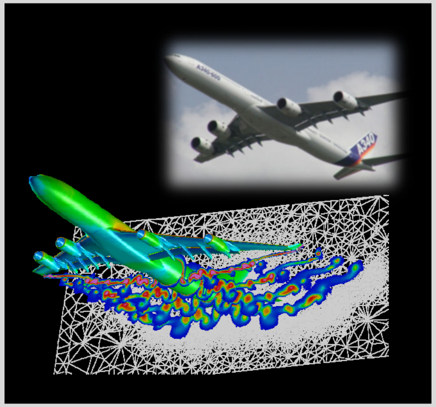


High Performance Computing

ADVANCED SCIENTIFIC COMPUTING

Prof. Dr. – Ing. Morris Riedel
Adjunct Associated Professor
School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland
Research Group Leader, Juelich Supercomputing Centre, Forschungszentrum Juelich, Germany

@Morris Riedel @MorrisRiedel @MorrisRiedel

SHORT LECTURE 11

Scientific Visualization & Scalable Infrastructures

November 18, 2019
Webinar

UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES
FACULTY OF INDUSTRIAL ENGINEERING, MECHANICAL ENGINEERING AND COMPUTER SCIENCE

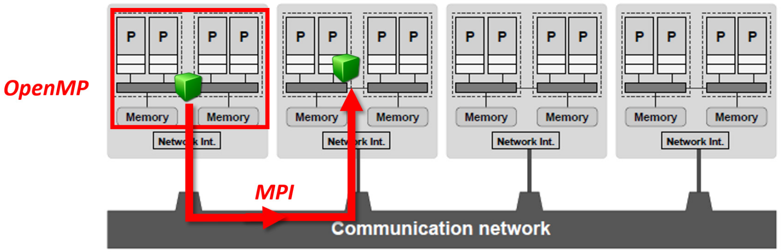
JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

DEEP Projects

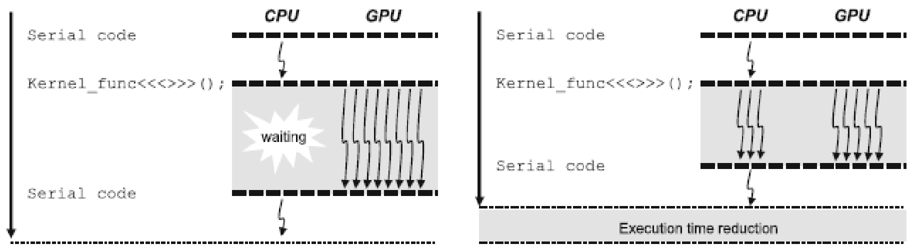
HELMHOLTZ RESEARCH FOR GRAND CHALLENGES

HAICU HELMHOLTZ ARTIFICIAL INTELLIGENCE COOPERATION UNIT

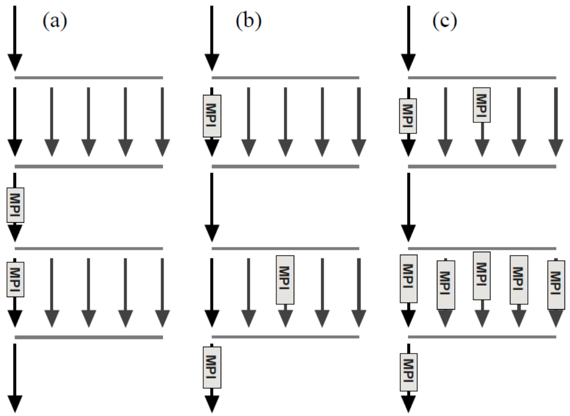
Review of Lecture 10 – Hybrid Programming & Patterns



(traditional hybrid programming using OpenMP & MPI)

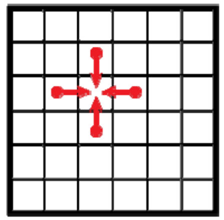


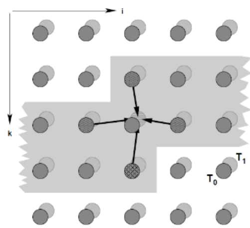
(emerging hybrid programming using CPU & GPU at the same time)

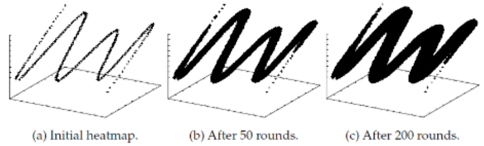


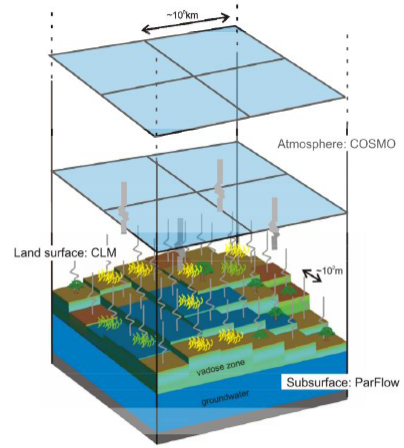
(vector mode recommended for beginners)

	Vector mode	Task mode
Improved/easier load balancing	✔	✔
Additional levels of parallelism	✔	✔
Reliable overlapping of communication and computation	✘	✔
Improved rate of convergence	✔	✔
Re-use of data in shared caches	✔	✔
Reduced MPI overhead	✔	✔





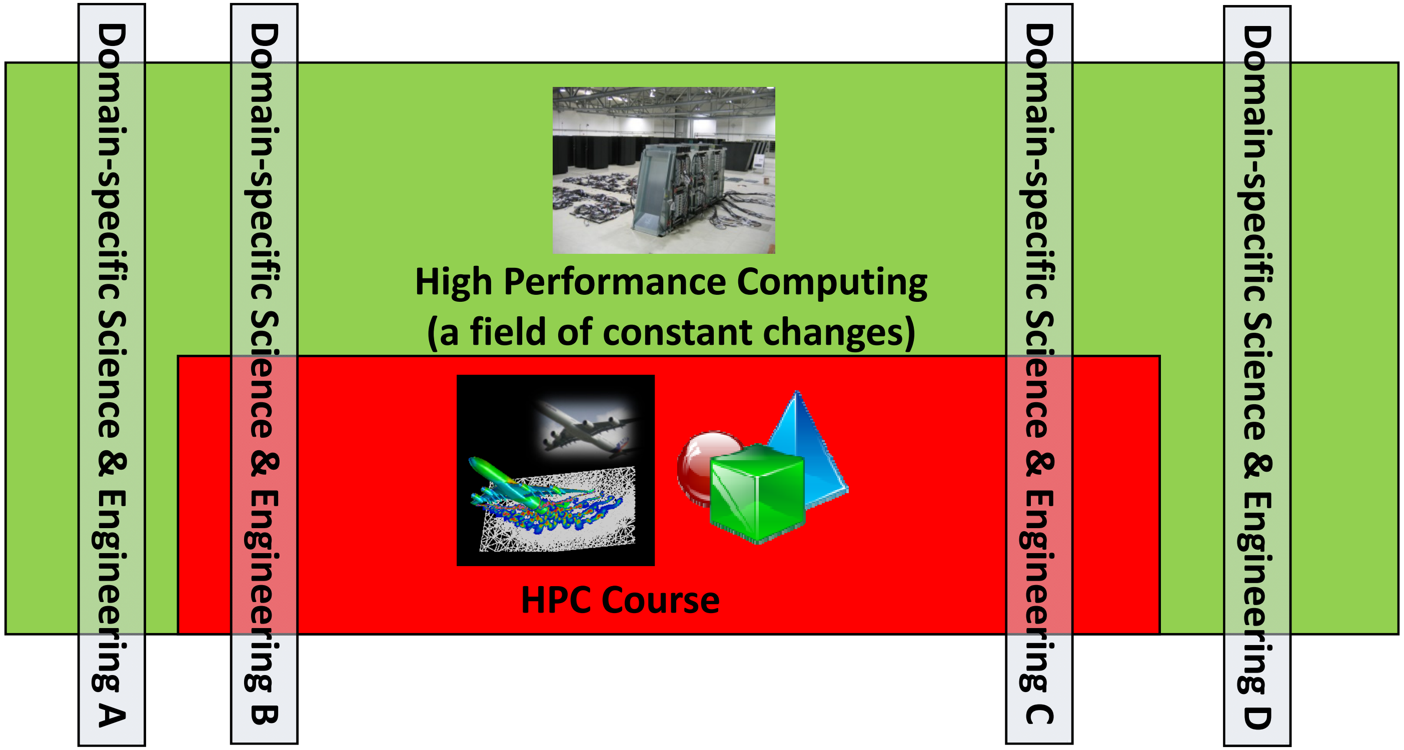




[1] Introduction to High Performance Computing for Scientists and Engineers [2] G. Hager [3] 'Boosting CUDA Applications with CPU-GPU Hybrid Computing' [4] CLM Web page

HPC-A[dvanced] Scientific Computing (cf. Prologue) – Second Part

- Consists of techniques for programming & using large-scale HPC Systems
  - Approach: Get a broad understanding what HPC is and what can be done
  - Goal: Train general HPC techniques and systems and selected details of domain-specific applications



Outline of the Course

1. High Performance Computing
2. Parallel Programming with MPI
3. Parallelization Fundamentals
4. Advanced MPI Techniques
5. Parallel Algorithms & Data Structures
6. Parallel Programming with OpenMP
7. Graphical Processing Units (GPUs)
8. Parallel & Scalable Machine & Deep Learning
9. Debugging & Profiling & Performance Toolsets
10. Hybrid Programming & Patterns
11. Scientific Visualization & Scalable Infrastructures
12. Terrestrial Systems & Climate
13. Systems Biology & Bioinformatics
14. Molecular Systems & Libraries
15. Computational Fluid Dynamics & Finite Elements
16. Epilogue

+ additional practical lectures & Webinars for our hands-on assignments in context

