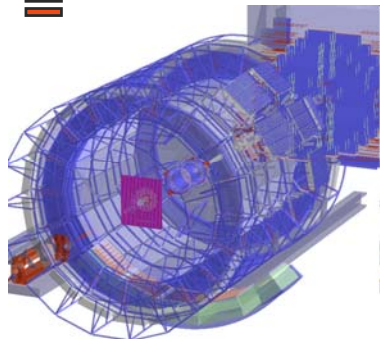
Event Summary Data (ESD)

Analysis Object Data (AOD)

Derived Physics Data (DPD)

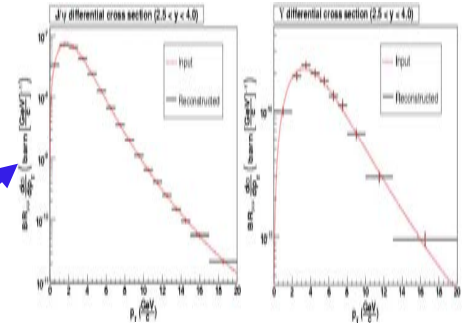


Event Simulation



Data reduced to essential information for analysis

Graphical Representation



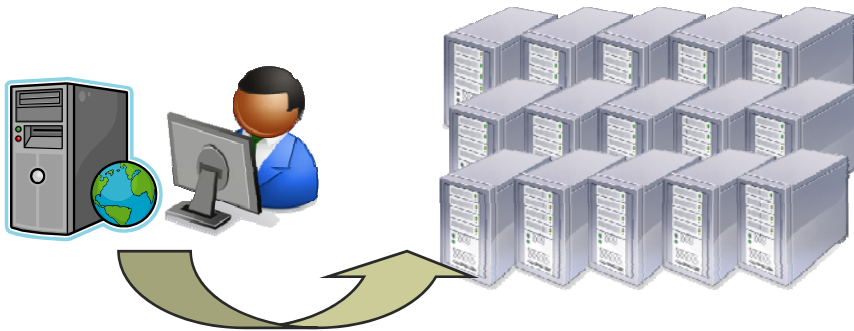
Used by physicists for final data analysis

LHC Analysis Stages

- **Scheduled analysis** for conversion of ESD into AOD



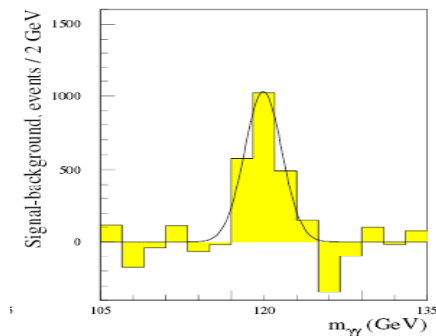
- For testing, **interactive user analysis** on small event samples on local computers or clusters.



- Individual **batch analysis** on full data sets.

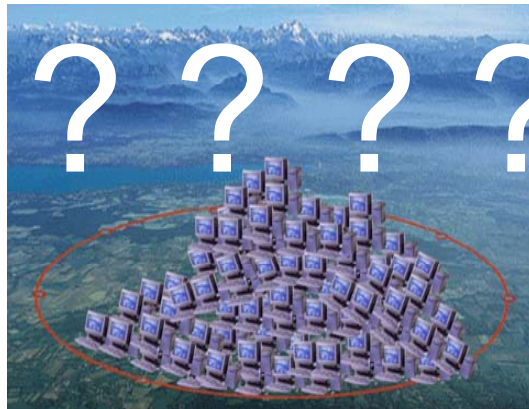


- Output sent back to user to be analyzed interactively. **Final plots** are produced.



[Analysis at the LHC]

- Analysis of LHC data
 - Huge data sets (typically 10^9 AOD events/year, 5kB – 1MB/event)
 - Where to store the data ?
 - How to get enough computing power to analyze all the data ?
 - How to guarantee data access for every physicist in all institutes and universities ?



