PHILIPS

Supporting e-Science

Philips Research Laboratories

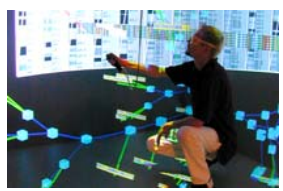
May 2007 - V1.1

Contents

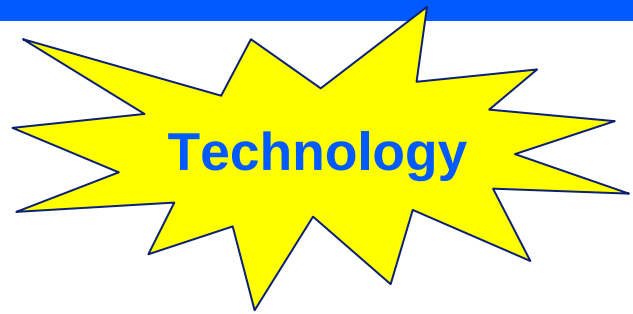
- Observations 1 – 4
- Vision
- Researcher workflow
- Success factors
- Examples 1 - 4

Observations 1

- In many application fields the way of working is changing "*from model driven to data driven experiments*"
  - High-energy Physics, Astronomy
  - Bioinformatics and Computational Medicine
  - ✓ A rapid increase in the number of (high-resolution) sensors used in equipment
  - ✓ Correlate data from different sources
  - ✓ Extract information from an overwhelming amount of data
- Need multi-physics and 3D calculation models
- Virtualization is a key element



Observations 2



- Network technology is a big enabler
  - WAN* bandwidth >> Local computer bus bandwidth
  - Point to point connections via optical links without routers and/or switches
    - High bandwidth and low latency
- Resources are pushed to the WAN
  - Access to specialized and shared resources
  - New business models are pending
- Need standards to access resources in the WAN
  - Grid standardizes the sharing of WAN resources
- Many compute problems can be parallelized!
  - Cluster and multi-core technology boost parallel jobs

* WAN: Wide Area Network

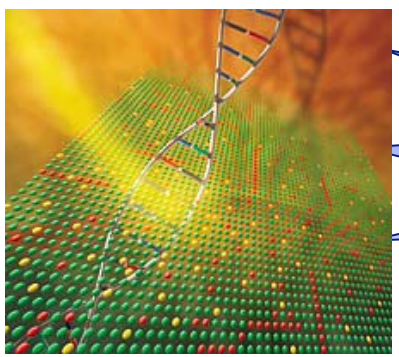
Observations 3

- New application types pop-up
  - Combination of very specific resources
- Real-time access to shared resources
  - Access to specialized external resources
    - E.g. data storage and parallel compute clusters
    - Shared in virtual teams
    - NIKHEF, SARA, Jülich
  - Rendering for virtual reality visualisation
    - Large high-density tiled displays
- Coping with very large and distributed data sets
  - Storage brokers for storing very large data sets.
  - Data grids for secure and fast exchange of data
  - Molecular Medicine and Bio-Science applications