- Practical Topics
- Theoretical / Conceptual Topics

Outline

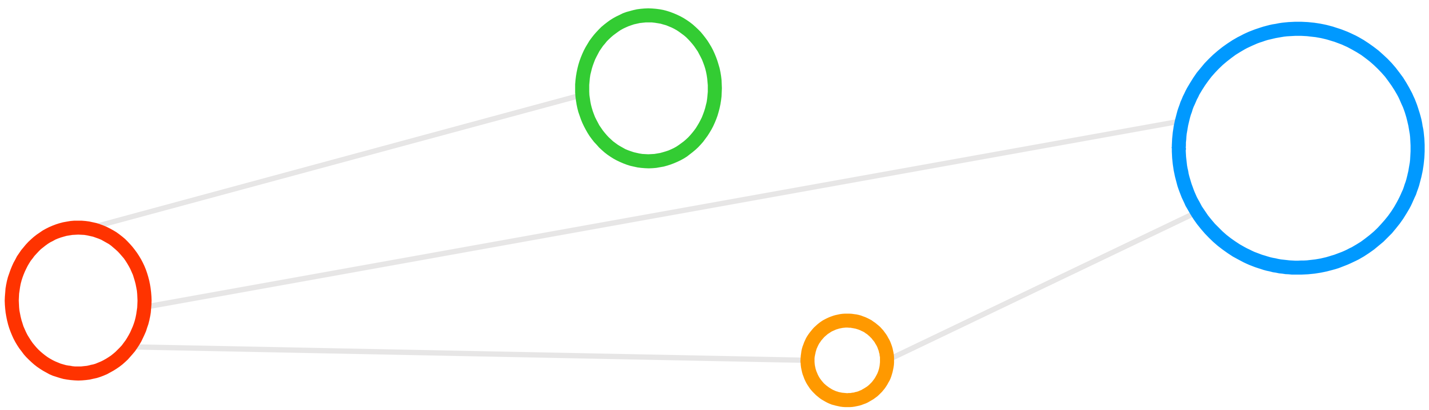
- Scientific Visualization & Computational Steering
  - Motivation & Objectives for Visualizations in Scientific Computing
  - Understanding different HPC Simulation Data & Data Types
  - Selected Visualization Tools & Technologies arranged in a Stack
  - Multi-scale Visualization & Interactive HPC with JupyterHub
  - Computational Steering & Scientific Wave Application Example

- Scalable Infrastructures
  - Large Scale HPC Infrastructures & PRACE Research Infrastructure
  - e-Science & Grid Computing Infrastructures for Resource Sharing
  - Single System View with Grid Middleware Systems & UNICORE Example
  - Scientific Workflows & Ice Dynamic Model Application Example
  - Collaborative Data & Cloud Infrastructures

- **Promises from previous lecture(s):**
- ***Lecture 1 & 5:*** **Lecture 11 will give in-depth details on scalable approaches in large-scale HPC infrastructures and how to use them with middleware**

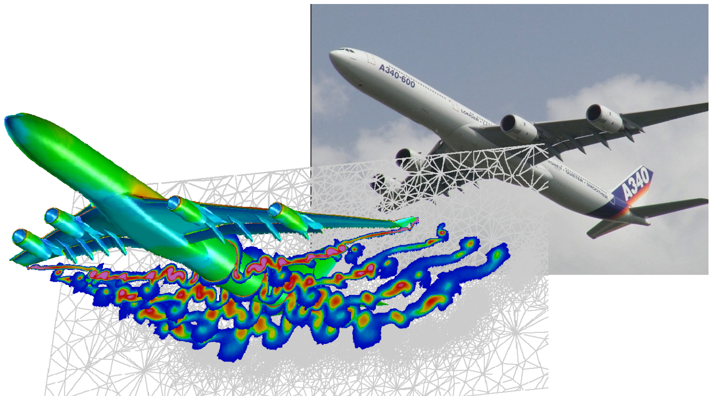
- **Note that this lecture is only a short lecture that usually needs a full course**
- **The goal is to understand selected HPC application fields & provide a few pointers to other advanced related university courses/topics/tutorials**

Scientific Visualization & Computational Steering

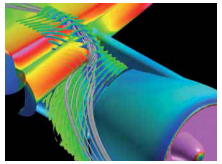


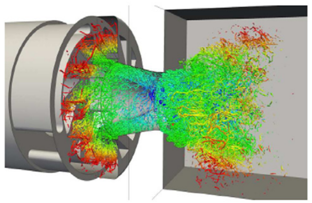
Scientific Visualization – Motivation

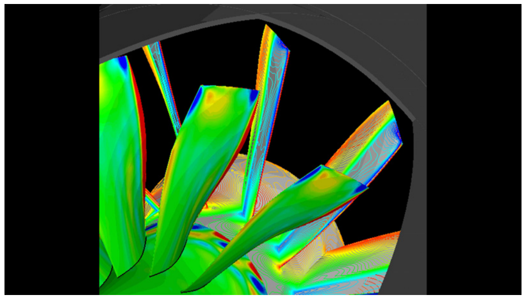
- Graphically illustrate scientific data based on HPC simulations
  - Enable scientists to understand and glean insights from simulation data & machine learning datasets
  - Provides feedback for simulation models derived from known physical laws or data measurements
  - Mostly applied in a post-processing step in simulations or visualization during run-time of simulations



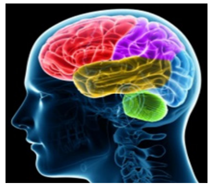
('a picture is worth 1000 ~~words~~ numbers')

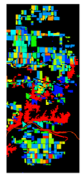


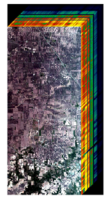


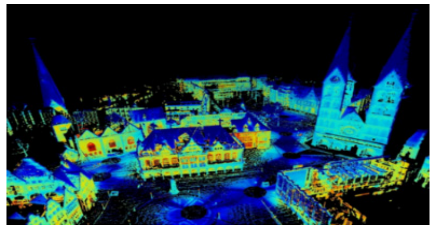


(visualization over time is crucial)







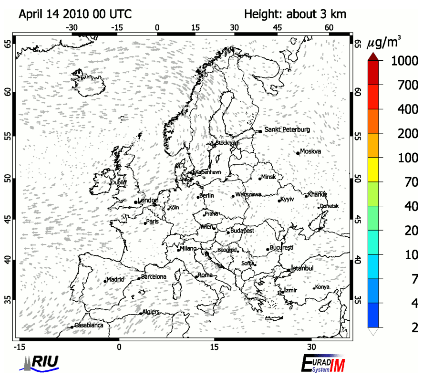


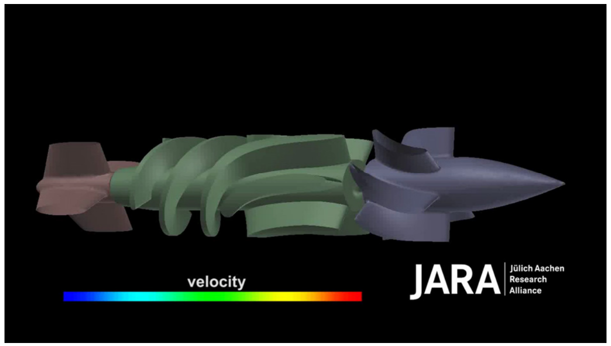
- **Scientific Visualization is an interdisciplinary branch of science and a research field of its own**
- **It is primarily concerned with the visualization of multi-dimensional phenomena where the emphasis is on realistic renderings of volumes, surfaces, etc. with a dynamic time component**

modified from [5] 'Scientific Visualization', Wikipedia

Scientific Visualization – Objectives in HPC

- Selected objectives
  - Analyze data and explore information: Easier to get patterns, regularity, and associations
  - Reduces time to understand complex data
  - Improve comprehension of complex phenomena and processes
  - Find new meanings and interpretations
  - Make visible the invisible
  - Check quality of HPC simulations & measures
  - Make effective presentation of information and results (brief & efficiently)



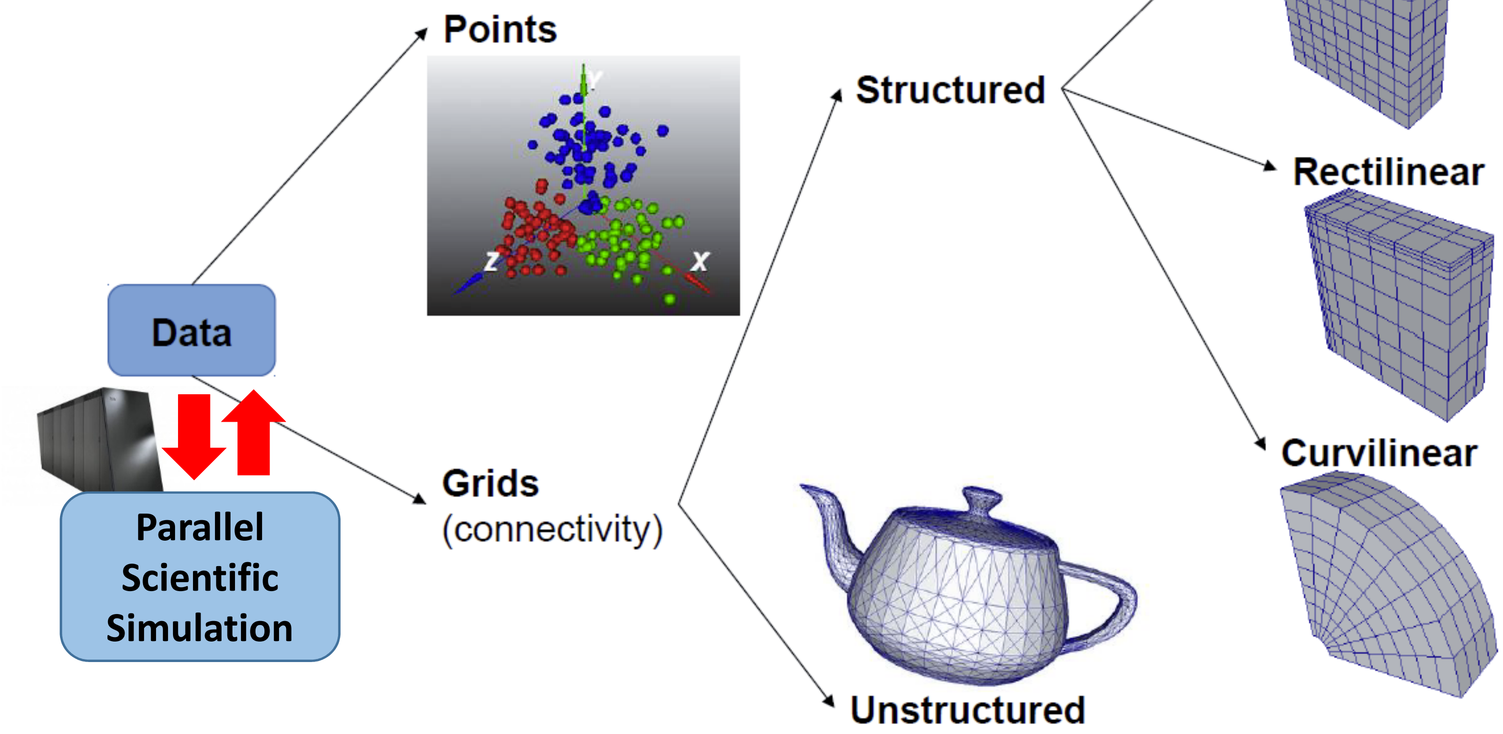


- **Key objectives of scientific visualization in HPC are to (a) analyse/explore & (b) present and communicate scientific data**

