Data Analysis on Handheld Devices Using Clarens

Tahir Azim
NUST

Background

- High Energy Physics (HEP) data storage requirements
  - ROOT files, HBOOK files, HEPREP files …
- Data not conclusive in its raw form
- Grid-enabled analysis
  - Graphical Visualizations (2D – 3D)
- Desktop Analysis Applications Available
  - Java Analysis Studio (histogramming and analysis package)
  - WWW Interactive Remote Event Display (detector geometry and event display)
  - ROOT (efficient file format)
- Requirement of efficient Grid-enabled analysis capabilities on handheld devices

The Grid For Handheld Devices…

- Handheld devices can enable ubiquity
- Portable access to data
- But they cannot provide all the computing power to do data analysis and processing on the fly
- Grid & distributed computing can provide the power of much more resourceful computers constantly on your handheld

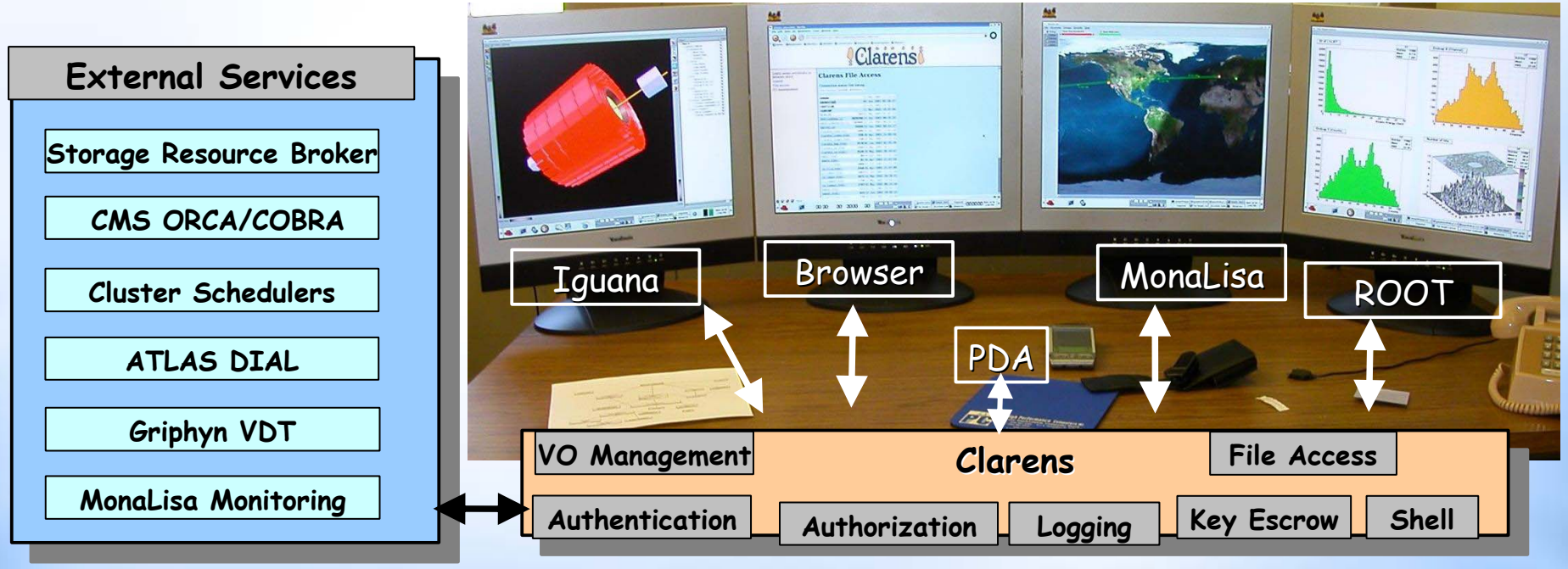
Challenges of Handheld Devices

- Relatively slow processors
- Small display screen
- Small storage space (excluding add-on memory cards)
- Intermittent, relatively low bandwidth network connections
- Lack of software support
  - Bugs in JVMs
  - Java 1.1 instead of Java 1.4
  - Embedded Visual C++ instead of MS Visual C++
  - Only a small subset of operating system libraries and APIs available
  - Relatively difficult installation procedures

