Fabric Infrastructure

LCG Review

November 18th 2003

Tony.Cass@CERN.ch

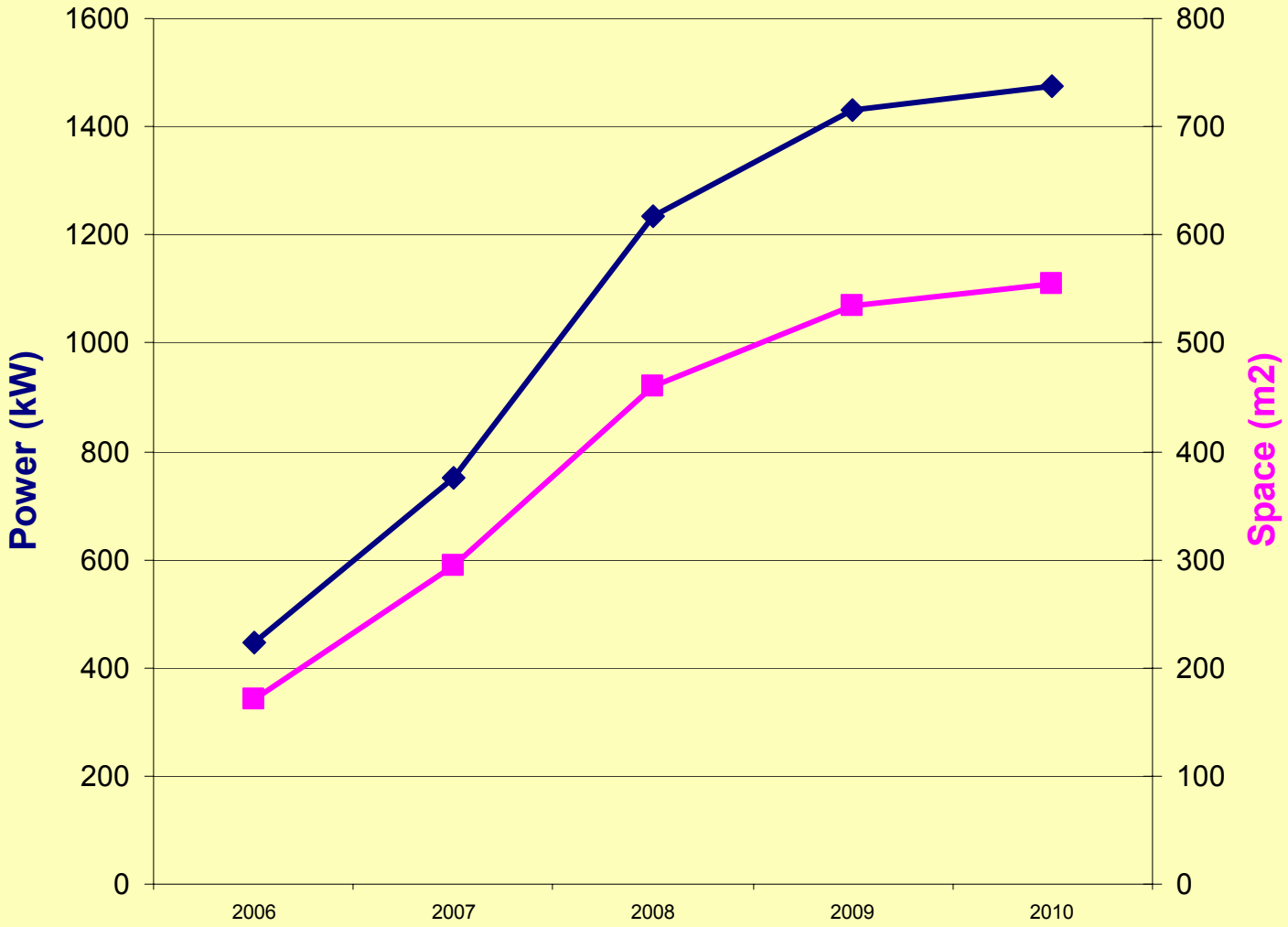
Agenda

- Space, Power & Cooling
- CPU & Disk Server Purchase
- Fabric Automation

Agenda

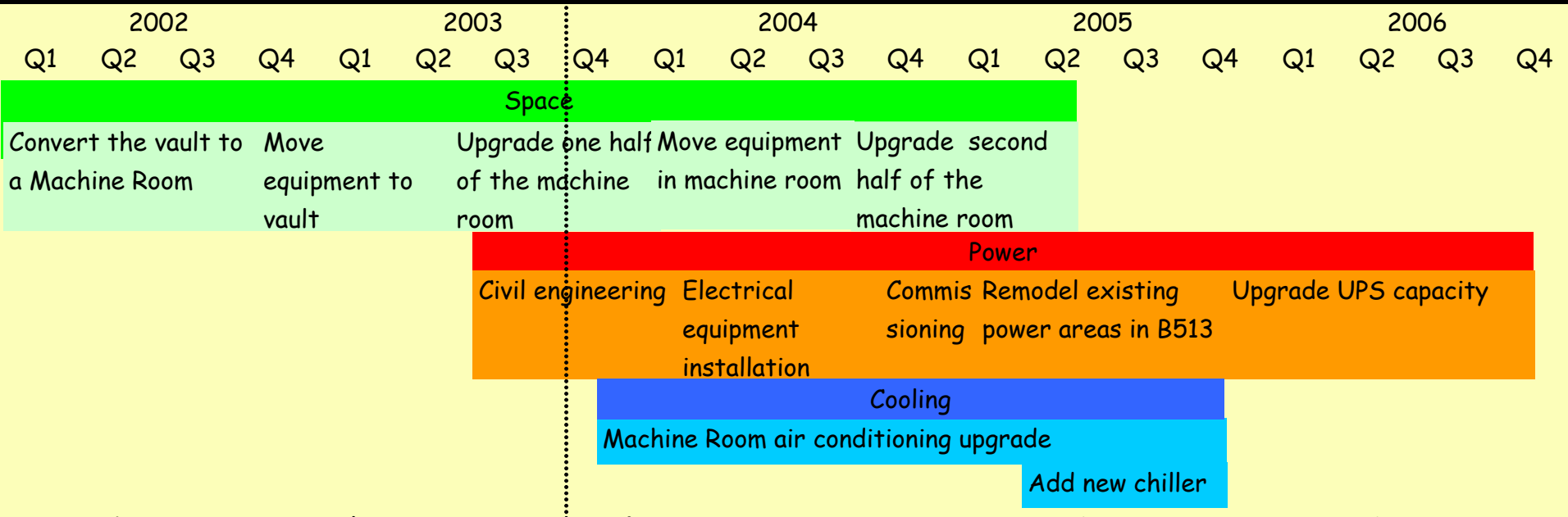
- **Space, Power & Cooling**
- CPU & Disk Server Purchase
- Fabric Automation

Space & Power Requirements



- CPU & Disk Servers only. Maximum power available in 2003: 500kW.