[6] CINECA – Scientific Visualization Training

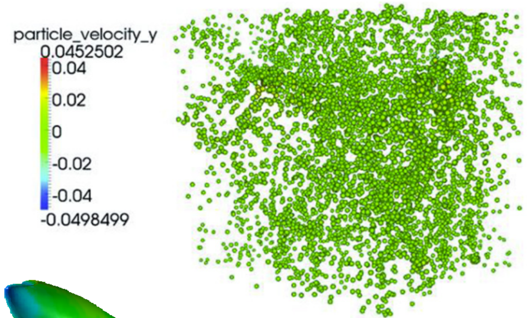
Scientific Visualization – Understanding HPC Data

- Simulation data can be simple Points or connected structured & unstructured Grids

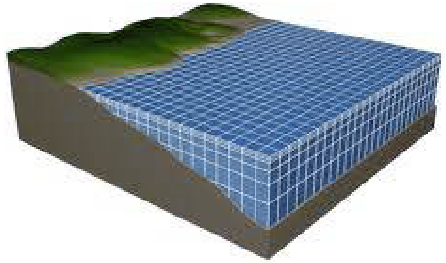


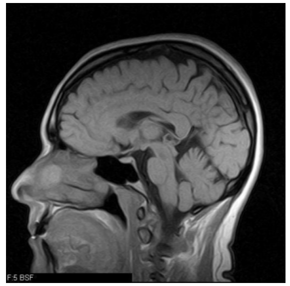
Scientific Visualization – Selected HPC Simulation Data Types

Points: e.g. in particle simulations

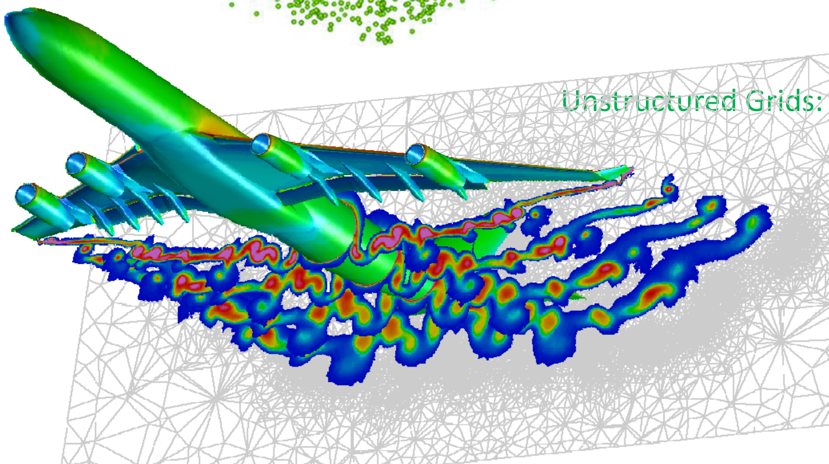


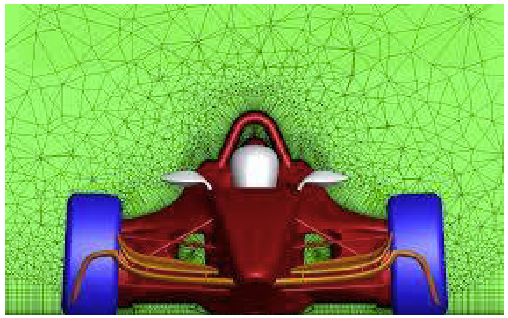
Rectilinear Grid: e.g. in oceanographic or medical simulations





Unstructured Grids: e.g. in computational fluid dynamics or aerodynamics

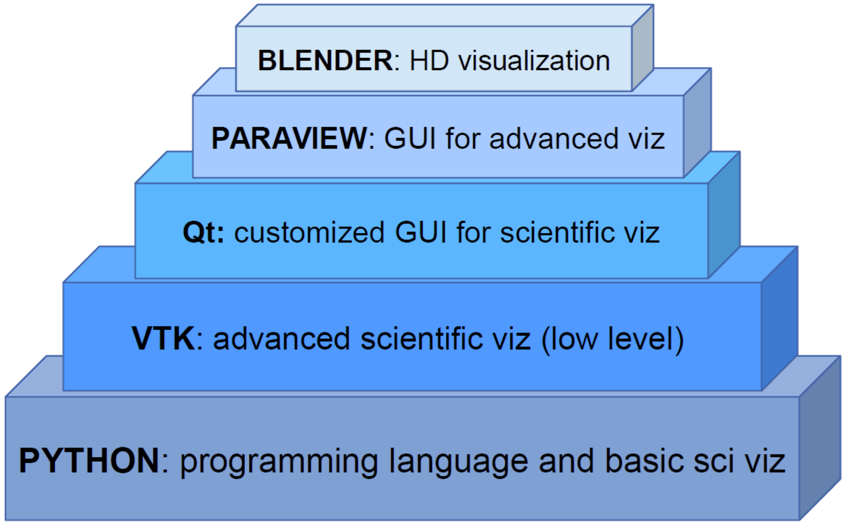




modified from [6] CINECA – Scientific Visualization Training

Visualization Tools & Technology – Different Stacks

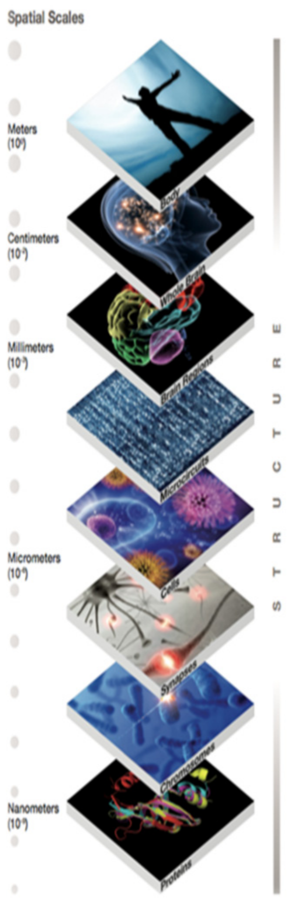
- A wide range of tools and technologies exist (small selection)
  - Examples stand for several different tools on the same level
  - Many free open source tools are available
  - Range of commercial tools offer massive amounts of visualization support
  - Supercomputing centres offer training courses and employ typically also some visualization experts

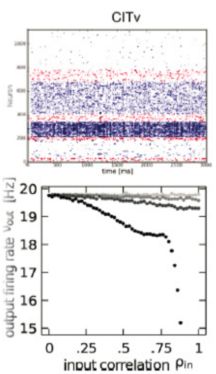


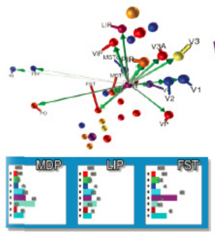
[6] CINECA – Scientific Visualization Training

- **A wide variety of tools & techniques exists for scientific visualization starting from low-level programming languages support and cutomizable GUIs to high level GUIs and HD visualization**

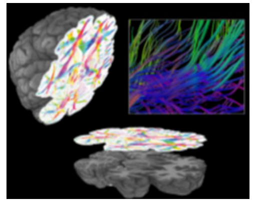
Multi-Scale Visualization



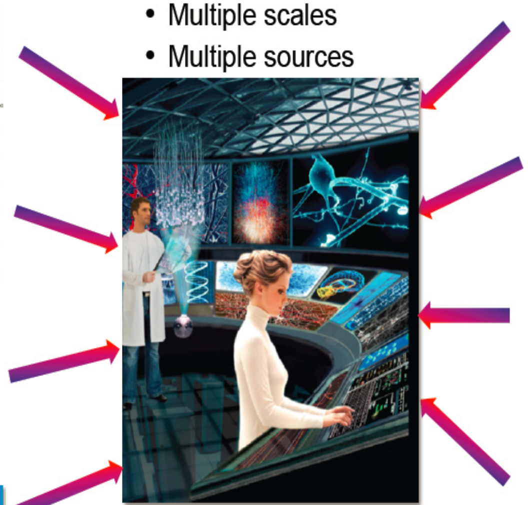


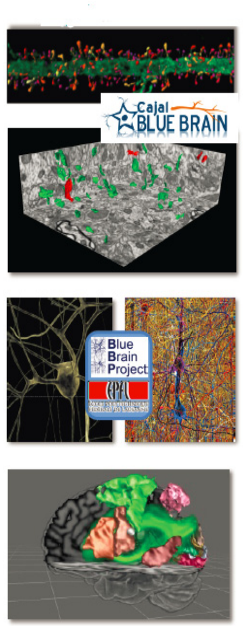


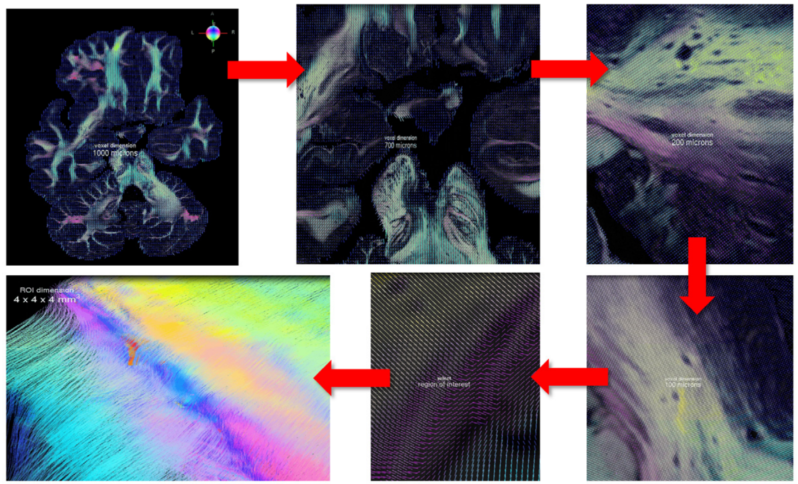
(data from reality
as guiding design
for simulations)



