A dynamic Provisioning-Concept for HPC-Services

- Status and Challenges -

Alfred Geiger, T-Systems Solutions for Research GmbH

· · **T** · · Systems ·

Outline

- Business-Objectives for the HPC-Services of T-Systems

- Service-Concept for the provisioning of HPC-Resources

- Dynamic Provisioning

  - The gap between actual service and customer requests

  - HPC and Clouds: The role of virtualisation

- Summary and Outlook

· · **T** · ·Systems· ·

HPC-Services by T-Systems:
Business-Objectives.

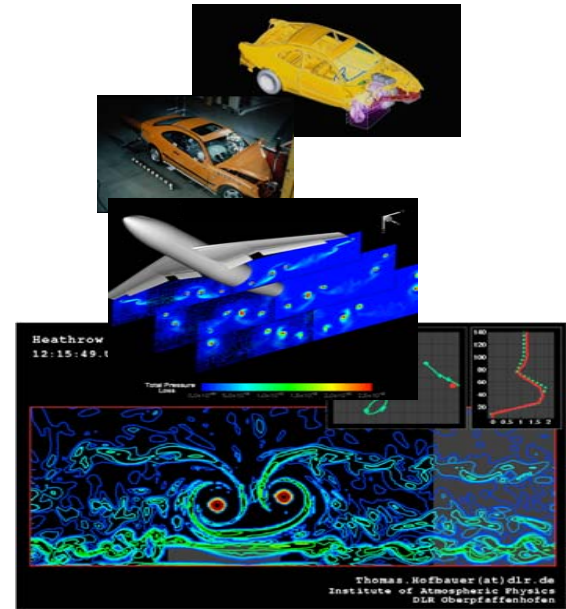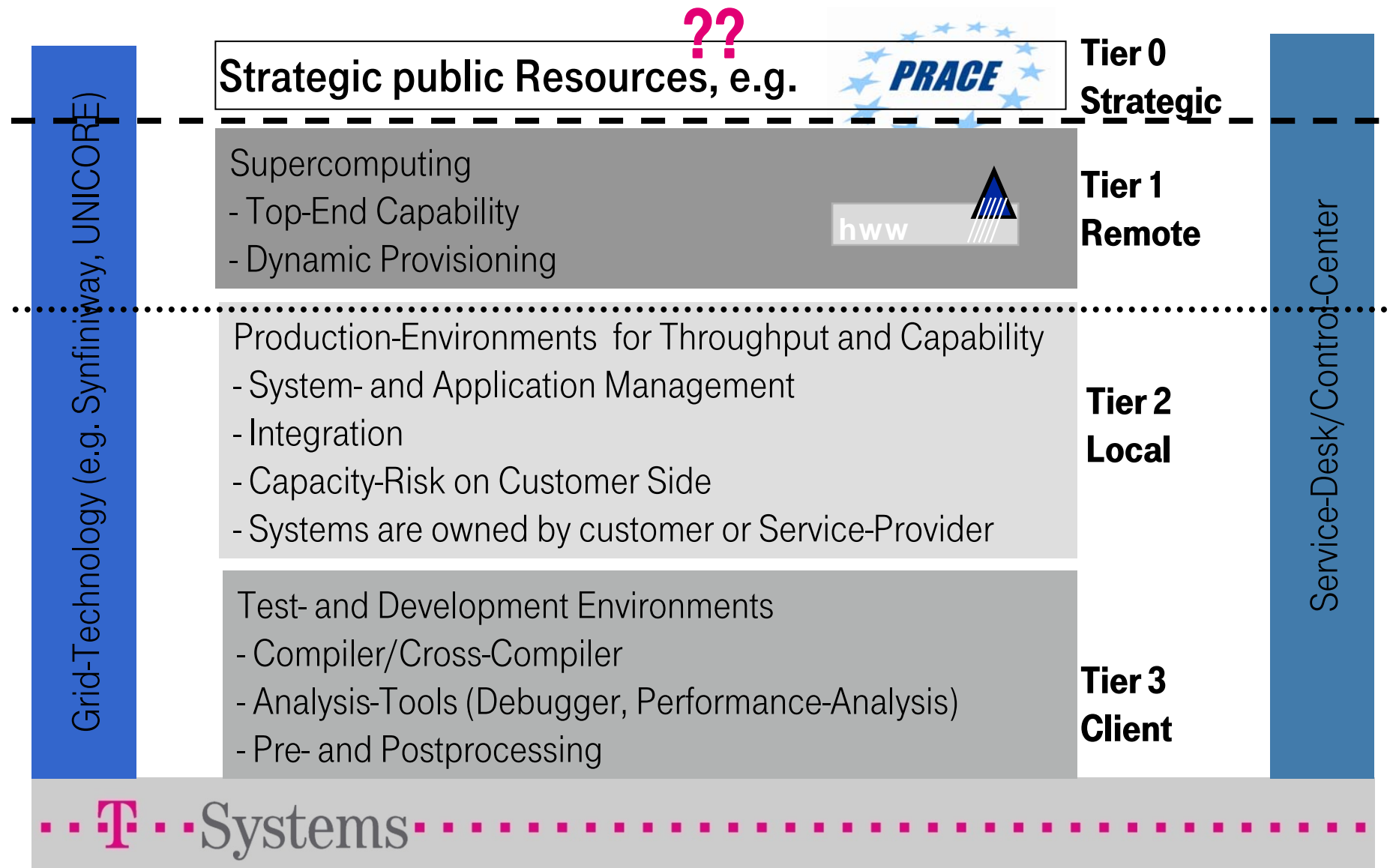- **Current Base**
  - Customers from
    - Automotive
    - Aerospace
    - Nuclear Engineering
    - Climate and Weather
    - Public and Industrial Research
    - Oil and Gas

