HPC & BD Services @ Uni.lu

Building up High Performance Computing & Big Data Competence Center to support national priorities

Dr. Sébastien Varrette

International Supercomputing Conference (ISC’18)
2nd Workshop on HPC Collaboration Between Europe and Latin America

Thursday, June 28th, 2018
Why HPC & BD?

- **HPC**: High Performance Computing
- **BD**: Big Data
- Essential tool for Science, Society and Industry

- All scientific disciplines are becoming computational today
  - modern scientific discovery requires very high computing power and handles huge volumes of data
  - cf. J. Rifkin report: "3rd Industrial Revolution Strategy for the Grand Duchy of Luxembourg"

- Industry and SMEs are increasingly relying on the power of supercomputers...
  - ... to invent innovative solutions while reducing cost and decreasing time to market

- HPC is part of a global race (recognized as a strategic priority) - EU is taking up the challenge
  - Ambitious plans from many countries (USA, China, Japan, Brazil & South America, India...) around HPC

To out-compete you must out-compute

Increasing competition, heightened customer expectations and shortening product development cycles are forcing the pace of acceleration across all industries.

Excellent research requires excellent research infrastructure
European HPC strategy and its implementation

- **EU HPC strategy** initiated in 2012
  - implementation within H2020 program

- Latest advances:
  - **IPCEI on HPC and Big Data (BD) Applications** (IPCEI-HPC-BDA) (Nov. 2015)
    - Luxembourg (leader), France, Italy & Spain
    - Testbed around Personalized Medicine, Smart Space, Industry 4.0 and Smart Manufacturing, New Advanced Materials, FinTech, Smart City...
  - **PRACE** (Partnership for Advanced Computing in Europe) transitioning to PRACE2.
    - Luxembourg: 25th country to join in Oct. 17th, 2017
      - Official Delegate/Advisor (P. Bouvry/S. Varrette) from Uni.lu
  - EU Member States sign the **EuroHPC initiative** and prepare its implementation (Mar. 2017)
    - A common effort to create and grow the European supercomputing ecosystem
    - Federation of national and regional HPC centers (see also PRACE/PRACE2)
      - Funding next-generation Peta-scale / Pre-exascale / Exascale systems
      - EuroHPC Joint Undertaking (JU) effective starting Jan. 2019
University of Luxembourg as a Jewel

● **Created in 2003, moved to Belval in 2015**
  ○ An international university serving its country

● **Ranked 12** among the young universities for the Times Higher Education Ranking in 2018

● **With regards to HPC, University of Luxembourg offers:**
  ○ **People**
    ■ Domain experts
    ■ Computational and data scientists
    ■ Specialists in parallel algorithmics
  ○ **Services**
    ■ HPC clusters and management team
    ■ IT team (IT department)
    ■ Infrastructure team (IT department) in collaboration with Fonds Belval
  ○ **Infrastructure**
    ■ Data center and a set high-end clusters
  ○ **Education & Training**
New Trends in HPC

- Continued scaling of scientific, industrial & financial applications
  - ... well beyond Exascale

- New trends changing the landscape for HPC
  - Emergence of Big Data analytics
  - Emergence of (Hyperscale) Cloud Computing
  - Data intensive Internet of Things (IoT) applications
  - Deep learning & cognitive computing paradigms
Toward Modular Computing

- Aiming at scalable, flexible HPC infrastructures
  - Primary processing on CPUs and accelerators
    - HPC & Extreme Scale Booster modules
  - Specialized modules for:
    - HTC & I/O intensive workloads
    - Data Analytics and AI
- 88%* of stakeholders will have multiple architectures
- Creates new adopters, targets also SME market

Next-generation HPC-BD platforms expected to increase modularity … and thus flexibility

* Source: Addison Snell, Intersect 360

Source: "Towards Modular Supercomputing: The DEEP and DEEP-ER take on Heterogeneous Cluster Architectures", Norbert Eicker, SC'16 (Nov. 2016)
HPC Facility @ Uni.lu

- Started in 2007, under responsibility of Prof. P. Bouvry & Dr. S. Varrette
  - expert UL HPC team
    - S. Varrette, V. Plugaru, S. Peter, H. Cartiaux, C. Parisot, among others
    - 8,173,747€ cumulative investment hardware (excl. Server rooms)
- Enables & accelerates scientific discovery & innovation
  - Largest HPC facility in Luxembourg w. GoodYear