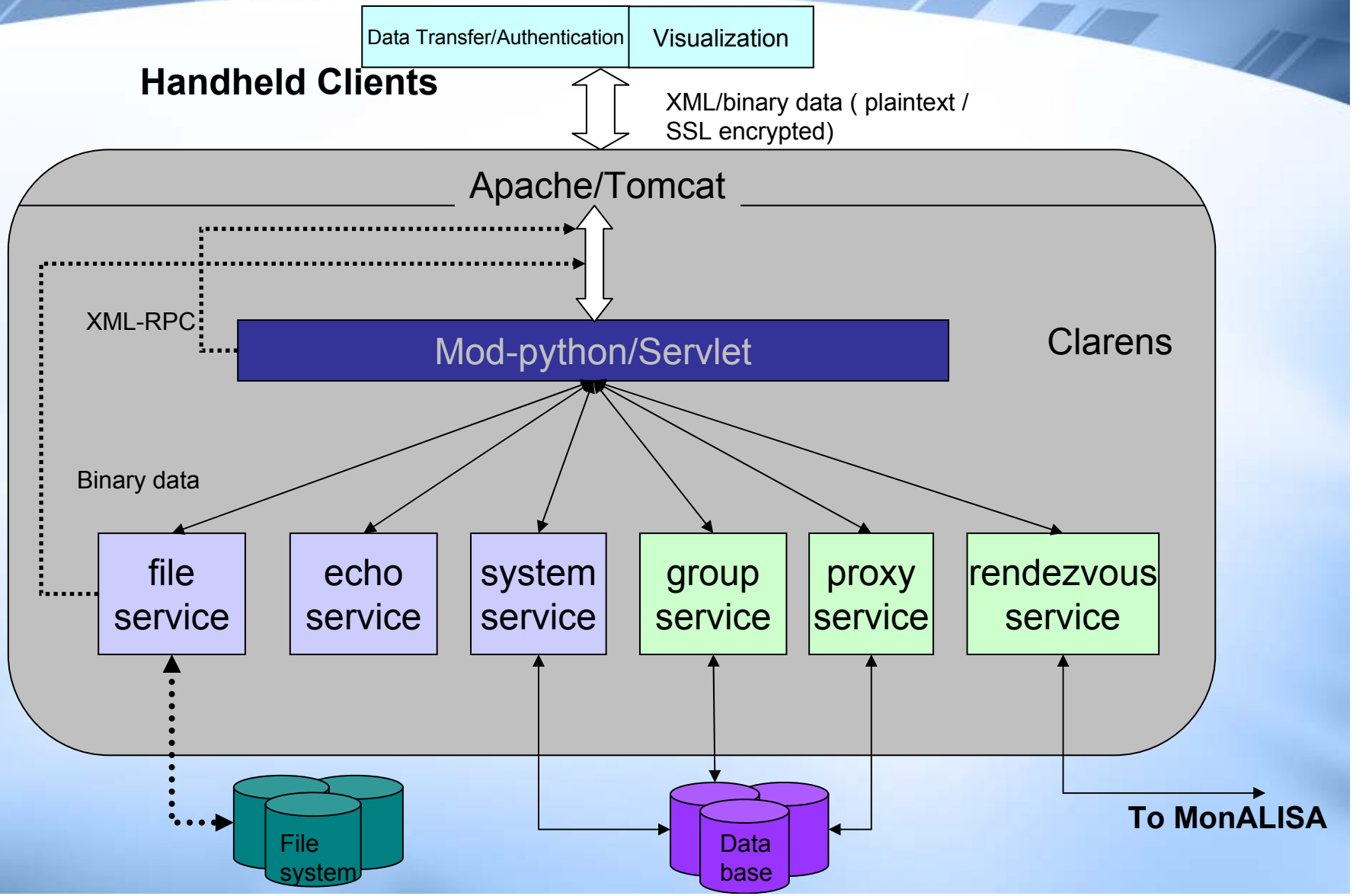
The Clarens Web Services Framework

- Clarens is a high-performance, Grid-enabled web service framework
- Supports lightweight XML-RPC protocol for accessing all kinds of tools and services
- Provides a uniform way of accessing widely differing (e.g. in terms of access protocols, running platforms) Grid tools
- Provides a remote file access service and a catalog service to look for distributed datasets
- A shell service to provide shell-like access to remote machines
- Clarens services can be designed to act as lightweight wrappers to powerful backend utilities and to other services
- Available in Python and Java implementations

View of a Collaborative Desktop



Architectural Overview

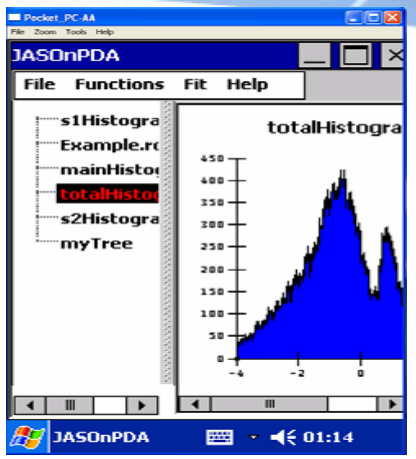