- **Goals**
  - Giving our Partners and Customers a competitive advantage
  - Provide a solution for a broad Range of Problems
  - Performance, Availability and Reliability
  - Flexibility with Respect to Customer´s Business
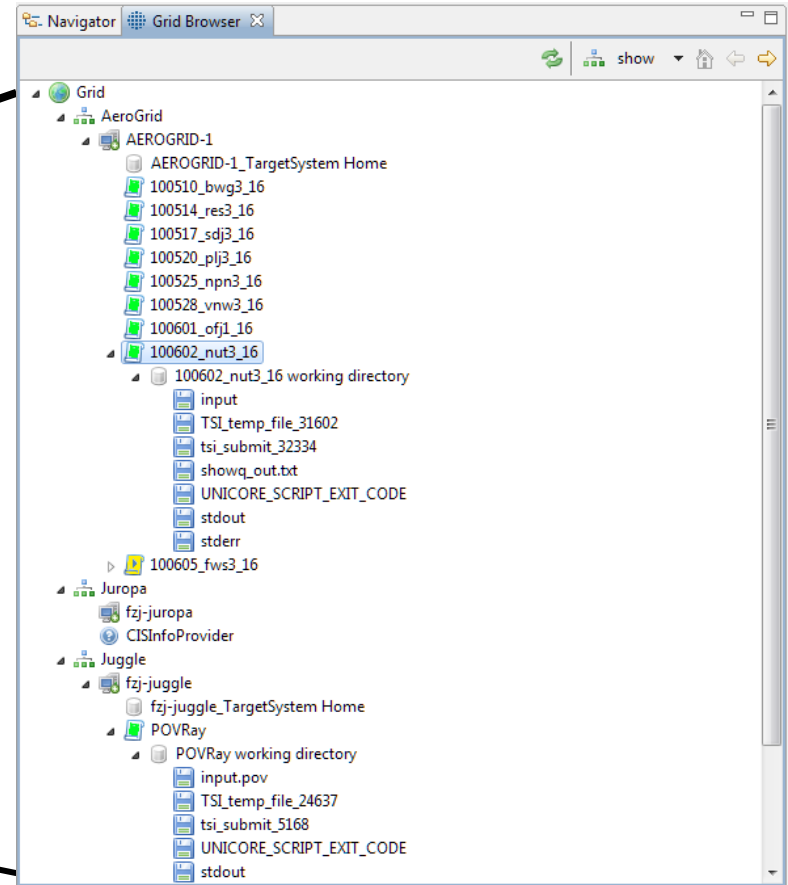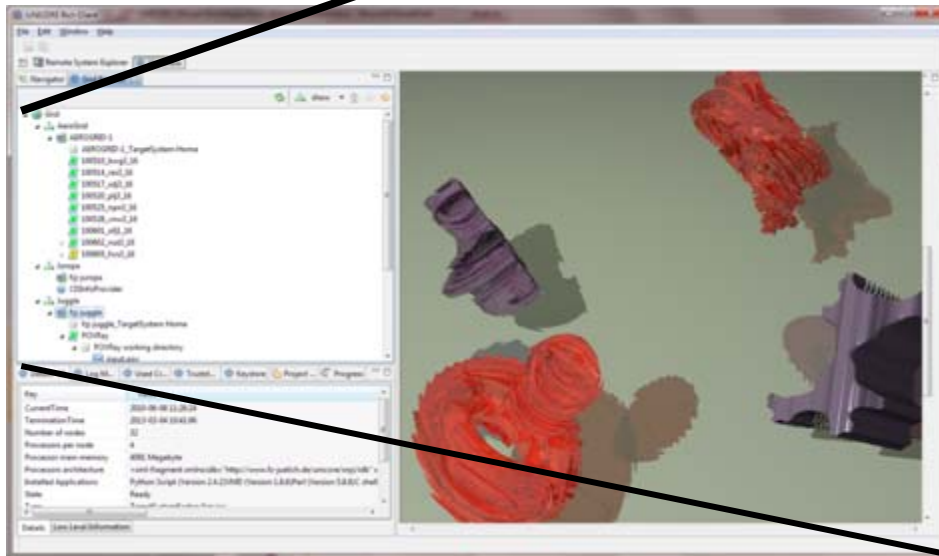 - Technical and economic Efficiency

