

# PAAD: Platforms for Scientific Data Analysis and HPC

Joachim J. Włodarz

University of Silesia, Faculty of Mathematics, Physics  
and Chemistry, Theoretical Chemistry Dept.

PL 40-007 Katowice, Bankowa 14; <jjw@us.edu.pl>

KU KDM'16, Zakopane, 16-18.03.2016



# PAAD: rationale

- research is *de facto* data processing
- data  $\Rightarrow$  information  $\Rightarrow$  knowledge (we hope :-)
- *Big Data* are commonplace
- 4V rule: Variety, Velocity, Veracity, Volume
- *HPC* needed for many tasks
- interactive scientific computing welcome
- scientific data archiving required (Nature, ...)



# PAAD: for biosciences

- high-speed genome sequencing
- morphogenesis modeling
- biological processes modeling
- image processing
- *Galaxy* bioinformatics interactive toolkit
- scientific data archival storage



# PAAD: for chemistry & physics

- laboratory data acquisition and processing
- quantum chemical calculations
- particle physics *Monte Carlo* calculations
- algebraic calculations and tests on CAS
- interactive computing: *Sage/Jupyter/IPython*
- calculations for the *LHCPhenoNet* project
- scientific data archival storage



# PAAD: for earth sciences

- *G/S* data processing
- geo-referenced data processing
- meteorological data processing
- geomorphological and explorational data
- simulations of geological processes
- scientific data archival storage



# PAAD: for any research

- Linux-based environment
- other environments in virtual machines
- universal *HPC/HA* setup, batch & interactive
- Python-based interactive computing
- numerical and symbolic calculations
- open source computing software (!)
- archival data storage

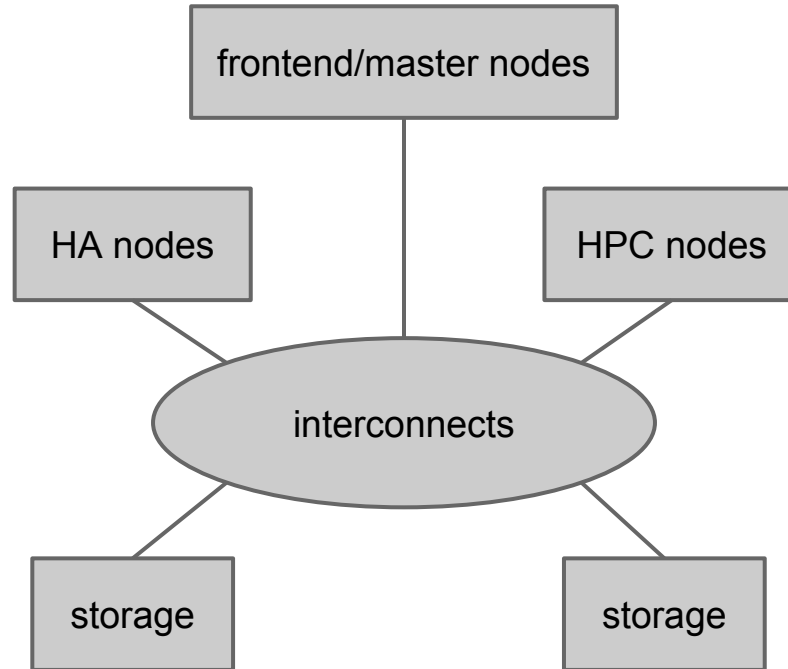


# PAAD: for education

- *Sage/Jupyter/IPython* CAS environments
- “living documents” paradigm (“notebook”)
- ready-to-use materials from *iCSE* project
- <http://icse.us.edu.pl/materialy-dydaktyczne/>
- from linear algebra to molecular modeling
- multiuser setup, browser-based access
- any decent browser supported



# PAAD: outline of the infrastructure





# PAAD: hardware dimensioning

- storage: 80 TB/yr \* 5 yr = 400 TB  $\Rightarrow$  700 TB
- memory: 8-32 GB/job  $\Rightarrow$  ~ 16 GB/job
- CPU: 4-16 C/job  $\Rightarrow$  ~ 8 C/job
- 3-4 grp \* 5-10 job/grp  $\Rightarrow$  40 jobs (?)
- HPC nodes:  $\Rightarrow$  ~ 40: 16 C, 128 GB  $\Rightarrow$  44
- HA nodes:  $\Rightarrow$  4: 16 C, 256 GB  $\Rightarrow$  4
- C/: 56 Gbps IB, storage: IB or 10 Gbps Eth



# PAAD: computing accelerators

- performance  $\Rightarrow$  **~100 TFLOPS** (DP)
- energy consumption  $\Rightarrow$  ~ 40-50 kW
- support for parallel processing
- support of GPU-accelerated software
- cost effective hardware
- $\Rightarrow$  12 nodes: +2x Xeon Phi 7120P (~Intel64)
- $\Rightarrow$  12 nodes: +2x Nvidia Tesla 40M



# PAAD: the machinery



# Acknowledgements



UNIA EUROPEJSKA  
EUROPEJSKI FUNDUSZ  
ROZWOJU REGIONALNEGO



KAPITAŁ LUDZKI  
NARODOWA STRATEGIA SPÓJNOŚCI

UNIA EUROPEJSKA  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY



Thank you for your attention

