Das PC²: Vom Transputer zur „intelligenten Steckdose“

PC²

PADERBORN
CENTER FOR
PARALLEL
COMPUTING

Gliederung

1. PC²: Was ist das ?

2. PC²: Die Hardware

3. PC²: Die Projekte

4. PC²: Wie geht‘s weiter ?

1. PC²: Was ist das ?

PC^2 - Paderborn Center for Parallel Computing

Organisationsform:

wissenschaftliches Institut der Universität-GH Paderborn

gegründet:

April 1991 (Vorstandsvorsitzender: Prof. Monien)

Aufgaben:

landesweites Kompetenzzentrum für parallele und verteilte Datenverarbetiung

Zielsetzung:

effiziente Nutzung paraller und verteilter Systeme

Leistungsspektrum:

- Forschung und Entwicklung (Tools, Algorithmen, Anwendungen)
- Rechenzentrum-Service für parallele Systeme
- Leistungsmessung innovativer Architekturen
- heterogene Systeme, Rechnerzugangs-Software
- Workshops, Tutorials, Schulungen

Nutzer:

~ 250 Anwender in 120 Projekten, 50% extern

MitarbeiterInnen:

- 11 wissenschaftliche
- 4 nichtwissenschaftliche, 1 Azubi

Rechnersysteme:

- massiv parallele Systeme mit bis zu 1024 Prozessoren
- innovativer SCI-Compute Cluster

PADERBORN CENTER FOR PARALLEL COMPUTING

Supporting Groups

HNI - Heinz Nixdorf Institute

interdisciplinary research institute for computer science and technique

~ 15 professors ~ 100 staff members

CS Department

~ 15 professors - 60 staff

C-Lab

Cooperative Computing & Communications Laboratory

~ 120 researchers

DFG Graduate Center

„Parallel Computer Networks in Production Technique“

~ 18 students

Parallel Computing in Paderborn

PC² - Paderborn Center for Parallel Computing

interdisciplinary institute for parallel & distributed computing

~ 100 members, 14 staff members

SFB 376

DFG Research Unit „Massively Parallel Computing“

~ 50 scientists

PADERBORN CENTER FOR PARALLEL COMPUTING

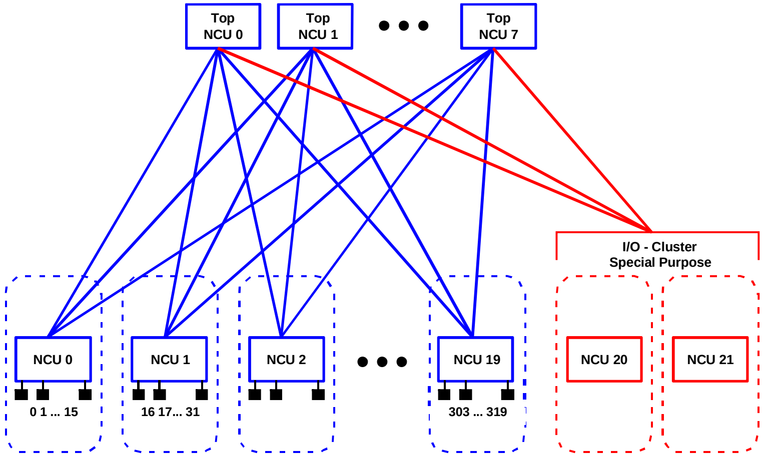
2. PC²: Die Hardware

1991

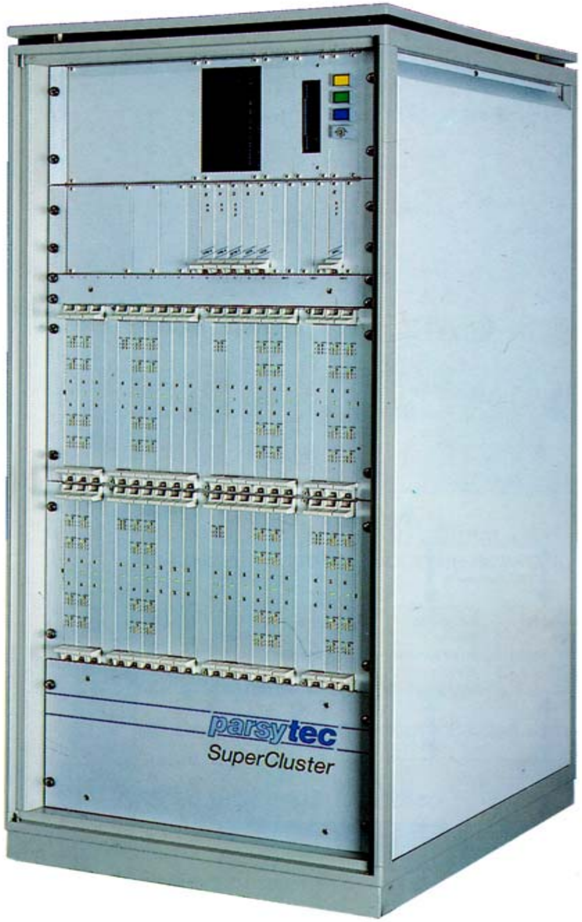
Parsytec Supercluster 320

Number of nodes		:	320
Type of processor		:	T800, 25 MHz
Memory per node		:	4 MB
Network		:	Clos
MFLOPS	(node)	:	3,6
MFLOPS	(system)	:	1 152
Bandwidth	(Parix)	:	1,1 MB / s
Latency	(Parix)	:	45 µs

Feature: Network configurable per software



NCU = Network Configuration Unit (96 x 96 crossbar)



PADERBORN CENTER FOR PARALLEL COMPUTING

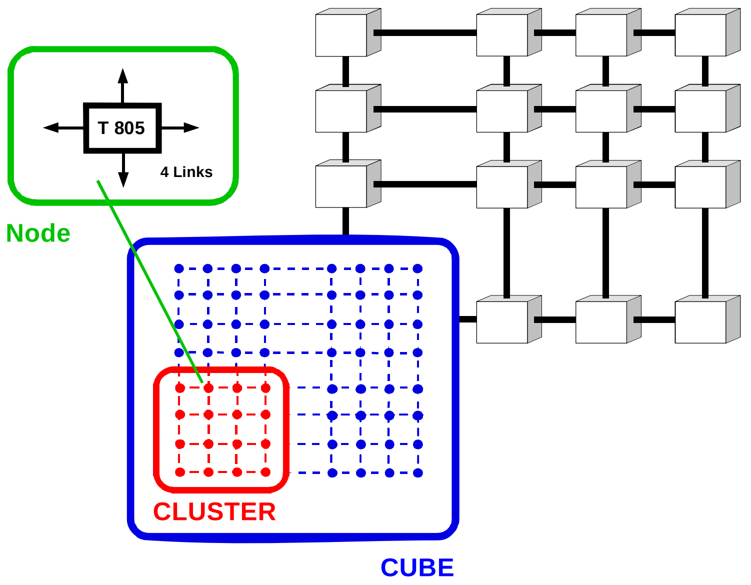
1992

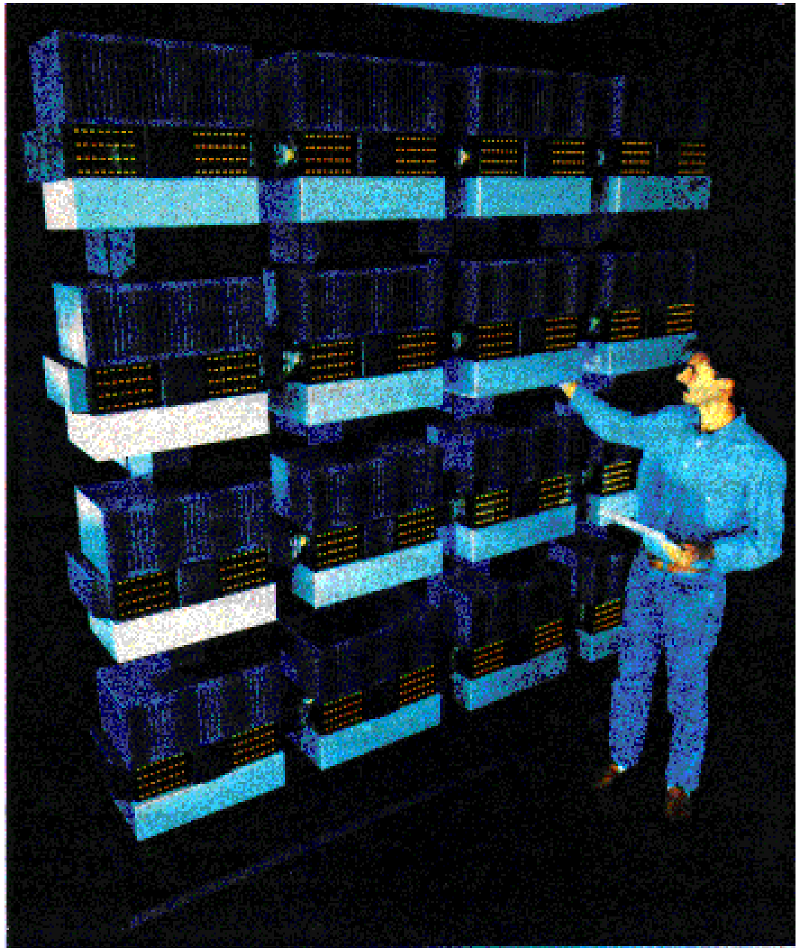
Parsytec GCel-1024

(Achieved rank #259 of the TOP 500 list)

Number of nodes		:	1024
Type of processor		:	T805, 30 MHz
Memory per node		:	4 MB
Network		:	32 x 32 Mesh
MFLOPS	(node)	:	4,3
MFLOPS	(system)	:	4 403
Linpack MFLOPS	(system)	:	971
Bandwidth	(Parix)	:	1,1 MB / s
Latency	(Parix)	:	45 µs

Feature: Scalable up to 1024 processors



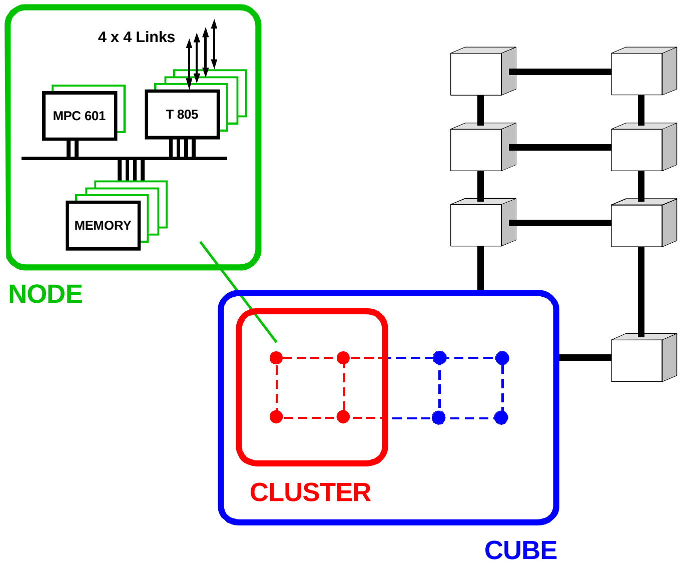


1994

Parsytec GC Power Plus

(Achieved rank #117 of the TOP 500 list,
8000 Linpack MFLOPS using 192 processors)

Number of nodes		:	32
Type of processor		:	MPC601, 80 MHz
Memory per node		:	64 MB
Network		:	8 x 8 Fat Mesh
MFLOPS	(node)	:	160
MFLOPS	(system)	:	10 200
Linpack MFLOPS	(system)	:	5 200
Bandwidth	(Parix)	:	3,3 MB / s
Latency	(Parix)	:	161 µs



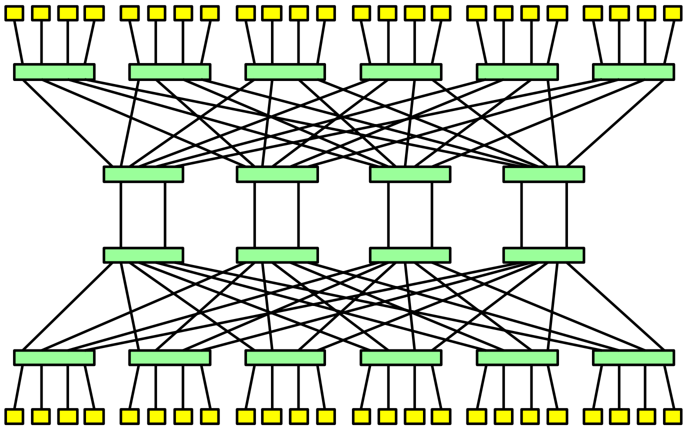
1996

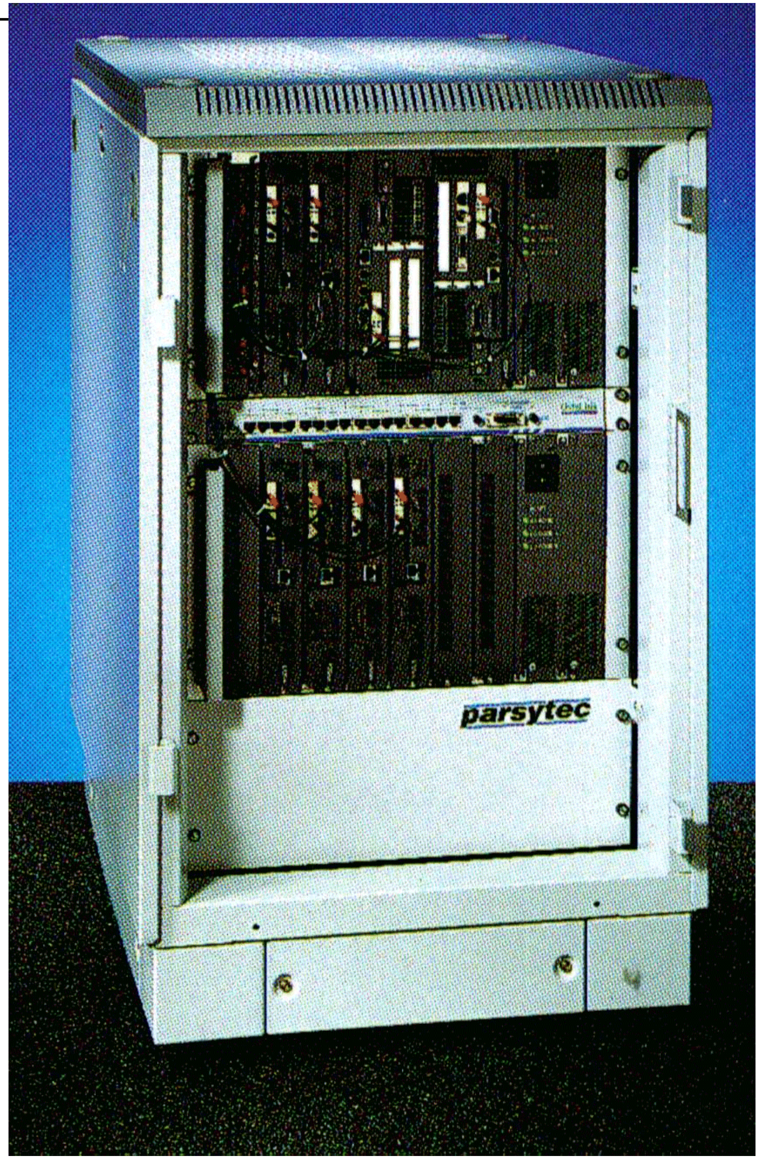
Parsytec CC-48

Number of nodes	:	48
Type of processor	:	MPC604, 133 MHz
Memory per node	:	64 MB
Network	:	Fat Mesh of Clos