Service-Concept for the Provisioning of HPC-Resources: 4-Tier Architecture.

??

Grid-Technology (e.g. Synfiniway, UNICORE)

Service-Desk/Control-Center

Strategic public Resources, e.g. **PRACE**

Tier 0 Strategic

Supercomputing
- Top-End Capability
- Dynamic Provisioning

hww

Tier 1 Remote

Production-Environments for Throughput and Capability
- System- and Application Management
- Integration
- Capacity-Risk on Customer Side
- Systems are owned by customer or Service-Provider

Tier 2 Local

Test- and Development Environments
- Compiler/Cross-Compiler
- Analysis-Tools (Debugger, Performance-Analysis)
- Pre- and Postprocessing
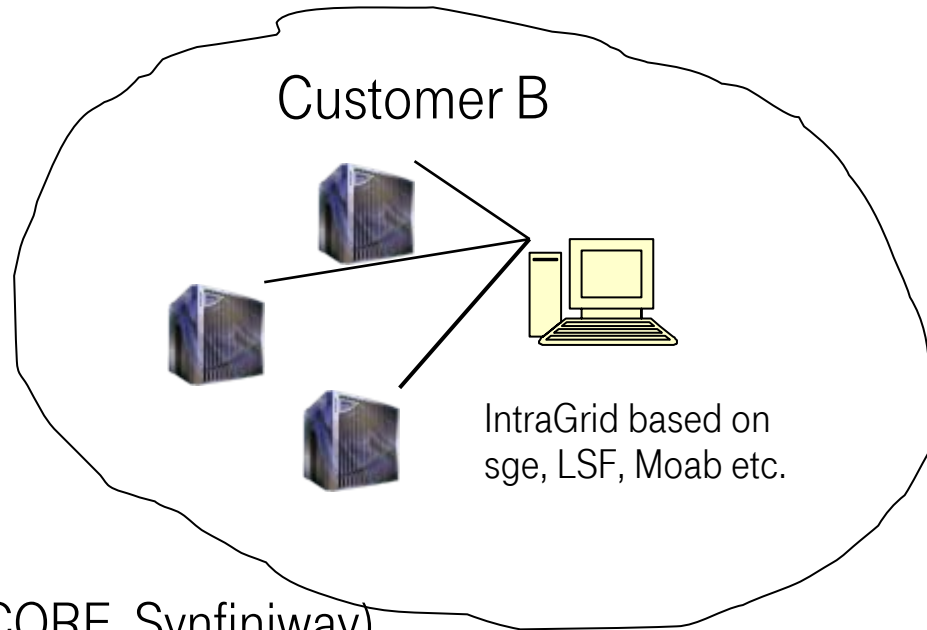
Tier 3 Client