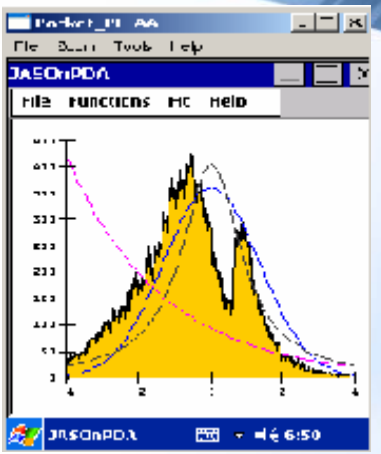
[Technologies for Development on Handheld Clients]

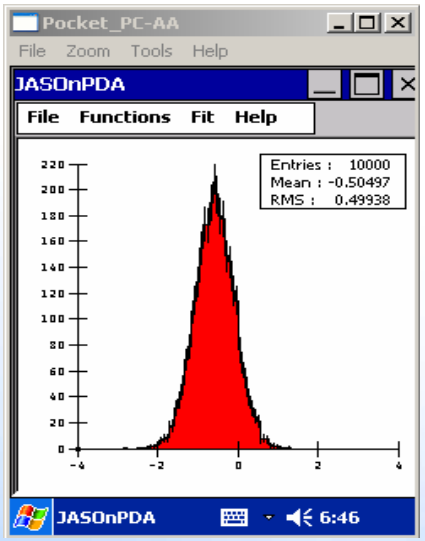
- Plethora of available technologies:
  - IBM WebSphere
  - MIDP
  - SuperWaba
  - Savaje
  - Personal Java
    - Insignia Jeode
    - CrEme

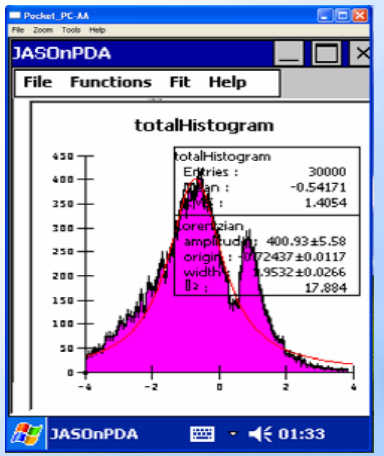
Pocket JAS

- Displays histograms and scatter plots from data in ROOT files and flat files
- ROOT IO rewritten to make it PDA-compatible
- Grid authentication and security built-in
- Uses Clarens ‘file’ services to look for and download ROOT files
- Supports X.509 as well as PKCS12 certificates
- Displays various statistics about the data sets and can fit function curves against the histogram data
- Handles unreliable, intermittent data transfers over wireless links

