E-Infrastructures
H2020-EINFRA-2016-2017

EINFRA-11-2016: Support to the next implementation phase of Pan-European High Performance Computing Infrastructure and Services (PRACE)

PRACE-5IP

PRACE Fifth Implementation Phase Project

Grant Agreement Number: EINFRA-730913

D7.4

Best Practices Guides for New and Emerging Architectures
Final

Version: 1.0.
Author(s): Volker Weinberg, BADW-LRZ; Sebastian Lührs, JUELICH; Wim Rijks, SURFsara; Ole W. Saastad, UiO; Terence Sloan, EPCC; Dominic Sloan-Murphy, EPCC; Caspar van Leeuwen, SURFsara
Date: 20.02.2019

Project and Deliverable Information Sheet

<table>
<tr><td rowspan="6">PRACE Project</td><td colspan="2">Project Ref. №: EINFRA-730913</td></tr>
<tr><td colspan="2">Project Title: PRACE Fifth Implementation Phase Project</td></tr>
<tr><td colspan="2">Project Web Site: http://www.prace-project.eu</td></tr>
<tr><td colspan="2">Deliverable ID: < D7.4></td></tr>
<tr><td colspan="2">Deliverable Nature: Report</td></tr>
<tr><td rowspan="2">Dissemination Level:
PU</td><td>Contractual Date of Delivery:
28 / 02 / 2019</td></tr>
<tr><td>Actual Date of Delivery:
28 / 02 / 2019</td></tr>
<tr><td colspan="3">EC Project Officer: Leonardo Flores Añover</td></tr>
</table>

* - The dissemination level are indicated as follows: **PU** – Public, **CO** – Confidential, only for members of the consortium (including the Commission Services) **CL** – Classified, as referred to in Commission Decision 2005/444/EC.

Document Control Sheet

<table>
<tr><td rowspan="5">Document</td><td colspan="2">Title: Best Practices Guides for New and Emerging Architectures</td></tr>
<tr><td colspan="2">ID: D7.4</td></tr>
<tr><td>Version: <1.0.></td><td>Status: Final</td></tr>
<tr><td colspan="2">Available at: http://www.prace-project.eu</td></tr>
<tr><td colspan="2">Software Tool: Microsoft Word 2013</td></tr>
<tr><td rowspan="4">Authorship</td><td colspan="2">File(s): D7.4.docx</td></tr>
<tr><td>Written by:</td><td>Volker Weinberg, BADW-LRZ; Sebastian Lührs, JUELICH; Wim Rijks, SURFsara; Ole W. Saastad, UiO; Terence Sloan, EPCC; Dominic Sloan-Murphy, EPCC; Caspar van Leeuwen, SURFsara</td></tr>
<tr><td>Contributors:</td><td>Sandra Aigner, TUM in collaboration with BADW-LRZ
Momme Allalen, BADW-LRZ
Valeriu Codreanu, SURFsara
Xu Guo, EPCC
Sandra Mendez, BADW-LRZ
Cristian Morales, BSC
Damian Podareanu, SURFsara
Anastasia Shamakina, HLRS
Andrew Turner, EPCC
Andreas Vroutsis, EPCC</td></tr>
<tr><td>Reviewed by:</td><td>Michal Białoskórski, CITASK
Veronica Teodor, JUELICH</td></tr>
<tr><td>Approved by:</td><td>MB/TB</td></tr>
</table>

Document Status Sheet

Version	Date	Status	Comments
0.1	01/02/2019	Draft	Initial version.
0.2	05/02/2019	Draft	Included updates by co-authors.
1.0	20/02/2019	Final version	Update after internal review.