Upgrade Timeline



- The power/space problem was recognised in 1999 and an upgrade plan developed after studies in 2000/1.
- Cost: 9.3MCHF, of which 4.3MCHF is for the new substation.
  - Vault upgrade was on budget. Substation civil engineering is overbudget (200KCHF), but there are potential savings in the electrical distribution.
  - Still some uncertainty on overall costs for air-conditioning upgrade.

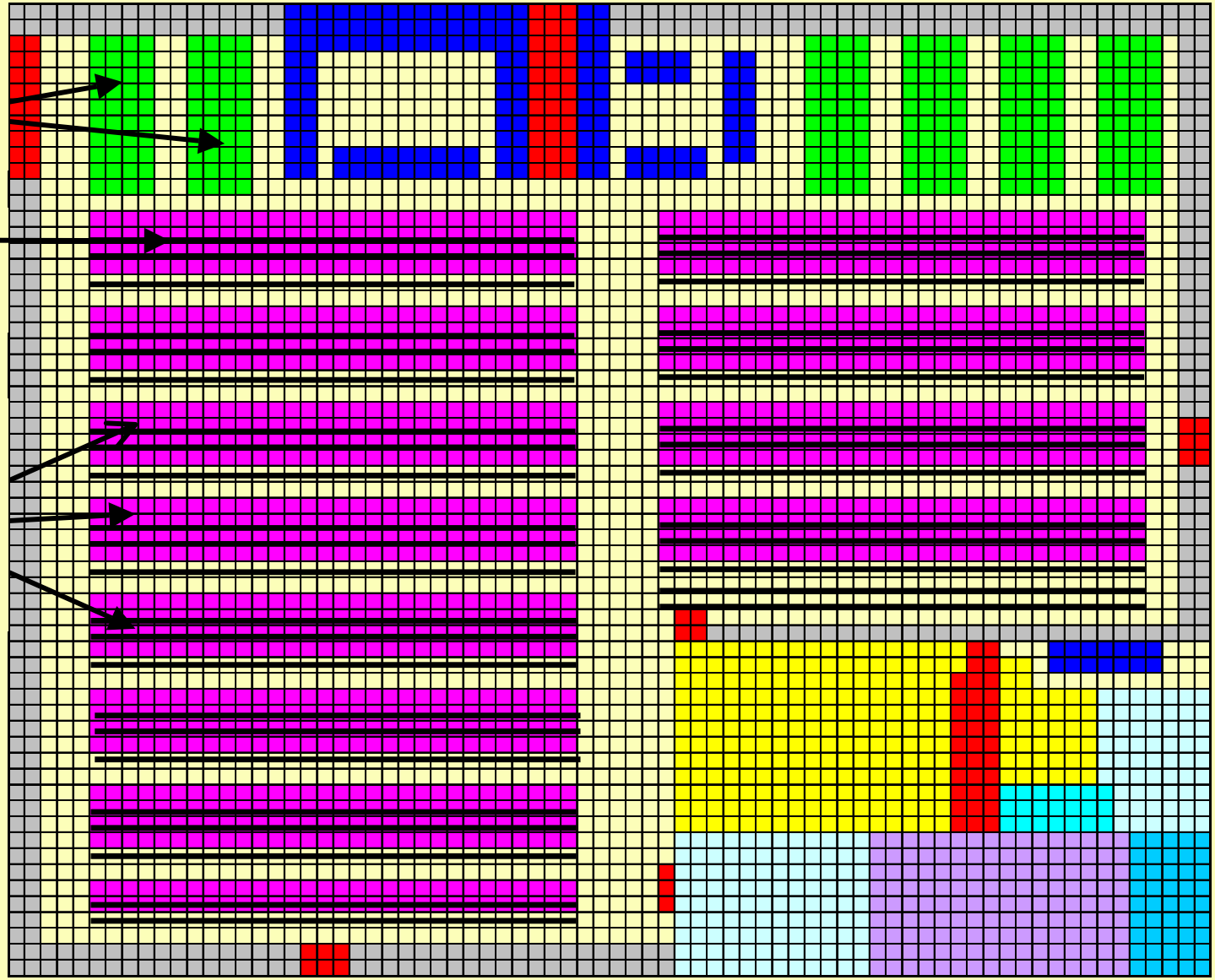
Future Machine Room Layout

9m double rows of racks for critical servers

Aligned normabarres

18m double rows of racks
12 shelf units
or 36 19" racks

528 box PCs
105kW
1440 1U PCs
288kW
324 disk servers
120kW(?)



Upgrade Milestones

Vault converted to machine room	01/11/02	✓ 01/11/02
Right half of machine room emptied to the vault	01/08/03	✓ 15/08/03
Substation civil engineering starts	01/09/03	✓ 15/08/03
Substation civil engineering finishes	01/03/04	
Substation electrical installation starts		
Substation commissioned	07/01/05	
Elec. distrib. on RH of machine room upgraded	02/02/04	
Left half of machine room emptied		
Elec. distrib. on LH of machine room u	01/06/05	
Machine room HVAC upgraded	01/03/05	
New 800kW UPS for physics installed		
Current UPS area rem		
2nd 800kW UPS added		
3rd 800kW UPS added	01/04/08	

