Swiss National Programs
ATIP Workshop on International Exascale and Next-Generation Computing Programs, SC17

Colin McMurtrie and Thomas Schulthess, CSCS

13th November 2017
High-Performance Computing Initiative (HPCN) in Switzerland
High-Performance Computing Initiative (HPCN) in Switzerland

- 2009: Monte Rosa
  - Cray XT5
  - 14'762 cores

- 2010: Hex-core upgrade
  - 22'128 cores

- 2011: Upgrade to Cray XE6
  - 47,200 cores


- 2015:

- 2016:

- 2017:
High-Performance Computing Initiative (HPCN) in Switzerland

- **Monte Rosa Cray XT5**
  - 14'762 cores
- **Hex-core upgrade**
  - 22'128 cores
- **Upgrade to Cray XE6**
  - 47,200 cores

- **Begin construction of new building**
  - 2010
  - New building complete
  - 2011

- **Upgrade timeline**
  - 2012:
    - 2012
  - 2013:
    - 2013
  - 2014:
    - 2014
  - 2015:
    - 2015
  - 2016:
    - 2016
  - 2017:
    - 2017
Monte Rosa
Cray XT5
14’762 cores

Hex-core upgrade
22’128 cores

Upgrade to
Cray XE6
47,200 cores

New building complete

Begin construction of new building

High-Performance Computing Initiative (HPCN) in Switzerland

Swiss HP2C Platform
High-risk & high-impact projects
(www.hp2c.ch)
High-Performance Computing Initiative (HPCN) in Switzerland

Three pronged approach of the HPCN Initiative

Monte Rosa
Cray XT5
14’762 cores

Hex-core upgrade
22’128 cores

Upgrade to
Cray XE6
47,200 cores

Begin construction
of new building

New building complete

Swiss HP2C Platform

High-risk & high-impact projects
(www.hp2c.ch)
High-Performance Computing Initiative (HPCN) in Switzerland

Three pronged approach of the HPCN Initiative

1. New, flexible, and efficient building

Monte Rosa
Cray XT5
14,762 cores

Hex-core upgrade
22,128 cores

Upgrade to
Cray XE6
47,200 cores

2009
Begin construction of new building

2010

2011

2012

2013

2014

2015

2016

2017

3. New, flexible, and efficient building

Swiss HP2C Platform
High-risk & high-impact projects
(www.hp2c.ch)
High-Performance Computing Initiative (HPCN) in Switzerland

Three pronged approach of the HPCN Initiative
1. New, flexible, and efficient building
2. Efficient supercomputers

**Monte Rosa**
- Cray XT5
- 14’762 cores

**Hex-core upgrade**
- 22’128 cores

**Upgrade to Cray XE6**
- 47,200 cores

**Cray XT5**
- 14’762 cores

**Swiss HP2C Platform**
High-risk & high-impact projects
(www.hp2c.ch)

**Begin construction of new building**

**New building complete**

**High-Performance Computing Initiative (HPCN) in Switzerland**

2009
- Begin construction of new building

2010
- New building complete

2011
- Upgrade to Cray XE6
- 47,200 cores

2012

2013

2014

2015

2016

2017
High-Performance Computing Initiative (HPCN) in Switzerland

Three pronged approach of the HPCN Initiative
1. New, flexible, and efficient building
2. Efficient supercomputers
3. Efficient applications

Monte Rosa
Cray XT5
14'762 cores

Hex-core upgrade
22'128 cores

Upgrade to
Cray XE6
47,200 cores

Begin construction
of new building

New building complete

2009
2010
2011
2012
2013
2014
2015
2016
2017
Three pronged approach of the HPCN Initiative:
1. New, flexible, and efficient building
2. Efficient supercomputers
3. Efficient applications

Monte Rosa
Cray XT5
14,762 cores

Hex-core upgrade
22,128 cores

Upgrade to
Cray XE6
47,200 cores

Application driven co-design
of pre-exascale supercomputing ecosystem

High-risk & high-impact projects
(www.hp2c.ch)