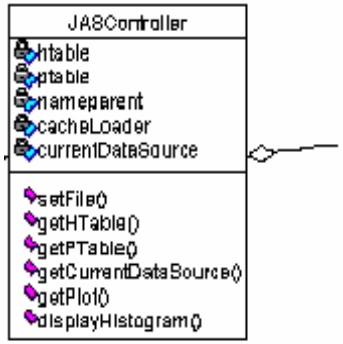






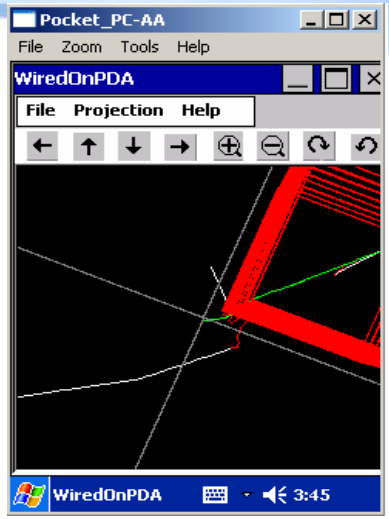


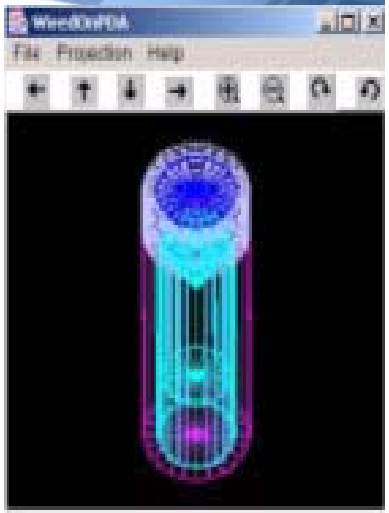
[Pocket JAS plug-in architecture]

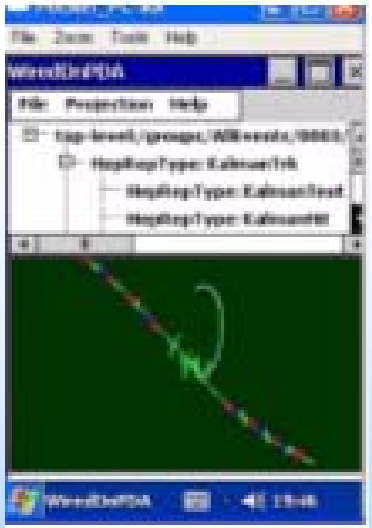


Pocket Wired

- 3D display of event data from HepRep and HepRep2 files
- Display of HepRep event tree
- Most functionality from WIRED ported including:
  - 2D Translation & Rotation
  - Scaling
  - 3D rotation
  - Projections
- Remarkable enhancement in speed by using CrEme JVM instead of Jeode