Observations II

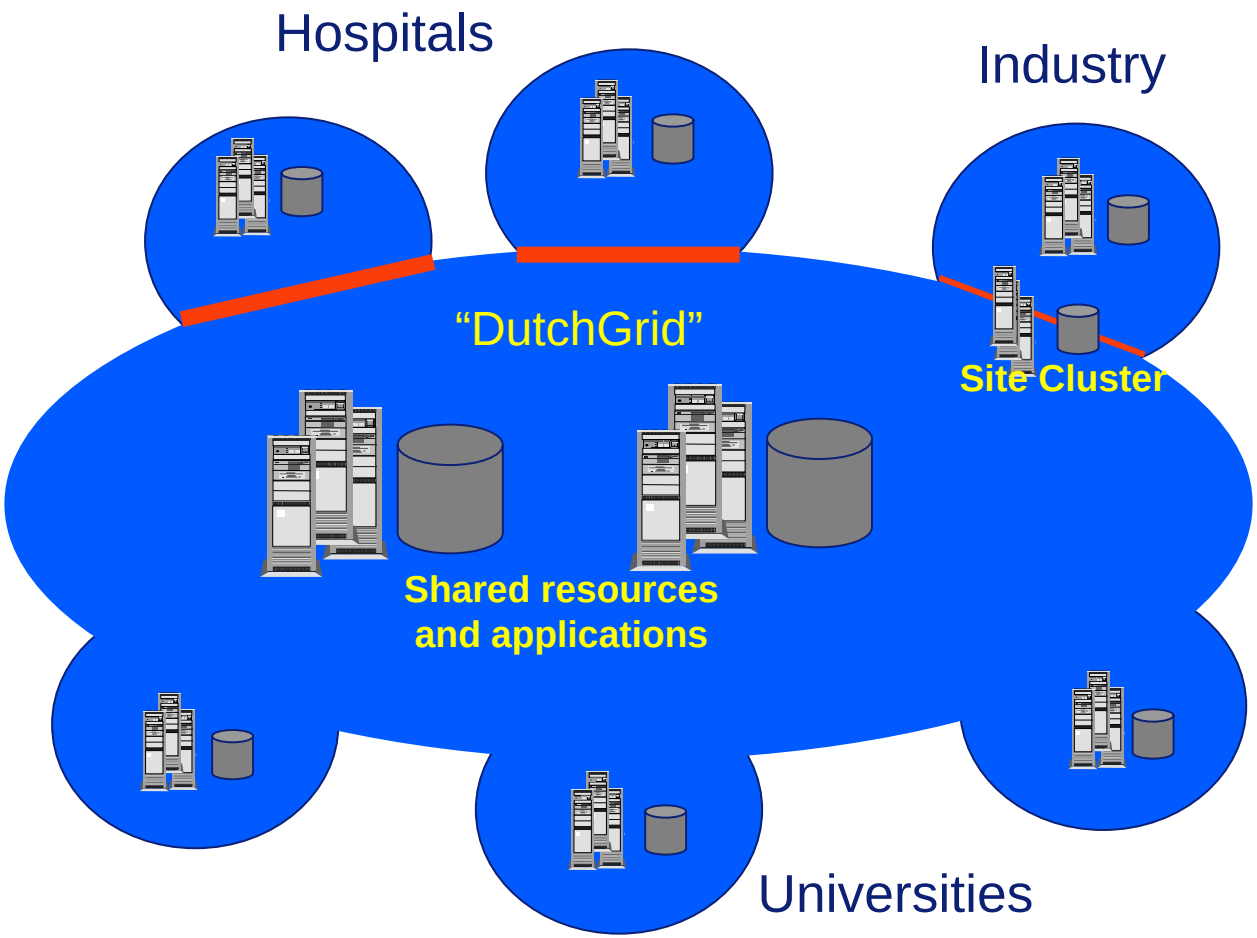
- Big growth in application complexity.
- Resource needs of some application fields may grow orders of magnitude
- Will lead to huge growth in data storage and compute capacity needs.

- And today's business innovation is done in an Open Innovation setting:
- Collaboration in virtual teams is the way of working.
(Industry, Universities & Institutes)
- Needs to be supported by applications and IT infrastructures.