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Best Practice Guide, AMD EPYC, ARM64, Deep Learning, HPC for Data Science on the Cray Urika, Modern Interconnects, Parallel I/O

Table of Contents

List of Figures

References and Applicable Documents

[1] http://www.prace-project.eu (identical to http://www.prace-ri.eu)

[2] Xu Guo and Ole W. Saastad (ed.), *Best Practice Guide – AMD EPYC*, produced by PRACE-5IP, February 2019. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/Best-Practice-Guide-AMD.pdf, http://www.prace-ri.eu/best-practice-guide-amd-epyc

[3] Xu Guo, Cristian Morales, Ole W. Saastad, Anastasia Shamakina, Wim Rijks (ed.), and Volker Weinberg (ed), *Best Practice Guide – ARM64*, produced by PRACE-5IP, February 2019. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/Best-Practice-Guide-ARM64.pdf, http://www.prace-ri.eu/best-practice-guide-arm64

[4] Damian Podareanu, Valeriu Codreanu, Sandra Aigner, Caspar van Leeuwen (ed.), and Volker Weinberg (ed.), *Best Practice Guide – Deep Learning in HPC*, produced by PRACE-5IP, February 2019. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/Best-Practice-Guide-Deep-Learning.pdf, http://www.prace-ri.eu/best-practice-guide-deep-learning

[5] Andreas Vroutsis, Sandra Mendez, Terence Sloan (ed.), and Volker Weinberg (ed.), *Best Practice Guide – HPC for Data Science on the Cray Urika*, produced by PRACE-5IP, January 2019. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/Best-Practice-Guide-Data-Science.pdf, http://www.prace-ri.eu/best-practice-guide-hpc-for-data-science-on-the-cray-urika

[6] Momme Allalen and Sebastian Lührs (ed.), *Best Practice Guide – Modern Interconnects*, produced by PRACE-5IP, February 2019. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/Best-Practice-Guide-Modern-Interconnects.pdf, http://www.prace-ri.eu/best-practice-guide-modern-interconnects

[7] Sandra Mendez, Sebastian Lührs, Dominic Sloan-Murphy (ed.), Andrew Turner (ed.), and Volker Weinberg (ed.), *Best Practice Guide – Parallel I/O*, produced by PRACE-5IP, February 2019. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/Best-Practice-Guide-Parallel-IO.pdf, http://www.prace-ri.eu/best-practice-guide-parallel-i-o/

[8] Jacques David, Jeroen Engelberts, Xu Guo, Florian Janetzko, and Walter Lioen, *Petascaling and Optimisation Guides for PRACE Systems*, PRACE-1IP D7.3, June 2012. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/d7.3_1ip.pdf

[9] Vegard Eide, Walter Lioen, Maciej Szpindler, and Volker Weinberg, *Petascaling and Optimisation for Tier-1 Architectures*, PRACE-2IP D7.3, May 2013. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/d7.3.pdf

[10] Nikos Anastopoulos, Maciej Cytowski, Mark Filipiak, Walter Lioen, Philippe Wautelet, and Volker Weinberg, *Best Practice Guides for New and Emerging Architectures*, PRACE-3IP D7.3.1, December 2013. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/d7.3.1_3ip.pdf

[11] Volker Weinberg, Alan Gray, and Ole W. Saastad, *Best Practice Guides for New and Emerging Architectures*, PRACE-4IP D7.6, January 2017. Available at the PRACE RI web site [1] as: http://www.prace-ri.eu/IMG/pdf/D7.6_4ip.pdf