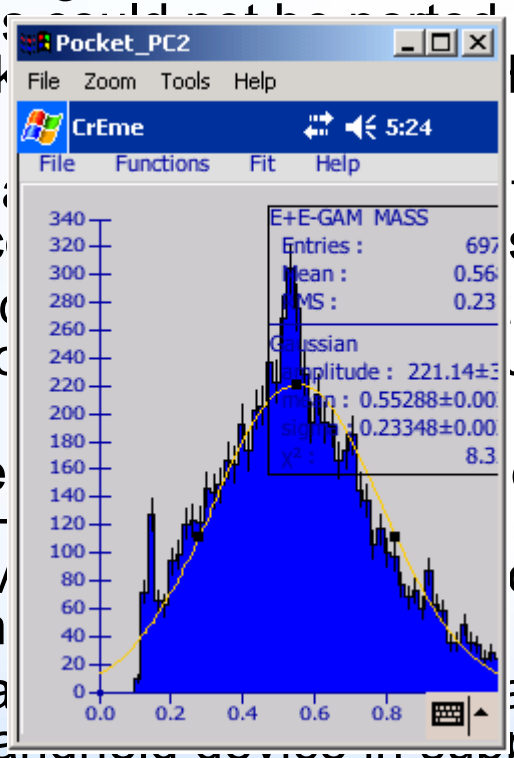


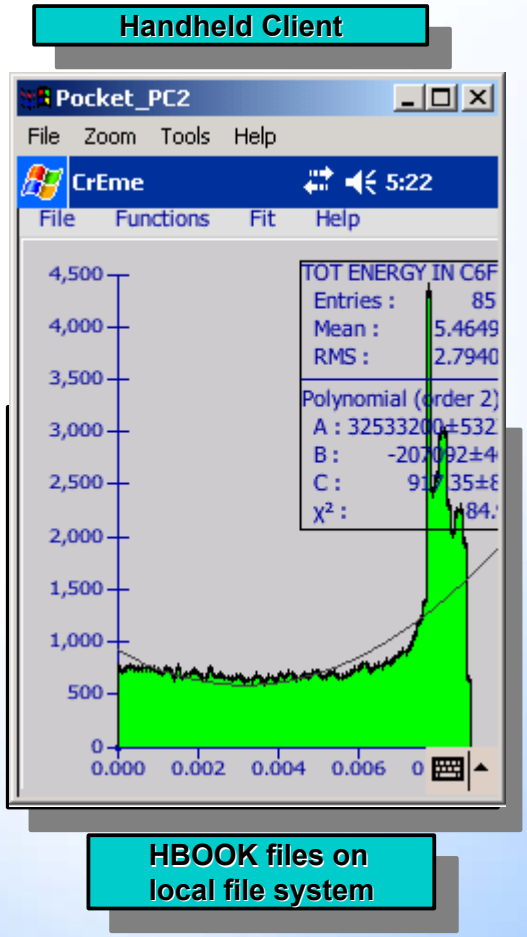




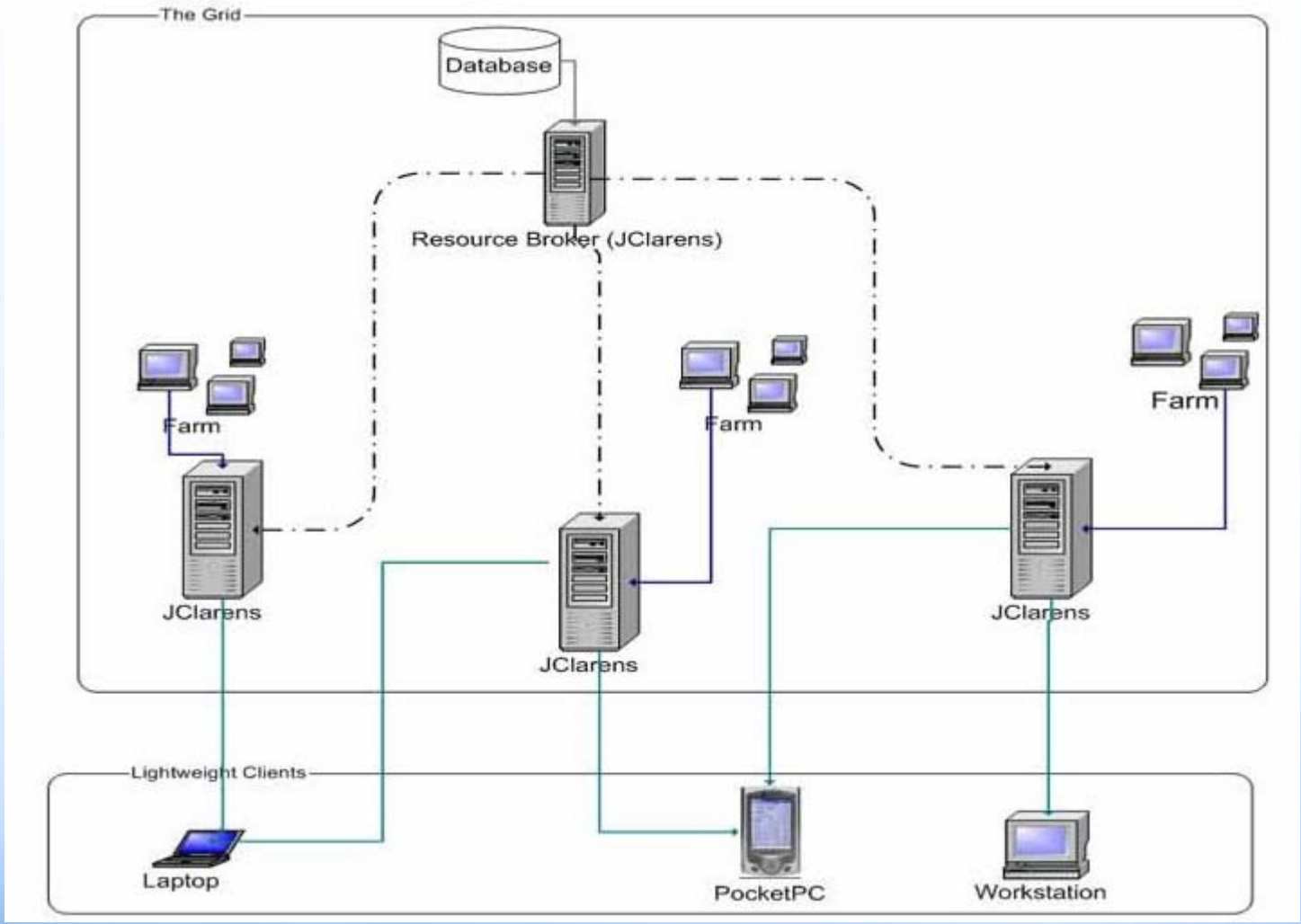
Using Clarens as the Data Processing Engine

