ATLAS Grid Planning

- ATLAS has used in “production mode” different Grids with simulation jobs
  - NorduGrid, US VDT like, EDG
- Similar use of the same Grids is in progress for the reconstruction
- ATLAS intends to use LCG-1 as much as possible as soon as it will be available (but the use of the other Grids will not disappear at once)
- The next DC (DC2) is foreseen for start 2004: a “usable” (75% effic?) LCG-1 with the agreed functionality (GDB WG1) should be available no later then September, to avoid running too much risks.

Layout

- DC1-2 Figures
- Work done and planned for each Grid flavor
  - Nordugrid, US Grid, EDG
- Production/Grid tools development status and plan
  - Magda (replica catalogue), AMI (Metadata DB), Chimera (VDC), GANGA
  - ATCOM :prod.scripts generation system, Magda, AMI interfaced
- Toward a Grid production (analysis) system

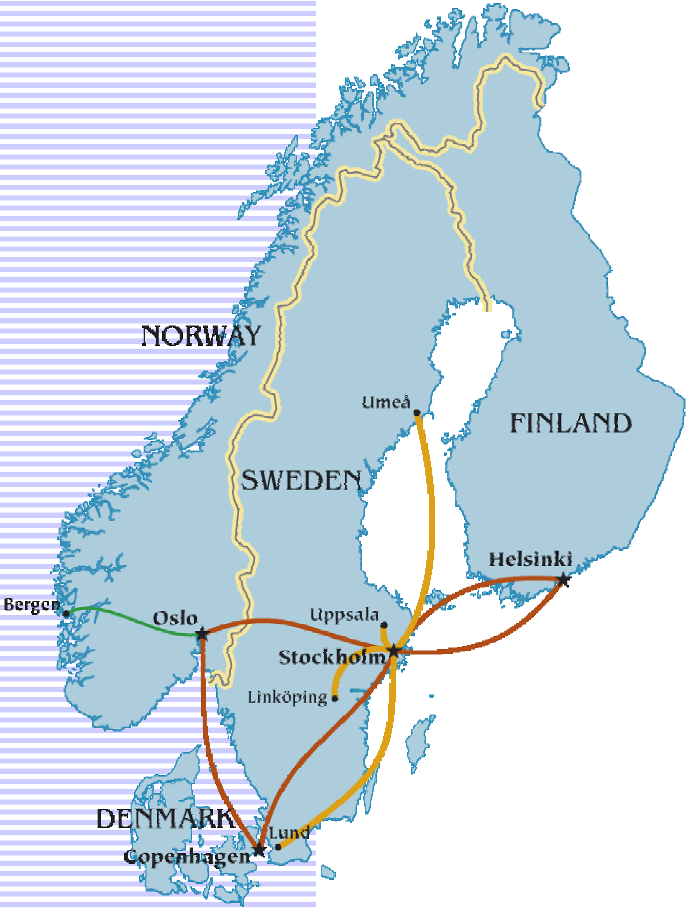
Figures for DC1 and beyond

- DC1 simulation
  - 10^7 events, 3 10^7 single particles: about 550 kSp2K months (100% effic.)
  - with pileup (10^{33}*2 & 10^{33}*10) 1.3 & 1.1 M events: about 40 kSp2K months (100% effic.)
- Reconstruction
  - Done till now 1 M (high prio. events) for each luminosity: about 50 kSp2K months (100% effic.): redo in the next few months, partly with Grids
  - At the some time reconstruct a fraction of the lower priority, partly with Grids too
- DC2 start in 2004, 2-3 times DC1 CPU, then full reconstruction
  - Use LCG-1 as much as possible, still some Grid activity foreseen outside LCG

Nordugrid in DC1 and beyond

- Fall 2002: NorduGrid is no longer considered a “test”, but rather a facility
  - Non-ATLAS users at times are taking over
  - Simulation of the full set of low ET dijets (1000 jobs about 25 hours each, 1 output partition each) *August 31 to September 10*
- Winter 2002-2003: running min. bias pile-up
  - Prevoius sample + 300 jobs dijets ET>17 GeV *Done by March 5th*
  - Some sites can not accommodate all the needed min. bias files, hence jobs are not really data-driven any longer
- As we are speaking: running reconstruction
  - The NorduGrid facilities and middleware are very reliable (people at times forget it’s actually a Grid setup)
  - Processing the data simulated above + other 1000 input files
  - No data-driven jobs
- The biggest challenge – to “generalize” the ATLAS software to suit everybody and to persuade big sites to install it
- These are **no tests**, but a **real** work, as there are no alternatively available conventional resources

