

R&D proposal from the session "Distributed Computing"

Armando Fella, INFN CNAF

Presentation outline

- 1) Session content summaries
- 2) Proposals: Parallel Grid
- 3) Proposals: Grid on demand
- 4) Conclusion

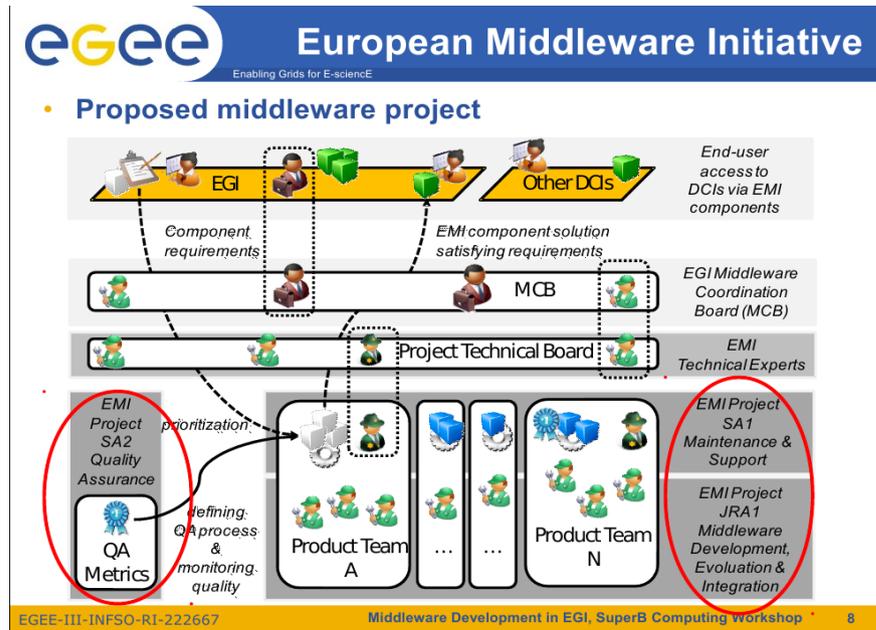
HEP experience

- The SuperB computing model for TDR will include many design elements HEP experiments already considered
- The cooperation with other experiments and/or Grid projects has to be considered as the key for R&D works

- **Join an existing infrastructure rather than building a new one**
 - It may actually not be your choice
- **Use as many services and tools provided by the infrastructure as possible rather than developing your own**
 - If there are missing functionalities, ask
 - If you're able to develop something else, try to do it as extensions to existing tools (and contribute them to the basic stack)
 - Current LHC (and pre-LHC) experiments have been built in parallel with the distributed infrastructures and often this was not possible
- **Concentrate on the high-level VO specific layers**
- **Plan for failures**
 - Modular architecture; avoid single points of failure
 - Plan for responsibility transfers (e.g. a site is lost for some time)
- **Think to end-user analysis since the beginning**
 - If you concentrate on production activities only, it will be much harder

European Grid Infrastructure

- EGI has the role of coordinating grid activities among National Grid Initiatives
- EGI project will last 4 years since May 1st 2010
- Plans: improve stability, reliability, simplification, standardization, virtualization --> cloud



- SuperB will refer to National Grid Initiatives in each country as first speaker (Eg.: IGI will be the Italian Grid Initiative, now INFN Grid)

Multi-thread and MPI enabling in Grid

- Within 2010: new MPI and multi-thread functionality will be deployed to the official Grid workload system and available for production usage
- SuperB is very interested in enabling many core resources exploitation at several levels within parallel paradigm, the Grid one will be fully usable within SuperB life time