MFLOPS	(node)	:	266
MFLOPS	(system)	:	12 768
Bandwidth	(EPX/nK)	:	26 MB / s
Latency	(EPX/nK)	:	45 µs

Feature: 19" Rack mounted standard components





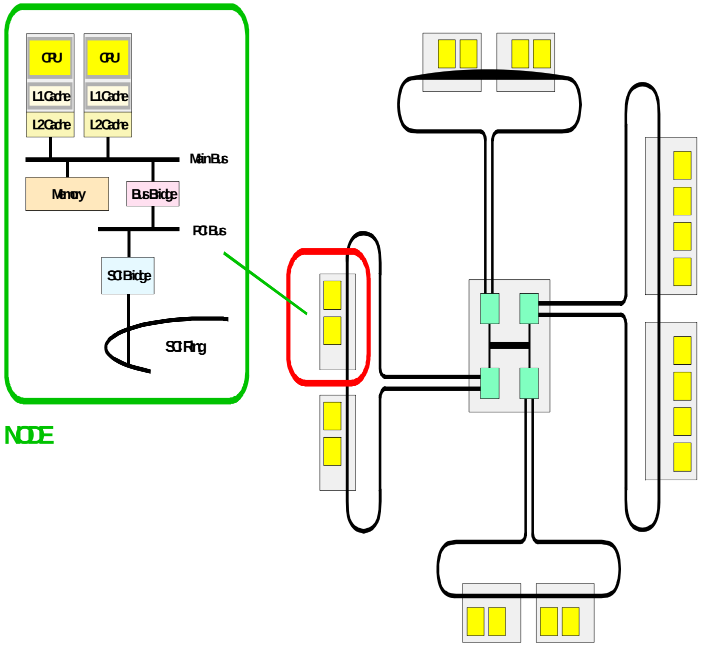
1997

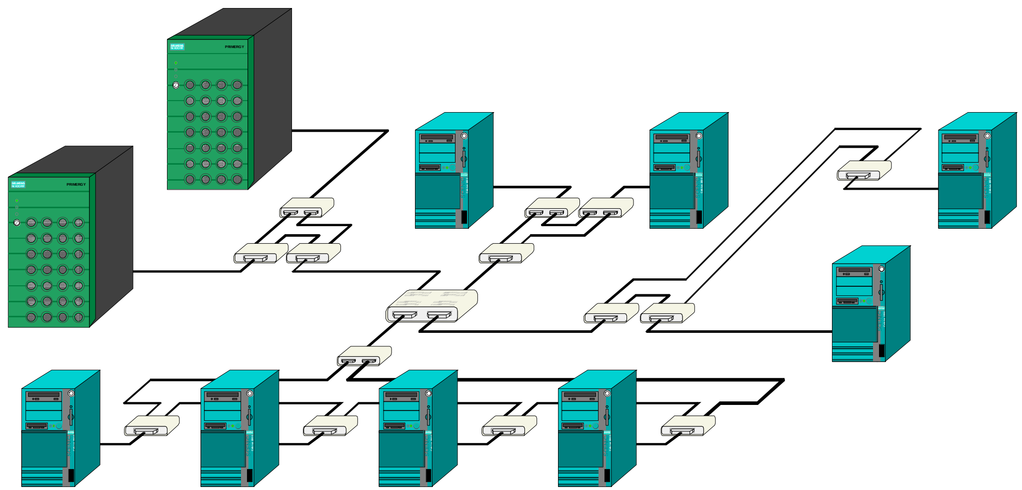
Heterogeneous Workstation Cluster

(consists of quad- and dual processor systems)

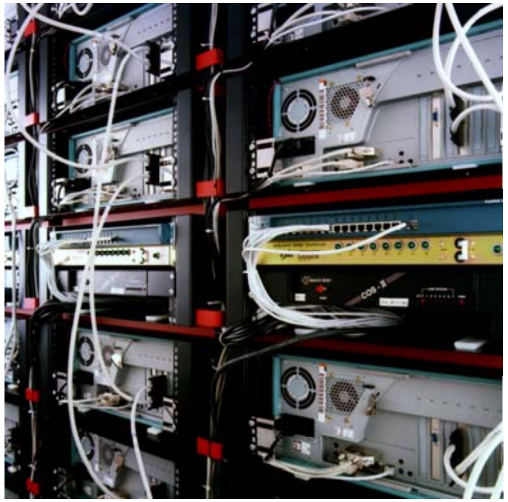
Number of nodes		:	8
Type of processor		:	Pentium Pro, 200 MHz
Memory per node		:	128 / 64 MB
Network		:	switched SCI rings
MFLOPS	(node)	:	800 / 400
MFLOPS	(system)	:	4 000
Bandwidth	(Linux)	:	70 MB / s
Latency	(Linux)	:	3 µs

Feature: Standard workstations connected by fast unidirectional links





1997 Paderborn SCI Cluster (PSC-64)



Manufacturer : Siemens - Nixdorf / Scali

System : 32 nodes, 19.2 GFlop/s peak,

8 GByte SDRAM

Node : 2 x Intel Pentium II 300 MHz, 300 MFLOPS, 256 Mbyte SDRAM

Interconnect : SCI (Scalable Coherent Interface): 500 MByte/s, 3µs

Topology : 8 x 4 Torus with distributed switches

OS : Solaris, Linux, (WindowsNT)

Bandwith : 45 MB/s (ScaMPI, sustained, unidirectional)

Latency : 19µs (ScaMPI)

PADERBORN CENTER FOR PARALLEL COMPUTING

1999 Siemens hpc*L*ine

System	: 96 nodes, 86,4 GFlop/s peak, 48 Gbyte RAM
Compute nodes	: Primergy Server with 2 x Intel Pentium II 450 MHz, 450 MFLOPS peak, 512 Mbyte DRAM
Interconnect	: SCI (Scalable Coherent Interface): 500 MByte/s, 3µs
Topology	: 8 x 12 torus with distributed switches
OS	: Solaris, Linux, (WindowsNT)
Bandwith	: 70 MB/s (ScaMPI, sustained, unidirectional)
Latency	: 17µs (ScaMPI)

PC² PADERBORN CENTER FOR PARALLEL COMPUTING

3. PC²: Die Projekte

PADERBORN
CENTER FOR
PARALLEL
COMPUTING

Projekte am PC²

Innovative Rechnerarchitekturen

- **1994: Simulation von Rechnerarchitekturen** ***(Parsytec)***
- **1993: Middleware für HPC-Betriebssysteme** ***(Parsytec, Hewlett Packard)***
- **1996: Basissoftware für SCI-Cluster** ***(Dolphin, Scali, Siemens)***

Zugangssysteme

- **1992: CCS: Computing Center Software**
- **1995: Forschungsverbund "NRW-Metacomputing"** ***(MSWWF NRW)***
- **1996: WAMM** ***(CNUCE, Pisa)***
- **1997: UNICORE** ***(BMBF)***
- **2000: UNICORE+** ***(BMBF)***

Heterogenes Höchstleistungsrechnen

- **1995: Forschungsverbund "NRW-Metacomputing"** ***(MSWWF NRW)***
- **1997: DYNAMITE: dynamische Task-Migration in LANs und WANs** ***(ESPRIT)***
- **1997: PLUS - „Program Linkage by Universal Software Interfaces"**

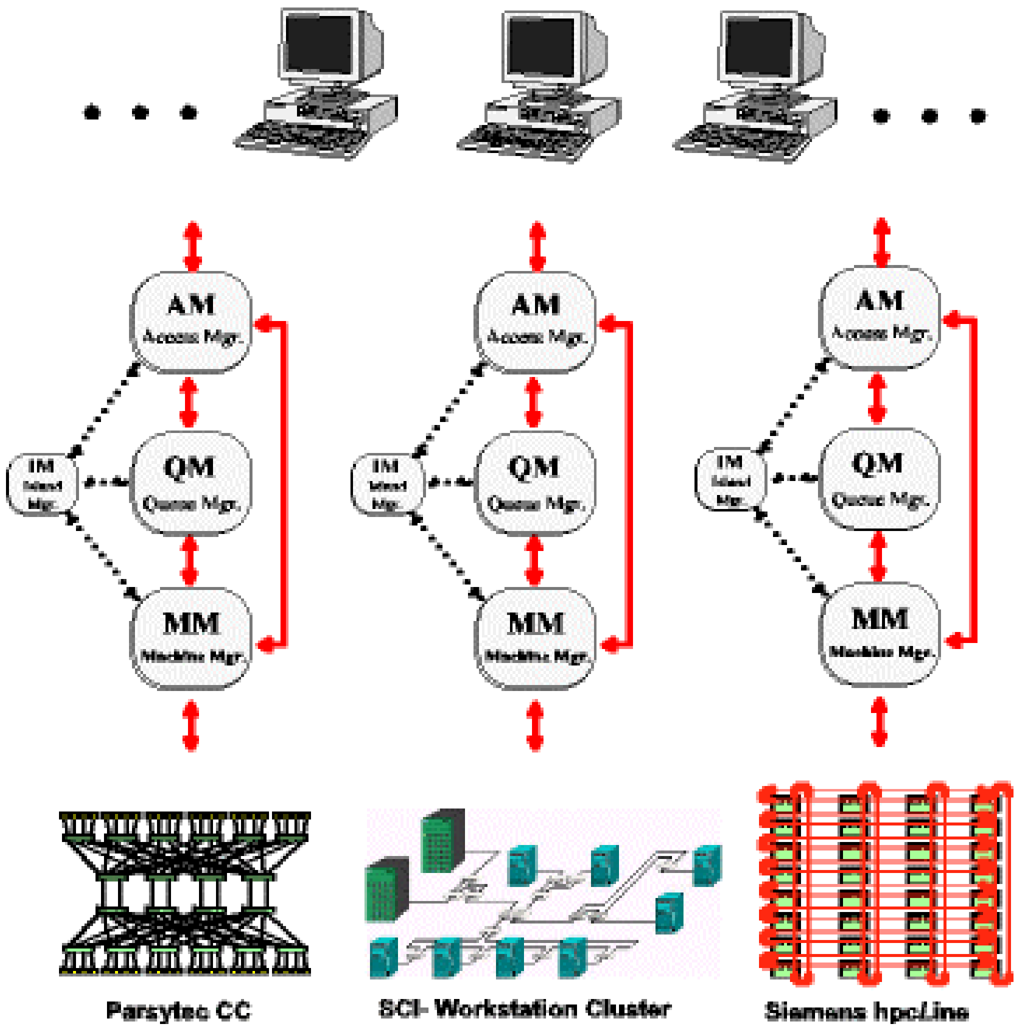
HPC-Anwendungen

- **1996: MICA: A Model for Industrial CFD-Applications** ***(ESPRIT)***
- **1996: PHASE: A Distributed Pharmaceutical Applications Server** ***(ESPRIT)***
- **1998: FEM Strukturmechanik** ***(HILTI AG)***
- **2000: Abwassersimulation** ***(DBU)***

PC² PADERBORN CENTER FOR PARALLEL COMPUTING

CCS : Computing Center Software

topology based resource management for networked hig-perfomance computers



PC²
PADERBORN
CENTER FOR
PARALLEL
COMPUTING

Characteristics of CCS

What is CCS?

- a management software
  for local **HPC systems** in a **single site**
- **interactive** “desktop supercomputing”
  and also for **batch** processing
- effective partitioning of **exclusive**
  resources
- powerful administrator interface

What is it not?