Nordugrid resources (O.Smirnova)

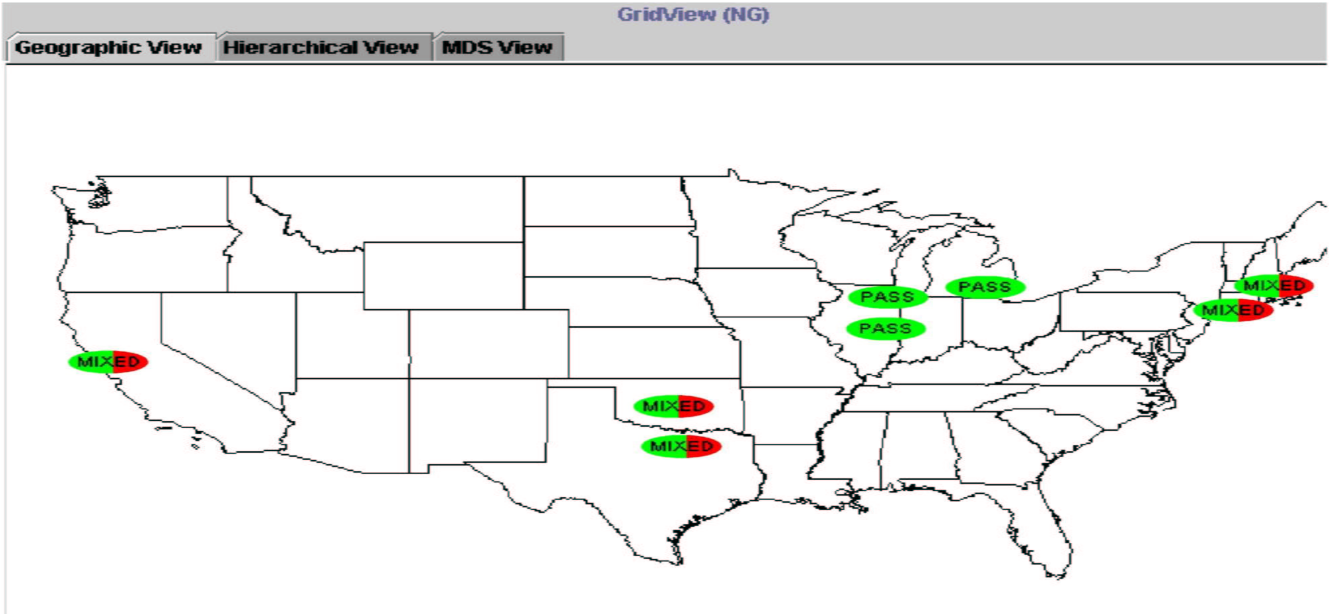


- Harnesses nearly everything the Nordic academics can provide:
  - 4 dedicated test clusters (3-4 CPUs)
  - Some junkyard-class second-hand clusters (4 to 80 CPUs)
  - Few university production-class facilities (20 to 60 CPUs)
  - Two world-class clusters in Sweden, listed in Top500 (200 – 300+ CPUs)
- Other resources come and go
  - Canada, Japan – test set-ups
  - CERN, Russia – clients
  - It's open, anybody can join or part
- People:
  - the "core" team grew to 7 persons
  - Sysadmins are only called up when [ATLAS] users need an upgrade

DC1 and GRID in U.S. (K.De mid-april)

- Dataset 2001: 10^6 jet_25
  - simulated at BNL using batch system
  - lumi10 pileup done using grid at 5 testbed sites
  - finishing lumi10 QC right now
  - reconstruction started using BNL batch system
  - grid reconstruction using Chimera starting soon
- Dataset 2002: 500k jet_55
  - simulated at BNL using batch system
  - 30% lumi02 piled-up using grid
  - to be finished after 2001 is completed
- Datasets 2107, 2117, 2127, 2137: 1 TeV single particles
  - simulated on grid testbed. Pile-up?
- Dataset 2328,2315: Higgs, SUSY
  - simulation completed, pile-up after dataset 2001