MPI and multi-thread support in EGEE : the future

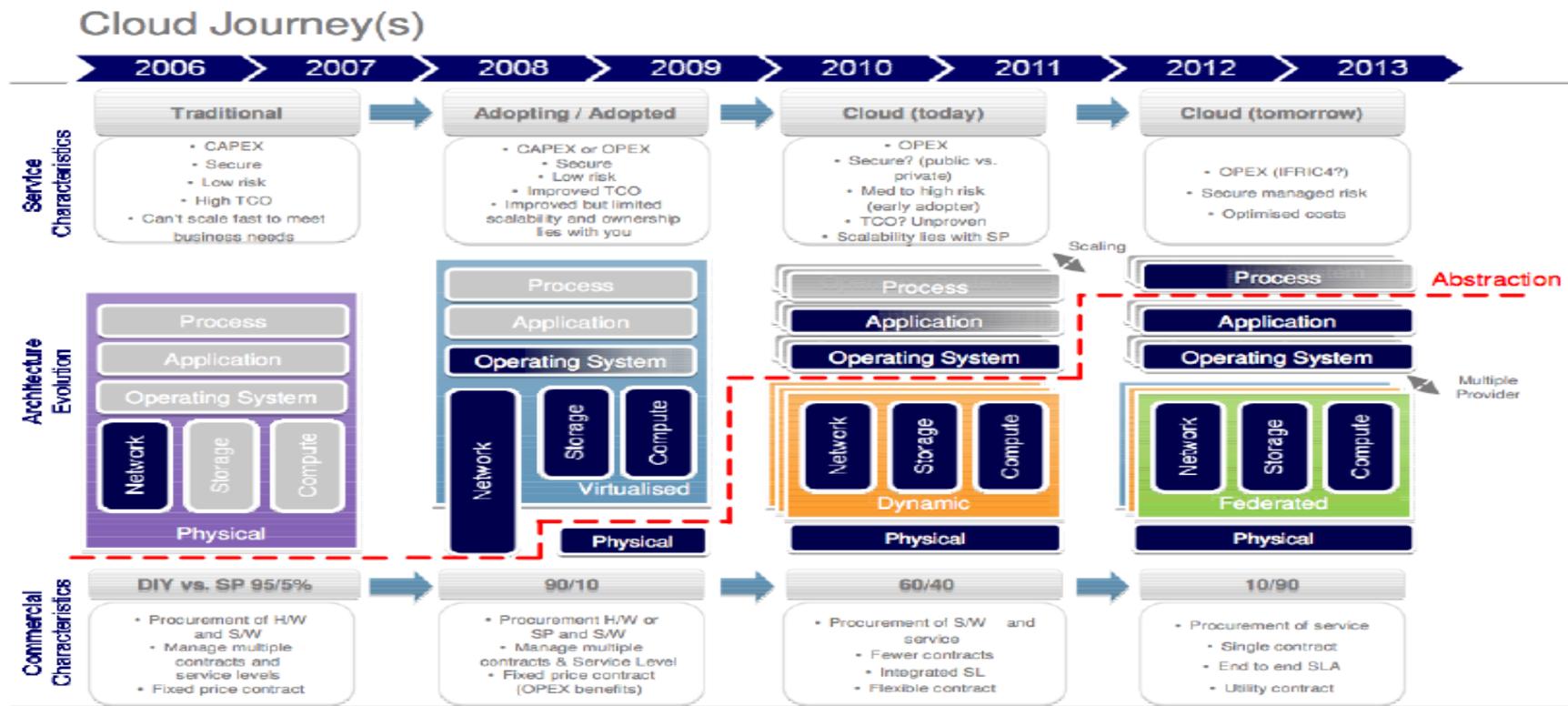
EGEE-III (2008-2010)

- Typical WN: 2-4 CPUs, 4-6 cores per CPU
- **Multi-thread programming** should be supported in EGEE to exploit the upcoming multi-core architectures.
- Survey for users and administrators in April 2009: **MPI is still scarcely used**
- Set-up of a new **MPI WG**. Recommendation document will be released soon.
- **New attributes will be introduced in the JDL** for multi-thread support.
- **Discussion ongoing** concerning semantics and implementation.

Attribute	Meaning
CPUNumber=P	Total number of required CPUs
SMPGranularity=C	Minimum number of cores per node
NodeNumber=N	Total number of required nodes
WholeNode=true	Reserve the whole node (all cores)

Grid on demand

- SuperB will face a distributed infrastructure scenario as a **merging of Cloud and Grid infrastructures**
- The degrees of freedom given by **on demand services escalation** should be included at all levels into R&D programs



Proposal: Grid on demand

- Evaluating the possibility of defining a set of SuperB Virtual machines per functionality (WNs VM, Storage Server VM, online VM, offline VM, ...)
- On demand philosophy, from Virtual Machine to Virtual Service:
 - evaluate the CM design choice of defining SuperB services as included in on-demand virtualized infrastructure
- Investigate/think about sharing of resource vs lease of resource
- Do we still **need or want a multi-tiered computing model**, with – say – $N > 1$ or $N > 2$ in Tier-N? Do we still need per-experiment, per-research, per-subgroup resources? Think also of overhead, critical mass, and expected availability/reliability of smaller sites.
- The computing/business world is undoubtedly showing, like always, a changing (and cloudy) scenario, e.g. for what regards service provider offerings. Can we **start exploring or asking for or testing "novel" requirements now?**

Proposal: parallel Grid

- Evaluating how to adapt the R&D projects regarding parallel paradigm to future distributed infrastructure

In particular:

- integration of **many core cpu** in distributed context (GRID)
- investigate about the Integration of **"different CPU"** (as GPU) and their exploitation in Grid environment

Conclusion

Current Grid Infrastructures (EGI, EMI...) are stable and consolidated so we need to verify if they offer all the services we need

Distributed Computing is non independent from other "computing", so R&D is strictly related with other new technologies R&D.

- evaluate the possibility of use/integrate the new persistency techniques into distributed environment.
- integration of new DB technologies into grid infrastructure
- Grid integration with Gfarm, new Distributed FS? new AAI ?