- **no metacomputing** environment
  but a component
- **not for high-throughput** computing

CCS General Features

User Management

- authentication
- authorization (project or user specific)
- accounting

Scheduling / Mapping

- space sharing
- deterministic & fair scheduling
- reservation, deadline scheduling
- interactive & batch usage
- optimal, topology dependant mapping

Resource Description

- describing the available resources
- specifying the required resources

Portability

- hardware independent
- modular & extensible

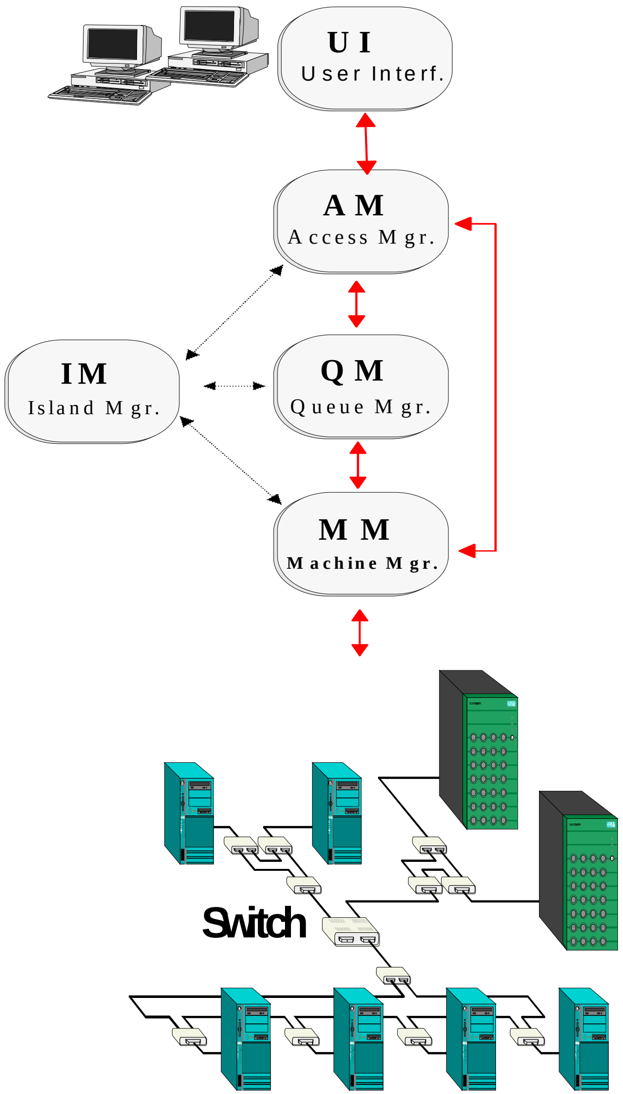
Reliability

- re-bind after network breakdown
- automatic recovery, migration

CCS Cluster Specific Features

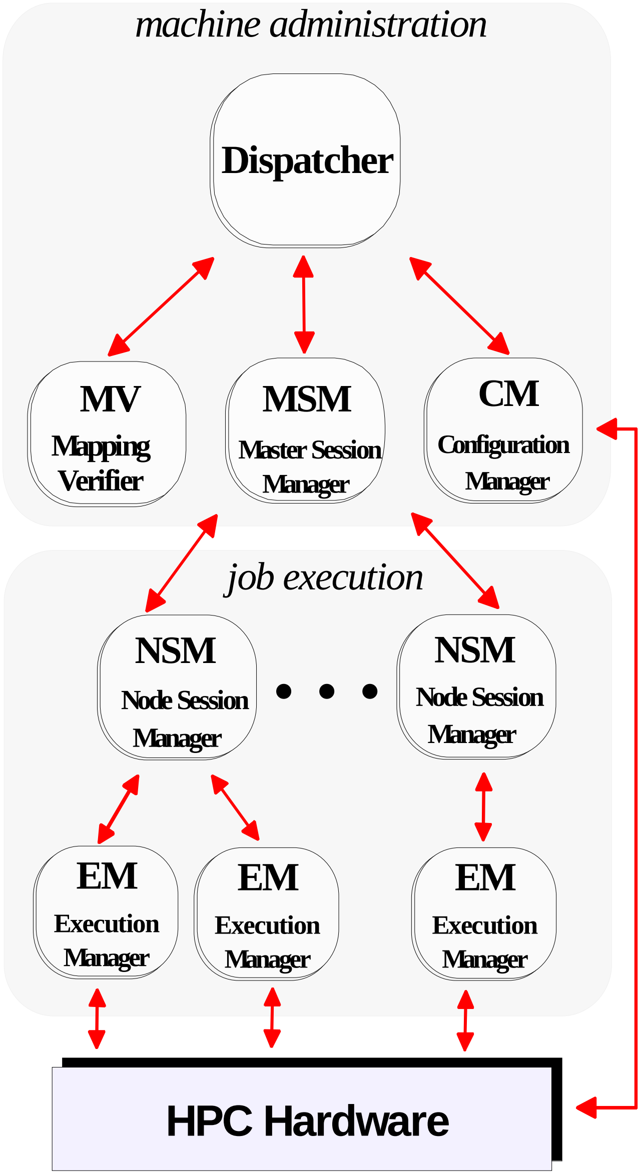
- Node checking
- Access / job control
- Worker specific pre / post processing
- Automatic disabling of defect nodes
- Monitoring tools

CCS Island Architecture



- **UI - User Interface**
  - *for X-Window- & ASCII access*
- **AM - Access Manager**
  - *for authorization, accounting*
- **QM - Queue Manager**
  - *for scheduling user requests*
- **MM - Machine Manager**
  - *for system control*
- **IM - Island Manager**
  - *name services, watchdog*
- **OS - Operator Shell**
  - *for operator interaction*

CCS Machine Manager Architecture



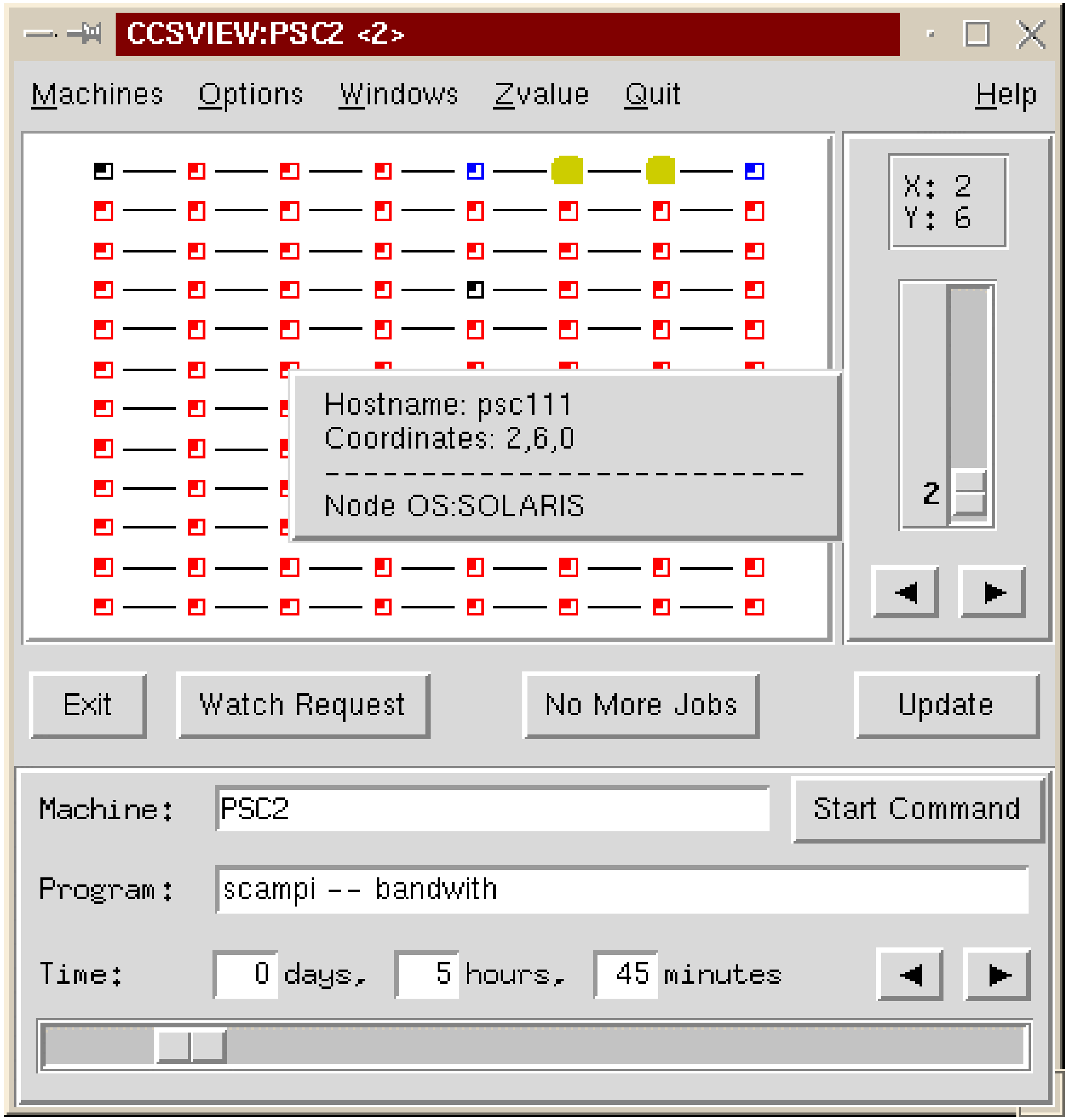
- **Dispatcher**
  - *event handling*
- **MV - Mapping Verifier**
  - *machine specific partitioning*
- **MSM - Master Session Manager**
  - *controls all jobs on the machine*
- **CM - Configuration Manager**
  - *interface to system OS*
- **NSM - Node Session Manager**
  - *controls all jobs on an entry node*
- **EM - Execution Manager**
  - *establishes user environment*

PC² PADERBORN CENTER FOR PARALLEL COMPUTING

CCS Commands

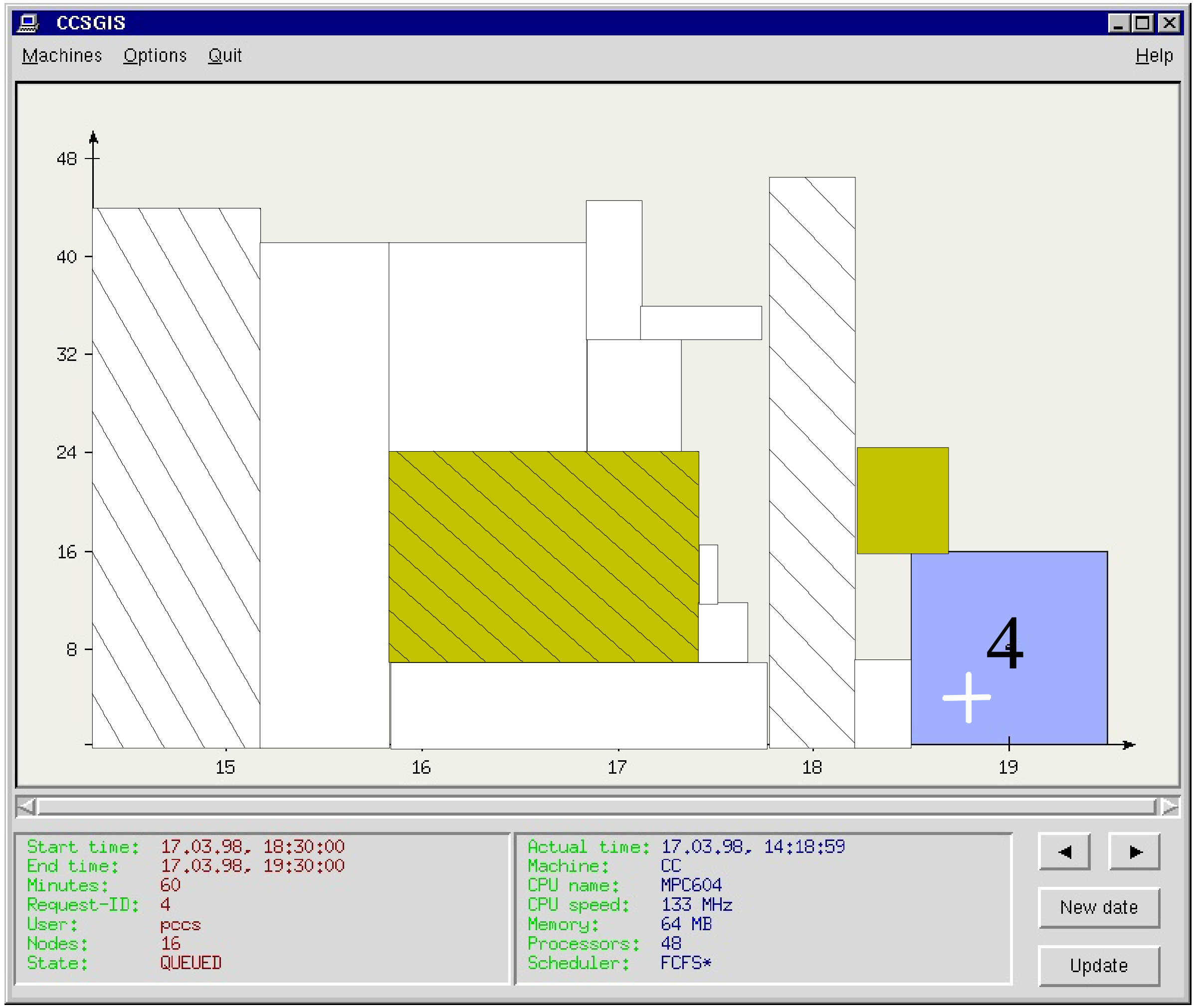
- **ccsalloc** — *allocate or reserve resources (or run an interactive job)*
  ccsalloc -m SCI_Cluster -p 64 -r 9:00:13.02.99 -t 2h
- **ccsrun** — *start a job on a previously reserved resource*
  ccsrun 7 pvm myjob
- **ccsbind** — *re-connect to a lost session (application)*
- **ccsinfo** — *information on job status, users, schedule, ...*
- **ccskill** — *kill a job, release resources*