U.S. ATLAS Grid Testbed (K.De)



- BNL - U.S. Tier 1, 2000 nodes, 5% ATLAS (100), 10 TB
- LBNL - pdsf cluster, 400 nodes, 5% ATLAS (20) , 1 TB
- Boston U. - prototype Tier 2, 64 nodes
- Indiana U. - prototype Tier 2, 32 nodes

- UT Arlington - 20 nodes
- Oklahoma U. - 12 nodes
- U. Michigan - 10 nodes
- ANL - test nodes
- SMU - 6 nodes
- UNM - new site

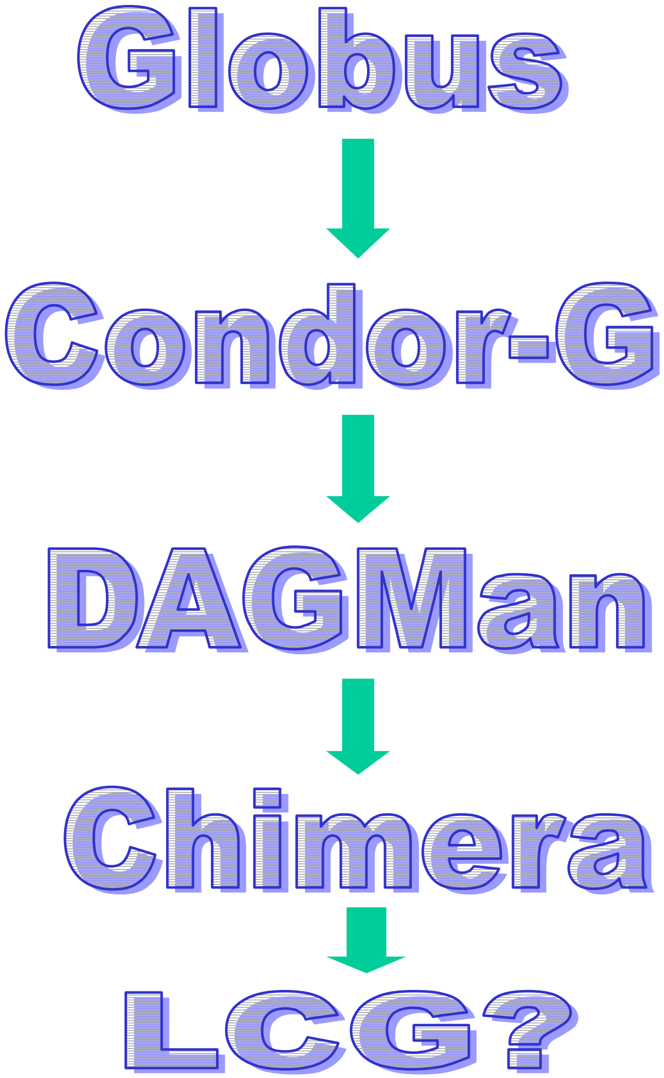
Grid Quality of Service (K.De)

- Anything that can go wrong, WILL go wrong
  - During 18 days of grid production (in August), every system died at least once
  - Local experts were not always be accessible
  - Examples: scheduling machines died 5 times (thrice power failure, twice system hung), Network outages multiple times, Gatekeeper died at every site at least 2-3 times
  - Three databases used - production, magda and virtual data. Each died at least once!
  - Scheduled maintenance - HPSS, Magda server, LBNL hardware, LBNL Raid array…
  - Poor cleanup, lack of fault tolerance in Globus
- These outages should be expected on the grid - software design must be robust
- We managed > 100 files/day (~80% efficiency) in spite of these problems!

GRAT Software (K.De)

- GRid Applications Toolkit
- Used for U.S. Data Challenge production
- Based on Globus, Magda & MySQL
- Shell & Python scripts, modular design
- Rapid development platform
  - Quickly develop packages as needed by DC
    - Single particle production
    - Higgs & SUSY production
    - Pileup production & data management
    - Reconstruction
- Test grid middleware, test grid performance
- Modules can be easily enhanced or replaced by Condor-G, EDG resource broker, Chimera, replica catalogue, OGSA… (in progress)

Middleware Evolution of U.S. Applications (K.De)



Globus

Condor-G

DAGMan

Chimera

LCG?