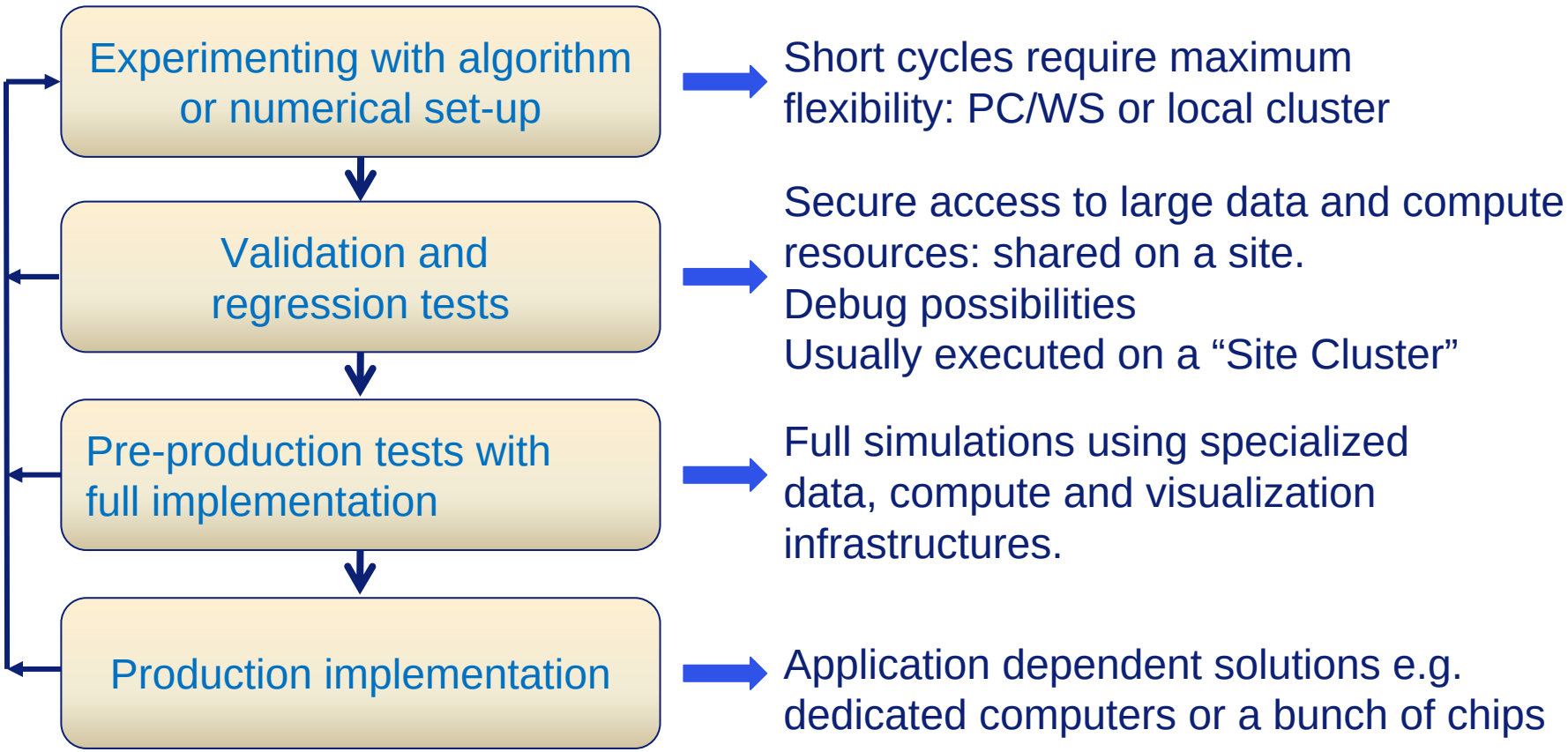
Cannot be done by a single entity
Must share and cooperate

