SRA – update, 
Workprogramme 2018-2020, 
European Cloud Initiative, 

cPPP meeting, HPC Summit Prague, May 11\textsuperscript{th} 2016
Strategic Research Agenda

SRA

a multi-annual roadmap towards Exascale High-Performance Computing Capabilities
Horizon 2020 WPs and SRAs
Priorities

• There is a demand for R&D and innovation in both extreme performance systems and mid-range HPC systems
  – Scientific domain and some industrial users want extreme scale
  – ISVs and part of the industry expect more usability and affordability of mid-range system

• The ETP4HPC HPC technology providers are also convinced that to build a sustainable ecosystem,
  – their R&D investments should target not only the exascale objective (too narrow a market)
  – an approach that aims at developing technologies capable of serving both the extreme-scale requirements and mid-market needs can be successful in strengthening Europe’s position.
4 dimensions of the SRA

- **HPC SYSTEM ARCHITECTURE**
- **SYSTEM SOFTWARE AND MANAGEMENT**
- **PROGRAMMING ENVIRONMENT**
  Including: Support for extreme parallelism
- **MATHEMATICS & ALGORITHMS FOR EXTREME SCALE HPC SYSTEMS**
  --- NEW ---

---

- **HPC SERVICES**
  INCLUDING: ISV support, End-user support
- **SME FOCUS**
- **EDUCATION AND TRAINING**

---

- **HPC USAGE EXPANSION**
- **HPC STACK ELEMENTS**
- **NEW HPC DEPLOYMENTS**
- **EXTREME SCALE REQUIREMENTS**

---

- **HPC USAGE MODELS**
  Including: Big data, HPC in clouds
- **IMPROVE SYSTEM AND ENVIRONMENT CHARACTERISTICS**
  Including: Energy efficiency, System resilience
- **BALANCE COMPUTE SUBSYSTEM, I/O AND STORAGE PERFORMANCE**

---

[ETP 4 HPC]
Transversal issues to be addressed

• Three technical topics:
  – Security in HPC infrastructures to support increasing deployment of HPDA
  – Resource virtualisation to increase flexibility and robustness
  – HPC in clouds to facilitate ease of access

• Two key element for HPC expansion
  – Usability at growing scale and complexity
  – Affordability (focus on TCO)
How was the SRA been built?

8 Workgroups covering the 8 technical focus areas:

**SRA 2015 technical focus areas**

- HPC System Architecture and Components
- Energy and Resiliency
- Programming Environment
- System Software and Management
- Big Data and HPC usage Models
- Balance Compute, I/O and Storage Performance
- Mathematics and algorithms for extreme scale HPC systems
- Extreme scale demonstrators

- 48 ETP4HPC member orgs/companies involved in these workgroups
- Members named 170 individual experts to contribute, 20-30 per working group
Other interactions

• Feedback sessions with end-users and ISVs at Teratec Forum
  – 20 end-users outline their deployment of HPC, future plans and technical recommendations
  – Very diverse set of priorities (performance & scale, robustness, ease of access, new workflows etc.)
  – No ‘One size fits all’ – approach possible

• Technical session with Big Data Value Association (BDVA) to understand architectural influences of HPDA
  – Technical dialogue started, much more to be done over next 1-2 years
  – BDVA has issued an update to their SRIA in Jan 2016
The technical domains and the ESD proposal

Trends and recommended research topics – a few examples
HPC System Architecture, Storage and I/O, Energy and Resiliency

• Major trends - a subset:
  – Increased use of accelerators (e.g. GPUs, many core CPUs) in heterogeneous system architectures
  – Compute node architectures efficiently integrate accelerators, CPUs with high bandwidth memory
  – Non volatile memory types open up new interesting memory and caching hierarchy designs
  – System networks to significantly scale up and cut latencies, introducing virtualisation mechanisms
  – Storage subsystems to become more ‘intelligent’ to better balance compute and I/O
  – Increased activities in object storage technologies with major architectural revamp in the next years
  – Focus on architectural changes to improve energy efficiency and reduce data movement

• Research topics to be addressed (examples)
  – Compute node deep integration with embedded fast memory and memory coherent interfaces
  – Silicon photonics and photonic switching in HPC system networks
  – Global energy efficiency increases with targets of 60kW/PFlops in 2018 and 35 kW in 2020
  – Active storage technologies to enable ‘in situ’ and ‘on the fly’ data processing
  – Research in methods to manage ‘energy to solution’
  – Prediction of failures and fault prediction algorithms
## HPC System Architecture, Storage and I/O: milestones