Used in current production software (GRAT & Grappa)

Tested successfully (not yet used for large scale production)

Under development and testing

Tested for simulation (may be used for large scale reconstruction)

Conclusion ATLAS US Grid(K.De)

- Large scale (>10k Cpu days, >10TB) grid based production was done by U.S. testbed
- Grid production is possible, but not easy right now - need to harden middleware, need higher level services
- Many tools are missing - monitoring, operations center, data management
- Requires iterative learning process, with rapid evolution of software design
- Pile-up was a major data management challenge on the grid - moved >0.5 TB/day
- Successful so far - but slower than plan
- Continuously learning and improving
- New Chimera based product being tested
- Many more challenges coming up!

ATLAS EDG

- ATLAS was the first experiment to test EDG in production mode (back to July 2002)
  - Almost 1000 simulation jobs (20-30 hours each) submitted over 8 months with evolving EDG releases: very valuable feedback provided
  - The last systematic test (130 jobs in 2 weeks end February):
    - Only < 5% problems traceable to EDG m/w
    - Still a lot of instability, most “local problems” (disk full, machine down, failed file transfer): **week 1 80% success, week 2 < 25% !!!**
  - Work started for partial production of ATLAS reconstruction (ATHENA) with EDG

ATLAS reconstruction on GRID

Why

- Check stability of grid for a real production with ATHENA (reconstruction phase of ATLAS DC1)

What has been done

- Test (few jobs, 5-6) at RAL, Lyon, CNAF. Only few technical (but time consuming) problems (WNs disks full…)

To be done: Real production

- install RH 7.3 and ATLAS 6.0.3 on the WNs (currently creating and testing LCFGng profiles, installation already done at Lyon where LCFG is not used)
- copy and register input files (from CERN & RAL)
- submit the jobs

ATLAS reconstruction on GRID

Involved sites:

Milan, Rome, Cambridge, CNAF, RAL, Lyon

Selected input data

sample of 20k QCD di-jets at different energies simulated at RAL and CERN (not high priority) 500 GB

Time expected to complete all the jobs ~5-6 days with 15-20 nodes

Activity on Grid tools

- Much work done:
  - MAGDA (US), AMI (Grenoble) used already on the current productions (independent from Grids): ATLAS intend to evolve them as thin layers for interface to LCG (but not exclusively)
  - Other tools in different stages of development and test, not all aimed at general Atlas use
    - GANGA (ATLAS-LHCb UK main effort,) is seen as a promising framework
    - Chimera (US) is aimed to exploit Virtual Data ideas
  - A coherent view of tool use and integration between themselves, with the Grid and with ATHENA is starting to emerge, but will need more work and thinking.

GANGA (K.Harrison)

- The Indian goddess Ganga descended to Earth to flow as a river (English: Ganges) that carried lost souls to salvation
- Ganga software is being developed jointly by ATLAS and LHCb to provide an interface for running Gaudi/Athena applications on the Grid
  - ⇒ Deal with all phases of a job life cycle: configuration, submission monitoring, error recovery, output collection, bookkeeping
  - ⇒ Carry jobs to the Grid underworld, and hopefully bring them back
- Idea is that Ganga will have functionality analogous to a mail system, with jobs having a role similar to mails
  - ⇒ Make configuring a Gaudi/Athena job and running it on the Grid as easy as sending a mail

Design considerations (K.Harrison)

- Ganga should not reproduce what already exists, but should make use of, and complement, work from other projects, including AtCom, AthASK, DIAL and Grappa in ATLAS

⇒ Should also follow, and contribute to, developments in Physicist Interface (PI) project of LCG

- The design should be modular, and the different modules should be accessed via a thin interface layer implemented using a scripting language, with Python the current choice
- Ganga should provide a set of tools that can be accessed from the command line (may be used in scripts), together with a local GUI and/or a web-based GUI that simplifies the use of these tools
- Ganga should allow access to local resources as well as to the Grid

Tentative Ganga architecture (K.Harrison)