·T··Systems

Service-Concept.
Tier 1: Service-Integration based on UNICORE or Synfiniway
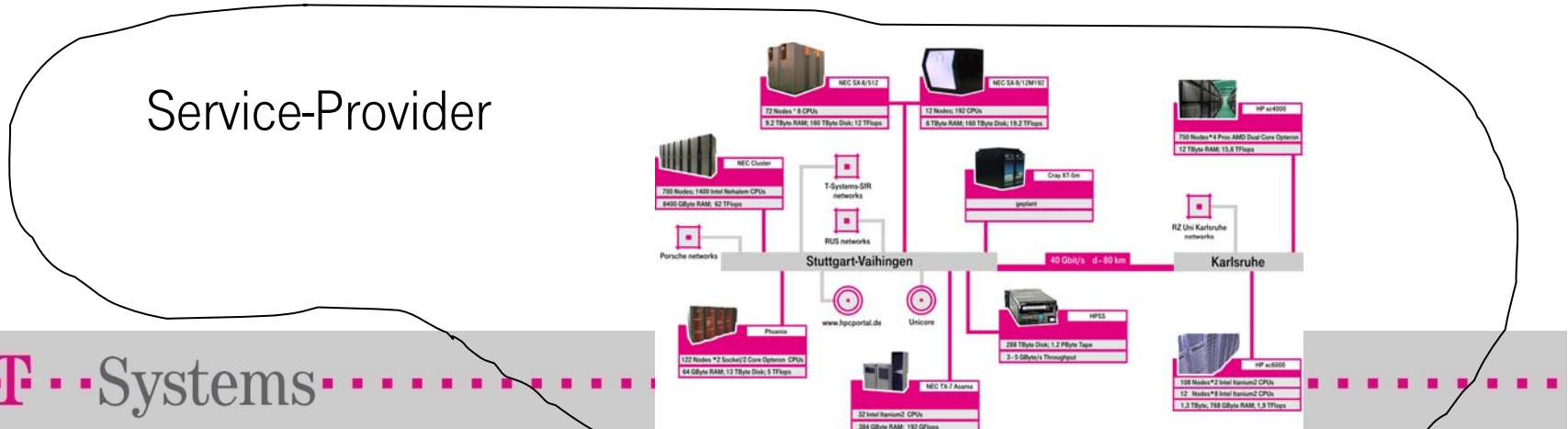
The UNICORE-Client shows all Systems reachable for
Users at a given organisation independent of their location and provider

Flexible and dynamic Provisioning-Concept.
What we offer today

Customer A

IntraGrid based on sge, LSF, Moab etc.

Customer B

IntraGrid based on sge, LSF, Moab etc.

Service-Grid (e.g. UNICORE, Synfiniway)