- Parsing libraries of certain file types could not be ported to Pock IBOOK files
- A Cl tten for proc s
- Hand he remo uired files
- Sele ed by an 'H conv ormat for h
- Data analysis to ha pported format





Distributed Analysis with a Condor-based Execution Engine



Features of the distributed analysis environment

- Finds convenient places for the users “job” (computing task) to be run
- Seeks to use the widely dispersed resources efficiently
- Deals with authentication to the different sites that the scientists will be using
- Runs the jobs
- Monitors progress
- Recovers from problems
- Tells you when the work is complete and transfers the result back

Monitoring Service

- Currently uses the end host monitoring agent (EMA) for monitoring
  - EMA is a plug-in of MonALISA developed at NUST
- Calculates an estimate of the load of each server
- Reports OS information, CPU Clock rate, load coefficient and other monitoring information and load coefficient to the resource broker

Job Submission Service

- Receives job requests from the handheld clients
- Decides the best site to submit the jobs based on monitoring information and OS specified
- Forwards the job requests to remote site
- Maintains a mapping of job IDs and submission sites
- Uses **CONDOR** for job submission; thereby enabling parallel job execution on farms and clusters
- Clients do not have to know where the job was finally submitted
- Provides status information to the handheld clients
- Enables the handheld clients to get back outputs of executed jobs

Results

- Handheld clients are able to access and utilize the system transparently
- Handhelds are able to do analysis with exceptionally good performance
- Analysis tasks on large number of data files can be started in parallel for faster results
- Using Clarens, lightweight handheld clients, ranging from PDAs to mobile phones, can make use of powerful computational services

Conclusion

- Work presented and used at major international workshops and exhibitions
  - Demonstration at 1st GAE Workshop at Caltech in June 2003
  - Work presented as part of “Grid Enabled Analysis” demo at ITU Telecom World 2003, carried out by Caltech, CERN, UERJ, UPB, KEK, Sinica and NUST
  - Various demonstrations in workshops and conferences in Pakistan
- Clarens can enable lightweight devices, like PDAs and mobile phones, to access Grid resources and distributed datasets
- Work demonstrates how the Grid can be used to provide powerful computational and processing capabilities on handhelds