CCS GUI

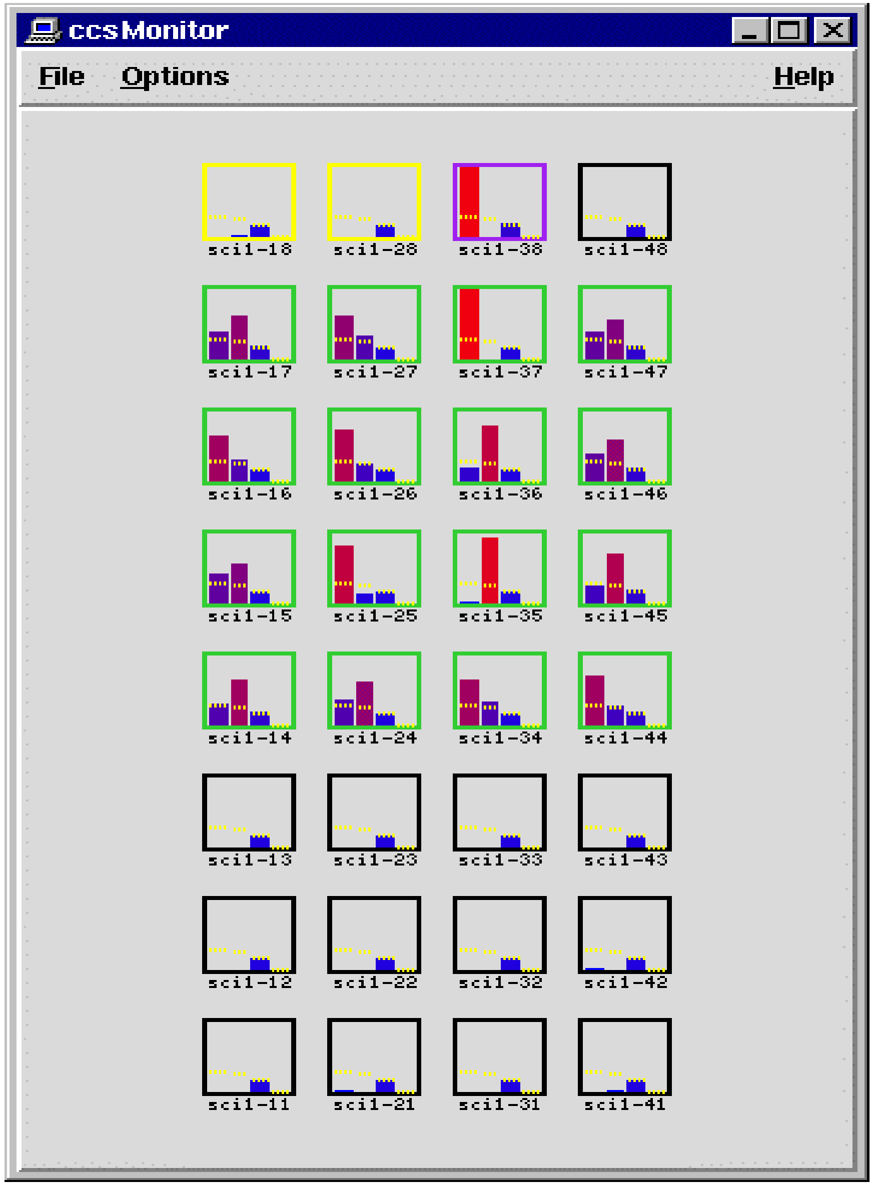


PC²
PADERBORN CENTER FOR PARALLEL COMPUTING

CCS Scheduler GUI

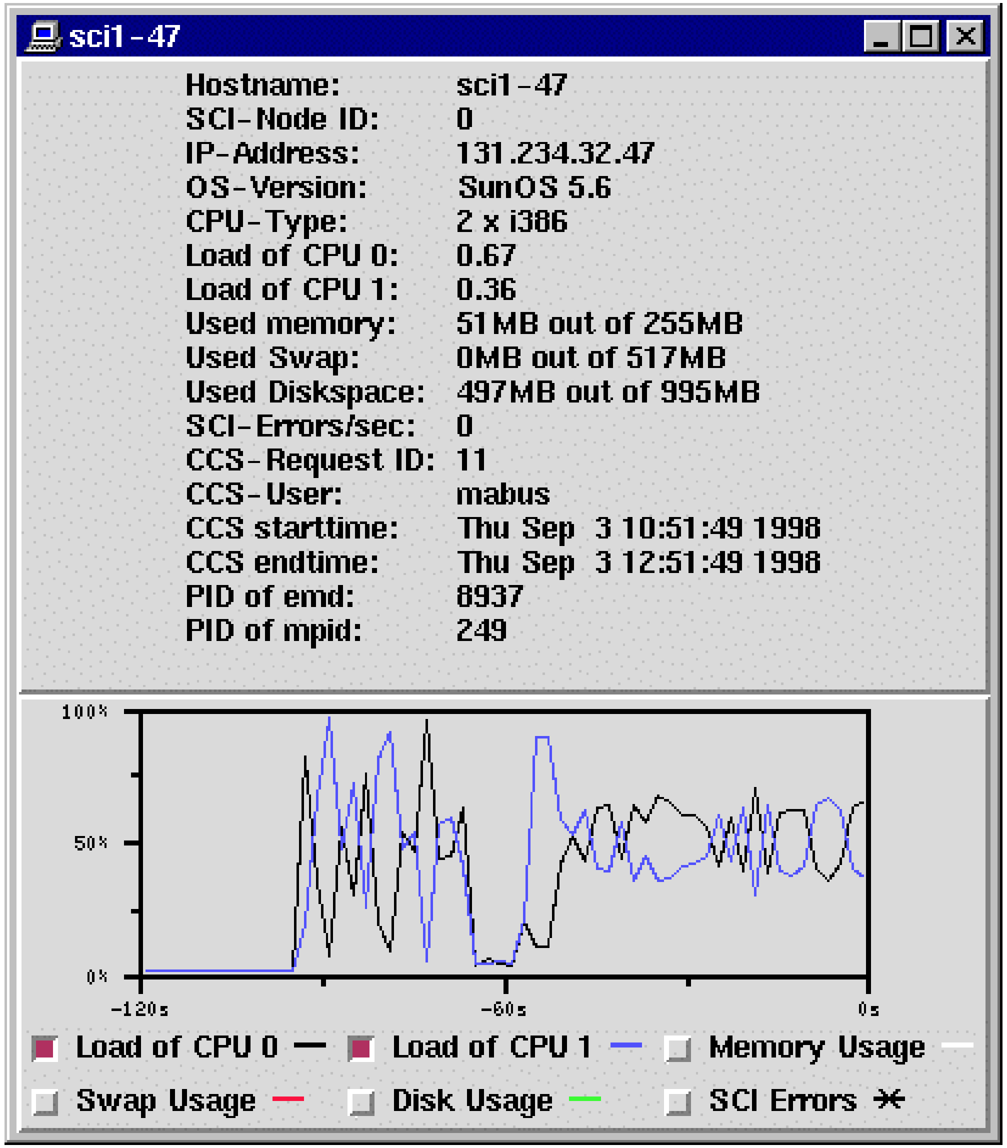


CCS Cluster Monitor



PADERBORN CENTER FOR PARALLEL COMPUTING

CCS Cluster Monitor

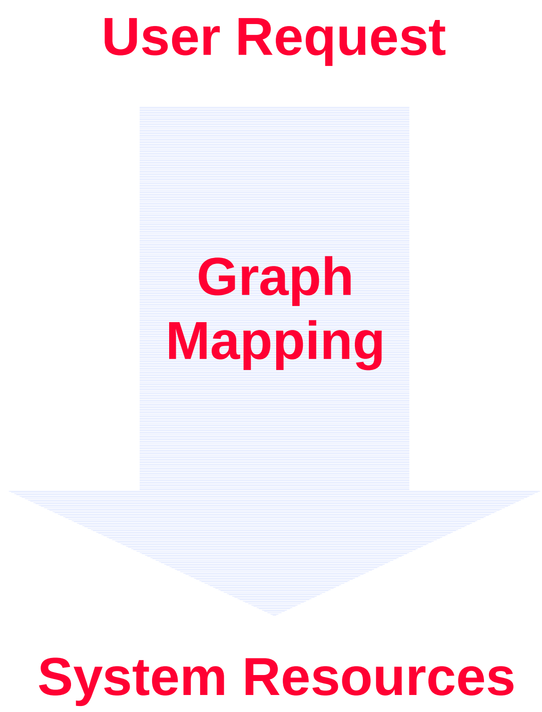


PADERBORN CENTER FOR PARALLEL COMPUTING

RSD - Resource and Service Description

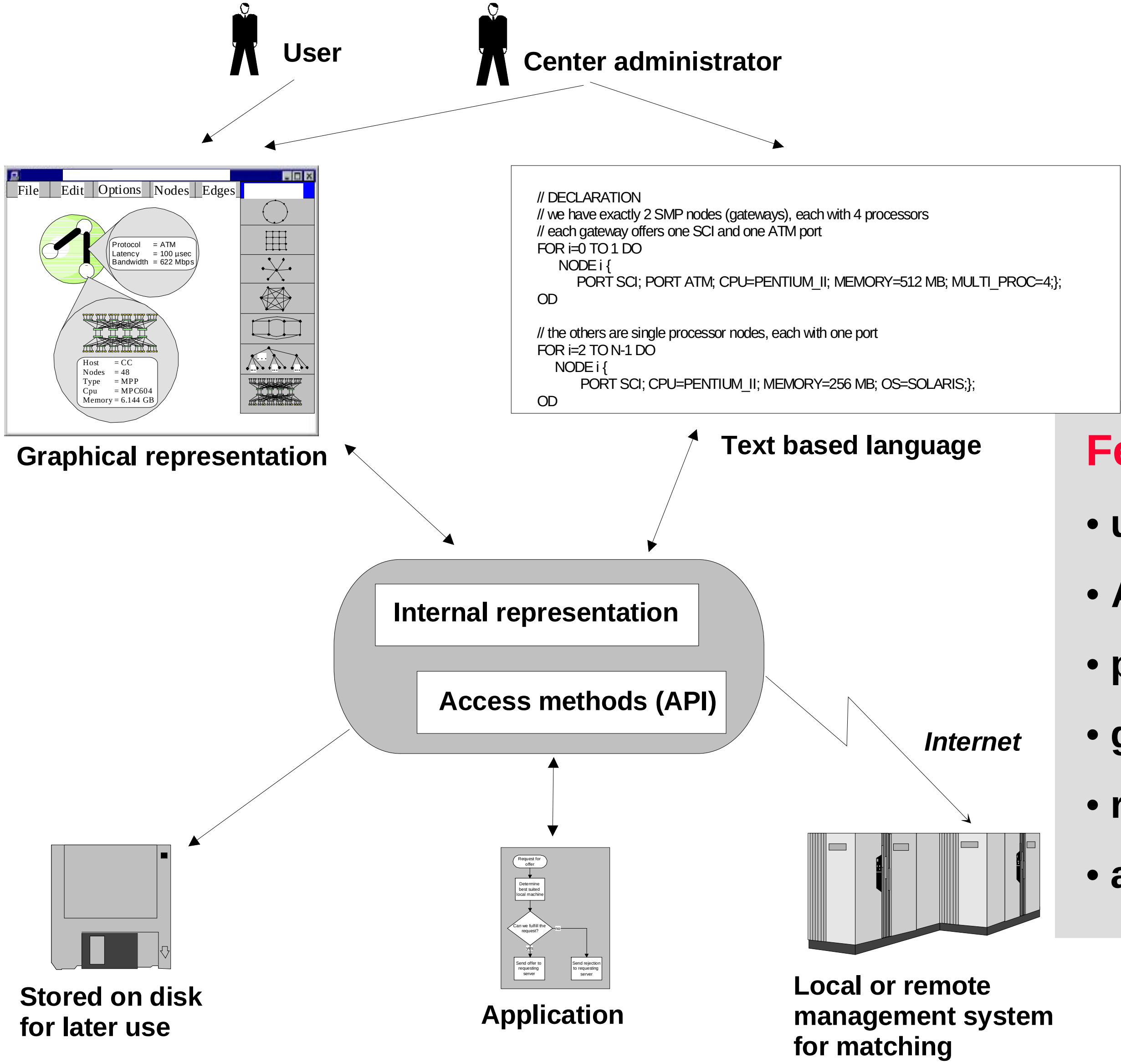
RSD deals with three levels:

- **user level:** description of a resource request (= task graph)
- **machine level:** description of internal structure (nodes, links, software, ...)
- **site level:** description of available systems, networks, & services