HPC/Computing Capacity
423 TFlops → 1035,8 TFlops by eoy

HTC/Storage Capacity
9,8 PB shared
HPC Facility @ Uni.lu / Computing

- 5 clusters, 2 sites
- **Total computing capacity: 423 TFlops**
  - 10,130 CPU cores + 120,704 GPU cores
  - Planned extension to **1035,8 TFlops** by Q4 2018 (RFP 180027)
  - Fast interconnect based on Infiniband
    - ... typically over a non-blocking Fat-tree network topology
HPC Facility @ Uni.lu / Storage

Total shared storage capacity: 9.8 PB

- 3.24 PB on GPFS/SpectrumScale
- 1.94 PB on Lustre
- 4.67 PB on other FS (OneFS...)
  - incl. backups
Case Study 1: Material Science & Engineering

- **Companies & Research centers in Luxembourg**
  - GoodYear, IEE, Delphi, ArcelorMittal, ProNewTech, Prosciens, CrmClouder.com, UL, LIST...

- **Application domains**
  - Physics and Chemistry (materials design, new insights), Finite Element Analysis (FEA), Computational Fluid Dynamics (CFD), Optimization, Visualisation...

- **Computing infrastructure answering these needs**
  - Traditional (CPU only) or Hybrid (CPU + Accelerators)
Case Study 2: Biomedical Industry

- **Companies & Research centers in Luxembourg**
  - ITTM, IBBL, UL, LIH...

- **Potential to attract external companies**
  - Edico Genome (US), Fabric Genomics (US), Swarm64 (DE)...

- **Application domains**
  - System Bio-medicine, BD Analytics, Pharmacology, Personalised Medicine

- **Computing infrastructure answering these needs**
  - Traditional (CPU only), HTC, Hybrid (CPU + Accelerators)
  - [High Performance] Data Storage

Visualization of biological knowledge: Bi-level optimization/ Parkinson disease map. CSC/LCSB (UL)
Case Study 3: Deep Learning - Cognitive Computing

- Companies & Research centers in Luxembourg
  - Churchill Frank, SES, Aiva Technologies, IEE, UL...

- Potential to attract external companies
  - Amazon (US), Google (US), Uber (US), Tesla (US), Deepsense.AI (US)...

- Application domains
  - Data Mining, Self-Driving cars, Satellite & Communications, Big Data Analytics

- Computing infrastructure answering these needs
  - Traditional (CPU only), HTC, Hybrid HPC (CPUs + Accelerators)
Case Study 4: Data science, IoT and FinTech

- **Companies & Research centers in Luxembourg**
  - Big Four (E&Y, Deloitte, PwC, KPMG), ExaMotive, AXA, BIL, BCEE, UL...

- **Potential to attract external companies**
  - Amazon (US), NEST (US), Somfy (FR), Google (US)...

- **Application domains**
  - Risk & Asset Management, Data intensive IoT, Smart City

- **Computing infrastructure answering these needs**
  - Cloud Computing, HTC, HPC

Example 1: Financial Risk Analysis
Using Monte Carlo method On GPGPU

Chaos-enhanced mobility models for multilevel swarms of UAVs (PCOG)

FinTech ecosystem, from 2016 PwC report
“How FinTech is shaping Financial Services in Luxembourg”
HPC Facility @ Uni.lu usage

Statistics extracted on HPC active users (submitted at least one job in the last year)
HPC Facility @ Uni.lu usage

Statistics extracted on HPC active users (submitted at least one job in the last year)
HPC Facility @ Uni.lu usage

Statistics extracted on HPC active users (submitted at least one job in the last year)
Uni.lu CDC (Centre de Calcul): Toward Energy-Efficient HPC enabling DLC
Uni.lu CDC (Centre de Calcul): Toward Energy-Efficient HPC enabling DLC

- 2x500 m² deployed since 2015
  - Electric energy, ventilation and chilled water produced by Fonds Belval power plant (located basement MNO)
  - CDC managed by SIU, one floor dedicated to HPC developments
- 4 server rooms ready and operational in 2018
  - One room (CDC S-02-005) holds our (current) flagship HPC facility (iris)