Swiss HP2C Platform

CSCS
Platform for Advanced Scientific Computing

Platform for Advanced Scientific Computing

Begin construction of new building

New building complete

2009

2010

2011

2012

2013

2014

2015

2016

2017
High-Performance Computing Initiative (HPCN) in Switzerland

Three pronged approach of the HPCN Initiative
1. New, flexible, and efficient building
2. Efficient supercomputers
3. Efficient applications

2009
- Begin construction of new building

2010
- New building complete

2011
- Hex-core upgrade
- 22,128 cores
- Upgrade to Cray XE6
- 47,200 cores

2012
- Development & procurement of petaflop/s scale supercomputer(s)

2013
- Application driven co-design of pre-exascale supercomputing ecosystem

2014

2015

2016

2017

Monte Rosa
Cray XT5
14,762 cores

Swiss HP2C Platform
High-risk & high-impact projects (www.hp2c.ch)

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

Platform for Advanced Scientific Computing

Upgrade to
Cray XE6
47,200 cores

Platform for Advanced Scientific Computing
High-Performance Computing Initiative (HPCN) in Switzerland

Monte Rosa
Cray XT5
14’762 cores

Hex-core upgrade
22’128 cores

Upgrade to
Cray XE6
47,200 cores

Aries network & multi-core

Development & procurement of petaflop/s scale supercomputer(s)

Three pronged approach of the HPCN Initiative
1. New, flexible, and efficient building
2. Efficient supercomputers
3. Efficient applications

High-risk & high-impact projects
(www.hp2c.ch)

Application driven co-design
of pre-exascale supercomputing ecosystem

Begin construction of new building
New building complete

2011

2010

2009

2013

2014

2015

2016

2017
High-Performance Computing Initiative (HPCN) in Switzerland

Three pronged approach of the HPCN Initiative
1. New, flexible, and efficient building
2. Efficient supercomputers
3. Efficient applications

Phase I
- 2010: Begin construction of new building
- 2011: Aries network & multi-core

Phase II
- 2013: Development & procurement of petaflop/s scale supercomputer(s)
- 2014: K20X based hybrid
- 2015: New building complete
- 2016: Phase II
- 2017: Upgrade to Cray XE6

Hex-core upgrade
- 2009: 14,762 cores
- 2010: 22,128 cores

Swiss HP2C Platform
High-risk & high-impact projects (www.hp2c.ch)

CSCS
Centre Svizzer di Calcolo Scientifico
Swiss National Supercomputing Centre
High-Performance Computing Initiative (HPCN) in Switzerland

Three pronged approach of the HPCN Initiative
1. New, flexible, and efficient building
2. Efficient supercomputers
3. Efficient applications

Monte Rosa
Cray XT5
14’762 cores

Hex-core upgrade
22’128 cores

Phase I
2010
Begin construction of new building

Phase II
2011
Aries network & multi-core

Upgrade to Cray XE6
47,200 cores

Phase III
2014
K20X based hybrid

2015
Development & procurement of petaflop/s scale supercomputer(s)

2016
Application driven co-design of pre-exascale supercomputing ecosystem

2017
Pascal based hybrid

Monte Rosa
Cray XT5
14’762 cores

Upgrade to Cray XE6
47,200 cores

Three pronged approach of the HPCN Initiative
1. New, flexible, and efficient building
2. Efficient supercomputers
3. Efficient applications

High-risk & high-impact projects
(www.hp2c.ch)