- As simulation and reconstruction, analysis requires computing resources exceeding the capacities of local computing centers.
- LHC experiments rely on distributed analysis on large computing grids (WLCG).

[Analysis on the Grid]

- The important new element here is of human nature:
 - Analysis is performed by many individuals, possibly the whole community of physicists.
- The challenge is to give similar opportunities to everybody while maximizing the discovery potential of the whole collaboration.

Analysis on the Grid: *The Grid perspective*

- To alleviate this problem standard analysis can also be performed using scheduled productions just like reconstruction and event simulation.
 - Mainly used in the ESD→AOD processing
 - So called *Analysis Trains* represent a compromise between individual and scheduled analysis. The wagons of the train are individual user tasks which are executed in a sequence during frequently scheduled runs of the train over a full data set.

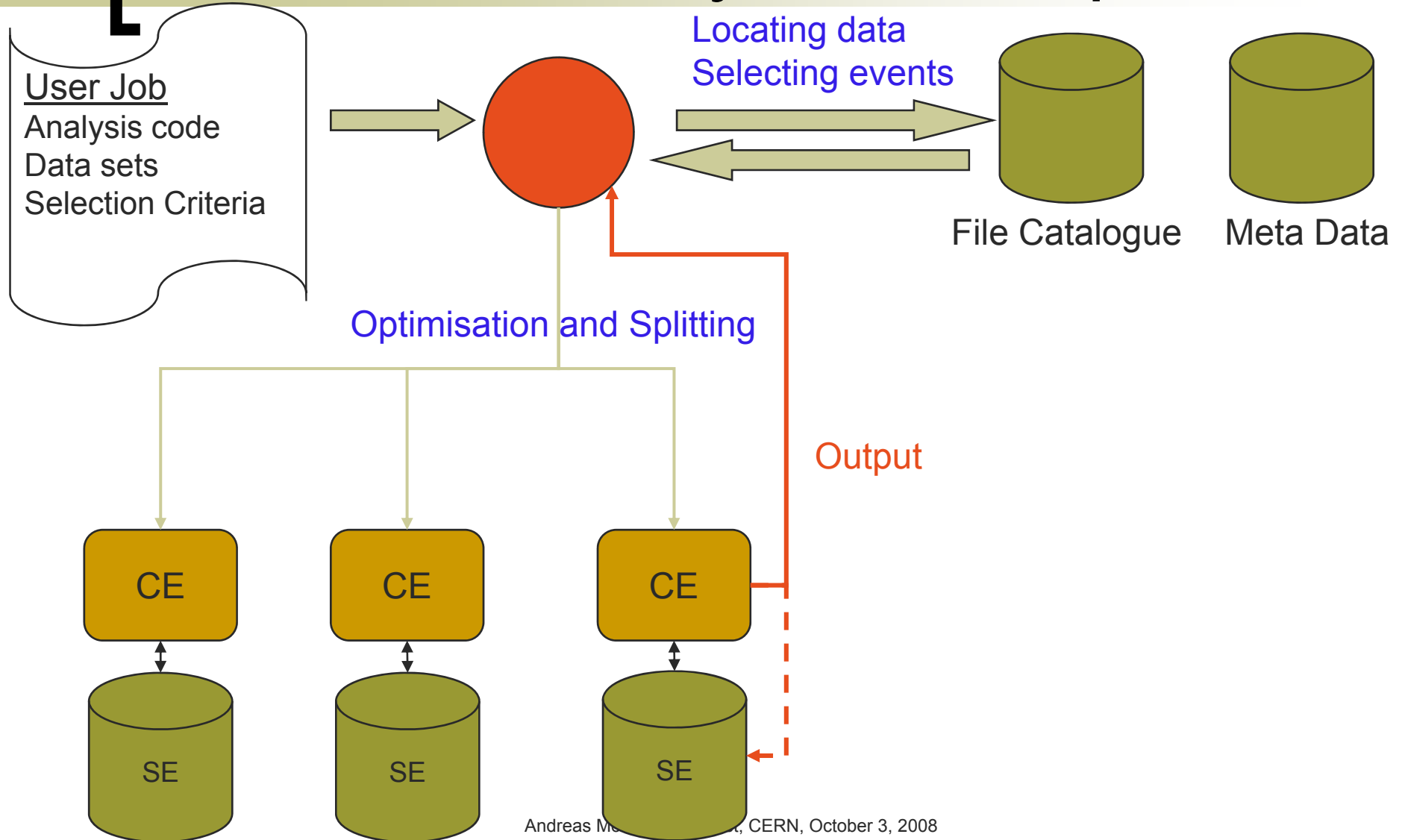
Analysis on the GRID: *The user perspective*