PADERBORN CENTER FOR PARALLEL COMPUTING

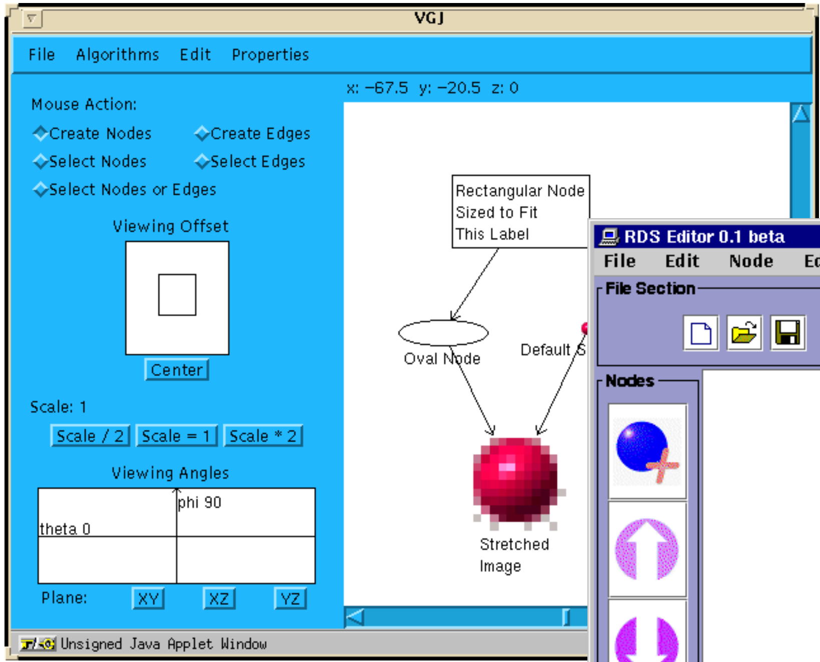
RSD - Resource and Service Description

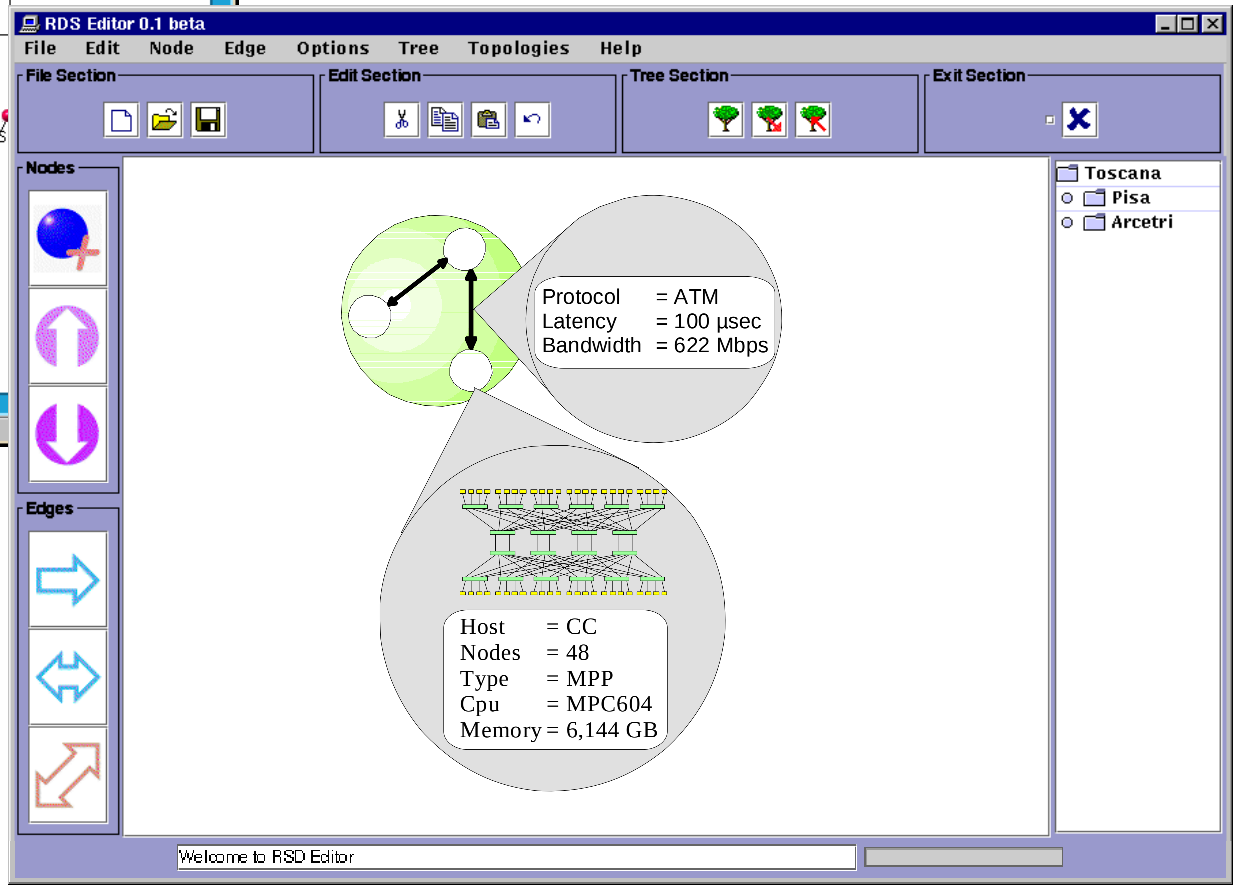


Features:

- user-friendly interface
- API
- portable
- graph based
- recursive structures
- attributed objects

RSD User Interface and Editor



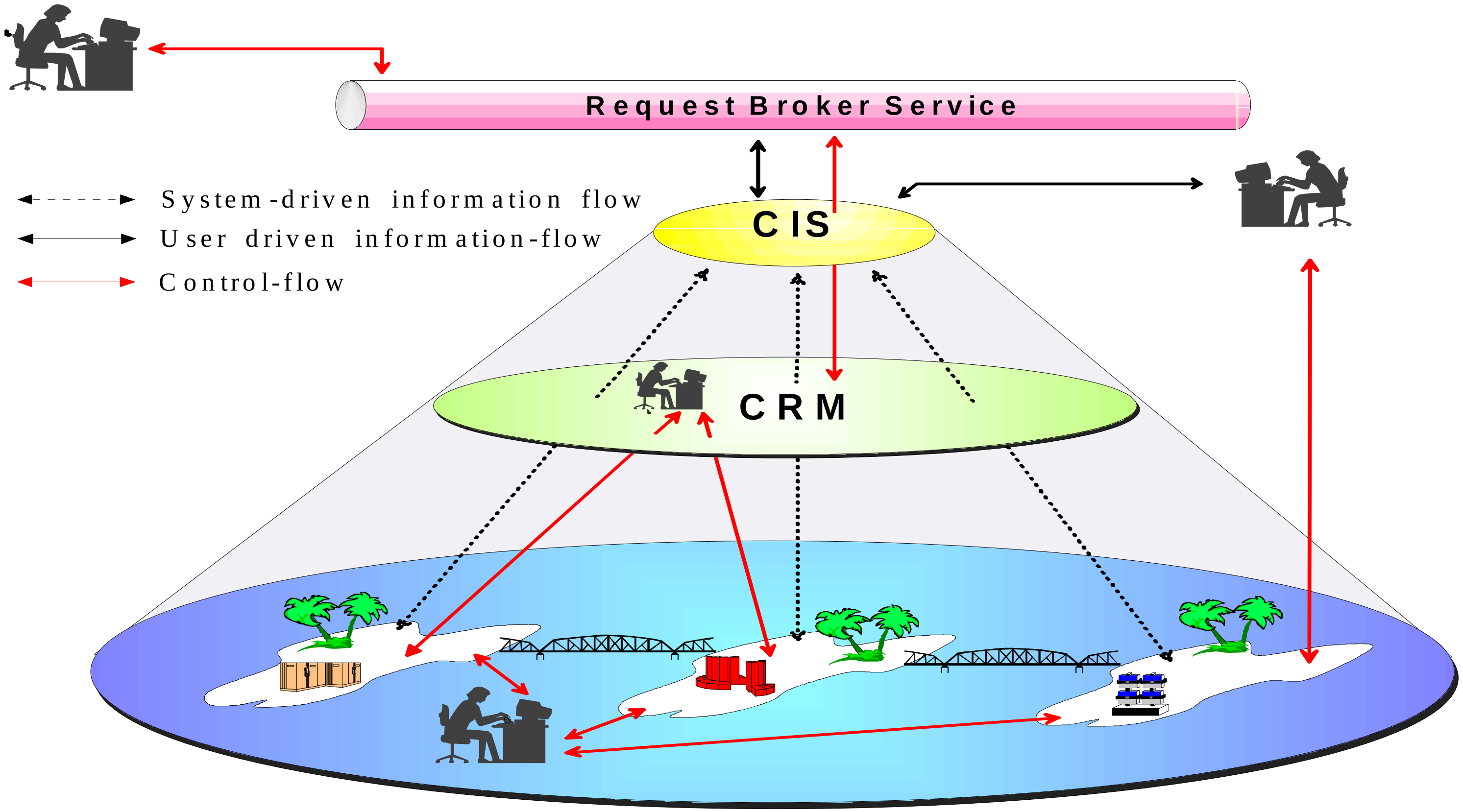


- based on the ideas in VGJ

  (visualizing graphs with Java)
- implemented in JavaSwing

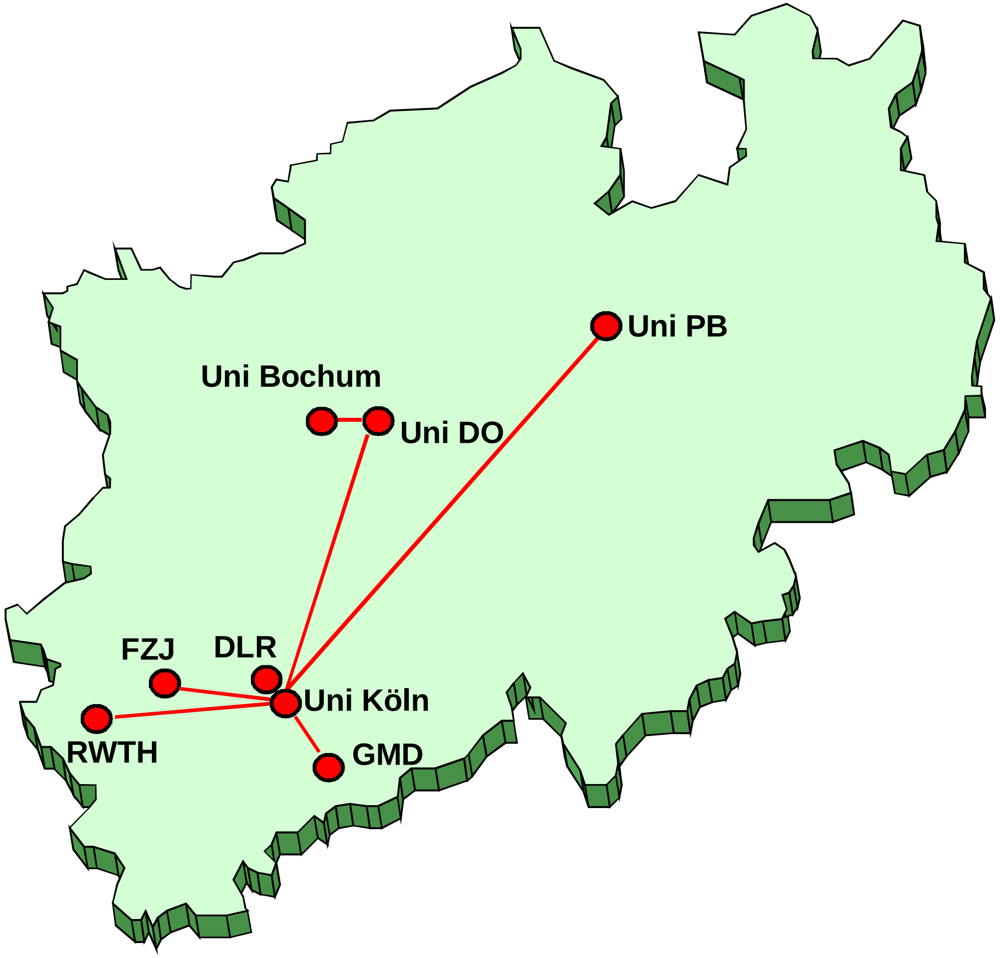
PC²
PADERBORN
CENTER FOR
PARALLEL
COMPUTING

Das Ziel



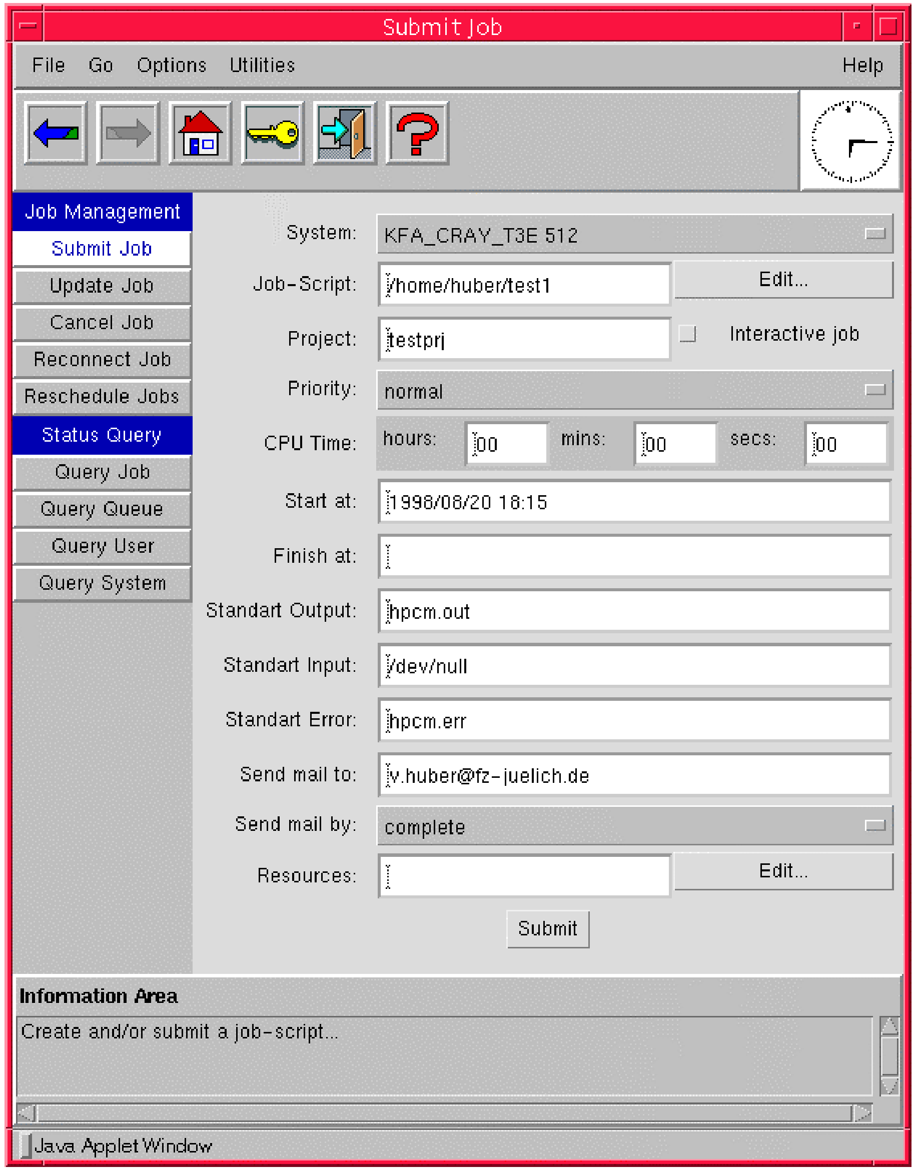
Dynamische Ressourcen Anforderung
Multi-site Applikationen
Migration
Kopplung an andere RMS