Swiss HP2C Platform

Platform for Advanced Scientific Computing

CSCS
Centro Svizzero di Calcolo Scientifico
Swiss National Supercomputing Centre

T. Schulthess
2
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

Model Cray XC40/Cray XC50

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hybrid Compute Nodes</td>
<td>5,320</td>
</tr>
<tr>
<td>Number of Multicore Compute Nodes</td>
<td>1,431</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Hybrid Node</td>
<td>4.761 Teraflops Intel Xeon E5-2690 v3/Nvidia Tesla P100</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Multicore Node</td>
<td>1.210 Teraflops Intel Xeon E5-2695 v4</td>
</tr>
<tr>
<td>Theoretical Hybrid Peak Performance</td>
<td>25.326 Petaflops</td>
</tr>
<tr>
<td>Theoretical Multicore Peak Performance</td>
<td>1.731 Petaflops</td>
</tr>
<tr>
<td>Hybrid Memory Capacity per Node</td>
<td>64 GB; 16 GB CoWoS HBM2</td>
</tr>
<tr>
<td>Multicore Memory Capacity per Node</td>
<td>64 GB; 128 GB</td>
</tr>
<tr>
<td>Total System Memory</td>
<td>437.9 TB; 83.1 TB</td>
</tr>
<tr>
<td>System Interconnect</td>
<td>Cray Aries routing and communications ASIC, and Dragonfly network topology</td>
</tr>
<tr>
<td>Sonexion 3000 Storage Capacity</td>
<td>6.2 PB</td>
</tr>
<tr>
<td>Sonexion 3000 Parallel File System Theoretical Peak Performance</td>
<td>112 GB/s</td>
</tr>
<tr>
<td>Sonexion 1600 Storage Capacity</td>
<td>2.5 PB</td>
</tr>
<tr>
<td>Sonexion 1600 Parallel File System Theoretical Peak Performance</td>
<td>138 GB/s</td>
</tr>
</tbody>
</table>

http://www.cscs.ch/publications/fact_sheets/index.html
Euclid Flagship Simulation 2016

Full sky map of the dark matter structure at ½ the age of the Universe. This structure will distort the shapes of more distant galaxies due to weak gravitational lensing.

2 trillion particles using all of available memory on Piz Daint and observing about 25 billion virtual galaxies (*)


(* this catalogue is being used to calibrate the experiments on board the Euclid satellite that will be launched in 2020 with the objective of investigating the nature of dark matter and dark energy)
Nucleon spin and momentum decomposition


- Lattice QCD simulation on “Piz Daint”
- Nucleons, such as protons or neutrons are made of quarks and gluons
- Only 1% of mass is due to mass of the three quarks; 99% is attributed to strong nuclear force
- First time a computation included gluons to determine total angular momentum of nucleon (broken down by different contributions)

Illustration: courtesy Brookhaven National Laboratory
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

Model Cray XC40/Cray XC50

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hybrid Compute Nodes</td>
<td>5,320</td>
</tr>
<tr>
<td>Number of Multicore Compute Nodes</td>
<td>1,431</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Hybrid Node</td>
<td>4.761 Teraflops Intel Xeon E5-2690 v3/Nvidia Tesla P100</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Multicore Node</td>
<td>1.210 Teraflops Intel Xeon E5-2695 v4</td>
</tr>
<tr>
<td>Theoretical Hybrid Peak Performance</td>
<td>25.326 Petaflops</td>
</tr>
<tr>
<td>Theoretical Multicore Peak Performance</td>
<td>1.731 Petaflops</td>
</tr>
<tr>
<td>Hybrid Memory Capacity per Node</td>
<td>64 GB; 16 GB CoWoS HBM2</td>
</tr>
<tr>
<td>Multicore Memory Capacity per Node</td>
<td>64 GB; 128 GB</td>
</tr>
<tr>
<td>Total System Memory</td>
<td>437.9 TB; 83.1 TB</td>
</tr>
<tr>
<td>System Interconnect</td>
<td>Cray Aries routing and communications ASIC, and Dragonfly network topology</td>
</tr>
<tr>
<td>Sonexion 3000 Storage Capacity</td>
<td>6.2 PB</td>
</tr>
<tr>
<td>Sonexion 3000 Parallel File System Theoretical Peak Performance</td>
<td>112 GB/s</td>
</tr>
<tr>
<td>Sonexion 1600 Storage Capacity</td>
<td>2.5 PB</td>
</tr>
<tr>
<td>Sonexion 1600 Parallel File System Theoretical Peak Performance</td>
<td>138 GB/s</td>
</tr>
</tbody>
</table>

http://www.cscs.ch/publications/fact_sheets/index.html
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