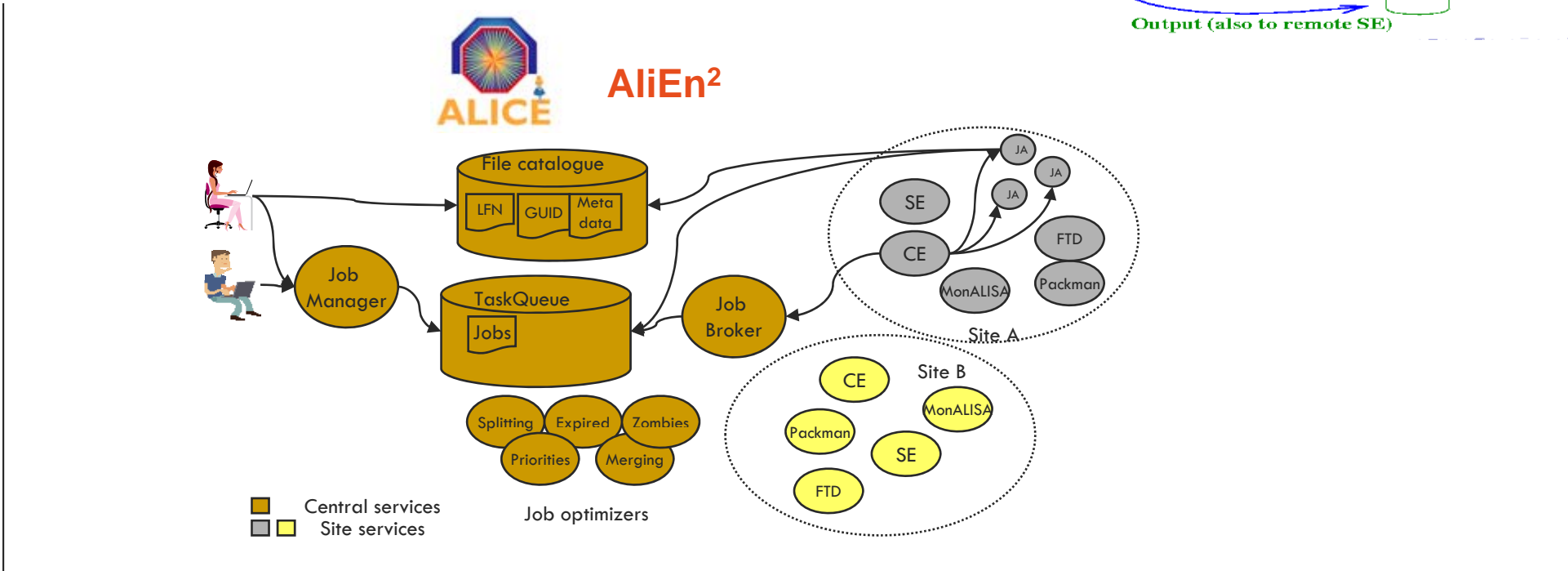
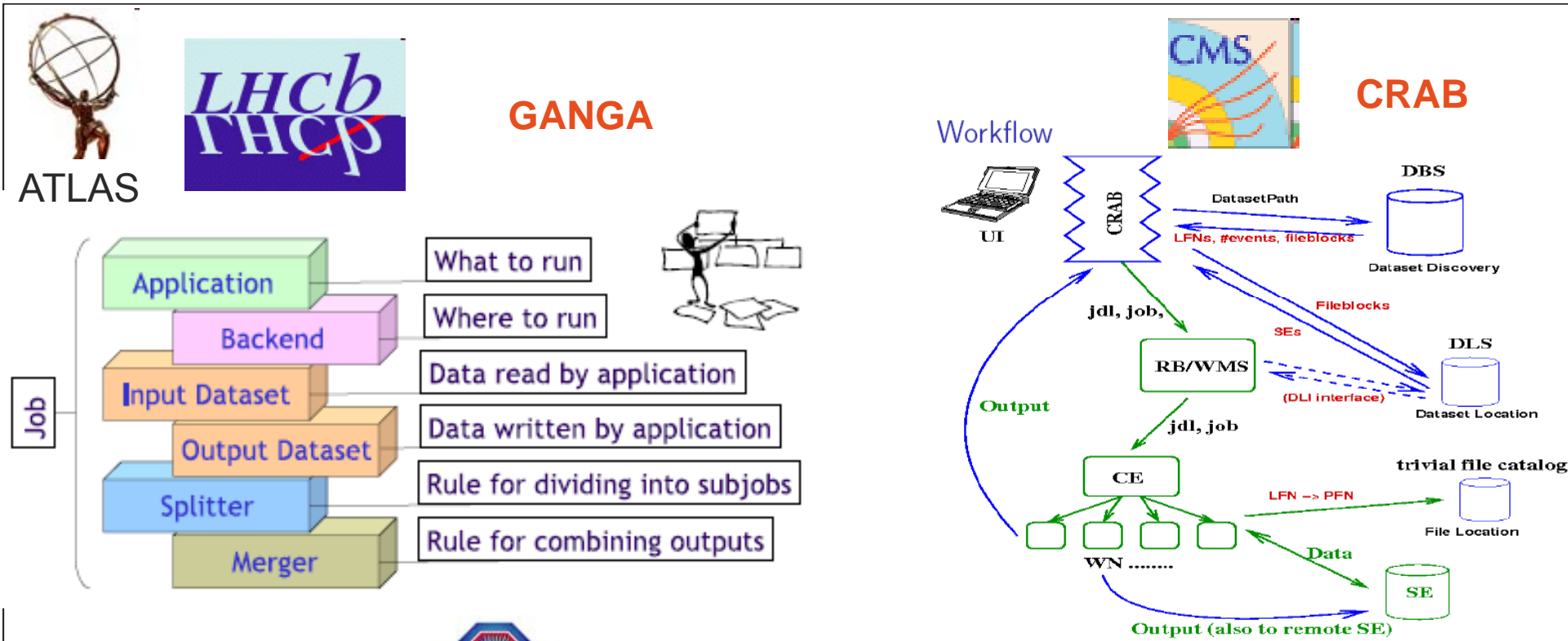
- End users have a wide spectrum of computing skills and physics interests.
 - LHC experiments have developed **simple grid entry points and interfaces** for the non-expert user.
 - Shield the user from the complexity of the grid
 - Provide transparent data access
 - Analysis frameworks protect the investment made in development of analysis code against **changes in the computing infrastructure**.
 - Format of analysis objects and analysis algorithms are bound to change rapidly as understanding of data proceeds and new discoveries are made.
 - **Frameworks are designed for changes.**

[GRID Analysis Tools]

- The LHC experiments have designed and implemented Analysis Frameworks and Grid Analysis Tools tailored according to their specific needs.

Distributed Analysis: Principle





[Summary]

- LHC experiments rely on the Grid for Data Analysis
- The “Grid for Everybody” has been realized by implementing simple Grid entry points and user interfaces.
- Hundreds of physicists are already using the Grid for analysis.