Service-Provider

· · **T** · · Systems ·

21.06.2010

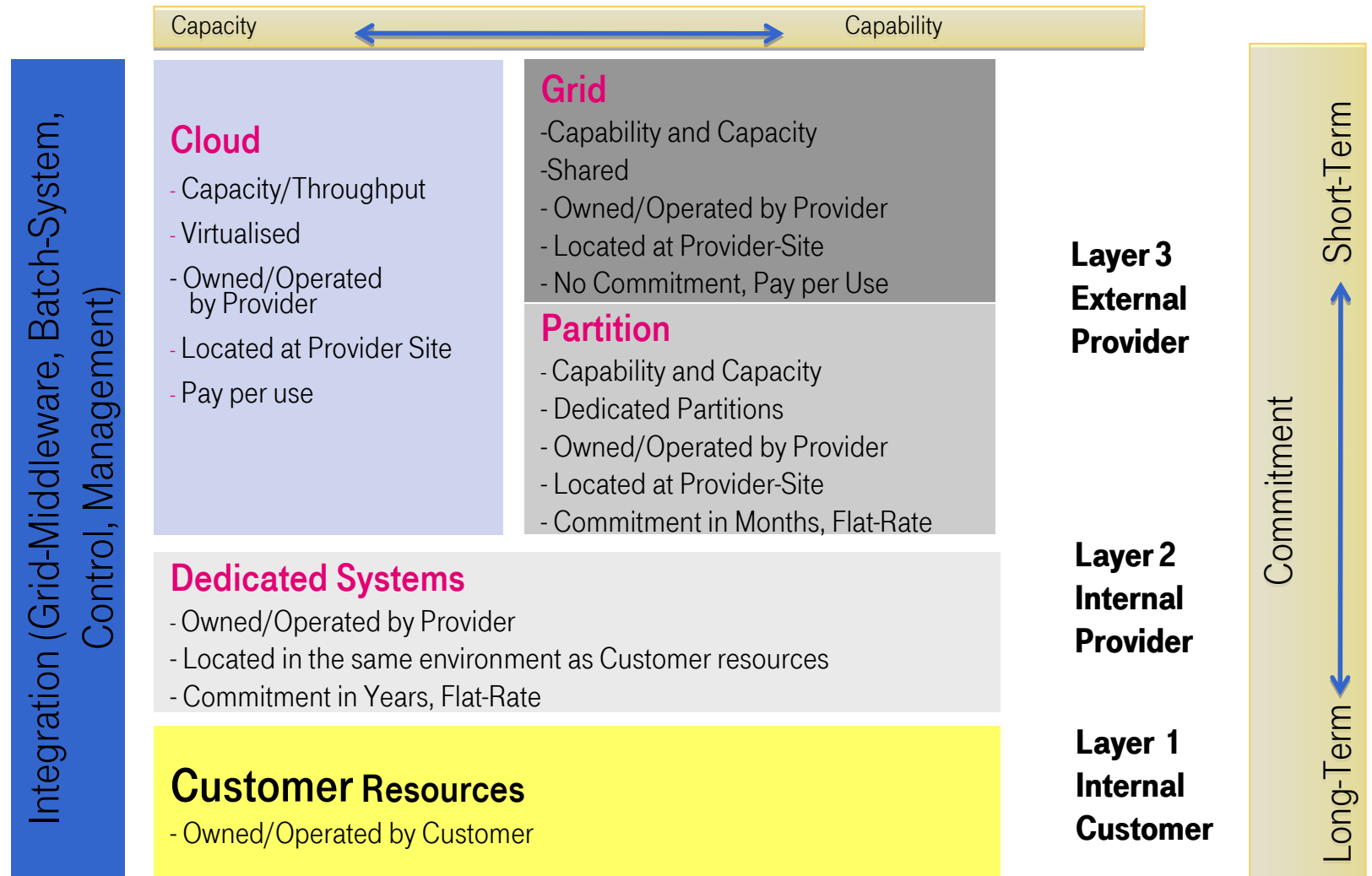# Flexible and dynamic Provisioning-Concept.
Experiences

- Technical Experiences

    - Absolutely stable

    - Security-Concept accepted

    - Administrative environment and accounting fully integrated

- Business Experiences

    - High acceptance for collaborative lighthouse projects

    - Accepted by SMEs for the sporadic use of HPC-resources if the appropriate software licenses are available on a pay per use base (www.hpcportal.de).

    - General use not growing significantly

    - → Requires further analysis

· · **T** · ·Systems· ·

Flexible and dynamic Provisioning-Concept.

What our customers request

Customer A

IntraGrid based on sge, LSF, Moab etc.

Customer B

IntraGrid based on sge LSF, Moab etc.

Service-Provider

Flexible and dynamic Provisioning-Concept.
Customer-Survey

Capacity	⟷	Capability

Integration (Grid-Middleware, Batch-System, Control, Management)

Cloud
- Capacity/Throughput
- Virtualised
- Owned/Operated by Provider
- Located at Provider Site
- Pay per use

Grid
- Capability and Capacity
- Shared
- Owned/Operated by Provider
- Located at Provider-Site
- No Commitment, Pay per Use

Partition
- Capability and Capacity
- Dedicated Partitions
- Owned/Operated by Provider
- Located at Provider-Site
- Commitment in Months, Flat-Rate

Dedicated Systems
- Owned/Operated by Provider
- Located in the same environment as Customer resources
- Commitment in Years, Flat-Rate

Customer Resources
- Owned/Operated by Customer

Layer 3 External Provider

Layer 2 Internal Provider

Layer 1 Internal Customer

Commitment

Short-Term ⟷ Long-Term

··T··Systems·

Flexible and dynamic Provisioning-Concept.
Customer-Survey - Results

- Dedicated Systems are the (static) base of a provisioning concept, independent from whether they are owned and operated by the customer-organisation or a service-provider.

- Shared services in general are only accepted for collaboration-projects, in research or within the customer-organisation (Intra-Grid).