Vision: Shared Resources and Infrastructure



Example: Researchers Workflow

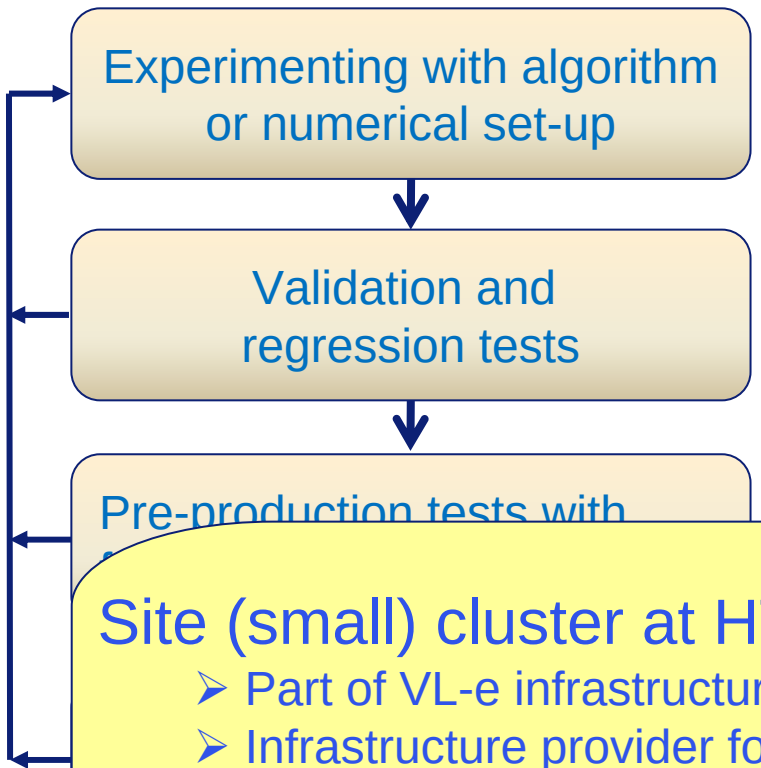
Workflow View



- ***E.g. medical imaging, bioinformatics simulation and search, 3D multi-physics modeling, system in package simulation, etc.***
- ***Many projects do not follow the complete flow, but stay at one or more stages.***

Example: Researchers Workflow

Workflow View



Short cycles require maximum flexibility: PC/WS or local cluster

Secure access to large data and compute resources: shared on a site.
Debug possibilities
This “Site Cluster” is part of “DutchGrid”!

lations using specialized
nd visualization

dent solutions e.g.
ers or a bunch of chips

ch,
n, etc.

Site (small) cluster at HTC Eindhoven

- Part of VL-e infrastructure
- Infrastructure provider for VL-e
- Maximum control and fast feedback
- Priority setting and security can be guaranteed

- ***Many projects do not follow the complete flow, but stay at one or more stages.***

Vision: Shared Resources and Infrastructure

End-user view

- Applications “flow through infrastructure” and can be run and used anywhere:
  - ➔ boosts sharing and collaboration
- An end-user chooses to keep data in internal secure environment or to put in shared space
- Applications might be split in secure and non-secure parts

Vision: Shared Resources and Infrastructure

Success factors

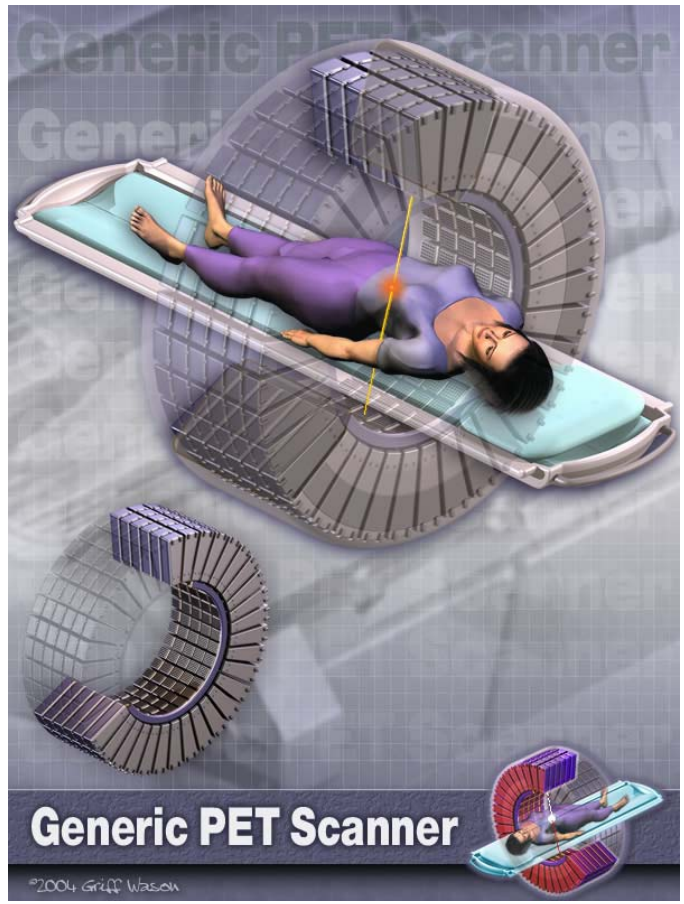
- **End-user guidance and support is essential!!**
- **Supply workbenches for application fields**
- **Continuous development, be a leader**