Model Cray XC40/Cray XC50

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hybrid Compute Nodes</td>
<td>5,320</td>
</tr>
<tr>
<td>Number of Multicore Compute Nodes</td>
<td>1,431</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Hybrid Node</td>
<td>4.761 Teraflops Intel Xeon E5-2690 v3/Nvidia Tesla P100</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Multicore Node</td>
<td>1.210 Teraflops Intel Xeon E5-2695 v4</td>
</tr>
<tr>
<td>Theoretical Hybrid Peak Performance</td>
<td></td>
</tr>
<tr>
<td>Theoretical Multicore Peak Performance</td>
<td></td>
</tr>
<tr>
<td>Hybrid Memory Capacity per Node</td>
<td></td>
</tr>
<tr>
<td>Multicore Memory Capacity per Node</td>
<td></td>
</tr>
<tr>
<td>Total System Memory</td>
<td></td>
</tr>
<tr>
<td>System Interconnect</td>
<td></td>
</tr>
<tr>
<td>Sonexion 3000 Storage Capacity</td>
<td></td>
</tr>
<tr>
<td>Sonexion 3000 Parallel File System</td>
<td></td>
</tr>
<tr>
<td>Sonexion 1600 Storage Capacity</td>
<td></td>
</tr>
<tr>
<td>Sonexion 1600 Parallel File System Theoretical Peak Performance</td>
<td>138 GB/s</td>
</tr>
</tbody>
</table>

Customers using Piz Daint

http://www.cscs.ch/publications/fact_sheets/index.html
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

<table>
<thead>
<tr>
<th>Model Cray XC40/Cray XC50</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hybrid Compute Nodes</td>
<td>5 320</td>
</tr>
<tr>
<td>Number of Multicore Compute Nodes</td>
<td>1 431</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Hybrid Node</td>
<td>4.761 TFlops Intel Xeon E5-2690 v3/Nvidia Tesla P100</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Multicore Node</td>
<td>1.210 TFlops Intel Xeon E5-2695 v4</td>
</tr>
<tr>
<td>Theoretical Hybrid Peak Performance</td>
<td>25.326 PFlops</td>
</tr>
<tr>
<td>Theoretical Multicore Peak Performance</td>
<td>1.731 PFlops</td>
</tr>
<tr>
<td>Hybrid Memory Capacity per Node</td>
<td>64 GB; 16 GB CoWoS HBM2</td>
</tr>
<tr>
<td>Multicore Memory Capacity per Node</td>
<td>64 GB, 128 GB</td>
</tr>
<tr>
<td>Total System Memory</td>
<td>437.9 TB; 83.1 TB</td>
</tr>
<tr>
<td>System Interconnect</td>
<td>Cray Aries routing and communications ASIC, and Dragonfly network topology</td>
</tr>
<tr>
<td>Sonexion 3000 Storage Capacity</td>
<td>6.2 PB</td>
</tr>
<tr>
<td>Sonexion 3000 Parallel File System Theoretical Peak Performance</td>
<td>112 GB/s</td>
</tr>
<tr>
<td>Sonexion 1600 Storage Capacity</td>
<td>2.5 PB</td>
</tr>
<tr>
<td>Sonexion 1600 Parallel File System Theoretical Peak Performance</td>
<td>138 GB/s</td>
</tr>
</tbody>
</table>

Customers using Piz Daint

- User Lab (including PRACE Tier-0 allocations)

http://www.cscs.ch/publications/fact_sheets/index.html
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

