Testing storage on the GRID

Klaus Wiebe, University of Mainz

Sep, 18th 2006

DESY summer student program

www.students.uni-mainz.de/wiebe/desy/grid/

Part I

Introduction to the GRID

DESY Summer Student Program 2006,
Klaus Wiebe

The Grid idea

- **Definition**

  „A computational grid is a hardware and software infrastructure that provides dependable, consistent, pervasive, and inexpensive access to (distributed) high-end computational capabilities”

  (Ian Foster and Carl Kesselmann, 1998)

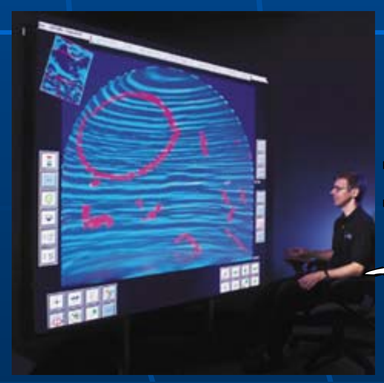
- **The name‘s origin**

  In analogy to the power grid a computational grid should provide an easy-to-use user interface hiding complex internal processes.

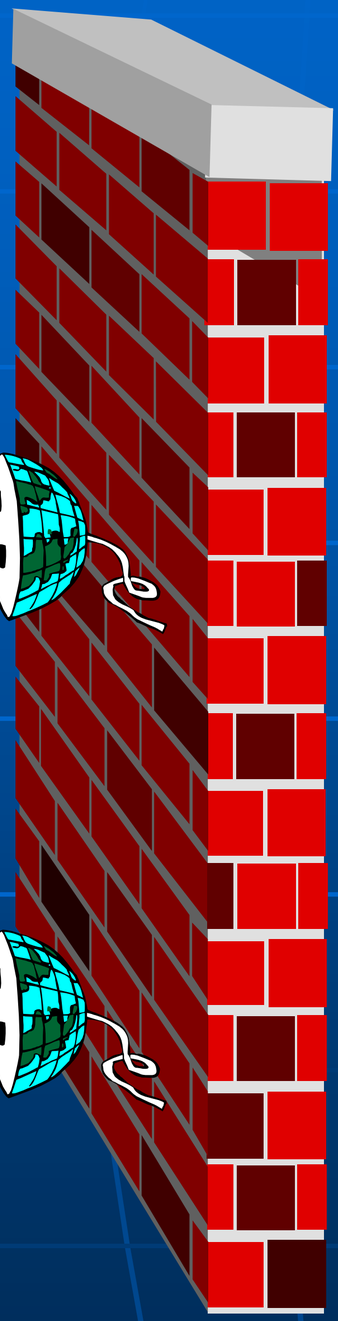
The Grid dream

Mobile Access

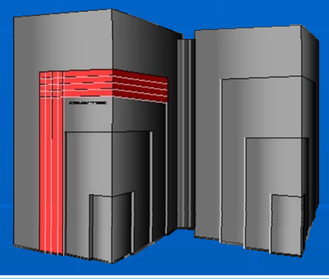
Desktop

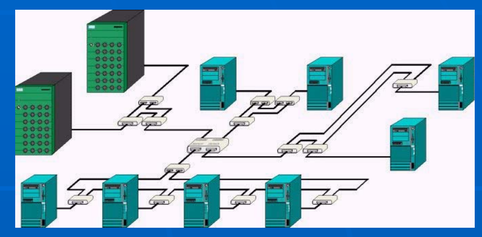


Visualizing

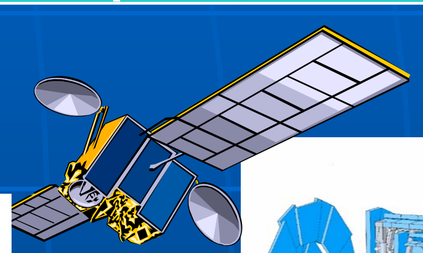


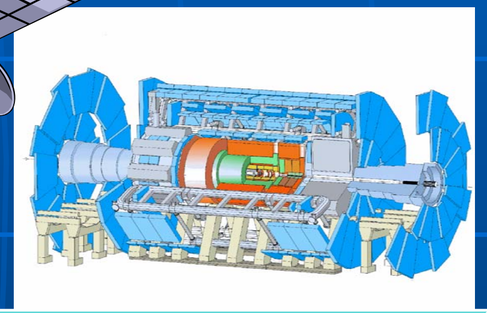
GRIDMIDDLEWARE



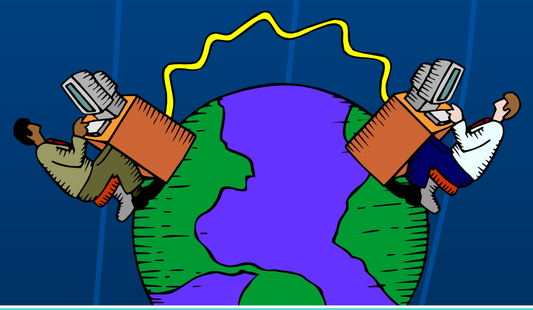


Supercomputer, PC-Cluster





Data Storage, Sensors, Experiments



Internet, Networks

What is the Grid ?

A Three Point checklist (Ian Foster, 2002)

A grid is a system that...

- coordinates resources which are not subject to centralized control

  integration and coordination of resources and users of different domains
  versus local management systems (batch systems)

- uses standard, open, general-purpose protocols and interfaces

  standard and open multi-purpose protocols
  versus application specific system

- delivers nontrivial qualities of service

  coordinated use of resources (delivering QoS, e.g. response time, throughput, availability, security, ...)
  versus uncoordinated approach (world wide web)