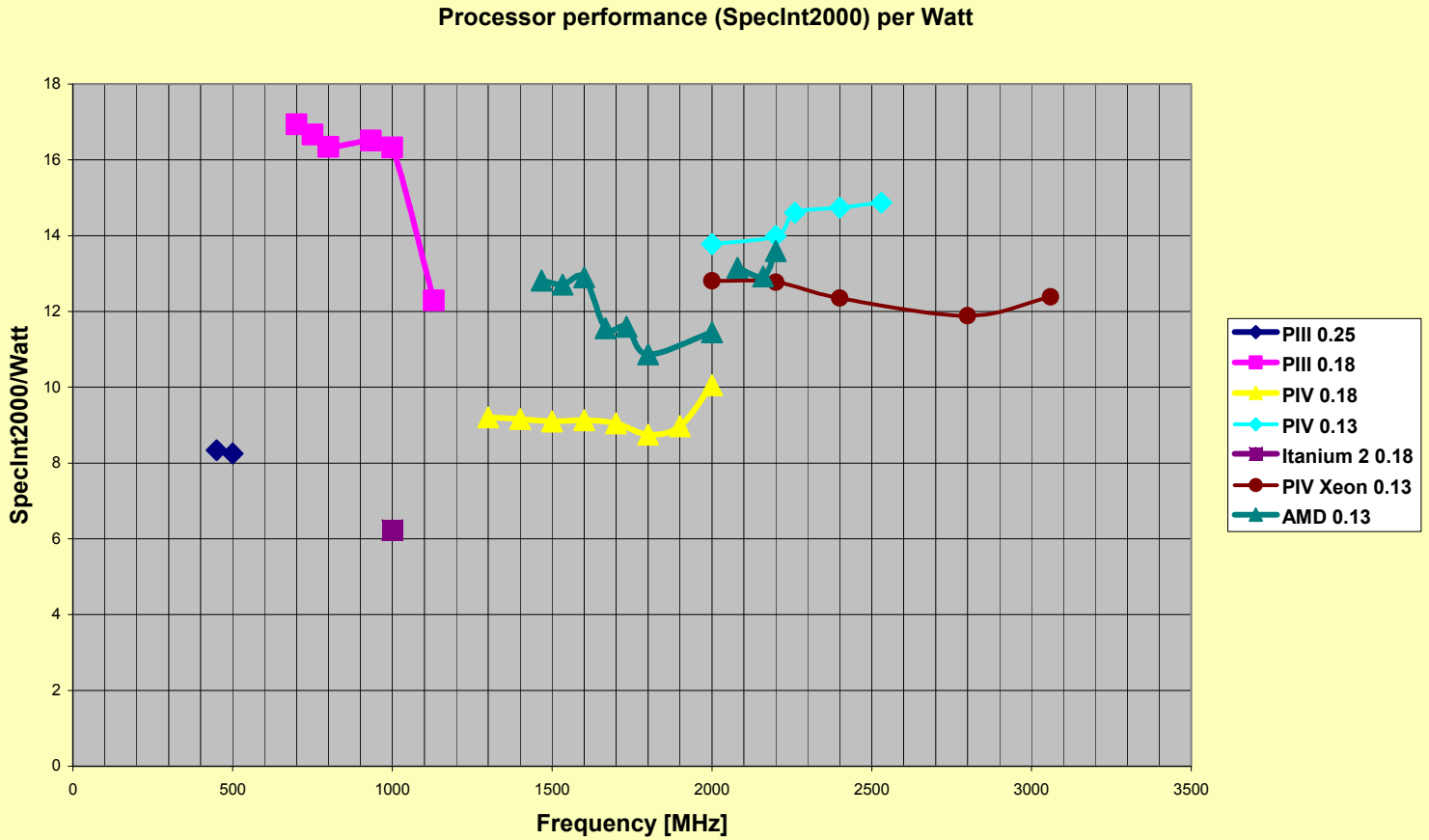
On Schedule

Progress acceptable

Capacity will be installed to meet power needs.

… and what are those needs?

- Box power has increased from ~100W in 1999 to ~200W today.
  - And, despite promises from vendors, electrical power demand seems to be directly related to Spec power:



Space and Power Summary

- Building infrastructure will be ready to support installation of production offline computing equipment from January 2006.
- The planned 2.5MW capacity will be OK for 1st year at full luminosity, but there is concern that this will not be adequate in the longer term.
- Our worst case scenario is a load of 4MW in 2010.
  - Studies show this can be met in B513, but more likely solution is to use space elsewhere on the CERN site.
  - Provision of extra power would be a 3 year project. We have time, therefore, but still need to keep a close eye on the evolution of power demand.

Agenda

- Space, Power & Cooling
- **CPU & Disk Server Purchase**
- Fabric Automation

The challenge

- Even with the delayed start up, large numbers of CPU & disk servers will be needed during 2006-8:
  - At least 2,600 CPU servers
    - » 1,200 in peak year; c.f. purchases of 400/batch today
  - At least 1,400 disk servers
    - » 550 in peak year; c.f. purchases of 70/batch today
  - Total budget: 24MCHF
- Build on our experiences to select hardware with minimal total cost of ownership.
  - Balance purchase cost against long term staff support costs, especially for
    - » System management (see next section of this talk), and
    - » Hardware maintenance.

Acquisition Milestones — I

- Agree CPU and Disk service architecture by 1st June 2004 (WBS 1.2.1.3.3 & 1.2.4.1).
- i.e. make the decisions on which hardware minimises the TCO:
  - CPU: White box vs 1U vs blades; install or ready packaged
  - Disk: IDE vs SAN; level of vendor integration
- and also the operation model
  - Can we run IDE disk servers at full nominal capacity?
    - » Impact on disk space available or increased cost/power/…
- Total Cost of Ownership workshop organised by the openlab, 11/12th November.

Acquisition Milestones — II

- Agreement with SPL on acquisition strategy by December (WBS 1.2.6.2).
  - Essential to have early involvement of SPL division given questions about purchase policy—likely need to select multiple vendors to ensure continuity of supply.
  - A little late, mostly due to changes in CERN structure.
- Issue Market Survey by 1st July 2004 (1.2.6.3)
  - Based on our view of the hardware required, identify potential suppliers.
  - Input from SPL important in preparation of the Market Survey to ensure adequate qualification criteria for the suppliers.
    - » Overall process will include visits to potential suppliers.
- Finance Committee Adjudication in September 2005 (1.2.6.6)

Acquisition Summary

- Preparation for Phase II CPU & Disk server acquisition is a complex process with a tight schedule.
- Significant effort will be required to ensure that equipment meeting our specifications is installed on the required timescale.
- The major risk is that equipment does not meet our specifications.
  - H/W quality is problem today, especially for disk servers.
  - Careful definition of the qualification criteria for suppliers is essential.