- Multiple modalities
- Multiple scales
- Multiple sources







modified from [13] M.Axer

modified from [8] T.Kuehlen

Computational Steering of (Iterative) Parallel Algorithms via MPI (cf. Lecture 5)

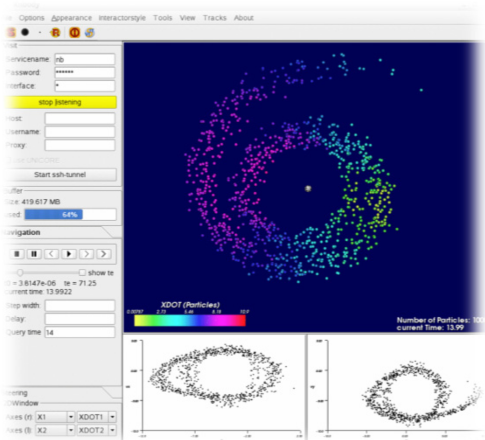
- Particle Simulations using PEPC library (see above)
  - E.g. research star cluster dynamics in astrophysics or particle acceleration simulations via laser pulses
  - E.g. Iterations over time using nbody6++ parallel algorithm
  - Steering: changing parameters during the run-time of simulation

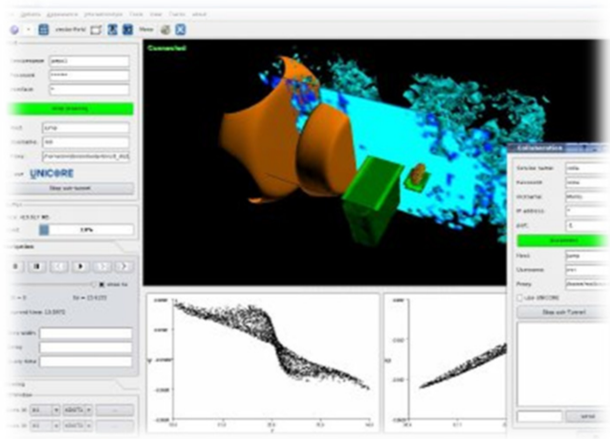
[11] M. Riedel et al., computational steering, 2007

```
call flvisit_nbody2_steering_recv(
& VISITDPARM1,VISITDPARM2,VISITDPARM3,
& VISITDPARM4,VISITIPARM1,VISITIPARM2,...)
...
if(LVISIT_ACTIVE.eq.1) Then
 VDISTANCE=VISITDPARM4
 write(*,*) 'VISCON: VDISTANCE=',VDISTANCE
endif
...
IF(VISITDPARM2.gt.0) THEN
 DTADJ = VISITDPARM2
END IF
IF(VISITDPARM3.gt.0) THEN
 DELTAT = VISITDPARM3
END IF
...
CALL MPI_BCAST(DTADJ,1,MPI_DOUBLE_PRECISION,
   0,& MPI_COMM_WORLD,ierr)
CALL MPI_BCAST(VISITDPARM3,1,
   MPI_DOUBLE_PRECISION,0,& MPI_COMM_WORLD,...)
```

change parameters interactively

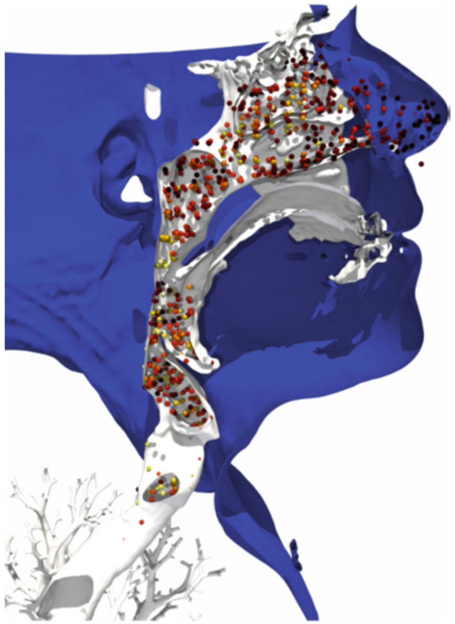
visualize status

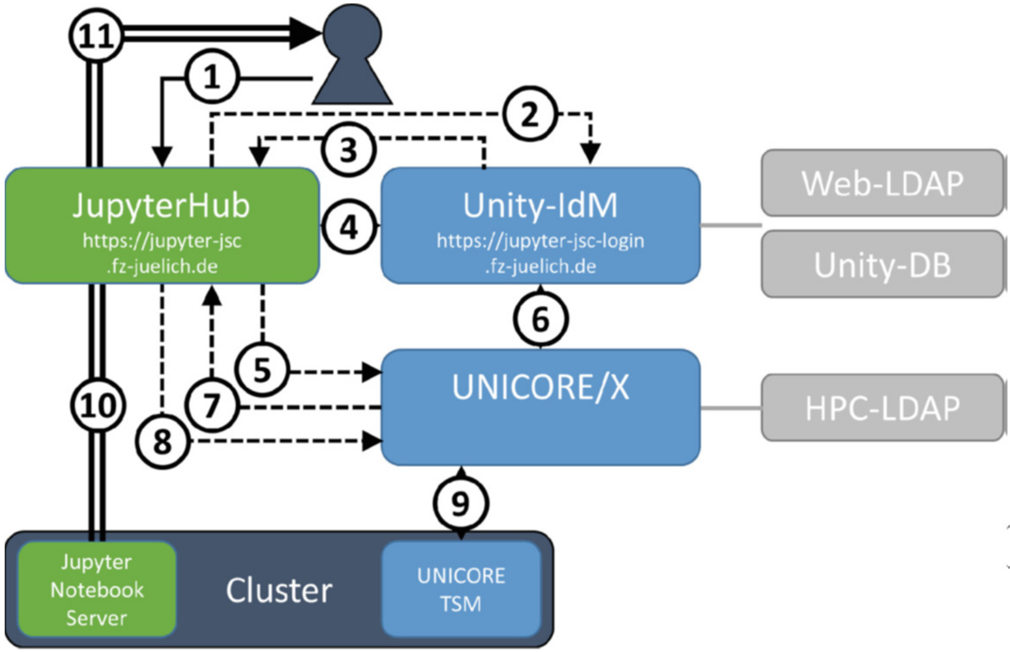




- **Computational steering is the technique of manually intervening with an HPC simulation in order to change its outcome by the manipulation of certain parameters computed**
- **It requires the visualization during runtime (online) in order to properly steer parameters**
- **Computational steering is an old term, recently more used is 'interactive simulations'**

Scheduling vs. Emerging Interactive HPC Applications (cf. Lecture 1)





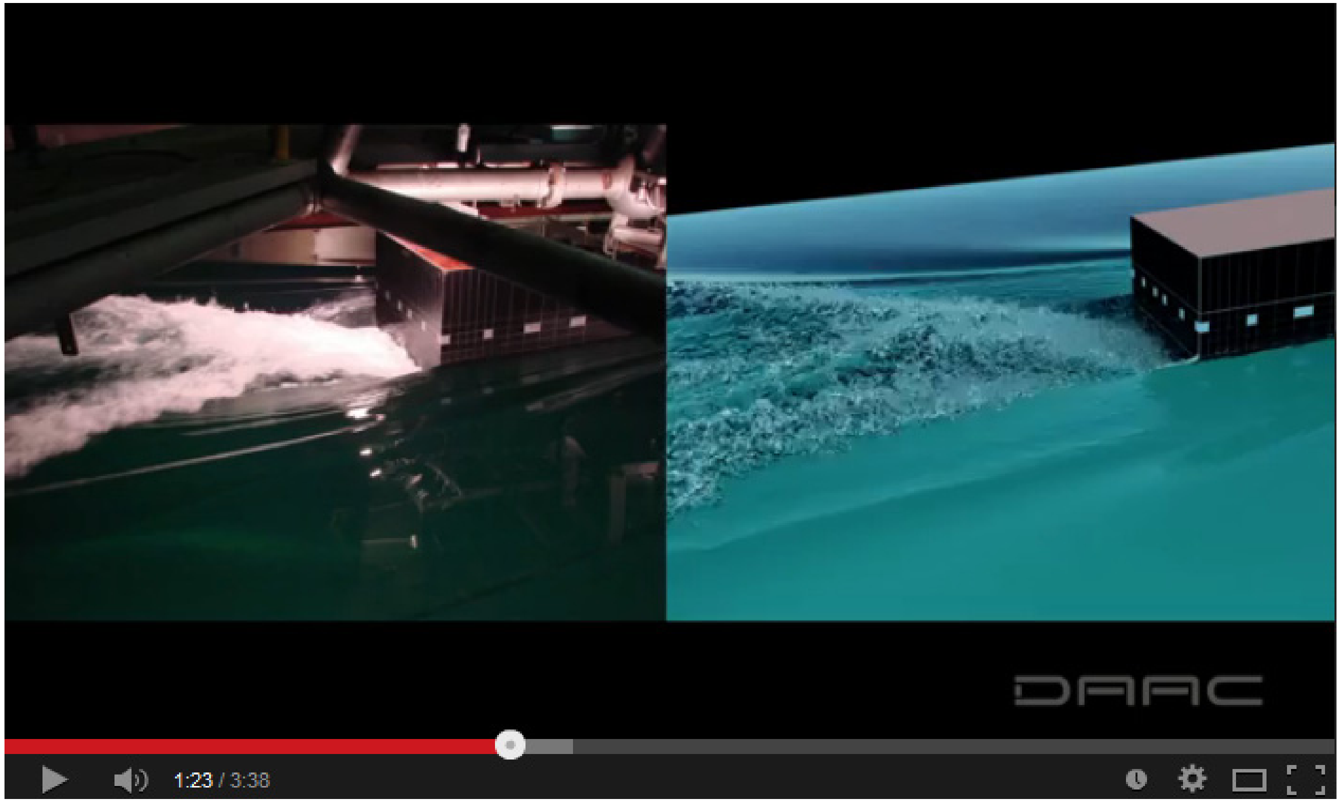
- JupyterHub is a multi-user version of the notebook designed for companies, classrooms and research labs