Model Cray XC40/Cray XC50

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hybrid Compute Nodes</td>
<td>5320</td>
</tr>
<tr>
<td>Number of Multicore Compute Nodes</td>
<td>1431</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Hybrid Node</td>
<td>4.761 Pflops Intel Xeon E5-2690 v3/Nvidia Tesla P100/P100</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Multicore Node</td>
<td>1.210 Pflops Intel Xeon E5-2695 v4</td>
</tr>
<tr>
<td>Theoretical Hybrid Peak Performance</td>
<td>25.326 Pflops</td>
</tr>
<tr>
<td>Theoretical Multicore Peak Performance</td>
<td>1.731 Pflops</td>
</tr>
<tr>
<td>Hybrid Memory Capacity per Node</td>
<td>64 GB; 16 GB CoWoS HBM2</td>
</tr>
<tr>
<td>Multicore Memory Capacity per Node</td>
<td>64 GB, 128 GB</td>
</tr>
<tr>
<td>Total System Memory</td>
<td>437.9 TB; 83.1 TB</td>
</tr>
<tr>
<td>System Interconnect</td>
<td>Cray Aries routing and communications ASIC, and Dragonfly topology</td>
</tr>
<tr>
<td>Sonexion 3000 Storage Capacity</td>
<td>6.2 PB</td>
</tr>
<tr>
<td>Sonexion 3000 Parallel File System Theoretical Peak Performance</td>
<td>112 GB/s</td>
</tr>
<tr>
<td>Sonexion 1600 Storage Capacity</td>
<td>2.5 PB</td>
</tr>
<tr>
<td>Sonexion 1600 Parallel File System Theoretical Peak Performance</td>
<td>138 GB/s</td>
</tr>
</tbody>
</table>

Customers using Piz Daint

- User Lab (including PRACE Tier-0 allocations)
- University of Zurich, USI, PSI, EMPA

http://www.cscs.ch/publications/fact_sheets/index.html
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

Model Cray XC40/Cray XC50
- Number of Hybrid Compute Nodes: 5,320
- Number of Multicore Compute Nodes: 1,431
- Theoretical Peak Floating-point Performance per Hybrid Node: 4.761 Teraflops Intel Xeon E5-2690 v3/Nvidia Tesla P100
- Theoretical Peak Floating-point Performance per Multicore Node: 1.210 Teraflops Intel Xeon E5-2695 v4
- Theoretical Hybrid Peak Performance: 25.326 Petaflops
- Theoretical Multicore Peak Performance: 1.731 Petaflops
- Hybrid Memory Capacity per Node: 64 GB; 16 GB CoWoS HBM2
- Multicore Memory Capacity per Node: 64 GB, 128 GB
- Total System Memory: 437.9 TB; 83.1 TB
- System Interconnect: Cray Aries routing and communications ASIC, and Dragonfly topology
- Sonexion 3000 Storage Capacity: 6.2 PB
- Sonexion 3000 Parallel File System Theoretical Peak Performance: 112 GB/s
- Sonexion 1600 Storage Capacity: 2.5 PB
- Sonexion 1600 Parallel File System Theoretical Peak Performance: 138 GB/s

Customers using Piz Daint
- User Lab (including PRACE Tier-0 allocations)
- University of Zurich, USI, PSI, EMPA
- NCCR MARVEL

http://www.cscs.ch/publications/fact_sheets/index.html
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

### Model Cray XC40/Cray XC50

<table>
<thead>
<tr>
<th>Feature</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hybrid Compute Nodes</td>
<td>5 320</td>
</tr>
<tr>
<td>Number of Multicore Compute Nodes</td>
<td>1 431</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance</td>
<td>Intel Xeon E5-2690 v3/Nvidia Tesla P100</td>
</tr>
<tr>
<td>Theoretical Hybrid Peak Performance</td>
<td>Intel Xeon E5-2695 v4</td>
</tr>
<tr>
<td>Theoretical Multicore Peak Performance</td>
<td></td>
</tr>
<tr>
<td>Hybrid Memory Capacity per Node</td>
<td></td>
</tr>
<tr>
<td>Multicore Memory Capacity per Node</td>
<td></td>
</tr>
<tr>
<td>Total System Memory</td>
<td></td>
</tr>
<tr>
<td>System Interconnect</td>
<td></td>
</tr>
<tr>
<td>Sonexion 3000 Storage Capacity</td>
<td></td>
</tr>
<tr>
<td>Sonexion 3000 Parallel File System</td>
<td></td>
</tr>
<tr>
<td>Sonexion 1600 Storage Capacity</td>
<td></td>
</tr>
<tr>
<td>Sonexion 1600 Parallel File System Theoretical Peak Performance</td>
<td>138 GB/s</td>
</tr>
</tbody>
</table>