[12] http://www.docbook.org/

List of Acronyms and Abbreviations

aisbl	Association International Sans But Lucratif (legal form of the PRACE-RI)
BCO	Benchmark Code Owner
CoE	Center of Excellence
CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture (NVIDIA)
DARPA	Defense Advanced Research Projects Agency
DEISA	Distributed European Infrastructure for Supercomputing Applications EU project by leading national HPC centres
DoA	Description of Action (formerly known as DoW)
EC	European Commission
EESI	European Exascale Software Initiative
EoI	Expression of Interest
ESFRI	European Strategy Forum on Research Infrastructures
GB	Giga (= 2^{30} ~ 10^9) Bytes (= 8 bits), also GByte
Gb/s	Giga (= 10^9) bits per second, also Gbit/s
GB/s	Giga (= 10^9) Bytes (= 8 bits) per second, also GByte/s
GÉANT	Collaboration between National Research and Education Networks to build a multi-gigabit pan-European network. The current EC-funded project as of 2015 is GN4.
GFlop/s	Giga (= 10^9) Floating point operations (usually in 64-bit, i.e. DP) per second, also GF/s
GHz	Giga (= 10^9) Hertz, frequency =10^9 periods or clock cycles per second
GPU	Graphic Processing Unit
HET	High Performance Computing in Europe Taskforce. Taskforce by representatives from European HPC community to shape the European HPC Research Infrastructure. Produced the scientific case and valuable groundwork for the PRACE project.
HMM	Hidden Markov Model
HPC	High Performance Computing; Computing at a high performance level at any given time; often used synonym with Supercomputing
HPL	High Performance LINPACK
ISC	International Supercomputing Conference; European equivalent to the US based SCxx conference. Held annually in Germany.
KB	Kilo (= 2^{10} ~10^3) Bytes (= 8 bits), also KByte
LINPACK	Software library for Linear Algebra
MB	Management Board (highest decision making body of the project)
MB	Mega (= 2^{20} ~ 10^6) Bytes (= 8 bits), also MByte
MB/s	Mega (= 10^6) Bytes (= 8 bits) per second, also MByte/s
MFlop/s	Mega (= 10^6) Floating point operations (usually in 64-bit, i.e. DP) per second, also MF/s
MOOC	Massively open online Course
MoU	Memorandum of Understanding.
MPI	Message Passing Interface
NDA	Non-Disclosure Agreement. Typically signed between vendors and customers working together on products prior to their general availability or announcement.
PA	Preparatory Access (to PRACE resources)
PATC	PRACE Advanced Training Centres

PRACE	Partnership for Advanced Computing in Europe; Project Acronym
PRACE 2	The upcoming next phase of the PRACE Research Infrastructure following the initial five year period.
PRIDE	Project Information and Dissemination Event
RI	Research Infrastructure
TB	Technical Board (group of Work Package leaders)
TB	Tera (= 2^{40} ~ 10^{12}) Bytes (= 8 bits), also TByte
TCO	Total Cost of Ownership. Includes recurring costs (e.g. personnel, power, cooling, maintenance) in addition to the purchase cost.
TDP	Thermal Design Power
TFlop/s	Tera (= 10^{12}) Floating-point operations (usually in 64-bit, i.e. DP) per second, also TF/s
Tier-0	Denotes the apex of a conceptual pyramid of HPC systems. In this context the Supercomputing Research Infrastructure would host the Tier-0 systems; national or topical HPC centres would constitute Tier-1
UNICORE	Uniform Interface to Computing Resources. Grid software for seamless access to distributed resources.

List of Project Partner Acronyms

BADW-LRZ	Leibniz-Rechenzentrum der Bayerischen Akademie der Wissenschaften, Germany (3rd Party to GCS)
BILKENT	Bilkent University, Turkey (3rd Party to UYBHM)
BSC	Barcelona Supercomputing Center - Centro Nacional de Supercomputacion, Spain
CaSToRC	Computation-based Science and Technology Research Center, Cyprus
CCSAS	Computing Centre of the Slovak Academy of Sciences, Slovakia
CEA	Commissariat à l'Energie Atomique et aux Energies Alternatives, France (3rd Party to GENCI)
CESGA	Fundacion Publica Gallega Centro Tecnológico de Supercomputación de Galicia, Spain, (3rd Party to BSC)
CINECA	CINECA Consorzio Interuniversitario, Italy
CINES	Centre Informatique National de l'Enseignement Supérieur, France (3 rd Party to GENCI)
CITASK	Centre of Informatics - Tricity Academic Supercomputer & networK, Poland (3rd party to PNSC)
CNRS	Centre National de la Recherche Scientifique, France (3 rd Party to GENCI)
CSC	CSC Scientific Computing Ltd., Finland
CSIC	Spanish Council for Scientific Research (3rd Party to BSC)
CYFRONET	Academic Computing Centre CYFRONET AGH, Poland (3rd party to PNSC)
EPCC	EPCC at The University of Edinburgh, UK
ETHZurich (CSCS)	Eidgenössische Technische Hochschule Zürich – CSCS, Switzerland
FIS	FACULTY OF INFORMATION STUDIES, Slovenia (3rd Party to ULFME)
GCS	Gauss Centre for Supercomputing e.V., Germany
GENCI	Grand Equipement National de Calcul Intensiv, France
GRNET	Greek Research and Technology Network, Greece