[9] A. Lintermann & M. Riedel et al., 'Enabling Interactive Supercomputing at JSC – Lessons Learned'

[10] A. Streit & M. Riedel et al., 'UNICORE 6 – Recent and Future Advancements'

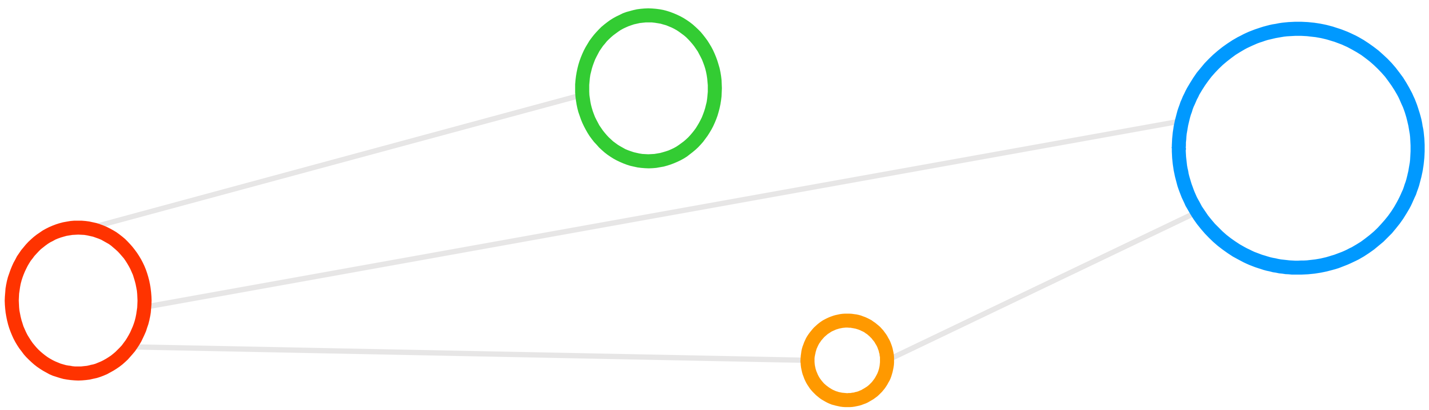
[12] Project Jupyter Web page

[Video] Scientific Visualization with Wave Application Example



[7] YouTube Video, WaveAnimations

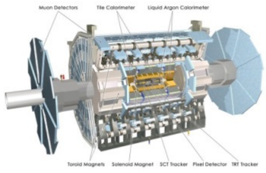
Scalable Infrastructures



The Power of 'Scalable Infrastructures'

- Parallel computing can be done within one or across n computers
  - Perform calculations, visualizations, and data processing…
  - … at an incredible, ever-increasing speed… as part of infrastructures
- Solutions of (scientific) problems often also require 'resources' that are not locally available
  - Academic: Global collaborations that jointly perform research
  - Industry: Companies with different branches share (customer) information

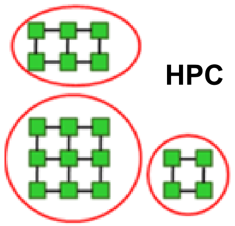


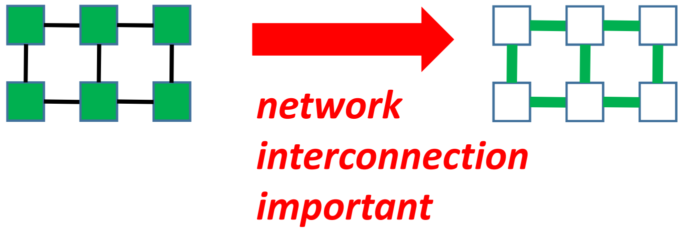



- **A resource is a specific hardware or software system such as a parallel computer, a disk or tape storage, 3D display capabilities, or a (scientific) measurement instrument like a telescope**
- **Parallel computing infrastructures enable the parallel use of such resources with many others**

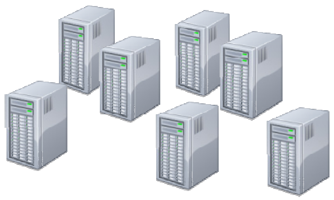
Different Computing Paradigms Drive Infrastructure Design

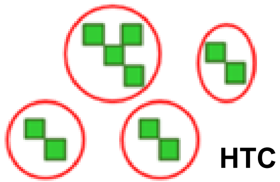
- **A High Performance Computing (HPC) – driven infrastructure is based on computing resources that enable the efficient use of parallel computing techniques through specific support with dedicated hardware such as high performance cpu/core interconnections**

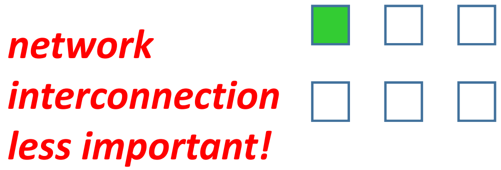




- **A High Throughput Computing (HTC) – driven infrastructure is based on commonly available computing resources such as commodity PCs and small clusters that enable the execution of 'farming jobs' without providing a high performance interconnection between the cpu/cores**



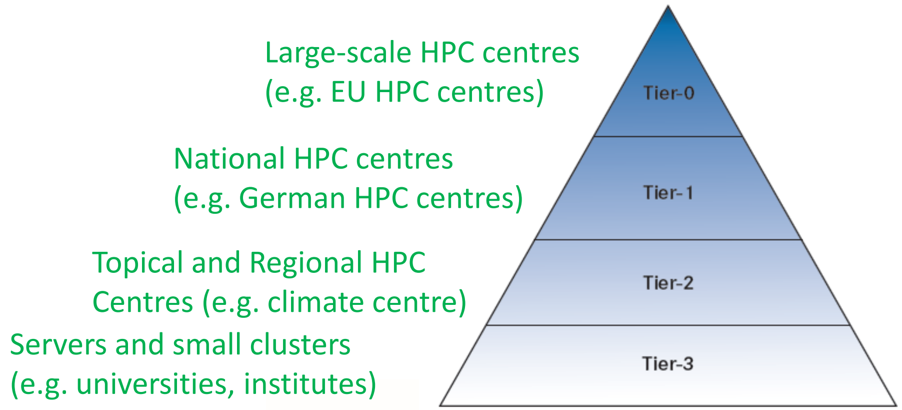




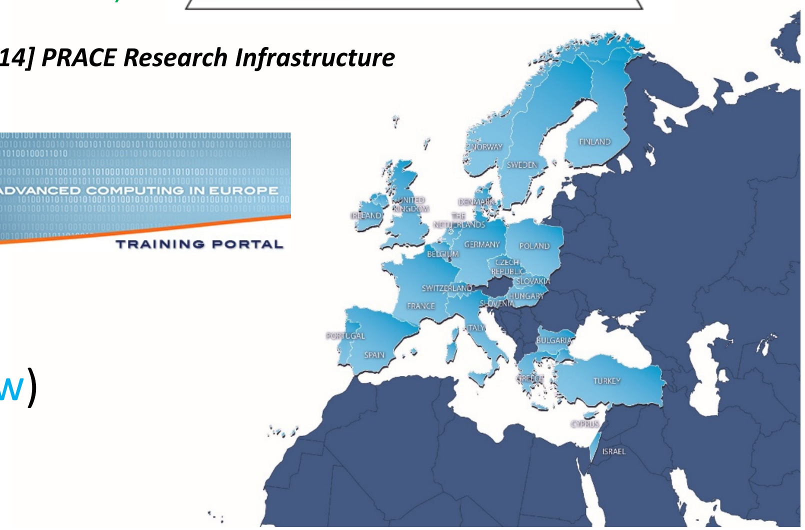
➢ **The complementary Cloud Computing & Big Data – Parallel Machine & Deep Learning Course focusses on High Throughput Computing**

Large Scale HPC Infrastructure – Major Concepts & PRACE Tier-0 Example

- Tier-X design
  - Based on different levels of computational power
  - Needs to be synchronized with EU, national, or regional funding streams and partner as HPC machines are costly resources
- Scientific Peer-review
  - Roundtable of known experts judge proposals to receive computing time
  - Achieves 'greatest minds get most valuable and costly resources'
- Common Tools & Training Approach
  - Idea of a common basic set of MPI/Open/application libraries
  - Emerging same set of modules for data sciences (e.g. TensorFlow)
  - Idea 'common production environment' (e.g. change systems)

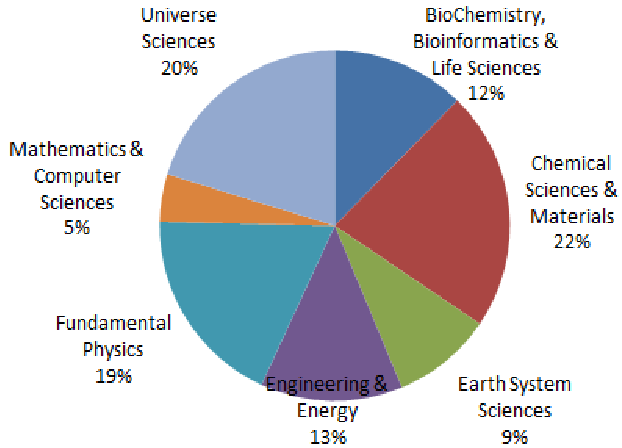


[14] PRACE Research Infrastructure

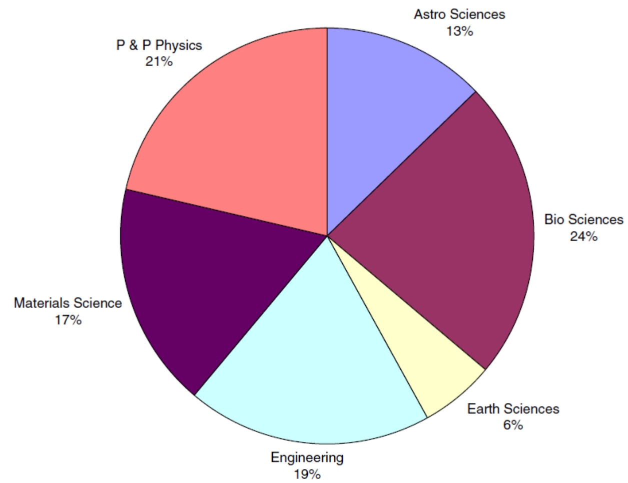


Large Scale HPC Infrastructure – PRACE Example

- Partnership for Advanced Computing in Europe (PRACE)
  - European multi-disciplinary HPC-driven infrastructure
  - Many European countries are members of this research infrastructure
  - Provides computational resources to a wide range of scientific disciplines
- Different ‘time grant‘ models
  - Preparatory access (prepare projects)
  - Project access (after peer-review)
  - Multi-year access (for large communities)
- Follow-on to earlier infrastructure efforts
  - E.g. Distributed European Infrastructure for Supercomputing Applications (DEISA)