---

**Customers using Piz Daint**

- User Lab (including PRACE Tier-0 allocations)
- University of Zurich, USI, PSI, EMPA
- NCCR MARVEL
- CHIPP (since Aug. 2017)

---

http://www.cscs.ch/publications/fact_sheets/index.html
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

Model Cray XC40/Cray XC50

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Hybrid Compute Nodes</td>
<td>5 320</td>
</tr>
<tr>
<td>Number of Multicore Compute Nodes</td>
<td>1 431</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Hybrid Node</td>
<td>4.761 Tflops Intel Xeon E5-2690 v3/Nvidia Tesla P100</td>
</tr>
<tr>
<td>Theoretical Peak Floating-point Performance per Multicore Node</td>
<td>1.210 Tflops Intel Xeon E5-2695 v4</td>
</tr>
<tr>
<td>Theoretical Hybrid Peak Performance</td>
<td>25.326 Petaflops</td>
</tr>
<tr>
<td>Theoretical Multicore Peak Performance</td>
<td>1.731 Petaflops</td>
</tr>
<tr>
<td>Hybrid Memory Capacity per Node</td>
<td>64 GB; 16 GB CoWoS HBM2</td>
</tr>
<tr>
<td>Multicore Memory Capacity per Node</td>
<td>64 GB, 128 GB</td>
</tr>
<tr>
<td>Total System Memory</td>
<td>437.9 TB; 83.1 TB</td>
</tr>
<tr>
<td>System Interconnect</td>
<td>Communications ASIC, and Dragonfly</td>
</tr>
<tr>
<td>Sonexion 3000 Storage Capacity</td>
<td>6.2 PB</td>
</tr>
<tr>
<td>Sonexion 3000 Parallel File System</td>
<td></td>
</tr>
<tr>
<td>Sonexion 1600 Storage Capacity</td>
<td></td>
</tr>
<tr>
<td>Sonexion 1600 Parallel File System Theoretical Peak Performance</td>
<td>138 GB/s</td>
</tr>
</tbody>
</table>

Customers using Piz Daint

- User Lab (including PRACE Tier-0 allocations)
- University of Zurich, USI, PSI, EMPA
- NCCR MARVEL
- CHIPP (since Aug. 2017)
- Others (exploratory)

http://www.cscs.ch/publications/fact_sheets/index.html
“Piz Daint” 2017 fact sheet

~5’000 NVIDIA P100 GPU accelerated nodes
~1’400 Dual multi-core socket nodes

Customers using Piz Daint

- User Lab (including PRACE Tier-0 allocations)
- University of Zurich, USI, PSI, EMPA
- NCCR MARVEL
- CHIPP (since Aug. 2017)
- Others (exploratory)
Moving Tensorflow to “Piz Daint”

Test-case setting

• simple neural network learning
• Standard model: LevNet-5-like convolutional MNIST model
• Written with Tensorflow/Python

Testbed environment

• Standard desktop with Intel Broadwell (4c)
• “Piz Daint” multi-core node with Intel Broadwell (2x18c)
• “Piz Daint” hybrid node with Intel Haswell (12c) and NVIDIA Pascal (P100)

Remark: this is a simple standard example, with complex models even more speedup expected

Source: Marcel Schöngens (schoengens@cscs.ch)
Moving Tensorflow to “Piz Daint”

Test-case setting

- simple neural network learning
- Standard model: LevNet-5-like convolutional MNIST model
- Written with Tensorflow/Python

Testbed environment

- Standard desktop with Intel Broadwell (4c)
- “Piz Daint” multi-core node with Intel Broadwell (2x18c)
- “Piz Daint” hybrid node with Intel Haswell (12c) and NVIDIA Pascal (P100)