- Standardisation of middleware layer (e.g. gLite)
- QoS and AAA are key in e-Science production infrastructures
- Basic building blocks are data storage and (low latency) compute clusters

Example 1

Grid Demonstrator – Aachen ➔ Eindhoven

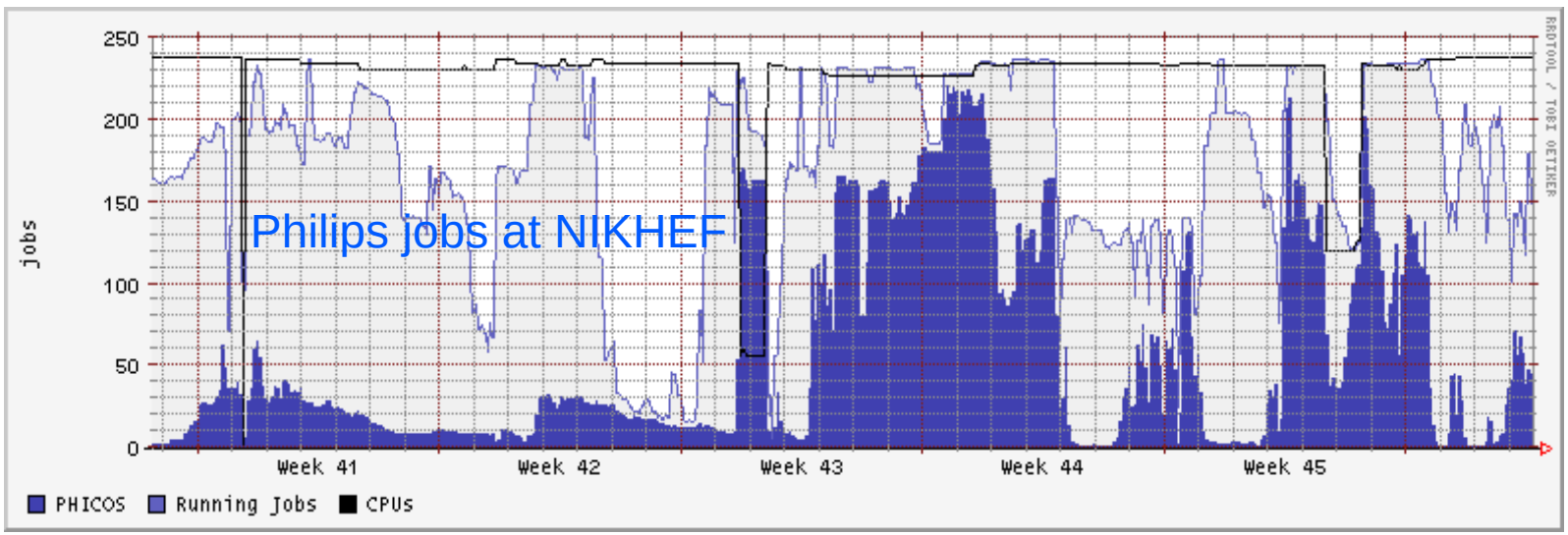
- PET System Simulation
- Runs at Eindhoven “Site Cluster”
- Monte Carlo simulation of the scanner up to the detection of scintillation light
  - ~32 CPU days on 2.8 GHz Xeon
  - 100 x 1 GB output
- System-level simulation of the detector electronics (Mona/Lisa)
  - 2 phase post processing
- Simulation can be highly parallelized



Example 2

SPECT Simulation

- 20,000 jobs of ~3.2 CPU hours each
- Runs at NIKHEF and SARA VLe clusters
- Access via LCG-2 Grid middleware software
- Single Photon Emission Computed Tomography