### Data center ready for DLC with a dedicated high temperature network for energy efficiency

<table>
<thead>
<tr>
<th>Location</th>
<th>Cooling</th>
<th>Usage</th>
<th>Max Capacity [kW]</th>
<th>Max Storage [m²]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDC S-02-001</td>
<td>Airflow</td>
<td>Future extension</td>
<td>280 kW</td>
<td>(120 m²)</td>
</tr>
<tr>
<td>CDC S-02-002</td>
<td>Airflow</td>
<td>Storage / Traditional HPC /Cloud/FPGA</td>
<td>280 kW</td>
<td>(88 m²)</td>
</tr>
<tr>
<td><strong>CDC S-02-003</strong></td>
<td><strong>DLC</strong></td>
<td><strong>High Density/Energy efficient HPC</strong></td>
<td><strong>1050 kW</strong></td>
<td>(90 m²)</td>
</tr>
<tr>
<td>CDC S-02-004</td>
<td>DLC</td>
<td><strong>High Density/Energy efficient HPC</strong></td>
<td><strong>1050 kW</strong></td>
<td>(92 m²)</td>
</tr>
<tr>
<td>CDC S-02-005</td>
<td>Airflow</td>
<td>Storage / Traditional HPC (iris cluster)</td>
<td>300 kW</td>
<td>(128 m²)</td>
</tr>
</tbody>
</table>
Education & Training

- **COST NESUS Winter School** on Data Science and Heterogeneous Computing
  - Last edition in January 2018 in Zagreb, Croatia

- **PRACE**
  - Last edition PRACE Days in May 2018: European HPC Summit Week 2018 in Ljubljana
  - H2020-INFRAEDI-2018-2020 - PRACE-6IP

- **Bachelor degree BICS/BINFO: AI & Middleware**

- **Master degree MICS: Parallel and Grid Computing**
  - Master in Information and Computer Sciences - see misc.uni.lu
  - Master 2 courses, lectured since 2008
  - Evaluation based on projects evaluated on the UL HPC platform. Ex:
    - Fault Tolerance and Performance of MPI toolchains over HPC Applications
    - Measurement and Optimization of Energy Consumption in HPC Applications
    - Parallelization of the Barnes-Hut N-Body Simulation Algorithm
    - Evaluation of Big Data Framework (Hadoop, Spark...)

- **Lifelong-learning Smart-ICT for Business Innovation certificate**: Cloud Computing/Big Data/IoT
### Education & Training

- **SC-Camp 2017 (Cadiz)**
- **Bi-annual HPC School @ Uni.lu** (Part of the doctoral program)
  - 7th edition June 12th - 13th, 2018 in Belval
  - Material publically available: [ulhpc-tutorials.rtfd.io](ulhpc-tutorials.rtfd.io)
  - Last edition joint event: UL-NSTDA workshop (EU-ASEAN E-READI)

#### Day 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Main Track (MSA 4,520)</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>9h00</td>
<td>PS1a: Getting Started (part I: SSH)</td>
<td>C. Parisot</td>
</tr>
<tr>
<td>10h00</td>
<td>Coffee break</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>11h45</td>
<td>Overview &amp; Challenges of the UL HPC Facility at the EuroHPC Horizon</td>
<td>H. Cartiaux</td>
</tr>
<tr>
<td>12h30</td>
<td>LUNCH</td>
<td>H. Cartiaux</td>
</tr>
<tr>
<td>13h30</td>
<td>PS2: HPC workflow with sequential jobs</td>
<td>H. Cartiaux</td>
</tr>
<tr>
<td>14h30</td>
<td>PS4a: UL HPC Monitoring in practice</td>
<td>H. Cartiaux</td>
</tr>
<tr>
<td>15h30</td>
<td>Coffee break</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>16h00</td>
<td>PS5: Parallel computations with OpenMP/MPI</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>17h30</td>
<td>PS6: User environment and storage data management</td>
<td>S. Peter</td>
</tr>
</tbody>
</table>