Remark: this is a simple standard example, with complex models even more speedup expected

Source: Marcel Schöngens (schoengens@cscs.ch)
Moving Tensorflow to “Piz Daint”

Test-case setting
- simple neural network learning
- Standard model: LevNet-5-like convolutional MNIST model
- Written with Tensorflow/Python

Testbed environment
- Standard desktop with Intel Broadwell (4c)
- “Piz Daint” multi-core node with Intel Broadwell (2x18c)
- “Piz Daint” hybrid node with Intel Haswell (12c) and NVIDIA Pascal (P100)

Remark: this is a simple standard example, with complex models even more speedup expected

Source: Marcel Schöngens (schoengens@cscs.ch)
Moving Tensorflow to “Piz Daint”

Test-case setting

- simple neural network learning
- Standard model: LevNet-5-like convolutional MNIST model
- Written with Tensorflow/Python

Testbed environment

- Standard desktop with Intel Broadwell (4c)
- “Piz Daint” multi-core node with Intel Broadwell (2x18c)
- “Piz Daint” hybrid node with Intel Haswell (12c) and NVIDIA Pascal (P100)

Remark: this is a simple standard example, with complex models even more speedup expected

Source: Marcel Schöngens (schoengens@cscs.ch)
“We develop algorithms, we don’t have time to deal with C/C++ or MPI”

—a well-known computer science colleague working in machine learning
… echoed by many scientists working with data
... echoed by many scientists working with data

Interactive Notebook

Import TensorFlow and start an interactive session

```
In [1]: import tensorflow as tf
   sess = tf.InteractiveSession()
```

Build a computation graph

```
In [2]: matrix = tf.constant([[1., 2.]])
   negMatrix = tf.neg(matrix)
```

Evaluate the graph

```
In [3]: result = negMatrix.eval()
   print(result)
   [[-1. -2.]]
```

Nishant Shukla (2017)
Architectural Developments – Traditional Architecture

1. CSCS User accesses External Login Access (ELA)
2. ELA provides access to Piz Daint Compute
3. Piz Daint Compute is accessible to CSCS User
4. /store is a part of the Piz Daint Compute system

CSCS: Centro svizzero di calcolo scientifico
Research Community

Data Flow

T. Schalthess
Architectural Developments – Improved Architecture Based on External Portal

Domain Specific Portal

Repository access
Workflow Manager

CSCS User

Data Flow

CSCS

External Login Access (ELA)
Piz Daint Login & Mgmt
Piz Daint

/store
Architectural Developments – Improved Architecture Based on External Portal

CSCS

Repository Access
Workflow Manager

Domain Specific Portal

External Login Access (ELA)
Piz Daint Login & Mgmt
Piz Daint

CSCS User

MARVEL
Swiss Institute of Particle Physics

ChiPP

HP

Does Not Scale
Architectural developments – Service Oriented Architecture (SOA)

Domain Specific Portal
- Repository access
- Workflow Manager

CSCS User

Infrastructure Services
- Authentication & authorization
- User Management
- Data Management
- Workflow Automation
- Capacity Management

IT Infrastructure
- DWH
- Networking & security
- OpenStack Services
- Archival Storage
- Active Storage
- HPC Services
... and the service should be up most of the time (like 99+ %)
Supporting Federation using SOA

Research Community

Domain Specific Portal
- Repository access
- Workflow Manager

Software services

Platform services

Infrastructure services
- Authentication & authorization
- User Management
- Data Management
- Workflow Management
- Capacity Management
- DWH
- Networking 
  security
- OpenStack Services
- Archival Storage
- Active Storage
- HPC Services

CSCS
User

CSCS
Research
Community

Domain Specific Portal

Repository access

Workflow Manager

Repository
access

Workflow
Manager

Infrastructure provider

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

CSCS

User

(905,307),(960,374)

T. Schulthess

14
Fenix Sites
Q&A

schulthess@cscs.ch

colin@cscs.ch