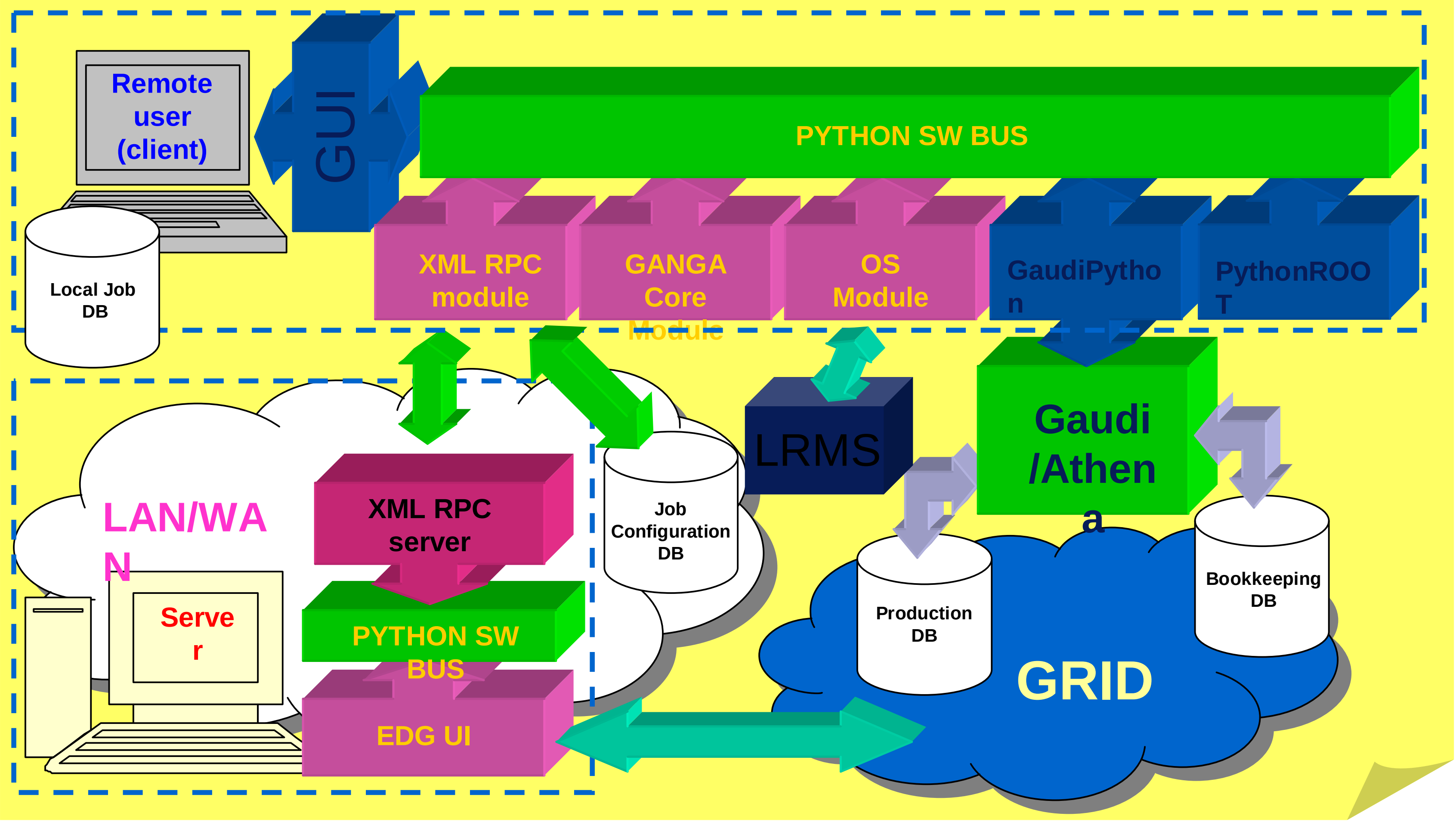