#### Day 1

<table>
<thead>
<tr>
<th>Time</th>
<th>Parallel Track (MSA 4,410)</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>10h30</td>
<td>Keynote in 4,520</td>
<td>V. Plugaru</td>
</tr>
<tr>
<td>13h30</td>
<td>PS3a: Advanced scheduling (SLURM, OAR)</td>
<td>V. Plugaru</td>
</tr>
<tr>
<td>14h30</td>
<td>PS3b: Software environment management using Easybuild</td>
<td>S. Peter</td>
</tr>
<tr>
<td>15h30</td>
<td>Coffee break (in 4,520)</td>
<td>V. Plugaru</td>
</tr>
<tr>
<td>16h00</td>
<td>PS4b: Performance engineering - HPC debugging and profiling</td>
<td>V. Plugaru</td>
</tr>
</tbody>
</table>

#### Day 2

<table>
<thead>
<tr>
<th>Time</th>
<th>Main Track (MSA 4,520)</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>9h00</td>
<td>PS7: Multi-Physics workflows: test cases on CFD/MD/Chemistry applications</td>
<td>V. Plugaru</td>
</tr>
<tr>
<td>10h30</td>
<td>Coffee break</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>11h00</td>
<td>Users’ session: UL HPC experiences</td>
<td>C. Parisot</td>
</tr>
<tr>
<td>12h40</td>
<td>LUNCH</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>13h30</td>
<td>PS9: [Basic + Advanced] Prototyping with Python</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>15h30</td>
<td>Coffee break</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>17h15</td>
<td>PS11: Big Data Applications</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>18h15</td>
<td>PS13: Machine / Deep learning (Pytorch, Tensorflow, Caffe2)</td>
<td>S. Varrette</td>
</tr>
</tbody>
</table>

- **Closing Keynote: Take Away Messages**

#### Day 2

<table>
<thead>
<tr>
<th>Time</th>
<th>Parallel Track (MSA 4,510)</th>
<th>Speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>9h00</td>
<td>PS8: Bio-informatics workflows and applications</td>
<td>S. Peter</td>
</tr>
<tr>
<td>10h30</td>
<td>Coffee break (in 4,520)</td>
<td>V. Plugaru</td>
</tr>
<tr>
<td>11h00</td>
<td>Users’ session: UL HPC experiences (in 4,520)</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>12h40</td>
<td>LUNCH</td>
<td>S. Varrette</td>
</tr>
<tr>
<td>13h30</td>
<td>PS10: Scientific computing using MATLAB</td>
<td>V. Plugaru</td>
</tr>
<tr>
<td>15h30</td>
<td>Coffee break</td>
<td>A. Ginei/hac</td>
</tr>
<tr>
<td>16h00</td>
<td>PS12: R – statistical computing</td>
<td>V. Plugaru</td>
</tr>
<tr>
<td>18h15</td>
<td>Closing Keynote in 4,520</td>
<td>S. Varrette</td>
</tr>
</tbody>
</table>
Next steps: National HPC-BD Competence Center

Built by ministerial, academic, industrial stakeholders

- **Comprehensive centre:**
  - HPC and data infrastructure
  - Expertise in technology
  - Domain knowledge in applications

- **More than just computing services:**
  - “Bring HPC and BD to the users”

- **Creates twofold innovation:**
  - Innovation in applications and
  - Innovation in technology

Credits: P. Bouvry, S. Varrette, V. Plugaru & UL HPC Team

Icons sourced from www.flaticon.com
Different Needs for Different Domains

Material Science & Engineering

Diagram showing the relationship between #Cores, Flops/Core, Storage Capacity, Network Latency, and I/O Performance.
Different Needs for Different Domains

Biomedical Industry / Life Sciences
Different Needs for Different Domains

**Deep Learning / Cognitive Computing**

- Network Bandwidth
- Network Latency
- Flops/Core
- Storage Capacity
- I/O Performance
- #Cores
Different Needs for Different Domains

IoT, FinTech

- Network Bandwidth
- Network Latency
- Flops/Core
- Storage Capacity
- I/O Performance
- #Cores
Different Needs for Different Domains

ALL Research Computing Domains

- #Cores
- Flops/Core
- Storage Capacity
- Network Bandwidth
- Network Latency
- I/O Performance