Forschungsverbund NRW-Metacomputing



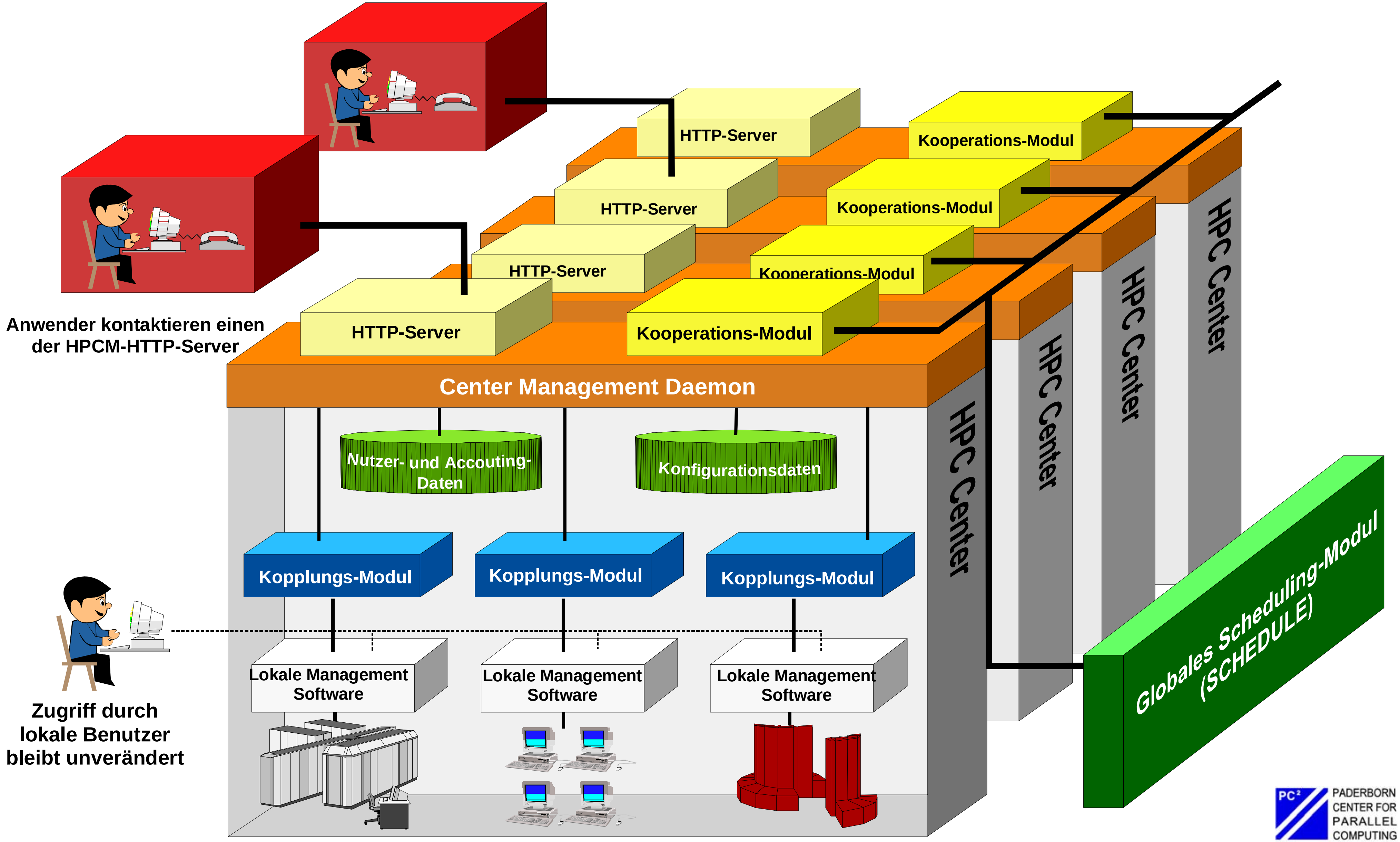
„...Bereitstellung einer Infrastruktur für Anwendungen des Metacomputings...“

Der NRW-Metacomputer



PC² PADERBORN CENTER FOR PARALLEL COMPUTING

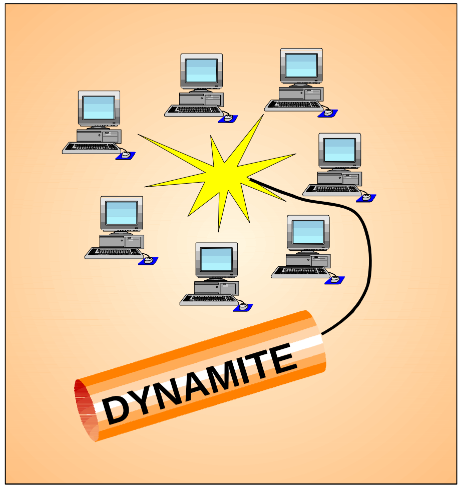
Aufbau des NRW-Metacomputers

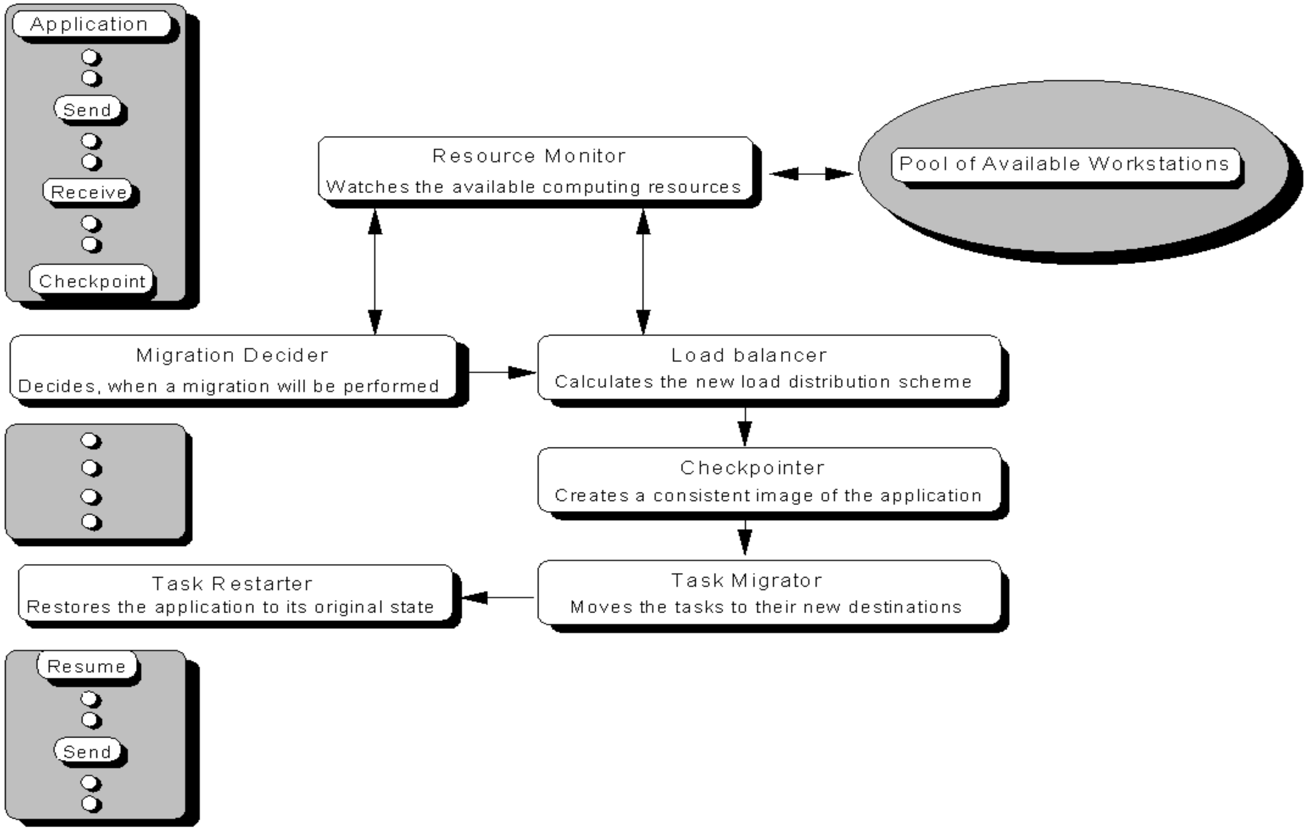


ESPRIT-Project “DYNAMITE”

Dynamic Task Migration Execution Environment

Partners:
ESI (France)
UvA (Netherlands)
GENIAS (Netherlands)
PC² (Germany)