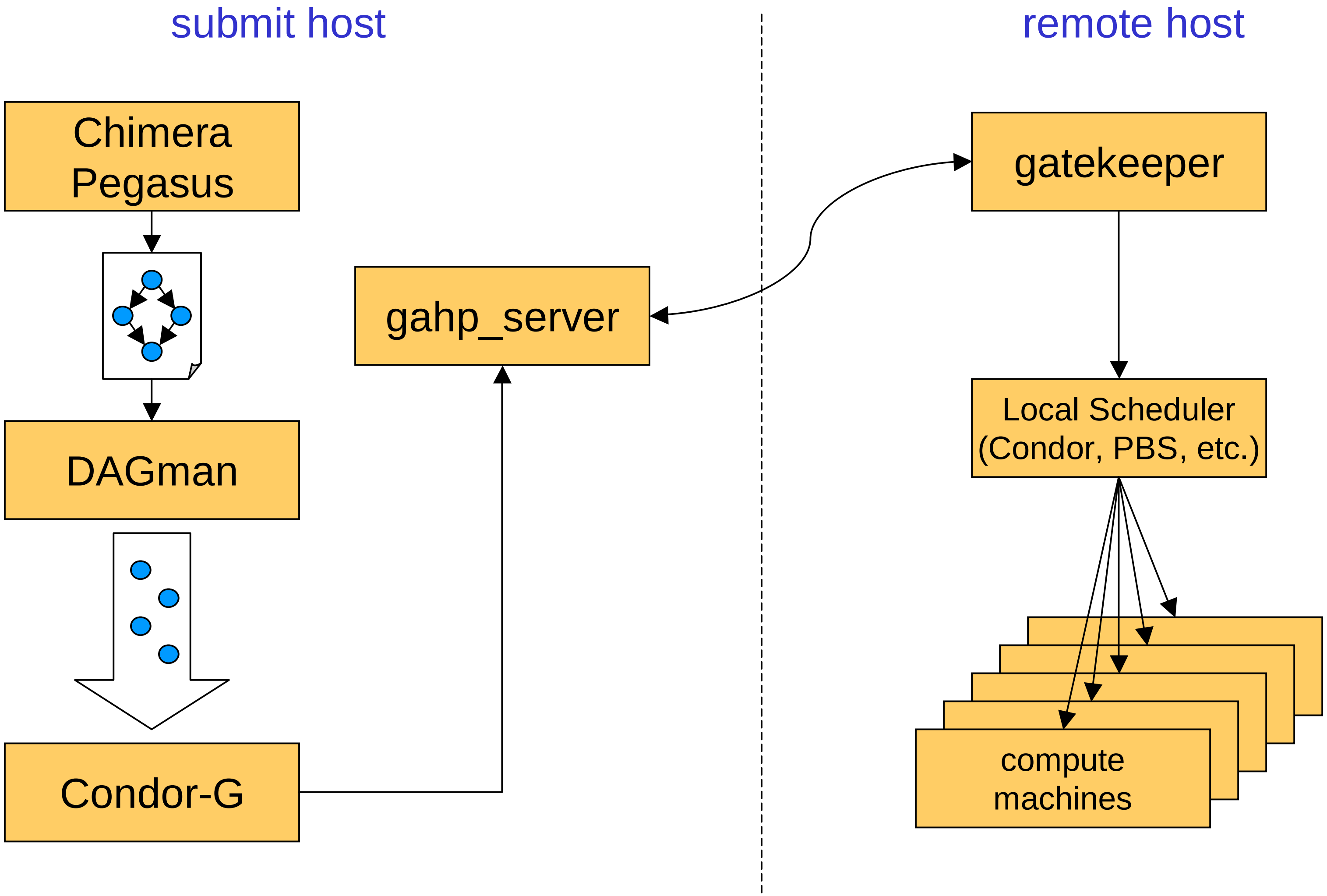