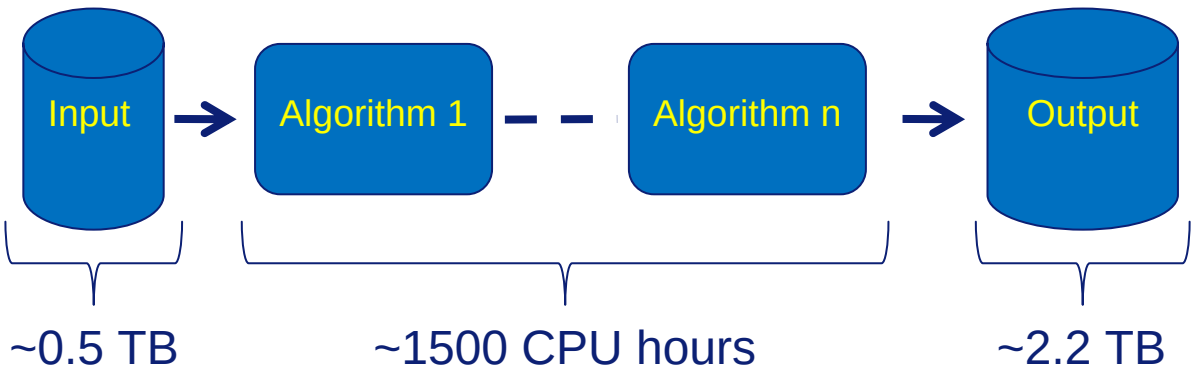


Example 3

Simulate Full HDTV Data Processing Chain

- Picture quality enhancement
- Picture rate-up conversion
- Pixel plus …
- Runs at NIKHEF and SARA VLe clusters

Original (Frame repetition)



Measured > 50 MB/s data transfer rates between Amsterdam ←→ Eindhoven over GigaPort connection

DIGITAL NATURAL MOTION

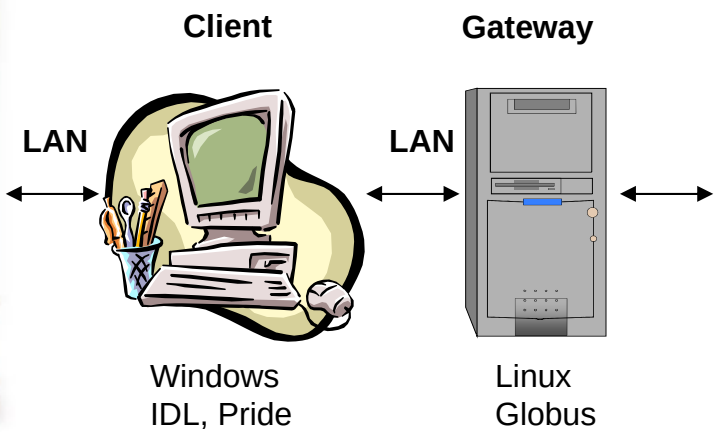
Example 4

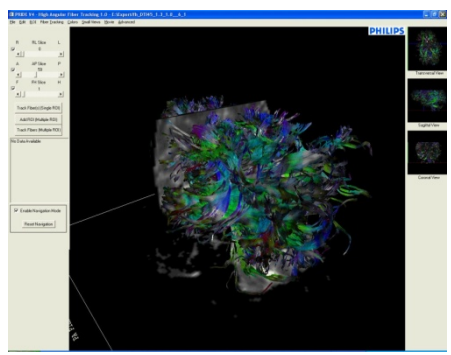
GAMA Research

Healthcare Systems Architecture research group

The GAMA architecture: computational Grids

- Architecture for solving compute-intensive medical applications
- Minimally invasive: Running on Grid as alternative, easy fall back to local versions
- Simultaneously provides different sets of services to multiple users and applications
- Adaptive to various healthcare applications
- **Example**: Brain imaging, the High Angular Fiber tracking (HAFT) application





Clinical trials of HAFT software at AMC are being executed at Eindhoven “Site Cluster”.


PHILIPS