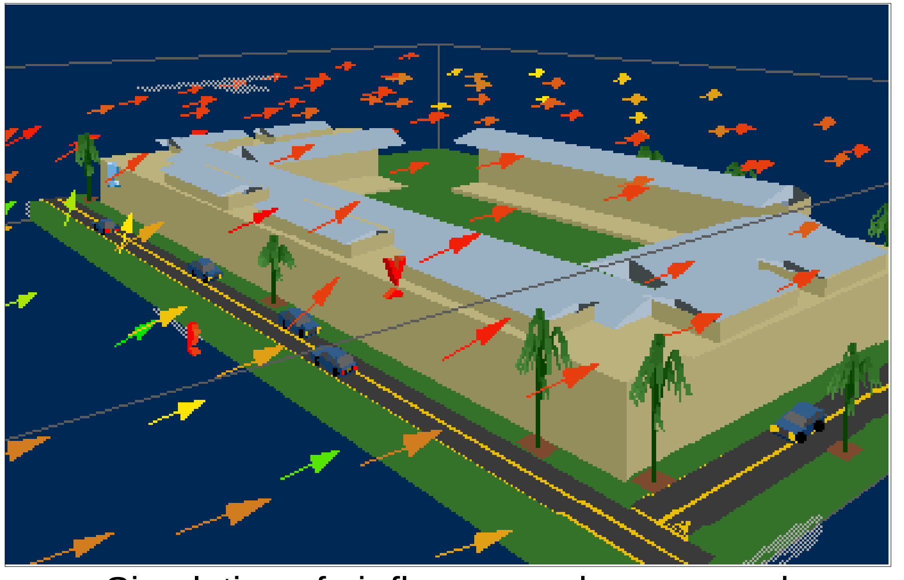


MICA

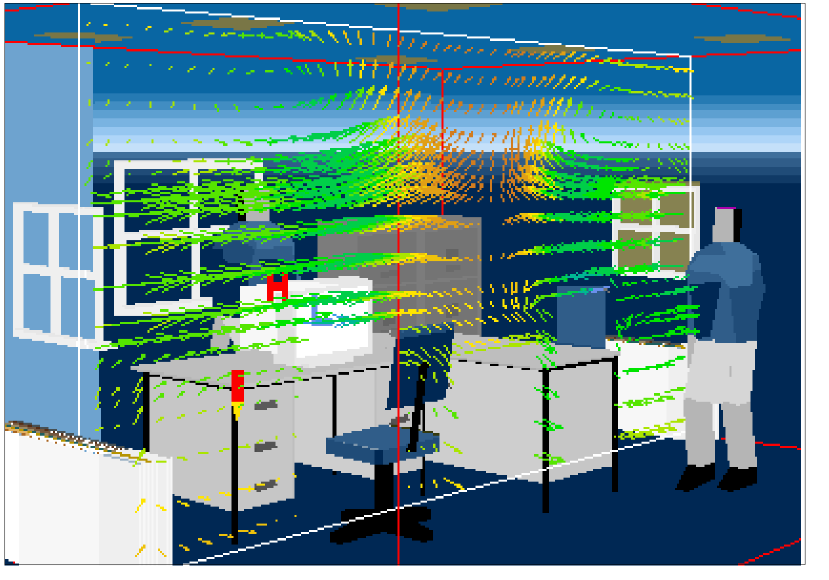
A Model for Industrial CFD Applications

http://www.uni-paderborn.de/pc2/projects/mica

PADERBORN CENTER FOR PARALLEL COMPUTING



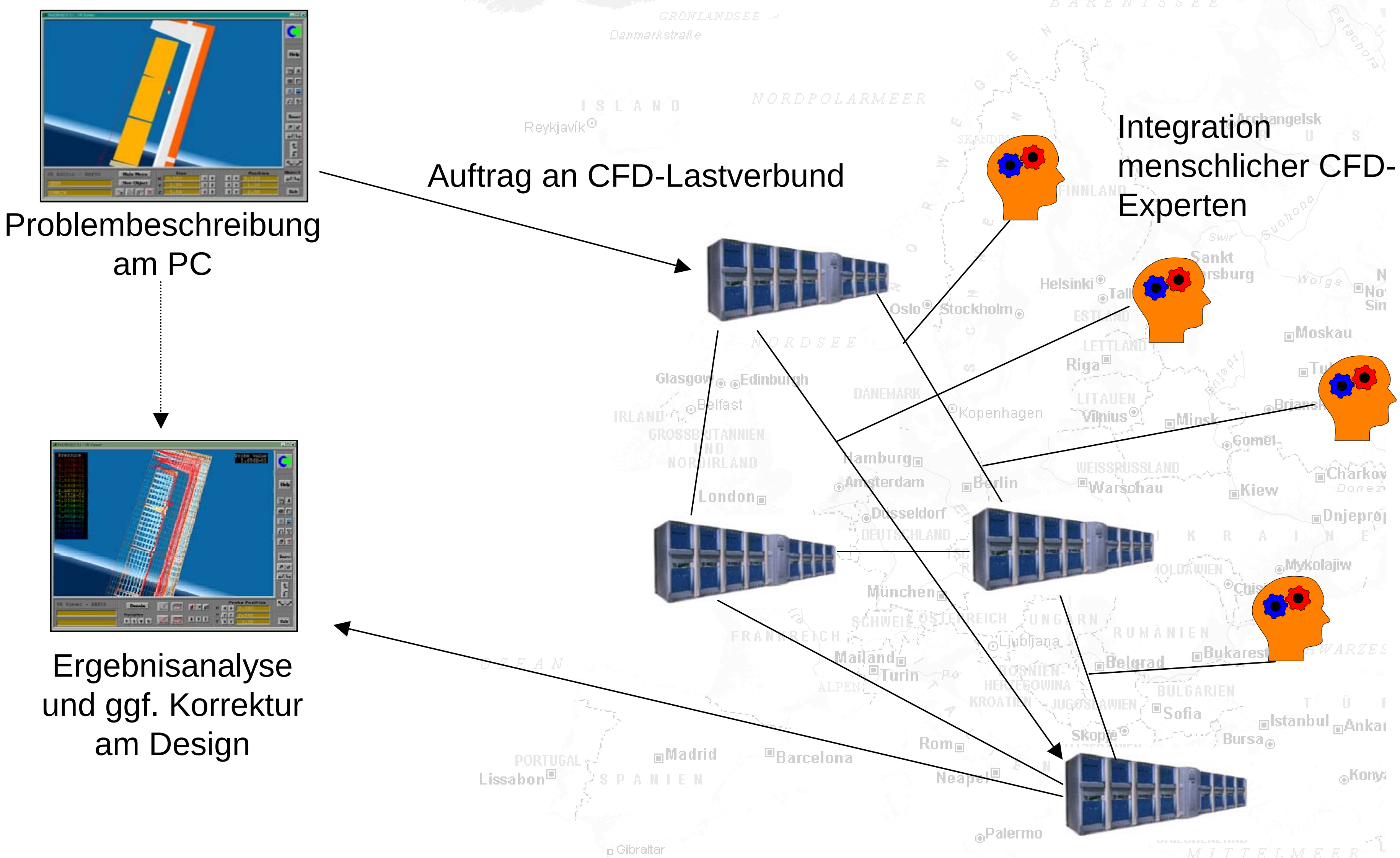
Simulation of air flow around a proposed stadium redevelopment



Air flow in a room

PADERBORN CENTER FOR PARALLEL COMPUTING

MICA: CFD für Endbenutzer



PC² PADERBORN CENTER FOR PARALLEL COMPUTING

ESPRIT 23486

PHASE

Distributed Pharmaceutical Applications Server

http://www.uni-paderborn.de/pc2/projects/phase

PC²
PADERBORN CENTER FOR PARALLEL COMPUTING

PHASE - Project Overview

ESPRIT 23486

Novo Nordisk

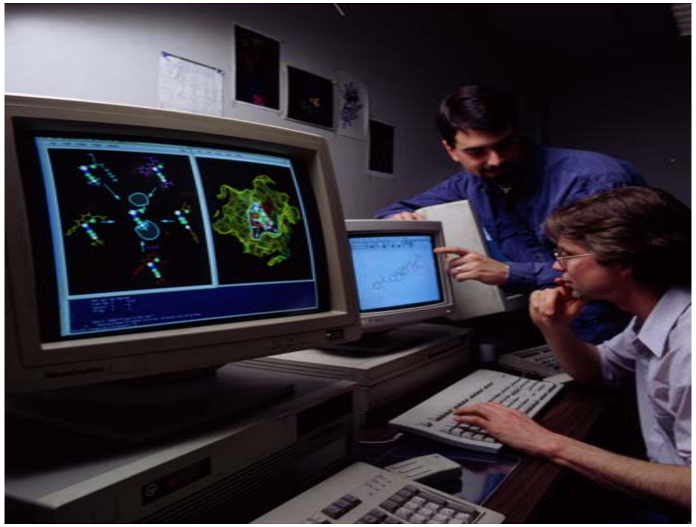
MERCK

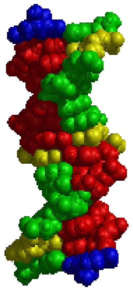
PC²

MRC Medical Research Council

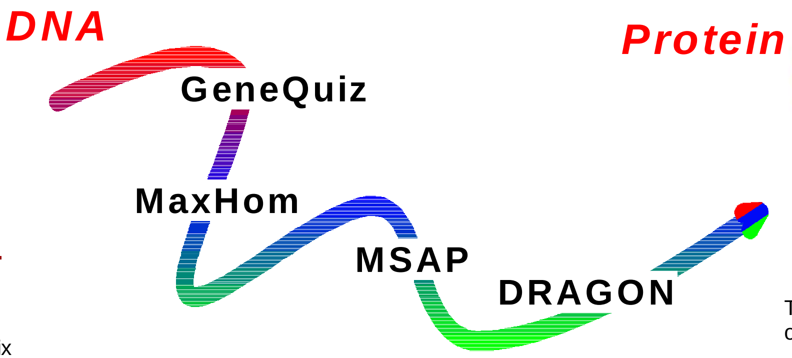
EMBL – European Bioinformatics Institute

Oxford Molecular
Solutions for Discovery Research





DNA Double Helix

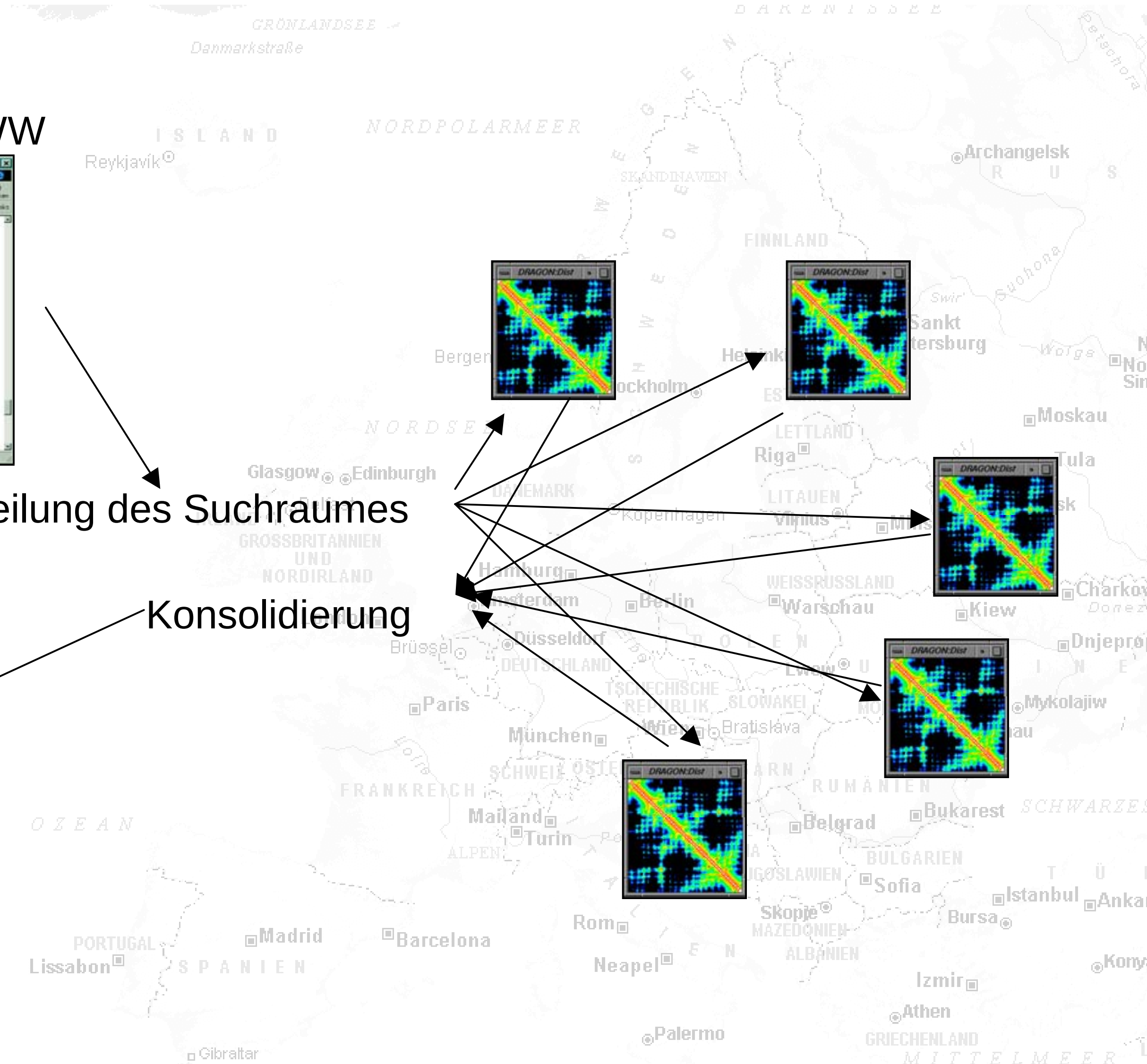


The 3-dimensional structure of a protein,Beta-amylase.

Interactive
Drug Target Design

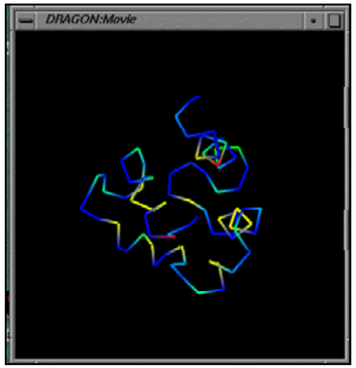
Beispiel: Protein-Modellierung

Eingabe der Rohdaten über WWW



Aufteilung des Suchraumes

Konsolidierung



Vorhergesagte Struktur

PC² PADERBORN CENTER FOR PARALLEL COMPUTING

UNICORE

Uniform Interface to Computer Resources

- Unifying access to HPC resources
- No metacomputing aspect
- Advanced security concept
- Support for “Chain-Jobs”
- May be combined with “NRW-Metacomputing”
  - *The metacomputer is accessed through UNICORE...*
  - *...and managed by NRW-Metacomputing*