Agenda

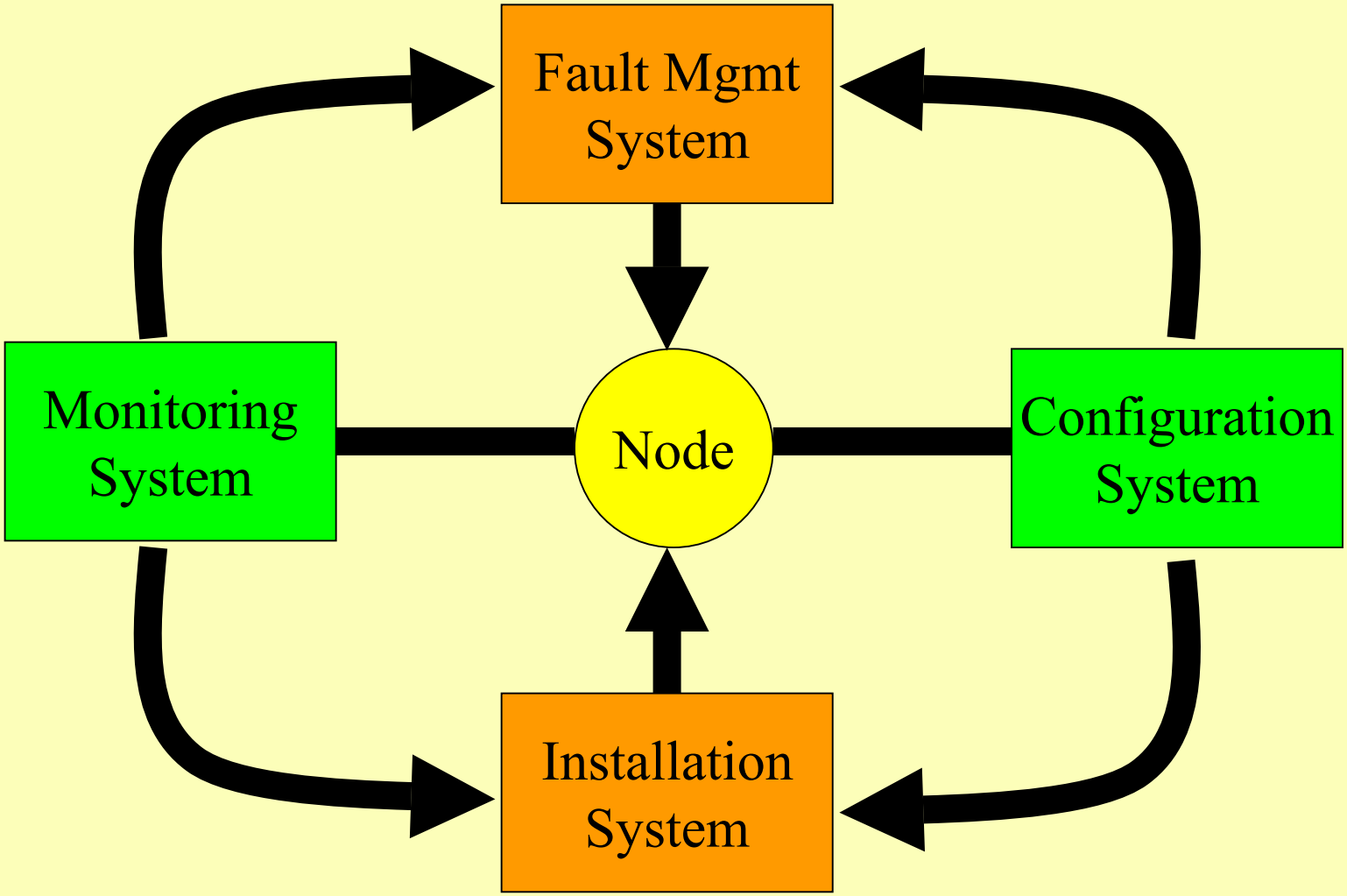
- Space, Power & Cooling
- CPU & Disk Server Purchase
- **Fabric Automation**

ELFms

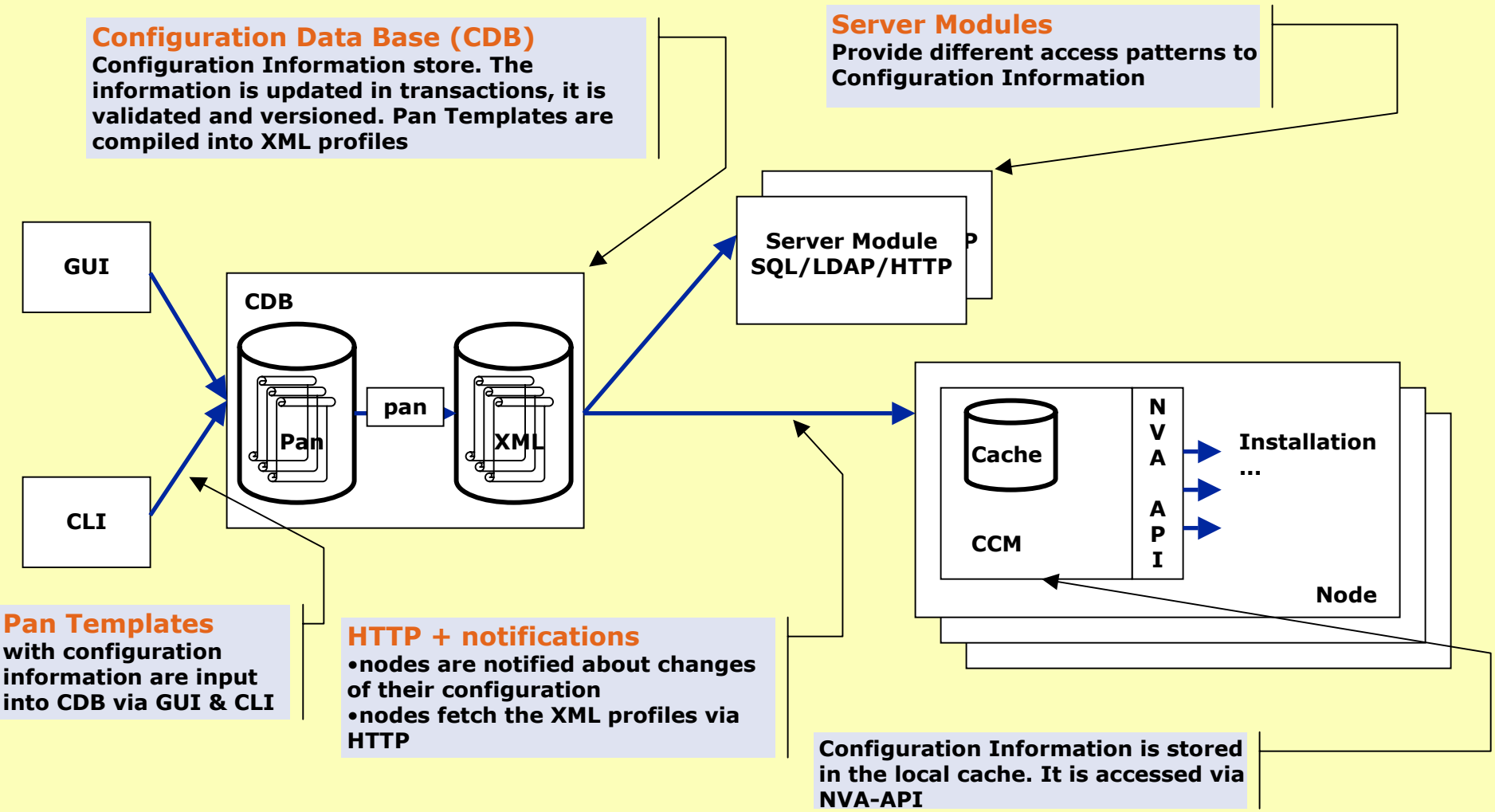
- The ELFms Large Fabric management system has been developed over the past few years to enable tight and precise control over all aspects of the local computing fabric.