<table>
<thead>
<tr>
<th>M-ARCH-1: New HPC processing units enable wide-range of HPC applications.</th>
<th>2018</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-ARCH-2: Faster memory integrated with HPC processors.</td>
<td>2018</td>
</tr>
<tr>
<td>M-ARCH-3: New compute nodes and storage architecture use NVRAM.</td>
<td>2017</td>
</tr>
<tr>
<td>M-ARCH-4: Faster network components with 2x signalling rate (rel. to 2015) and lower latency available.</td>
<td>2018</td>
</tr>
<tr>
<td>M-ARCH-5: HPC networks efficiency improved.</td>
<td>2018</td>
</tr>
<tr>
<td>M-ARCH-6: New programming languages support in place.</td>
<td>2018</td>
</tr>
<tr>
<td>M-ARCH-7: Exascale system energy efficiency goals (35kW/PFlops in 2020 or 20 kW/Pflops in 2023) reached.</td>
<td>2020-2023</td>
</tr>
<tr>
<td>M-ARCH-8: Virtualisation at all levels of HPC systems.</td>
<td>2018</td>
</tr>
<tr>
<td>M-ARCH-10: New components / disruptive architectures for HPC available.</td>
<td>2019</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>M-BIO-1: Tightly coupled Storage Class Memory IO systems demo.</th>
<th>2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-BIO-2: Common I/O system simulation framework established.</td>
<td>2017</td>
</tr>
<tr>
<td>M-BIO-3: Multi-tiered heterogeneous storage system demo.</td>
<td>2018</td>
</tr>
<tr>
<td>M-BIO-4: Advanced IO API released: optimised for multi-tier IO and object storage.</td>
<td>2018</td>
</tr>
<tr>
<td>M-BIO-5: Big Data analytics tools developed for HPC use.</td>
<td>2018</td>
</tr>
<tr>
<td>M-BIO-6: ‘Active Storage’ capability demonstrated.</td>
<td>2018</td>
</tr>
<tr>
<td>M-BIO-7: I/O quality-of-Service capability.</td>
<td>2019</td>
</tr>
<tr>
<td>M-BIO-8: Extreme scale multi-tier data management tools available.</td>
<td>2019</td>
</tr>
<tr>
<td>M-BIO-9: Meta-Data + Quality of Service exascale file i/o demo.</td>
<td>2020</td>
</tr>
<tr>
<td>M-BIO-10: IO system resiliency proven for exascale capable systems.</td>
<td>2021</td>
</tr>
</tbody>
</table>
Energy and resiliency: milestones

<table>
<thead>
<tr>
<th>Milestone Code</th>
<th>Description</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-ENR-MS-1</td>
<td>Quantification of computational advance and energy spent on it.</td>
<td>2017</td>
</tr>
<tr>
<td>M-ENR-MS-2</td>
<td>Methods to steer the energy spent.</td>
<td>2017</td>
</tr>
<tr>
<td>M-ENR-MS-3</td>
<td>Use of idle time to increase efficiency.</td>
<td>2018</td>
</tr>
<tr>
<td>M-ENR-AR-4</td>
<td>New levels of memory hierarchy to increase resiliency of computation.</td>
<td>2017</td>
</tr>
<tr>
<td>M-ENR-FT-5</td>
<td>Collection and Analysis of statistics related to failures.</td>
<td>2018</td>
</tr>
<tr>
<td>M-ENR-FT-6</td>
<td>Prediction of failures and fault prediction algorithms.</td>
<td>2019</td>
</tr>
<tr>
<td>M-ENR-FT-10</td>
<td>Application survival on unreliable hardware.</td>
<td>2019</td>
</tr>
<tr>
<td>M-ENR-AR-7</td>
<td>Quantification of savings from trade between energy and accuracy.</td>
<td>2018</td>
</tr>
<tr>
<td>M-ENR-AR-8</td>
<td>Power efficient numerical libraries.</td>
<td>2019</td>
</tr>
<tr>
<td>M-ENR-MS-9</td>
<td>Demonstration of a sizable HPC installation with explicit efficiency targets.</td>
<td>2019</td>
</tr>
</tbody>
</table>
Extreme-Scale Demonstrators

• Characteristics
  – Complete prototype HPC systems
  – high enough TRL to support stable production
  – using technologies developed in the previous projects
  – based on application – system co-design approach
  – large enough to address scalability issues (at least 5% of top performance systems at that time)