figure courtesy Adam Arbree, USCMS

Basic Chimera System

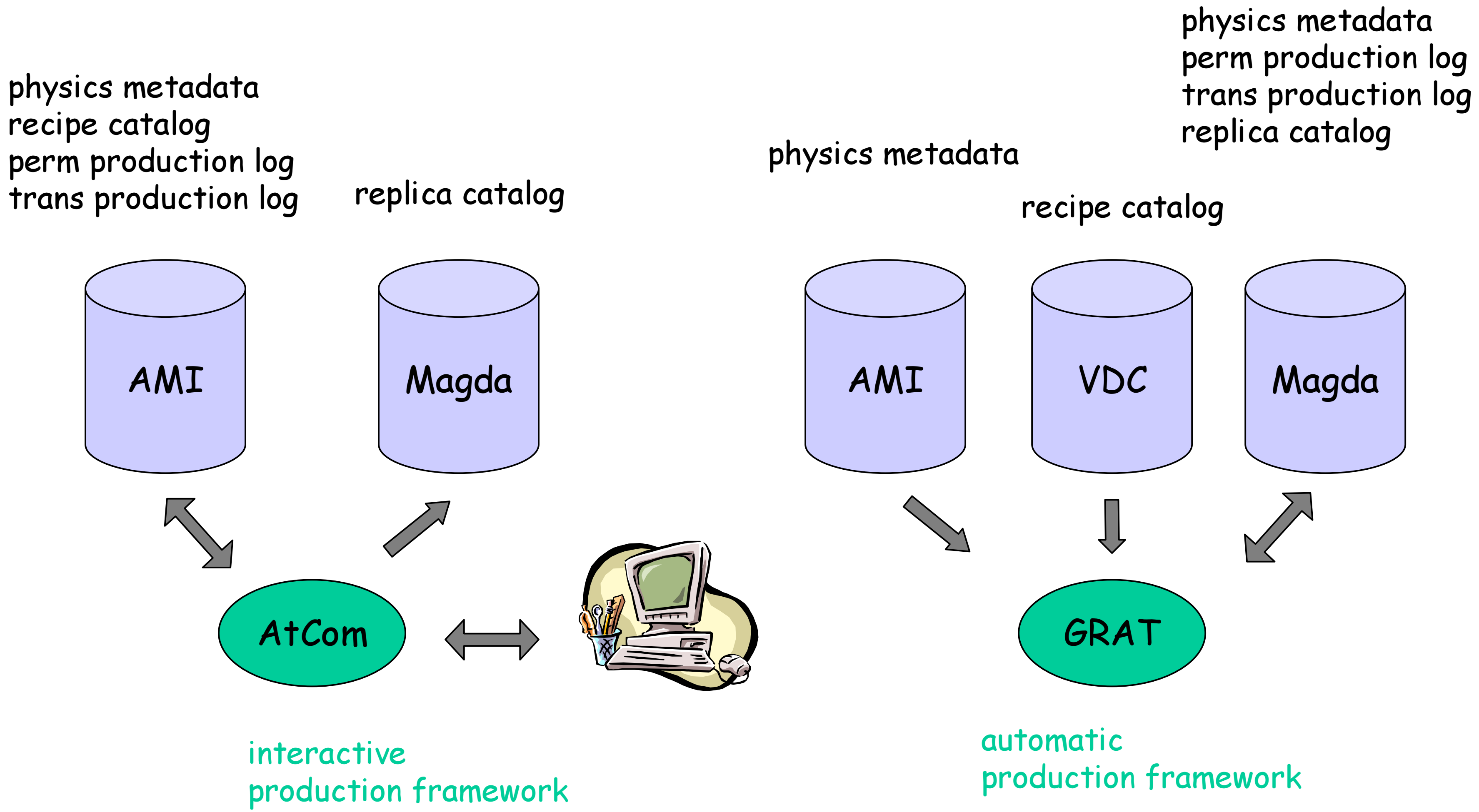


Typical CHIMERA functionality (R.Gardner)

- Condor DAGs are created which handle
  - **Data movement**: all steps needed to move files from the storage elements where they currently exist to the storage element where they are accessible to the execution nodes of the execution site
  - **Execution:** execute all derivations in the DAG
  - **Cataloging:** register all output data products in a replica catalog

Outline of CHIMERA Steps (R.Gardner)

- Define transformations and derivations
  - user scripts write VDLt
- Convert to XML description
- Update a VDC
- Request a particular derivation from the VDC
- Generate abstract job description, DAX
- Generate concrete job description, DAG
- Submit to DAGMan