- ELFms comprises
  - The EDG/WP4 quattor installation & configuration tools
  - The EDG/WP4 monitoring system, Lemon, and
  - LEAF, the LHC Era Automated Fabric system

ELFms architecture

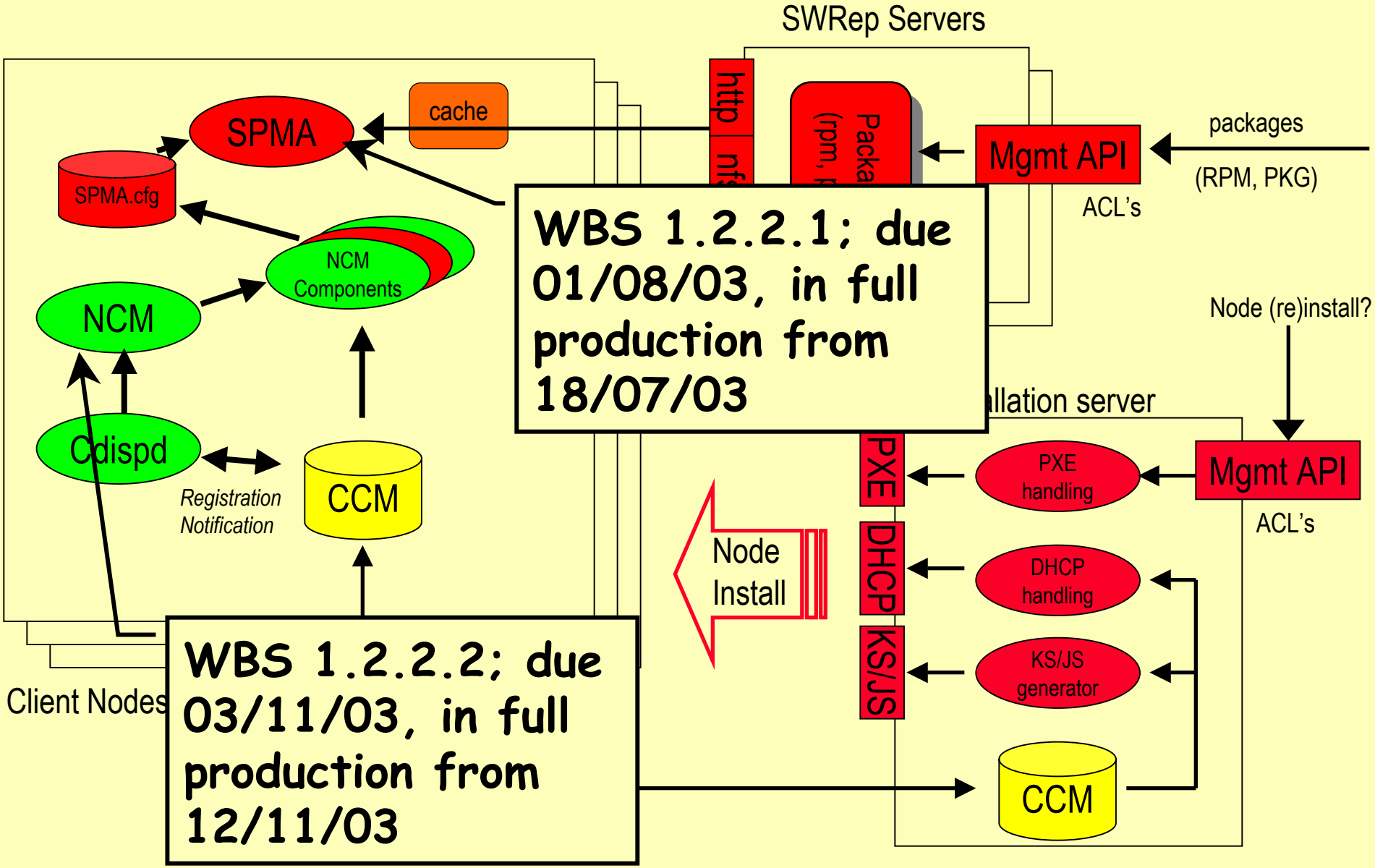


quattor configuration architecture



- All in production in spring for RH7 migration. Thus no milestones.

quattor installation architecture



quattor status

- quattor is in complete control of our farms.
- We are already seeing the benefits in terms of
  - ease of installation—10 minutes for LSF upgrade,
  - speed of reaction—ssh security patch installed across all lxplus & lxbatch nodes within 1 hour of availability, and
  - homogeneous software state across the farms.
- quattor development is not complete, but future developments are desirable features, not critical issues.
  - Growing interest from elsewhere—good push to improve documentation and packaging!
  - Ported to Solaris by IT/PS
- EDG/WP4 has delivered as required.