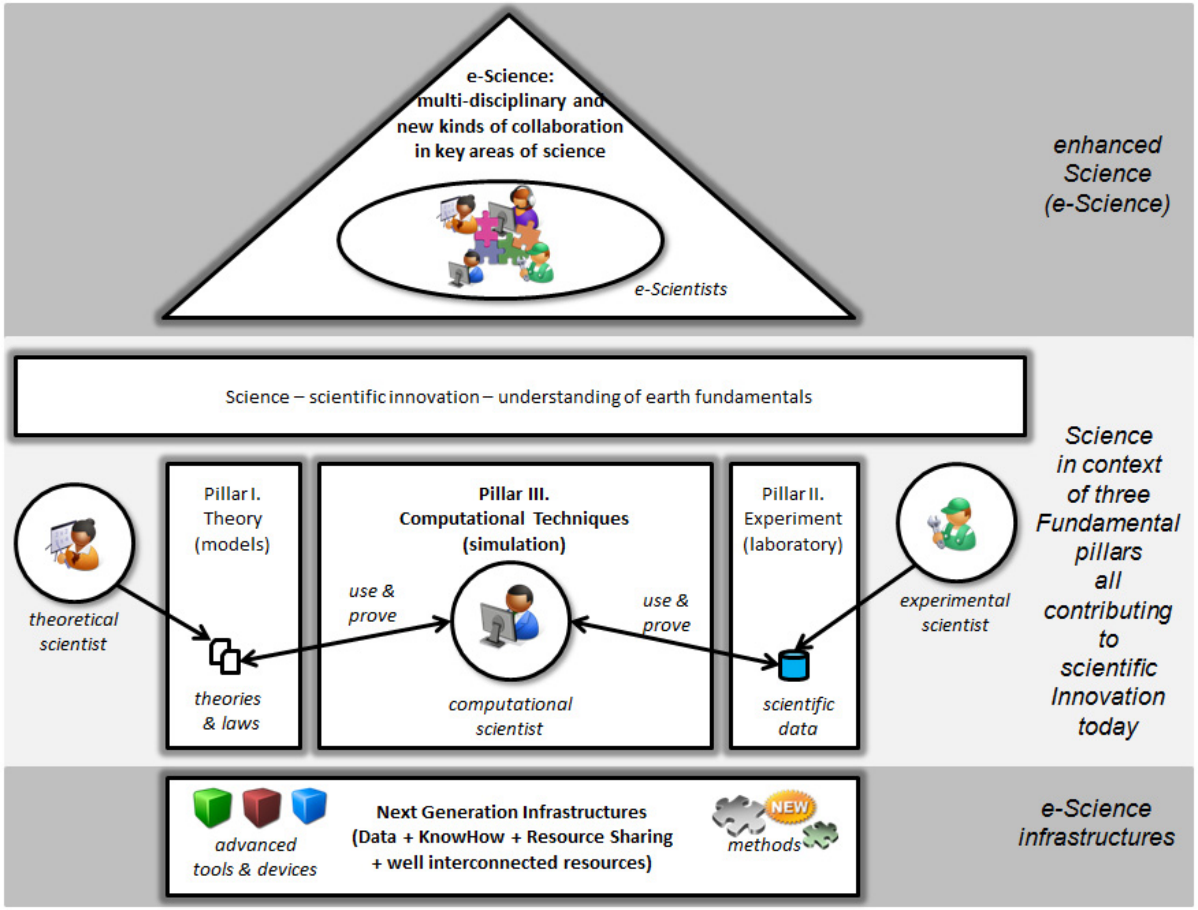


(diverse user communities)



[15] W. Gentzsch and M. Riedel et al., ‘DEISA – Distributed European Infrastructure for Supercomputing Applications’, Journal of Grid Computing,2011

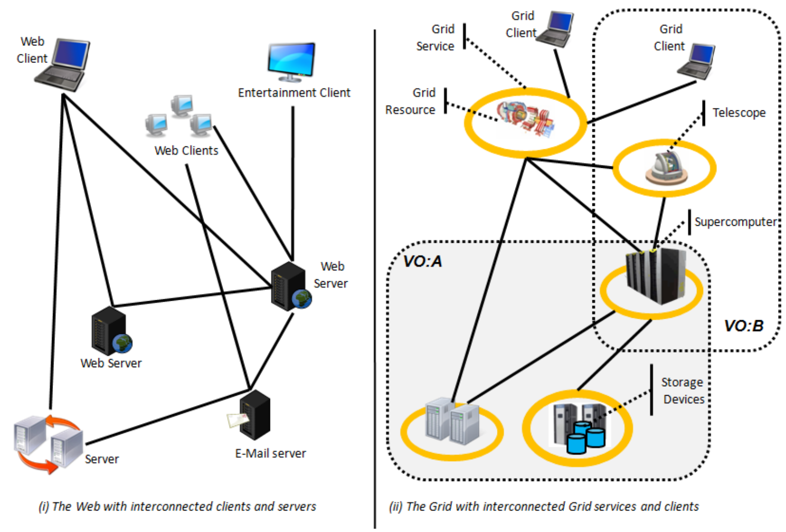
Next Generation 'e-Science' Infrastructures & Virtual Organizations



- *'e-Science is about collaboration in key areas of Science and the next generation infrastructure that will enable it.'*

[17] Taylor , 'Enhanced-Science (e-Science) Definition', 2000

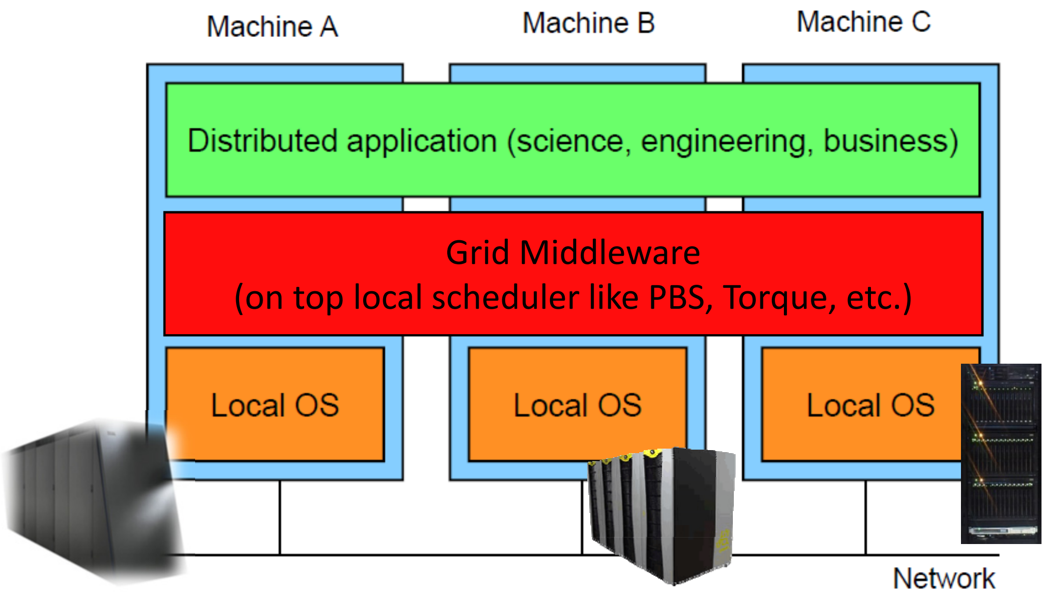
[18] I. Foster et al., 'The Anatomy of the Grid: Enabling Scalable Virtual Organizations', 2001



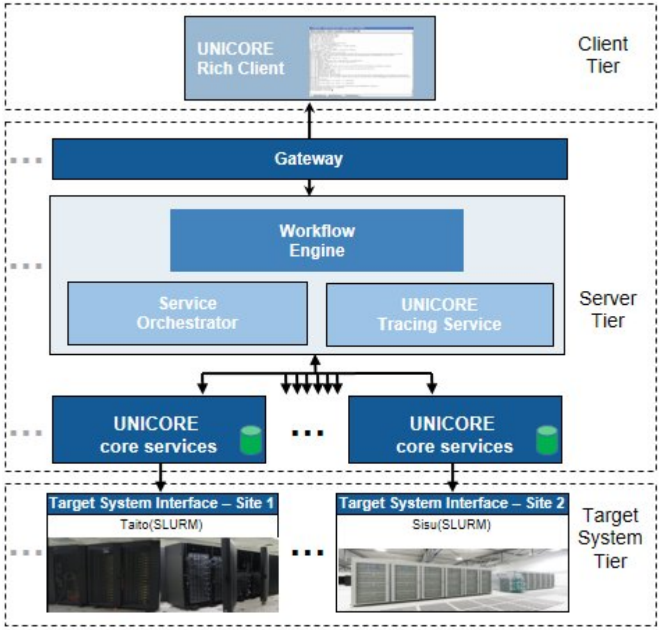
- **A virtual organization (VO) enables a secure sharing of a wide variety of geographically distributed resources across different organizational boundaries (e.g. time limited, dynamic add/remove)**

Grid Middleware to Access used in e-Science, Grid & HPC Infrastructures

- Specialization of a general 'middleware in distributed systems'
  - Establishes a 'single system view' for users including security
  - Example: UNICORE Grid Middleware System (open source)