• Two project phases:
  – phase A : development, integration (of results from R&D projects) and testing
  – phase B : deployment and use, code optimisation, assessment of the new technologies
Extreme scale Demonstrators call-integration-deployment schedule

WP 2014/15 (project execution)

WP 2016 (project execution)

EsD 1-2 projects

EsD 1

EsD 2

EsD call

4Q 2017

2Q 2018

EsD integration

1Q 2020

EsD deployment

2Q 2022

EsD 3-4 projects

EsD 3

EsD 4

EsD call

4Q 2018

2Q 2019

EsD integration

1Q 2021

EsD deployment

2Q 2023
SRA – next actions
Google

« Public Call for comments on SRA “

We will welcome your comments on the current SRA
http://www.etp4hpc.eu/strategic-research-agenda/
Next SRA-related events in 1H2016

• HPC summit – Extreme scale Demonstrator workshop - May 12th
  – focussed on the EsD definition (engage potential players, further implementation details)
  – at this event the three pillars for the EsD mission ( CoE, HPC centres and the FETHPC1 project speakers ) are invited. More than 80 registered participants!

• Participation in BDEC conference - June 16 & 17

• ISC16 – June 23rd
  – Scope: Feedback session on SRA directions, content and value to shape the next update
    (Invited are: End-users, ISVs and International HPC experts)
  – 2nd EsD workshop (follow-on to May 12th workshop)

• Level set with HPC application experts (EXDCI WP3) – September 21 & 22
• Technical workshop with Big Data Value Association (BDVA)  June/July
Workprogramme 2018-2020, European Cloud Initiative
Workprogramme 2018 – 2020 discussion - topics


• European Exascale: timing mismatch between expectations and investment so far

• ETP4HPC recommendation based on:
  – Diversity of System Architectures
  – Top priority: European system architectural leadership: Grow basic “know-how” and expertise
  – SME and start-ups require support for entry and participation in larger H2020 research projects

• Workprogramme elements:
  – Research in HPC Technology
  – Extreme-scale Demonstrators
  – Centres of Excellence for Computing Applications (CoEs)
  – Continuation and extension of support actions
Workprogramme 2018-2020 budget recommendations

<table>
<thead>
<tr>
<th>Area</th>
<th>Suggested funding volume (m€)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technology</strong></td>
<td></td>
</tr>
<tr>
<td>Focussed projects high TRL</td>
<td><strong>170 - 270</strong></td>
</tr>
<tr>
<td>Focussed projects low TRL</td>
<td>130 - 180</td>
</tr>
<tr>
<td>Extreme scale Demonstrators</td>
<td><strong>100 - 200</strong></td>
</tr>
<tr>
<td>Two asap, two incl. WP16 results</td>
<td></td>
</tr>
<tr>
<td><strong>Centres of Excellence</strong></td>
<td><strong>70</strong></td>
</tr>
<tr>
<td>Existing (after merge)</td>
<td>45</td>
</tr>
<tr>
<td>New</td>
<td>25</td>
</tr>
<tr>
<td><strong>Support actions</strong></td>
<td><strong>8</strong></td>
</tr>
<tr>
<td>HPC eco system development incl. Joint actions with Big Data and Cloud Computing</td>
<td>6</td>
</tr>
<tr>
<td>International cooperation</td>
<td>2</td>
</tr>
</tbody>
</table>
“European Cloud Initiative” discussion - topics

• From “European Cloud Initiative - Building a competitive data and knowledge economy in Europe”

• “......realising exascale supercomputers around 2022, based on EU technology, which would rank in the first 3 places of the world (p8)

• foster an HPC ecosystem capable of developing new European technology such as low power HPC chips (p9)

• The Commission and participating Member States should develop and deploy a large scale European HPC, data and network infrastructure, including (p10):
  – the acquisition of two co-designed, prototype exascale supercomputers and two operational systems which will rank in the top three of the world – as of 2018
  – the establishment of a European Big Data centre – as of 2016
Potential paths to commercial systems

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WP 2014 FET HPC 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP 2016 FET HPC 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP 2017 FET HPC 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP2018 ESD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP2019-2020 ESD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP2018 integration project</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WP2019-2020 additional low RTL research</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPC system acquisition mentioned in EC communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
THANK YOU!

For more information visit

www.etp4hpc.eu

contact: office@etp4hpc.eu