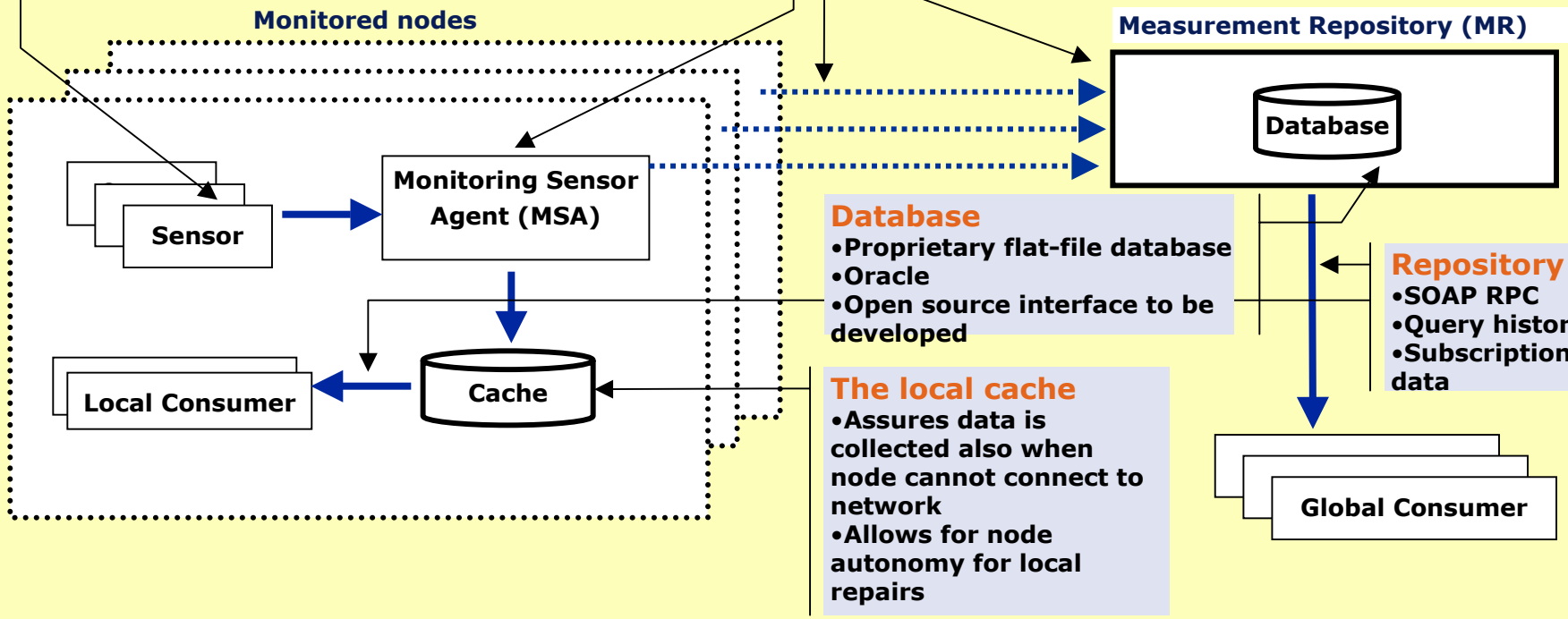
Lemon Overview

Monitoring Sensor Agent
- Calls plug-in sensors to sample configured metrics
- Stores all collected data in a local disk buffer
- Sends the collected data to the global repository

Transport
- Transport is pluggable.
- Two protocols over UDP and TCP are currently supported where only the latter can guarantee the delivery

Plug-in sensors
- Programs/scripts that implements a simple sensor-agent ASCII text protocol
- A C++ interface class is provided on top of the text protocol to facilitate implementation of new sensors

Measurement Repository
- The data is stored in a database
- A memory cache guarantees fast access to most recent data, which is normally what is used for fault tolerance correlations



Database
- Proprietary flat-file database
- Oracle
- Open source interface to be developed

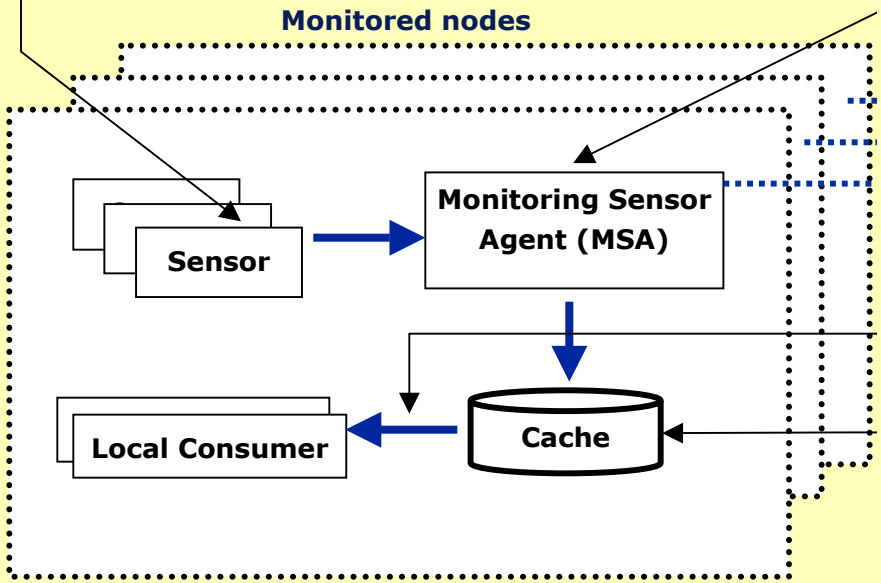
Repository API
- SOAP RPC
- Query history data
- Subscription to new data

The local cache
- Assures data is collected also when node cannot connect to network
- Allows for node autonomy for local repairs

Lemon Overview

Monitoring Sensor Agent
- Calls plug-in sensors to sample configured metrics
- Stores all collected data in a local disk buffer
- Sends the collected data to the global repository

Plug-in sensors
- Programs/scripts that implements a simple sensor-agent ASCII text protocol
- A C++ interface class is provided on top of the text protocol to facilitate implementation of new sensors



MSA in production for over 15 months, together with sensors for performance and exception metrics for basic OS and specific batch server items.