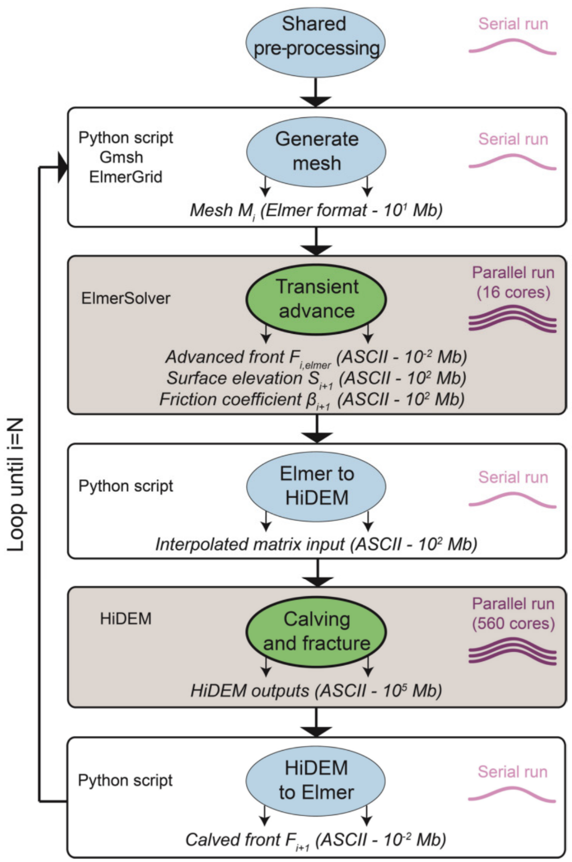


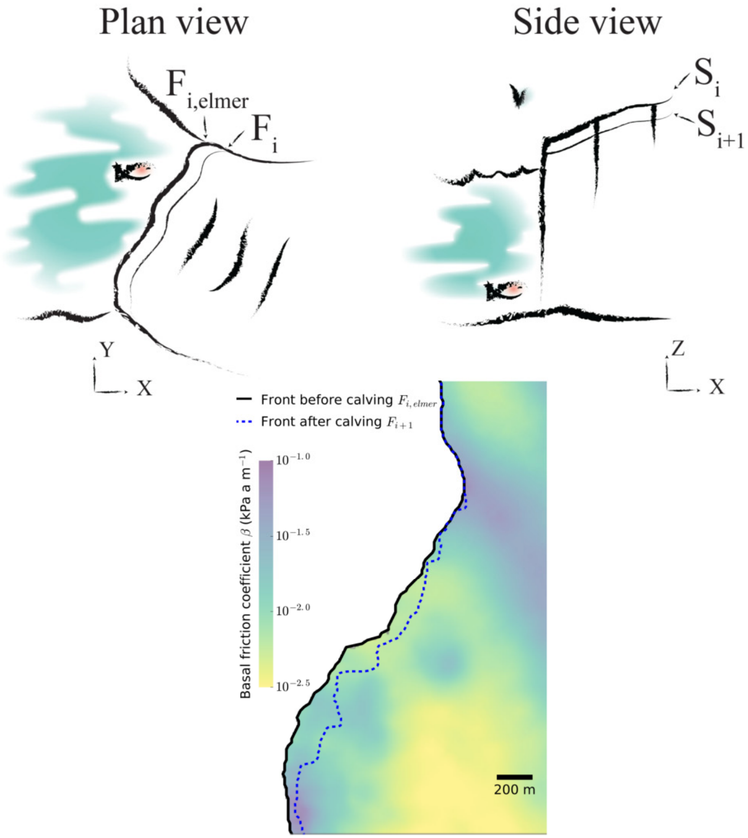
[19] A.Tanenbaum et al., 'Distributed Systems'

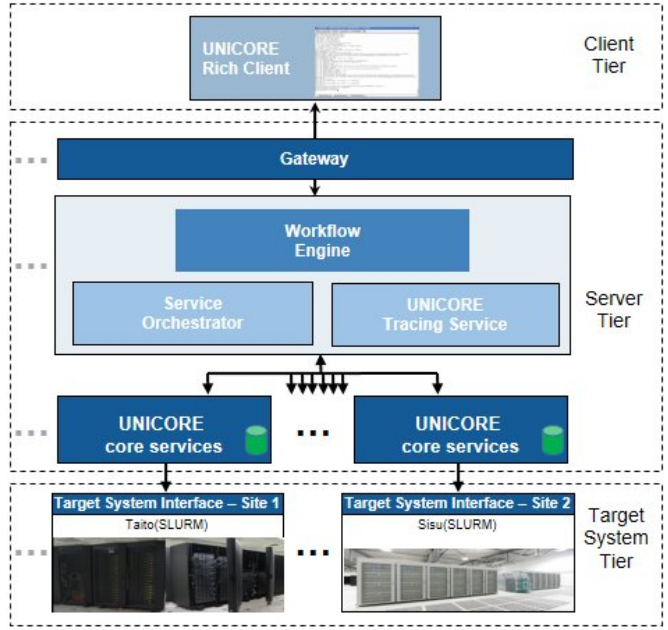


[20] M. Memon & M. Riedel et al., 'Scientific workflows applied to the coupling of a continuum (Elmer v8.3) and a discrete element (HiDEM v.1.0) ice dynamic model', 2019

Scientific Workflows using UNICORE Middleware on HPC Infrastructures

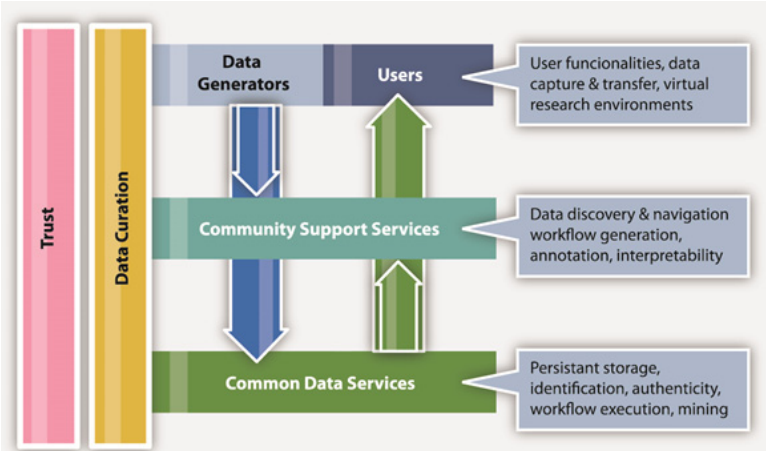






[20] M. Memon & M. Riedel et al., 'Scientific workflows applied to the coupling of a continuum (Elmer v8.3) and a discrete element (HiDEM v.1.0) ice dynamic model', 2019

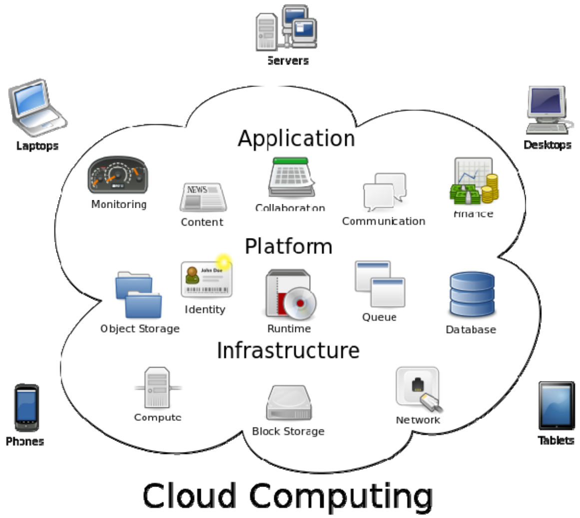
Collaborative Data & Cloud Infrastructures



[21] High Level Expert Group on Scientific Data, Riding The Wave – How Europe can gain from the rising tide of scientific data, 2010

- **A collaborative data infrastructure combines the massive amount of unique resources of large multi-disciplinary data and computing centers with strong domain-specific centers (e.g. climate)**

- **Grid computing can be seen as the major precursor of industry-driven Cloud computing : Inspired many technology approaches used in clouds & virtualization (e.g. in their backend infrastructures)**



[22] Wikipedia 'Cloud computing'

➢ **Complementary Cloud Computing & Big Data – Parallel Machine & Deep Learning Course offers more on Clouds & Data Infrastructures**

Large-scale Computing Infrastructures (cf. Lecture 1)

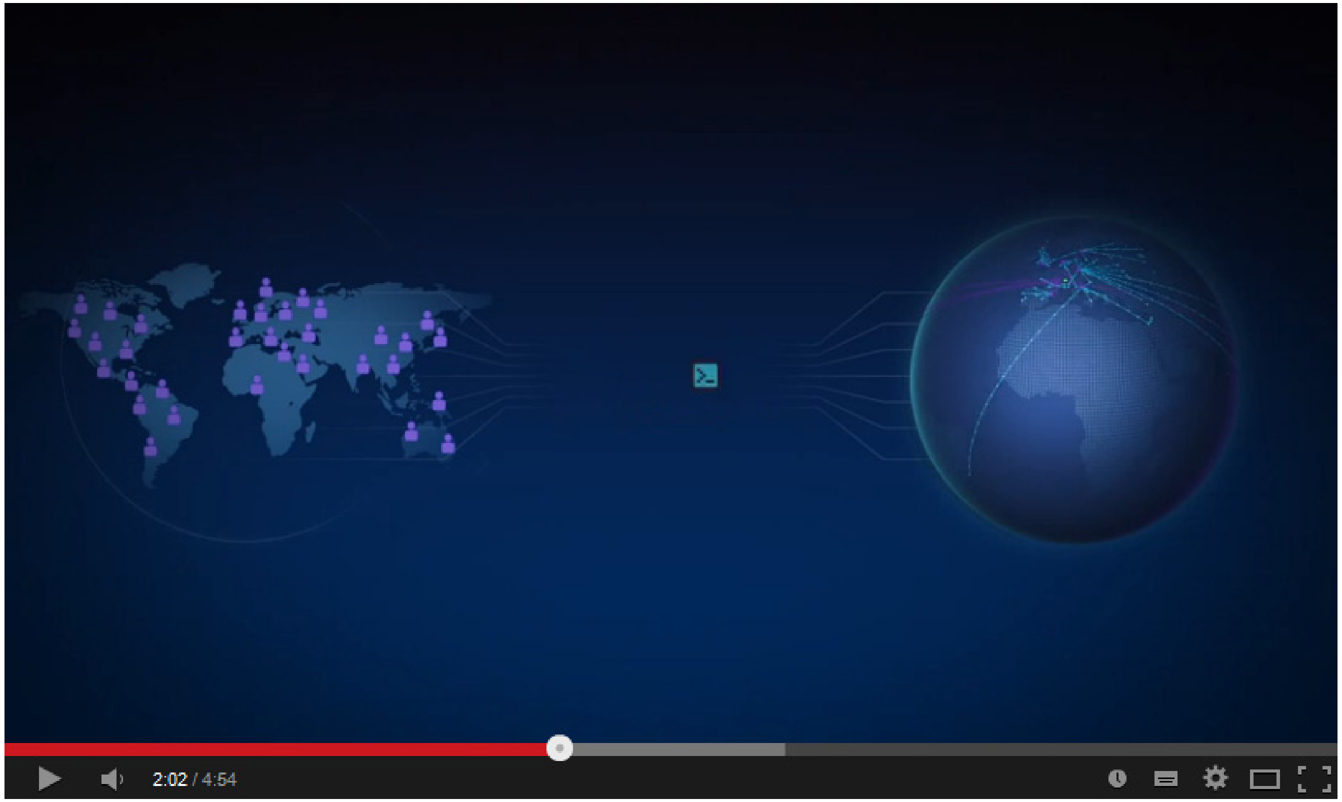
- Large computing systems are often embedded in infrastructures
  - Grid computing for distributed data storage and processing via middleware
  - The success of Grid computing was renowned when being mentioned by Prof. Rolf-Dieter Heuer, CERN Director General, in the context of the Higgs Boson Discovery:
- Other large-scale distributed infrastructures exist
  - Partnership for Advanced Computing in Europe (PRACE) → EU HPC
  - Extreme Engineering and Discovery Environment (XSEDE) → US HPC