physics metadata
recipe catalog
perm production log
trans production log

replica catalog

AMI

Magda

AtCom

interactive
production framework

physics metadata

recipe catalog

physics metadata
perm production log
trans production log
replica catalog

AMI

VDC

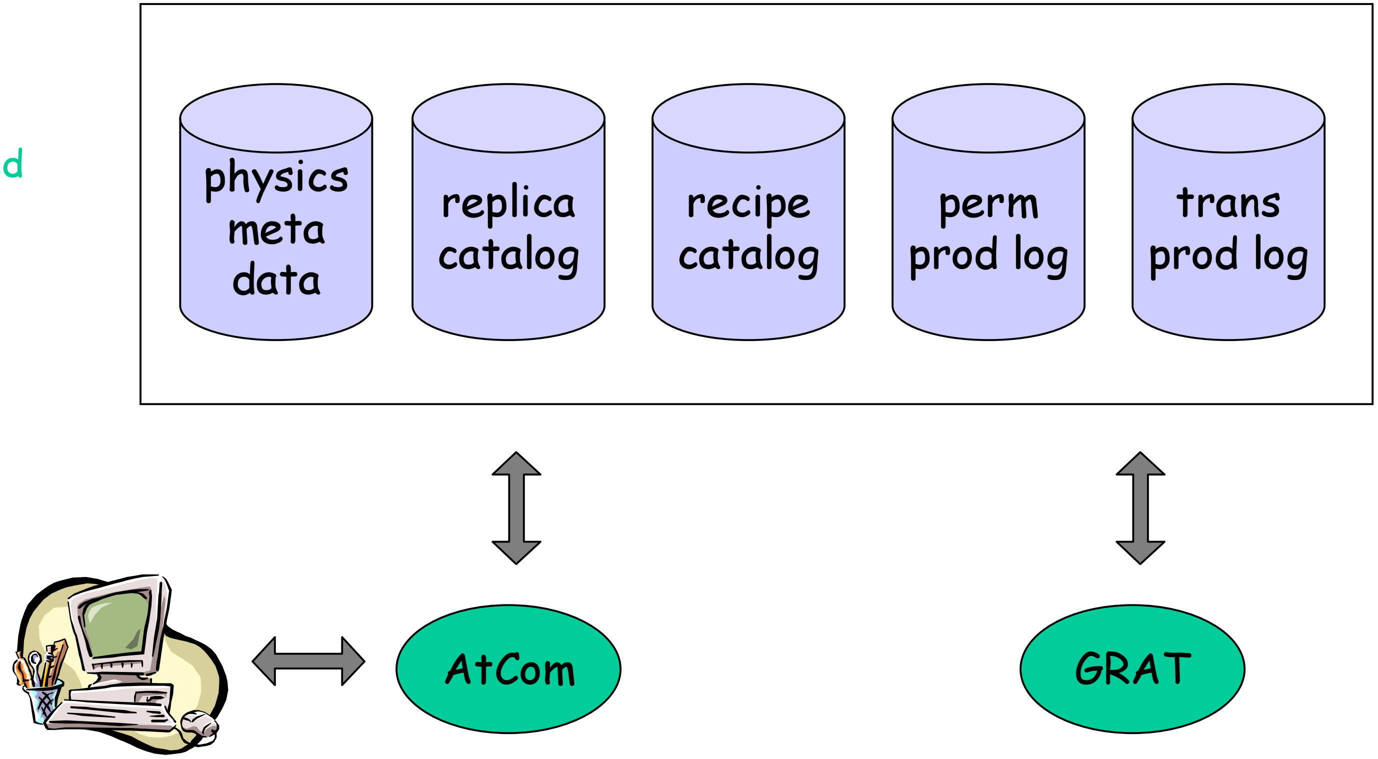
Magda

GRAT

automatic
production framework

a proposal ⇨ AMI | Magda | VDC | AMI | AMI, Magda

integrated database



physics meta data

replica catalog

recipe catalog

perm prod log

trans prod log

AtCom

GRAT

interactive
production framework

automatic
production framework

Toward an ATLAS Grid production and analysis system

- ATLAS has encouraged the development of diverse tools
  - often born with interface to one specific Grid flavor
- ATLAS has kept the general production system as simple as possible
  - Avoid building complex interfaces to a diverse and rapidly evolving m/w
    - **Provisional solutions in HEP risk to eternize themselves….**
  - Avoid ATHENA dependences from specific m/w
  - Foster m/w convergencies and common interfaces
- LCG has now to grant the framework for finally planning an ATLAS production and analysis system:
  - We expect to start with the m/w services decided in WG1-GDB (which EDG V2 is designed to implement)
    - well defined interfaces and agreed planes of evolution (EGEE …..)
  - **Fall-back interim solutions with severely descoped Grid functionality risk to be of limited interest for us**
  - **All the needed effort & support has to go in EDG V2**