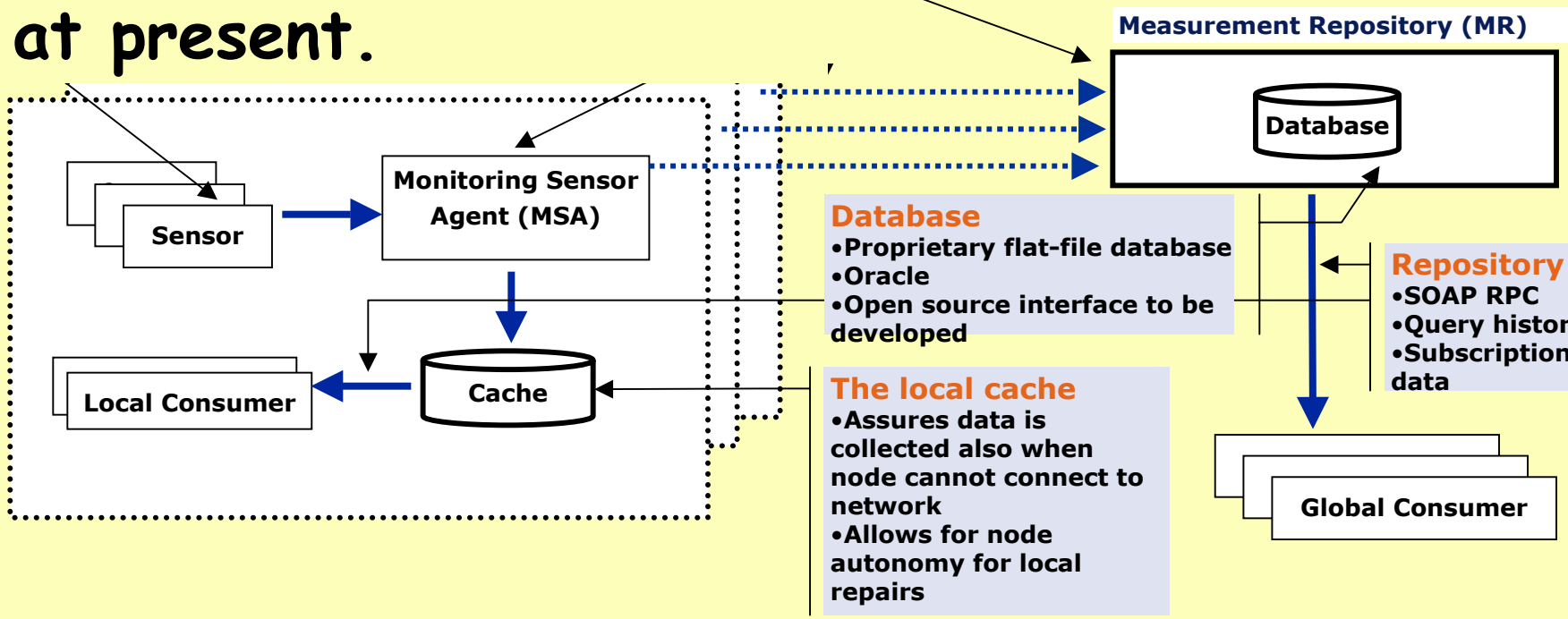
Focus now is on integrating existing monitoring for other systems, especially disk and tape servers, into the Lemon framework.

Lemon Overview

UDP transport in production. Up to 500metrics/s delivered with 0% loss; 185metrics/s recorded at present.

Transport
- Transport is pluggable.
- Two protocols over UDP and TCP are currently supported where only the latter can guarantee the delivery

Measurement Repository
- The data is stored in a database
- A memory cache guarantees fast access to most recent data, which is normally what is used for fault tolerance correlations



Database
- Proprietary flat-file database
- Oracle
- Open source interface to be developed

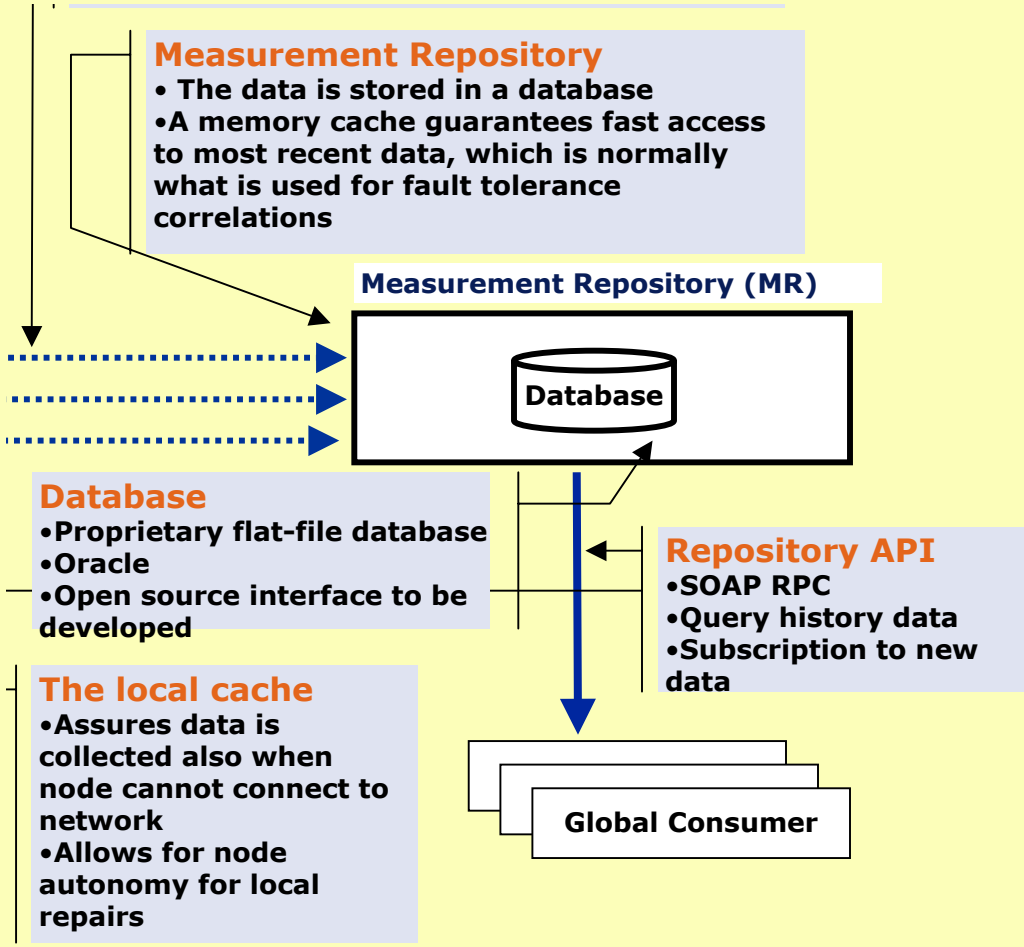
Repository API
- SOAP RPC
- Query history data
- Subscription to new data

The local cache
- Assures data is collected also when node cannot connect to network
- Allows for node autonomy for local repairs

Lemon Overview

RDBMS repository required to enable detailed analysis of performance issue over long periods.