- **'Results today only possible due to extraordinary performance of Accelerators – Experiments – Grid computing'**

[16] Grid Computing Video

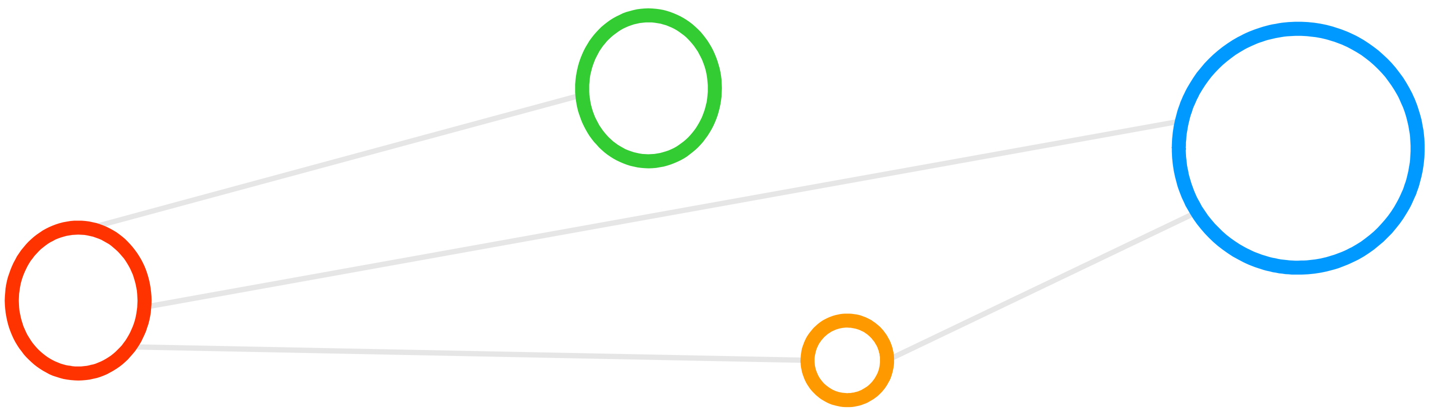
> **Complementary Cloud Computing & Big Data – Parallel Machine & Deep Learning Course uses Large-Scale Computing Infrastructures**

[Video] Grid Infrastructure Example Contributed to the Higgs Boson Discovery



[16] YouTube Video, Grid Computing

Lecture Bibliography



Lecture Bibliography (1)

- [1] Introduction to High Performance Computing for Scientists and Engineers, Georg Hager & Gerhard Wellein, ISBN 143981192X
- [2] G.Hager, MPI+OpenMP hybrid computing (on modern multicore systems), Online: http://www.speedup.ch/workshops/w39_2010/slides/hager.pdf
- [3] Changmin Lee, Won Woo Ro, Jean-Luc Gaudiot, 'Boosting CUDA Applications with CPU–GPU Hybrid Computing', Int J Parallel Prog (2014) 42:384–404, DOI 10.1007/s10766-013-0252-y
- [4] Community Land Model (CLM), Online: http://www.cgd.ucar.edu/tss/clm/
- [5] 'Scientific Visualization' on Wikipedia, Online: http://en.wikipedia.org/wiki/Scientific_visualization
- [6] 'CINECA – 13th Summerschool on Scientific Visualization', Online: http://www.hpc.cineca.it/content/training-material
- [7] YouTube Video, 'Interactive simulation of water flow and salinity transport using hands', Online: http://www.youtube.com/watch?v=kzFBZ8KaWF4
- [8] T. Kuhlen,'Automatisierte Datenanalyse und Experten-Beurteilung', 2013
- [9] A. Lintermann & M. Riedel et al., 'Enabling Interactive Supercomputing at JSC – Lessons Learned', ISC 2019, Frankfurt, Germany, Online: https://www.researchgate.net/publication/330621591_Enabling_Interactive_Supercomputing_at_JSC_Lessons_Learned_ISC_High_Performance_2018_International_Workshops_FrankfurtMain_Germany_June_28_2018_Revised_Selected_Papers
- [10] A. Streit & M. Riedel et al., 'UNICORE 6 – Recent and Future Advancements ', Online: https://www.researchgate.net/publication/225005053_UNICORE_6_-_recent_and_future_advancements
- [11] M. Riedel, Th. Eickermann, S. Habbinga, W. Frings, P. Gibbon, D. Mallmann, F. Wolf, A. Streit, Th. Lippert, Felix Wolf, Wolfram Schiffmann, Andreas Ernst, Rainer Spurzem, Wolfgang E. Nagel, 'Computational Steering and Online Visualization of Scientific Applications on Large-Scale HPC Systems within e-Science Infrastructures', pp.483-490, Third IEEE International Conference on e-Science and Grid Computing, 2007, Online: https://www.researchgate.net/publication/4309569_Computational_Steering_and_Online_Visualization_of_Scientific_Applications_on_Large-Scale_HPC_Systems_within_e-Science_Infrastructures

Lecture Bibliography (2)

- [12] Project Jupyter Web page, Online:
  https://jupyter.org/hub
- [13] M. Axer, BigBrain Presentation, INM-1, 2014
- [14] PRACE Research HPC Infrastructure, Online:
  http://www.prace-ri.eu/
- [15] W. Gentzsch, D. Girou, A. Kennedy, H. Lederer, J. Reetz, M. Riedel, A. Schott, A. Vanni, M. Vazquez, J. Wolfrat, '*DEISA - Distributed European Infrastructure for Supercomputing Applications'* Journal of Grid Computing, 9 (2011) 2, 2011, pp. 259 – 277, Online:
  https://www.researchgate.net/publication/220621391_DEISA-Distributed_European_Infrastructure_for_Supercomputing_Applications
- [16] How EMI Contributed to the Higgs Boson Discovery, YouTube Video, Online:
  http://www.youtube.com/watch?v=FgcoLUys3RY&list=UUz8n-tukF1S7fql19KOAAhw
- [17] J. Taylor, Enhanced-Science (e-Science) Definition, Online:
  http://www.e-science.clrc.ac.uk
- [18] Ian Foster et al., 'The Anatomy of the Grid: Enabling Scalable Virtual Organizations', *International Journal of High Performance Computing Applications, 2001, vol. 15, no. 3, pages 200-222,* Online:
  http://hpc.sagepub.com/content/15/3/200.full.pdf+html
- [19] A. S. Tanenbaum and M. van Steen. Distributed Systems - Principles and Paradigms. Prentice Hall International, ISBN 0132392275, 2006.
- [20] M. Memon & M. Riedel et al., ‚Scientific workflows applied to the coupling of a continuum (Elmer v8.3) and a discrete element (HiDEM v1.0) ice dynamic model', Geoscientific Model Development (GMD), Vol 12 (7), 2019, Online:
  https://www.researchgate.net/publication/326521081_ScientificWorkflows_Applied_to_the_Coupling_of_a_Continuum_Elmer_v83_and_a_Discrete_Element_HiDEM_v10_Ice_Dynamic_Model
- [21] High Level Expert Group on Scientific Data, ‚*Riding The Wave – How Europe can gain from the rising tide of scientific data*', 2010, Online:
  http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf
- [22] Wikipedia, ‚Cloud Computing', Online:
  http://de.wikipedia.org/wiki/Cloud-Computing


services
measurement
within
fusion
new
funding
Policy-based
concepts
forms
cross-disciplinary
resources
device
Centre
Analysis
climate
Computer
expertise
Approaches
construction
Fusion
year
like
Computational
Test
E.g
Computing
disciplines
results
Enable
different
Cross-Disciplinary
methods
High
big
nodes
PB
community
Environment
Modelling
EU
Science
centers
Key
scientific
important
files
etc
DKRZ
Use
storage
basis
Infrastructure
model
international
simulations
ENES
technologies
formats
brain
increasing
via
EUDAT
physical
IPCC
SMAQ
engineering
References
Big
Project
Data
MapReduce
create
processing
Common
TB
Methods
computing
Using
European
databases
codes
GB
Supercomputing
Work
images
analysis
Large
Summary
data-intensive
Service
Scientific
levels
often
stored
Resources
analyze
use
DLCL
Level
external
JSC
computational
Research
Matters
Concrete
performance
research
e.g
science
HPC
Juelich
system
datasets
acquisition
large-scale
compute
Landscape
management
System
Climate
modelling
AR5
Hardware
hundreds
Nutshell
Services
access
distributed
cores
Structure
current
network
DLCLs
Understanding
Technologies
Earth
General
Germany
Simulation
directory
project
Provide
NASA
Energy
systems
day
Processing
Health