Virtual Organizations

Virtual Organization

Same data,
same rules to analyze
and same access rights

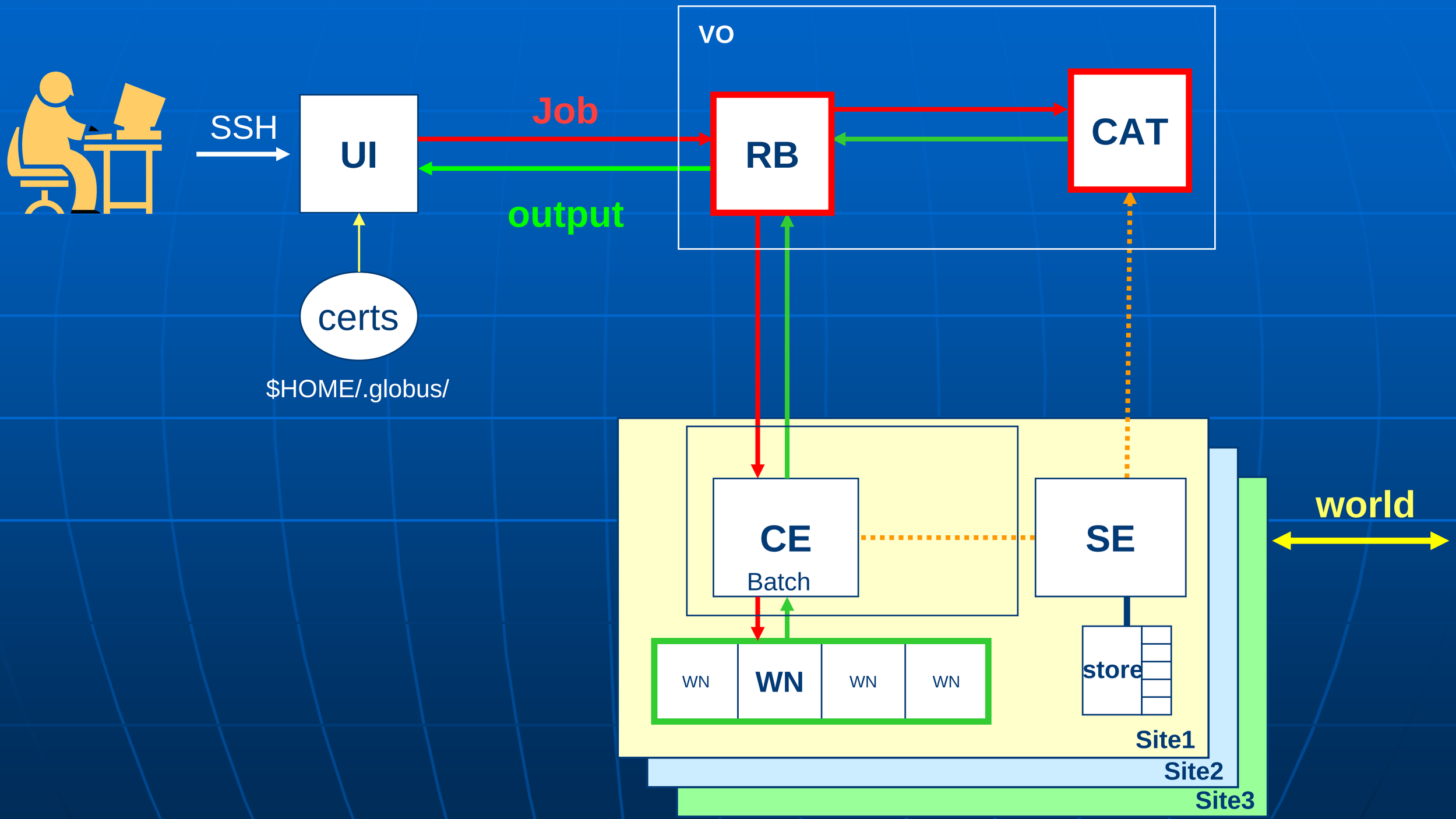
Virtual Organization

Same data,
same rules to analyze
and same access rights

“A set of individuals and/or institutions defined by such sharing rules is what we call a virtual organization.“

I. Foster, C. Kesselmann, S. Tuecke (2000)

Grid Infrastructure



The LHC Computing Grid (LCG)

- LHC experiments starting in 2007 produce large amounts of data (~15 PB per year) and will use Grid technology to realize computing and data storage

- EU supports scientific grids with the EGEE project

Part II

A test program for Storage Elements

DESY Summer Student Program 2006,
Klaus Wiebe

Motivation

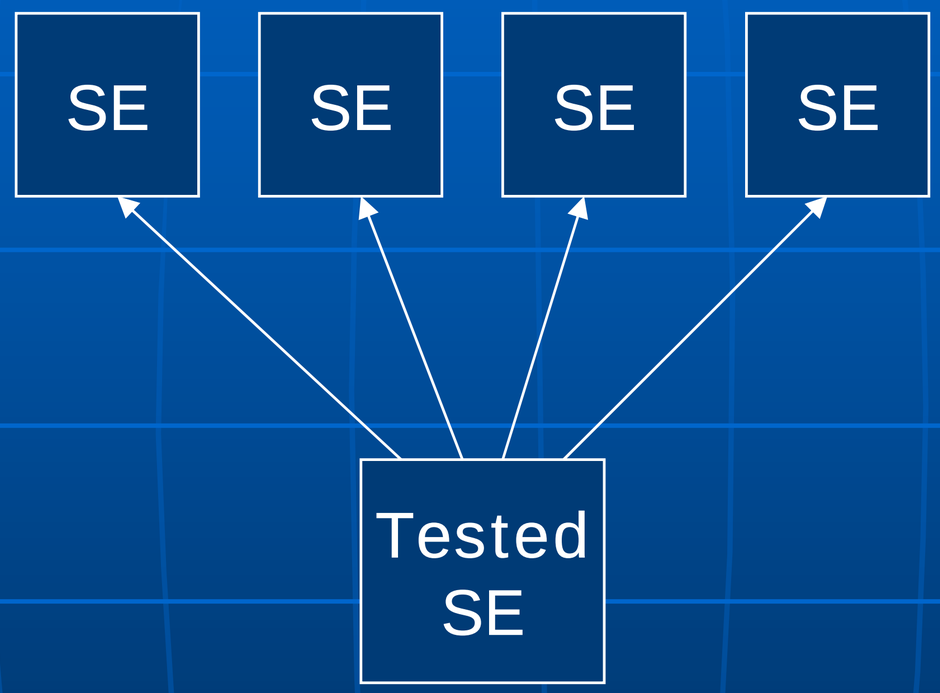
- In HEP data is very precious and contains all physics.
- Failed computing can be repeated, but corrupted storage will make data unavailable forever
- Data handling is not sophisticated yet, storage elements are the weak point in the Grid
- Availability and access to data needs to be tested

Realization: Replication

Error detection

GUI

Time measurement



Time-out detection

Step-by-Step

PHP

Access via web-browser

live demonstration

DESY Summer Student Program 2006,
Klaus Wiebe

Summary

Computing grids provide worldwide distributed computing and storage resources.
Grids are about virtualization, hiding complex technical processes from the user.
People using the grid are organized in so-called Virtual Organizations.
In HEP this would typically be research collaborations.

DATA is precious, contains the physics of HEP experiments.
Testing, availability and access to data is what my work is about.

Questions ?