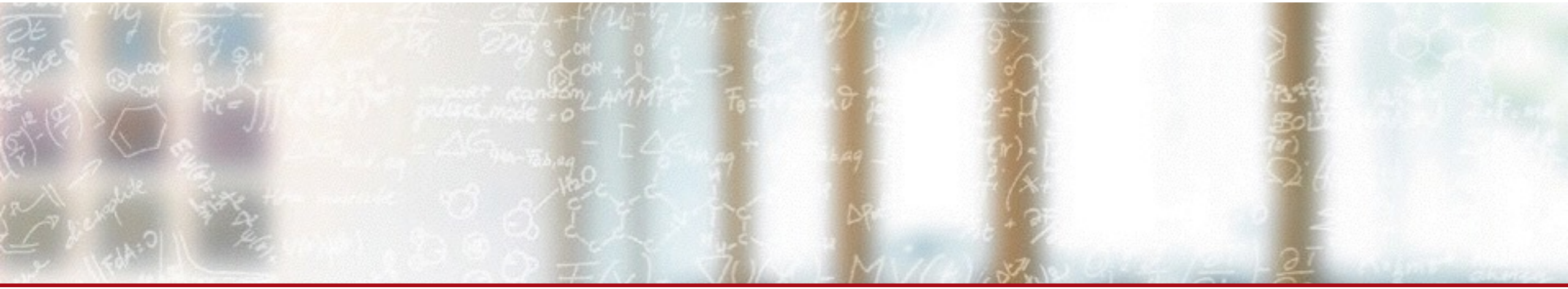
CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich



Data Science Services at CSCS

HPC-Forum: Data Science & Machine Learning

Marcel Schoengens, CSCS

May 18th 2017

Overview

- Software Services
  - Data Analytics
  - Classical Machine Learning
  - Deep Learning
- Infrastructure Services
  - Compute and Storage
  - Access Model: cscs2go

Software Services

Data Analytics

- Use case: clean-up, query, transform, apply statistics and visualize data

- Python Scientific Libraries: Pandas, Scipy, Numpy, Matplotlib
  - to use on a single node

- Apache Spark in combination with Python Scientific Libraries
  - to make use of multi-core, multi-node infrastructure

- Example
  - Analyze data from a distributed network simulation

Classical Machine Learning

- Use case: model calibration (inference) and generalized linear models

- Python Scientific Libraries: scikit-learn
  - to use on a single node

- Apache Spark ML
  - to make use of multi-core, multi-node infrastructure

- ABCpy: Library for approximate bayesian computation
  - runs in single- and multi-node mode

- Examples
  - Predict traffic in distributed networks (linear regression, SVM)
  - Infer parameters of a mechanistic social network models
  - Infer parameters of hydrological models

Deep Learning

- Use case: Train convolutional & recurrent neural networks

- Tensorflow and Theano
  - to use single node

- Examples
  - Predict treatments for patients learned from PubMed publications (NLP)
  - Detect events in high energy physics (classify event from noise)
  - Detect patterns in radio telescope images

Infrastructure Services

Compute & Storage

- Compute @ Piz Daint
  - Multicore (2 x Intel Broadwell CPU, 36 cores total)
  - Hybrid (Intel Haswell CPU, 12 cores & Nvidia P100 GPU)
  - 64 and 128 GB of main memory

- Storage
  - Home Directory, Project Storage, Data Mover
  - ~ 6PB shared Scratch storage

- Programming Environments
  - C++, Fortran, Python, …
  - debuggers and performance tools

Access Model: cscs2go

- Base Packages

	Start-up	Base	Medium	Large
# Users	1	3	5	10
Permanent Storage	1 TB	5 TB	20 TB	70 TB
Scratch Storage	~6PB shared			
Node hours	100	200	1'000	1'000
Yearly price (in CHF)	300	1'000	3'500	10'000

  - Included in every package are: 10 GB of storage on /home per user,
  - 12 months validity.

- Additional node hours

Number of node hours	Price per node hour
500 – 4'999	CHF 1.00
5'000 – 29'999	CHF 0.80
> 30'000	CHF 0.70

All prices are for academia only

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

ETH zürich

Thank you for your attention.