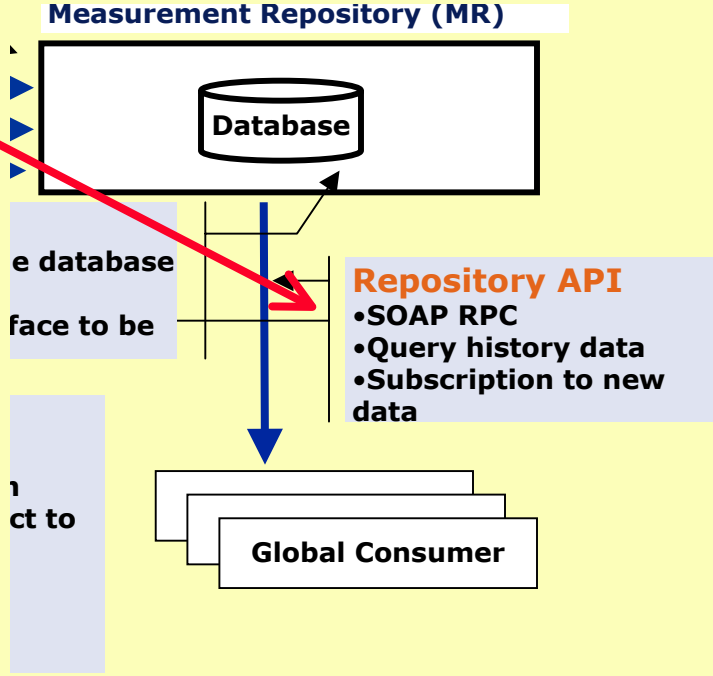
Extensive comparison of EDG/WP4 OraMon and PVSS with Oracle back end in 1H03. OraMon selected for production use in June and has been running in production since 1st September. Records 185 metrics/s (16M metrics per day).



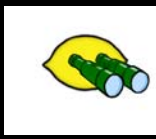
Lemon Overview

Lemon weakness is in the display area. We have no "system administrator" displays and operators still use the SURE system. Milestone 1.2.2.5 is late (systems monitor displays due 1st October), but this has no adverse impact on services.

The required repository API is now available (C, Perl, TCl & PHP), but we have chosen to concentrate effort on derived metrics which deliver service improvements (and relate to milestone 1.2.2.6, due 01/03/04).



Lemon Summary



- The sensors, MSA and OraMon repository are running in production, and running well.
  - Status of, e.g. lxbatch, nodes has much improved since initial sensor/MSA deployment.
- We are now working on *derived metrics*, created by a sensor which acts upon values read from the repository and, perhaps, elsewhere.
  - e.g. Maximum [system load|/tmp usage|/pool usage] on batch nodes running jobs for [alice|atlas|cms|lhcb].
  - Aim is to reduce independent monitoring of (and thus load on) our nodes & LSF system. Needs input from users.
- Lack of displays a concern, but just a mild concern. To be addressed in 2004 after EDG wrap-up.
- Again, much appreciation of solid work within EDG/WP4.

LEAF Components

- HMS Hardware Management System
- SMS State Management System
- Fault Tolerance

HMS

- HMS tracks systems through steps necessary for, e.g., installations & moves.
  - a Remedy workflow interfacing to ITCM, PRMS and CS group as necessary.
  - used to manage the migration of systems to the vault.
  - now driving installation of 250+ systems. So far, no engineer level time has been required. Aim is to keep it this way right through to production!
- As for quattor, no WBS milestones given status in Spring.
  - Important developments now to include more detail on hardware components, especially for disk servers.

SMS – WBS 1.2.2.2 & 1.2.2.3

- "Give me 200 nodes, any 200. Make them like this. By then." For example
  - creation of an initial RH10 cluster
  - (re)allocation of CPU nodes between lxbatch & lxshare or of disk servers.
- Tightly coupled to Lemon to understand current state and CDB—which SMS must update.
- Analysis & Requirements documents have been discussed.
  - Led to architecture prototype and define of interfaces to CDB.
- Work delayed a little by HMS priorities, but we still hope to see initial SMS driven state change by the end of the year (target was October).

Fault Tolerance

- We have started testing the Local Recovery Framework developed by Heidelberg within EDG/WP4.
- Simple recovery action code (e.g. to clean up filesystems *safely*) is available.
- We confidently expect initial deployment at this level by the end of the year (target was March 04).

Fault Tolerance Long Term

- **Real** Fault Tolerance requires actions such as reallocating a disk server to CMS "because one of theirs has failed and they are the priority users at present".
- Such a system could also
  - reallocate a disk server to CMS because this looks a good move given the LSF job mix.
  - reconfigure LSF queue parameters at night or weekend based on the queue.
- Essentially, this is just a combination of derived metrics and SMS.
  - Concentrate on building the basic elements for now and start with simple use cases when we have the tools.

HSM Operators Interface

- This ought to be a GUI on top of an intelligent combination of quattor, Lemon & LEAF.
- We want to be able to (re)allocate disk space (servers) to users/applications as necessary.
- These are CDB and SMS issues.
  - But there are strong requirements on Castor and the stager if we want to do this. Fortunately, many of the necessary stager changes are planned for next year.
- Initial development (milestone 1.2.2.10) has started with extension of CDB to cover definition of, e.g., staging pools.

Fabric Infrastructure Summary

- The Building Fabric will be ready for start of production farm installation in January 2006.
  - But there are concerns about a potential open ended increase of power demand.
- CPU & disk server purchase complex
  - Major risk is poor quality hardware and/or lack of adequate support from vendors.
- Computing Fabric automation is well advanced.
  - Installation and configuration tools are in place.
  - The essentials of the monitoring system, sensors and the central repository, are also in place. Displays will come. More important is to encourage users to query our repository and not each individual node.
  - LEAF is starting to show real benefits in terms of reduced human intervention for hardware moves.