



**BABEŞ-BOLYAI UNIVERSITY**  
Faculty of Mathematics and Computer Science



Master specialization

# High Performance Computing and Big Data Analytics

WHY

**HIGH PERFORMANCE COMPUTING**

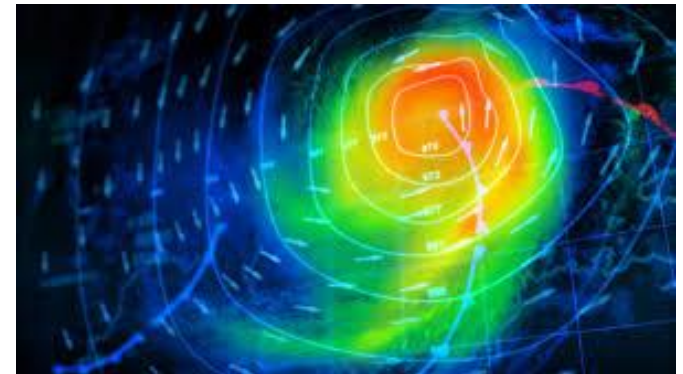
TOGETHER WITH

**BIG DATA ANALYTICS**

?

# HPC (always) uses *big data*

- ∞ For a long period *big data* has been synonymous with  
High Performance Computing (HPC)
- ∞ HPC -> **intensive computations** with a **big** volume of data
- ∞ HPC led to the progress and evolution in **High Performance Hardware**
- ∞ Simulation: An Original (*big data*) HPC domain
  - Weather forecasting/climate modeling
  - Physics - molecular dynamics
  - Engineering & design
  - Finance
  - Life sciences: genomics, drug discovery...
  - Security
  - Automotive
  - ...



# Analytics: a newer, complementary *big data* domain - BDA

- ∞ Analytics methods applied to established HPC domains in industry, government, academia ->
- ∞ High-end commercial analytics pushing up into HPC
- ∞ The path from science to industry and commerce can be relatively short:



# Closing a Gap between Big Data and Big Computing

## Ecosystems\*:

## Big Data

## Big Computing

- Orchestration **Crunch, Tez, Cloud Dataflow** →
- Libraries **MLib/Mahout, R, Python** →
- High-Level Programming **Pig, Hive, Drill** →
- Platform as a Service **App Engine, BlueMix, Elastic Beanstalk** →
- Languages **Java, Erlang, SQL, SparQL** →
- Streaming Parallel Runtime **Storm, Kafka, Kinesis** →
- Coordination **MapReduce** →
- Caching **Memcached** →
- Data Management **Hbase, Neo4J, MySQL** →
- Data Transfer **Sqoop** →
- Scheduling **Yarn** →
- File Systems **HDFS, Object Stores** →
- Formats **Thrift, Protobuf** →
- Virtualization **Openstack** →



- ← **Kepler, Pegasus**
- ← **Matlab, Eclipse, Apps**
- ← **Domain-specific Languages**
- ← **XSEDE Software Stack**
- ← **Fortran, C/C++**
- ← **MPI/OpnMP/OpenCL**
- ← **iRODS**
- ← **GridFTP**
- ← **Slurm**
- ← **Lustre**
- ← **FITS, HDF**
- ← **Docker, SR-IOV**

## Motivation:



**Spark**

## New Frontiers



**MPI**

## Leaders:

\*G. Fox et al. HPC-ABDC High Performance Computing Enhanced Apache Big Data Stack, CCGrid, 2015

# Hardware - flow of interest

Data-Intensive HPC

∞ **HPC vendors** are targeting commercial markets, driven by opportunity



Boundaries tend to dissolve

∞ **Commercial vendors** are moving up to HPC, driven by customers



High-End Commercial Analytics

# AI needs HPC and HPC needs AI

## ∞ Scaling brings AI progress!

- Best AI algorithms scale with systems!
- Make models bigger
- Use more data
- Reduce research cycle time

## ∞ Machine learning for *scientific applications* is a growing area!

- *In the high energy and astrophysics communities, simulation could mean examining images using deep learning*
- Machine Learning in HPC Environments 2019 - the 5<sup>th</sup> edition
  - held in conjunction with SuperComputing SC'19

# HPC <-> BDA

## ∞ HPC world

- realizes that there are more things in data storage than just files

## ∞ BDA world

- realizes that HPC can really speedup analytics

=>

- All major public Cloud services now have an HPC offering.
- Many academic HPC centers offer Cloud infrastructures and BDA tools.

## ∞ HPDA – High Performance Data Analytics

- using HPC for data-intensive challenges
- simulation + newer high-performance analytics



# HPC & BDA – master specialization

## ☞ Topics covered

Semester	High Performance Computing	Data Science
I	Distributed Operating Systems Concurrent Processes Modelling	Machine Learning Advanced Data Analysis
II	Parallel Programming Grid, Cluster and Cloud Computing	Data Visualization Formal Concept Analysis Data Streaming
III	Distributed & GPU Programming Optimisation models	Big Data Processing Data Mining Statistics
IV	Research Project Research Internship Dissertation Thesis	

## ☞ Computing infrastructure

Babeş-Bolyai University's HPC hybrid infrastructure (<http://hpc.cs.ubbcluj.ro>)

- HPC Cluster: IBM NextScale
- Private Cloud: IBM Flex System

# Sustained by Bosch starting from 2020 !!!



UNIVERSITATEA  
BABEŞ-BOLYAI



**BOSCH ACADEMIC**  
Brighten your future

**HPC&BDA**

## High Performance Computing and Big Data Analytics

**Master Program in:**

- Data Science and Big Data
- Parallel and Distributed Programming

Master Program organized by Babeş-Bolyai University, Faculty of Mathematics and Computer Science in partnership with Bosch Engineering Center Cluj

# Grants/Scholarships

- a) Susținerea studenților masteranzi din cadrul Programului la concursuri, congrese și conferințe studențești naționale și internaționale (taxă de participare, transport, diurnă, cazare, costuri aferente echipamentelor și materialelor necesare);
- b) Susținerea activităților de cercetare, practică și/sau internship în cadrul Centrului de Inginerie Bosch din Cluj-Napoca și în fabrica Bosch din Jucu
- c) implicarea în activitatea realizată de studenții masteranzi prin definirea de teme de cercetare și coordonare a activității studenților, împreună cu cadrele didactice;
- d) organizarea de stagii de cercetare, practică și/sau internship și pentru elaborarea lucrării de disertație;
- e) organizarea/susținerea de cursuri, seminarii sau ateliere de lucru pentru dezvoltarea competențelor și abilităților profesionale și transversale ale studenților masteranzi;
- f) asigurarea platii unui numar de taxe scolare pentru studentii nebugetari.

# International cooperation

## EUMaster4HPC

European Master For High Performance Computing

<https://eumaster4hpc.uni.lu/>

EUMaster4HPC is an HPC European consortium leading educational activities, funded by the [EuroHPC Joint Undertaking](#) to design and implement the first pan-European High Performance Computing (HPC) Master programme.