- Virtually dedicated external services in cloud-style are the preferred option for the flexible and dynamic expansion of internal resources. As long as the virtualisation-concepts do not fulfil all the requirements of HPC, temporarily dedicated partitions are acceptable.

··T··Systems···

Flexible and dynamic Provisioning-Concept.
The role of Virtualisation

Goals

- Consolidation/Fine-granular use of resources (granularity below e.g. 1 core of a CPU)
- Security /Safety
  - Integration of virtualized external resources into the customer security-domain
  - Snapshots: Fault-Tolerance/Resilience
  - System Level Checkpoint Restart
  - ´Clean´Computing/Consistency
- Adaptability/Heterogenity
  - The user decides on the complete software-stack (OS, Libraries, ...)
- Flexibility
  - Elasticity: Dynamic expansion of **customer**-resources (OPEX vs. CAPEX)
  - Dynamic Workload Management

Relevant for HPC

Requirements

- Dynamic virtualisation of all hardware and software components
- Federation of virtualized resources (Logical Cluster, Gang).

·T··Systems·

Clouds and HPC.
Where are we in Virtualisation1/2 ?

- **Compute-Server Virtualisation**
  - available (VMWare, Xen, KVM, Bare Metal (POWER, SPARC, ..))
  - SMP-performance degradation dramatic in hosted virtualisation (all x86....)

- **Virtualisation of Node-Interconnects**
  - available for Ethernet, but Latency +100%
  - Not available for Infiniband (except partitioning based e.g. on Voltaire UFM). In principle possible, however currently not enough requests from the market. Even worse in that sense: OS-bypass.

- **Virtualisation of Storage**
  - Global File System available (GFS), but not scalable in the sense of performance, not relevant for HPC
  - I/O Performance-degradation of 10-15% in non-scalable configurations
  - Parallel File-Systems (Lustre, GPFS): ....no idea yet, but discussions with RedHat etc.

- **Virtualisation of Access-Networks**
  - available (VPN)

available   partially available, in progress   unsolved

· · ·T· ·Systems·

Clouds and HPC.
Where are we in Virtualisation 2/2?

- **Virtualisation of Logical Clusters (Federation of virtualized Resources)**
  - Research in progress
    - xge, (InGrid, Univ. Marburg)
    - DGSI (D-Grid)

- **Virtualisation of Performance-Counters**
  - Not implemented so far

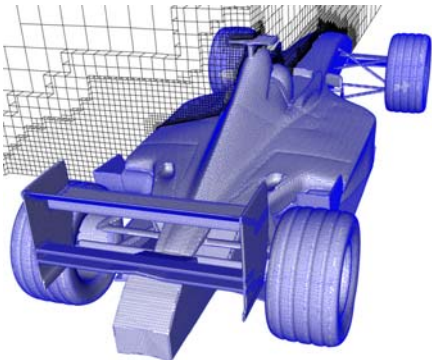- **´Virtualisation´ of Software-Licenses**
  - Research in progress (BEinGrid, SmartLM) related to flexible use of licenses → next slides
  - Other interesting topic: License-Scheduling in Clouds
    - License cost much higher than CPU cost → use fastest CPU for most expensive license
    - Status: Discussion of Ideas with RedHat and others