UNICORE

Forschungszentrum Jülich

Scientific Supercomputing Center Karlsruhe

ECMWF

RUS

HLRS

GENIAS Software

PC²

PADERBORN CENTER FOR PARALLEL COMPUTING

lrz

ZIB

FUJITSU

NEC

PALLAS

SiliconGraphics

SIEMENS NIXDORF

IBM®

DWD

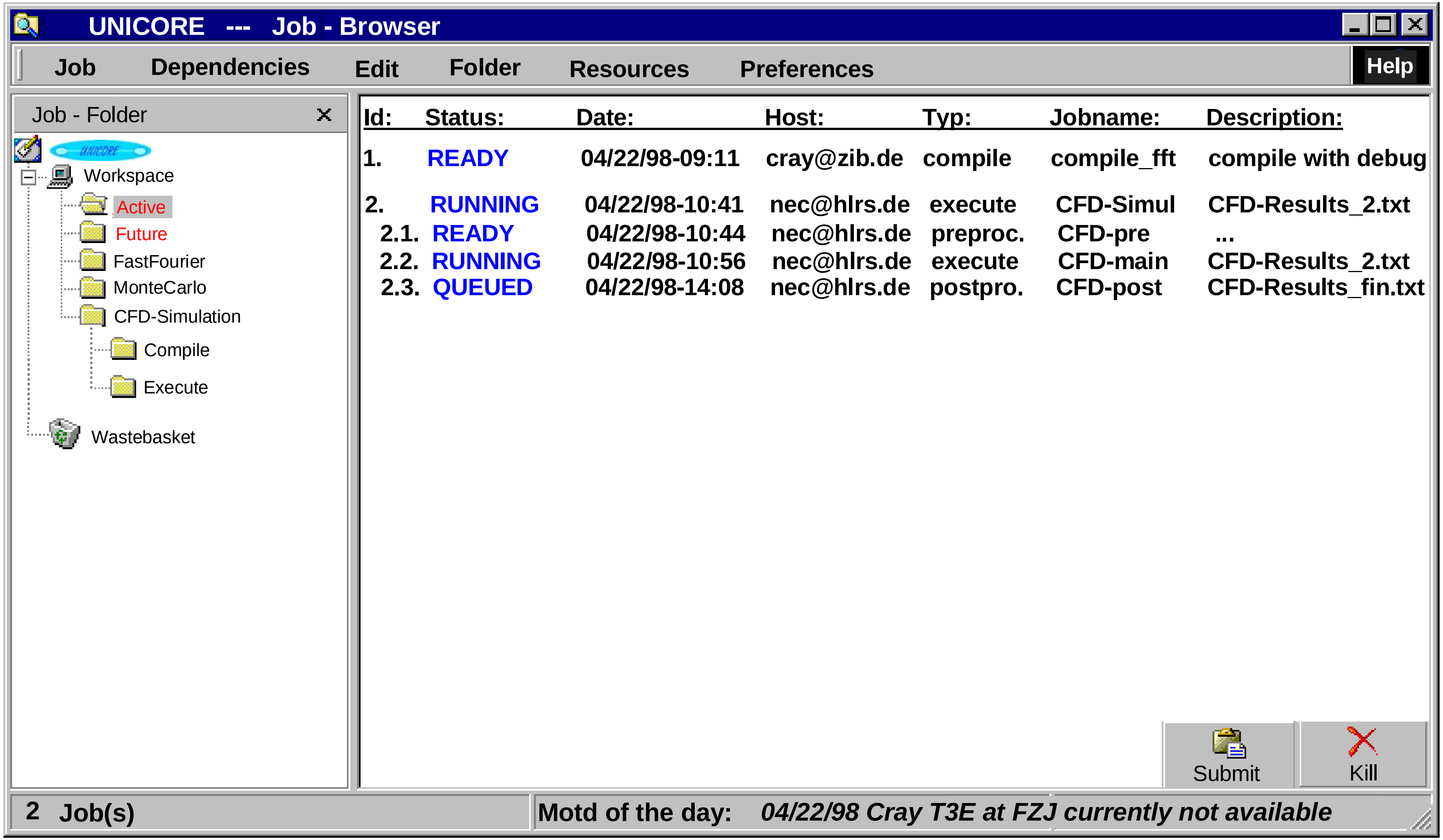
Deutscher Wetterdienst

CRAY RESEARCH

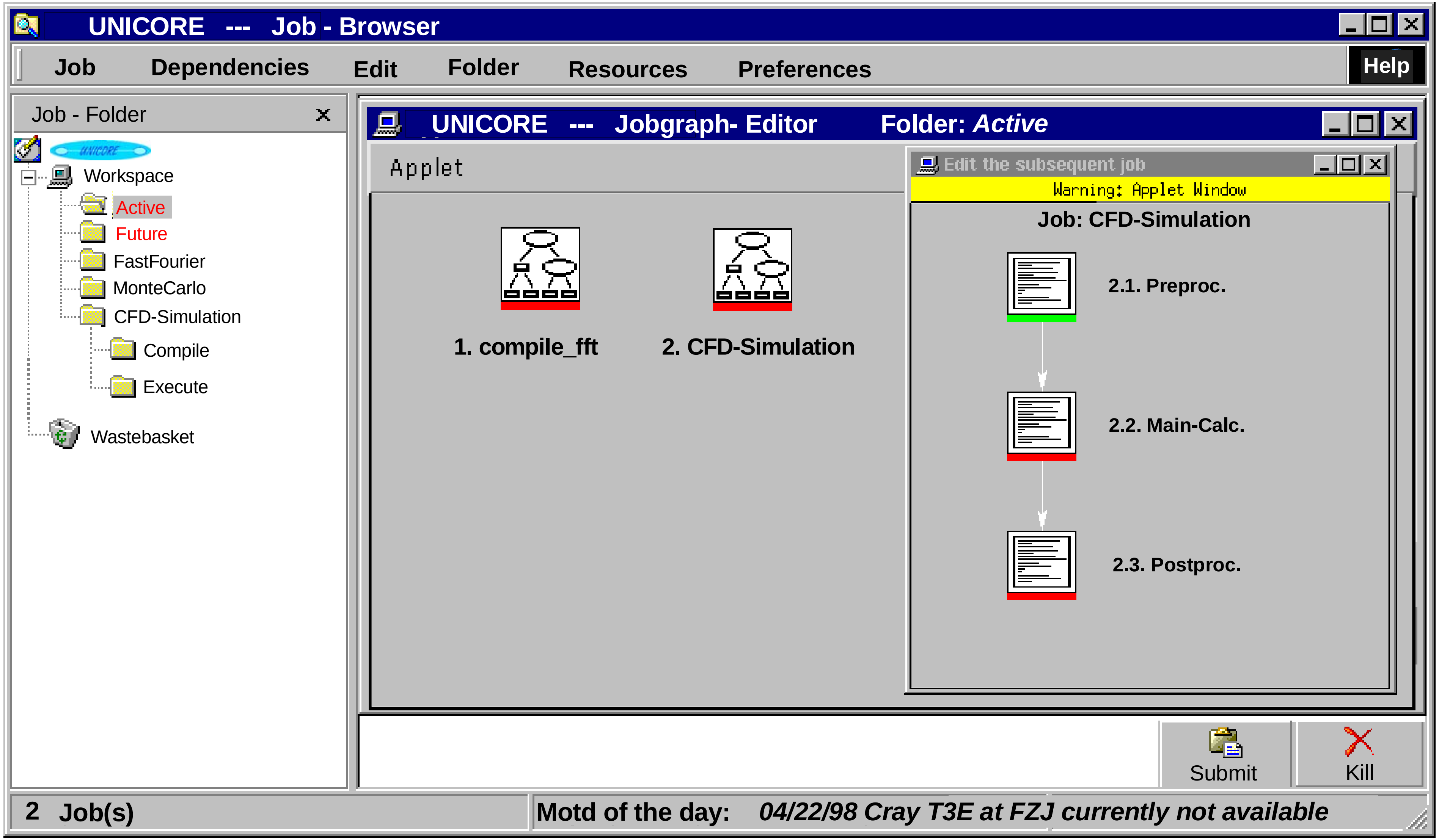
A Silicon Graphics Company

PADERBORN CENTER FOR PARALLEL COMPUTING

UNICORE - Job Browser



UNICORE - Jobgraph Editor



PADERBORN CENTER FOR PARALLEL COMPUTING

4. PC²: Wie geht‘s weiter ?

Tendenzen

Von eng gekoppelten parallelen Systemen

zu verteilten und lose gekoppelten Systemen

- Paralleles und verteiltes Rechnen konvergieren
- Neue Themen durch Internet Computing
- Methoden und Anwendungen werden vielfältiger
  - *nicht numerische Anwendungen werden wichtiger, OR, DB, MM*
- Internet2: Internet der nächsten Generation
  - *Gigabit Technologie, Quality of Services*

=> Konzepte des Parallel Computing sind relevant für Internet2 Systeme und Dienste:

- *Datenallokation (Replikation, Caching, Migration, ...)*
- *Kommunikationsscheduling auf Netzwerken*

Konsequenzen

- Integrierte Systeme: HPC und MM Information und Kommunikation
  - Zusammenführung bislang getrennter Systeme durch Internet Technologien
- Vom zentralen HPC zum vernetzten Computing
  - Rechenleistung on-demand und aus der „intelligenten Steckdose“
  - Rechenleistung gekoppelt mit Anwendungsleistung
- **Schlagwörter:**
  - Information Power Grid:
    - Integration von HPC- und Informationsservern
    - Transparenter Zugriff auf Systeme durch Internet Technologien
  - Science Portals:
    - Kopplung von Mechanismen zur Information, Kommunikation und HPC Anwendungen für einen Wissenschaftsbereich
    - „community builder“ für die Wissenschaft

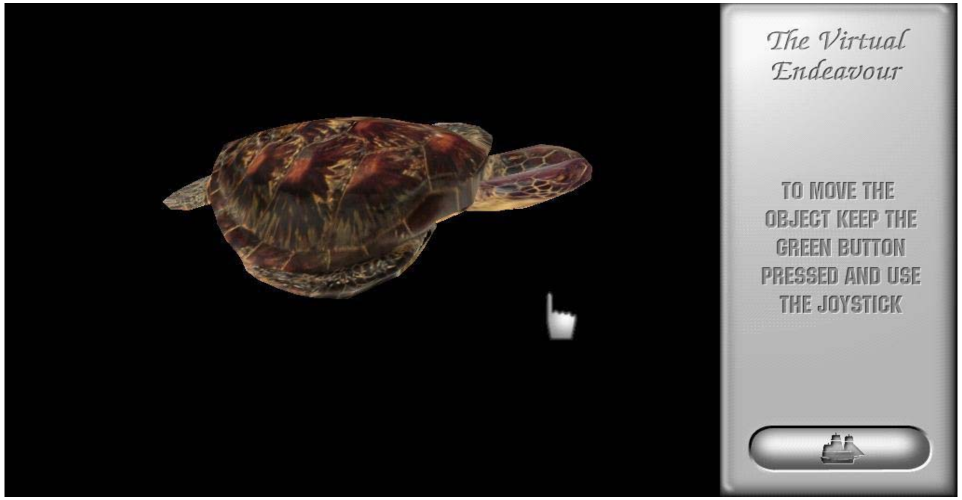
Beispiel 1: SICMA

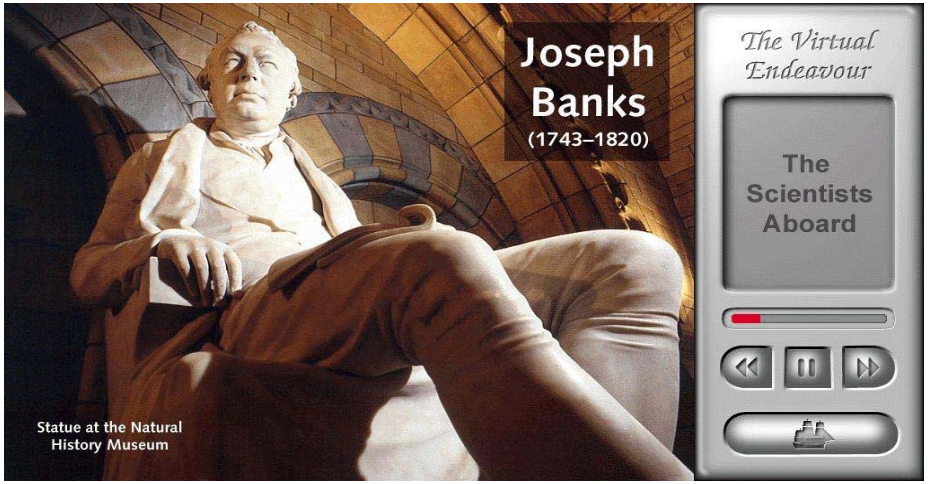
Scalable interactive continous media servers design and application

- SICMA Anwendung ‚Virtuelles Museum'
  - *VR System von Captain Cooks Endeavour*
  - *Integration von Audio/Video/Graphik*

PADERBORN CENTER FOR PARALLEL COMPUTING

Parallele Mediaserver - SICMA

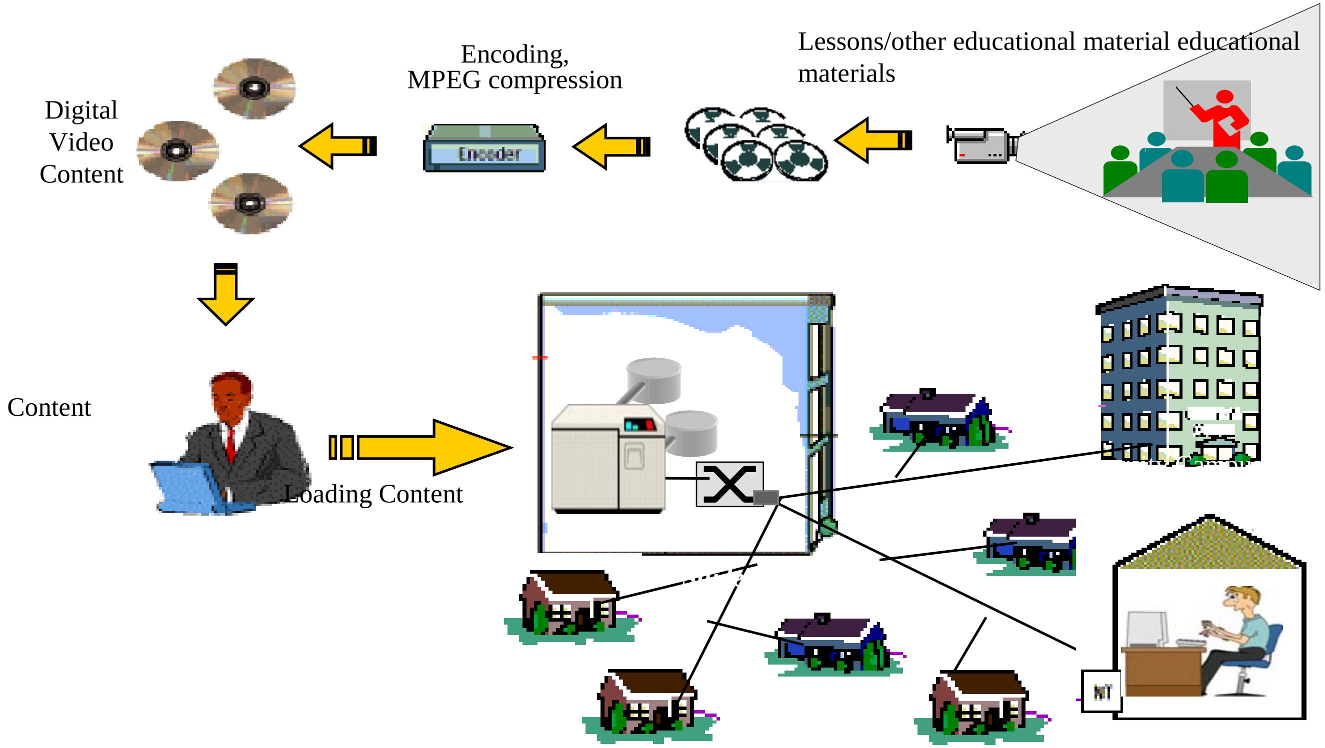




PC²
PADERBORN
CENTER FOR
PARALLEL
COMPUTING

Parallele Mediaserver - SICMA

- SICMA Anwendung ‚Teleteaching'
  - Teleteaching System in Limburg University
  - Auslieferung an Clients in Uni und über ADSL



Beispiel 2: HiQoS

High Performance Multimedia with Quality of Service

- Pixelpark: Pixelvision corporate information system
  - *Vorträge, Projekt Reviews, Meetings, ... werden mit Kameras aufgenommen und zeitunabhängig interessierten Mitarbeitern zur Verfügung gestellt*
  - *Zeitsynchrone Verknüpfung von Audio/Video mit Web Inhalten, Powerpoint Presentation, ...*
- Computer Based Training (CBT)
  - *Business TV auf firmeninternen Intranets*
- Rendering Services
  - *für die Visualisierung von Bauentwürfen*
  - *zur Generierung von Animations- und Videosequenzen in einer MM-Produktionsumgebung*

Beispiel 3: EU Projekt EPRI-COM

- Realisierung eines EU Informationsdienstes für Parlamentarier osteuropäischer Parlamente

- Lokale Mirror Server in Parlamenten
- Zentraler Content Server in Brüssel
- Verteilung der Media Assets über Satellit

Das PC²

Gestern: Die effiziente Nutzung massiv paraller Systeme

Morgen: Parallel Computing Know How für das Internet ?

PADERBORN
CENTER FOR
PARALLEL
COMPUTING