· · **T** · · Systems ·

´Virtualisation´ of Software-Licenses.
The technical side

The authorization of currently used client-server based license mechanisms relies on an IP-centric scheme.
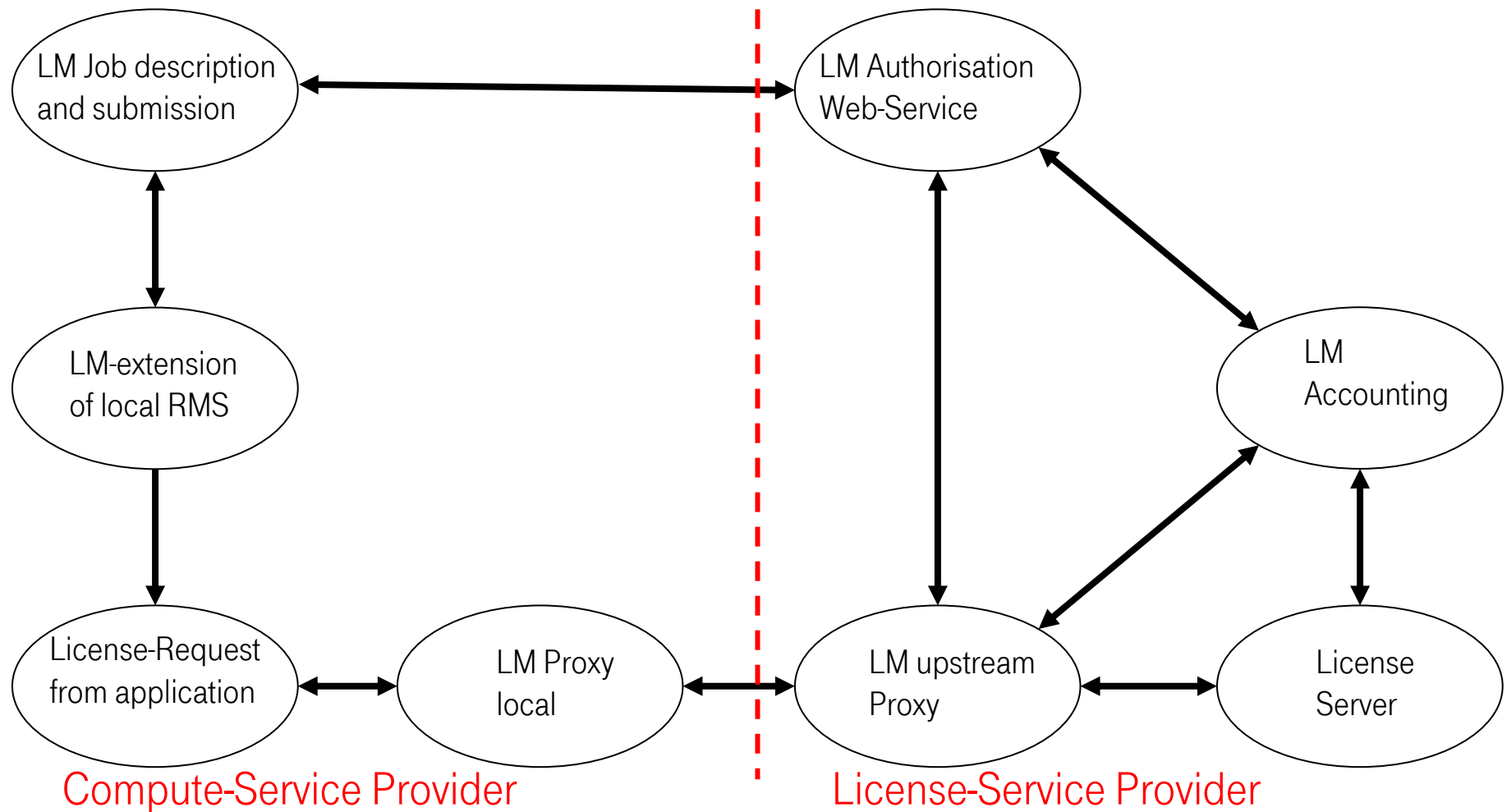
- Any user of shared (Grid) or virtualized resources (Cloud) may access any exposed license server.

- Secure and authorized access to a local or remote license server in cross-organisational environments has not been possible so far.

- The use of commercial ISV applications in such environments therefore was not possible so far. → Show-Stopper

- A disruptive approach will not be accepted by ISVs and end-users in the short term.

    - Existing schemes have to be extended

    - No change of Application-Software and LM-Service

    - Independence from specific middleware

· · ·**T**· ·Systems· ·

´Virtualisation´ of Software Licenses.
LM-Architecture based on PIN/TAN based Authorisation

LM Job description and submission

LM Authorisation Web-Service

LM-extension of local RMS

LM Accounting

License-Request from application

LM Proxy local

LM upstream Proxy

License Server

Compute-Service Provider

License-Service Provider

·· **T**·· Systems·

Summary and Outlook

- Cloud-like provisioning-models are requested by users and controllers

- In Technical and Scientific Computing, Cloud-Environments are usable for Throughput today

- Cloud Technology based on Virtualisation is not ready for HPC

- Cloud as a Business-Model is usable in HPC, if coarser granularity (days to weeks instead of minutes to hours) is acceptable.

- Projects to solve the open issues are partially in progress

- HPC and Cloud infrastructures are converging

Managed business flexibility

Thanks.