INRIA	Institut National de Recherche en Informatique et Automatique, France (3rd Party to GENCI)
IST	Instituto Superior Técnico, Portugal (3rd Party to UC-LCA)
IT4Innovations	IT4Innovations National supercomputing centre at VŠB-Technical University of Ostrava, Czech Republic
IUCC	INTER UNIVERSITY COMPUTATION CENTRE, Israel
JUELICH	Forschungszentrum Juelich GmbH, Germany
KIFÜ (NIIFI)	Governmental Information Technology Development Agency, Hungary
KTH	Royal Institute of Technology, Sweden (3rd Party to SNIC)
LiU	Linkoping University, Sweden (3rd Party to SNIC)
NCSA	NATIONAL CENTRE FOR SUPERCOMPUTING APPLICATIONS, Bulgaria
NTNU	The Norwegian University of Science and Technology, Norway (3rd Party to SIGMA)
NUI-Galway	National University of Ireland Galway, Ireland
PRACE	Partnership for Advanced Computing in Europe aisbl, Belgium
PSNC	Poznan Supercomputing and Networking Center, Poland
RISCSW	RISC Software GmbH
RZG	Max Planck Gesellschaft zur Förderung der Wissenschaften e.V., Germany (3 rd Party to GCS)
SIGMA2	UNINETT Sigma2 AS, Norway
SNIC	Swedish National Infrastructure for Computing (within the Swedish Science Council), Sweden
STFC	Science and Technology Facilities Council, UK (3rd Party to EPSRC)
SURFsara	Dutch national high-performance computing and e-Science support center, part of the SURF cooperative, Netherlands
TUM	Technical University of Munich, Germany
UC-LCA	Universidade de Coimbra, Labotatório de Computação Avançada, Portugal
UCPH	Københavns Universitet, Denmark
UHEM	Istanbul Technical University, Ayazaga Campus, Turkey
UiO	University of Oslo, Norway (3rd Party to SIGMA)
ULFME	UNIVERZA V LJUBLJANI, Slovenia
UmU	Umea University, Sweden (3rd Party to SNIC)
UnivEvora	Universidade de Évora, Portugal (3rd Party to UC-LCA)
UPC	Universitat Politècnica de Catalunya, Spain (3rd Party to BSC)
UPM/CeSViMa	Madrid Supercomputing and Visualisation Center, Spain (3rd Party to BSC)
USTUTT-HLRS	Universitaet Stuttgart – HLRS, Germany (3rd Party to GCS)
WCNS	Politechnika Wroclawska, Poland (3rd Party to PNSC)

Executive Summary

The Work Package 7 ‘Application Enabling and Support’ provides applications enabling services for HPC applications codes that are important for European academic and/or industrial researchers to ensure that these applications can effectively exploit current and future PRACE systems. Applications are selected for enabling via the PRACE Preparatory Access (PA) or the SHAPE programme. In addition to this enabling work on existing systems, the Work Package progresses the technical work needed to ensure that key applications are able to use future PRACE and EuroHPC (Pre-) Exascale systems, and investigates the tools and techniques needed to exploit such Exascale systems. Contribution to the development of numerical libraries for heterogeneous/hybrid architectures, and adoption of such libraries in community codes is a new goal in PRACE-5IP. Beyond directly working on improving applications and libraries, one of the main objectives of the Work Package is to support European HPC research communities through the provision of Best Practice Guides, benchmarks, and technical results in White Papers.

The successful series of Best Practice Guides has been already initiated in PRACE-1IP and has been continuously extended since then: PRACE-1IP provided four Best Practice Guides for PRACE Tier-0 systems (JUGENE, Curie, Cray XE and IBM Power) that cover programming techniques, compilers, tools and libraries (cf. [8]). PRACE-2IP added a generic guide about the x86 architecture and a Best Practice Guide for the SuperMUC system, together with a series of seven Best Practice Mini-Guides for other architectures which are important at Tier-1 to allow European researchers to make efficient use of these systems (cf. [9]). PRACE-3IP supplemented these with Best Practice Guides about Intel Xeon Phi, Blue Gene/Q and IBM Power 775 and provided updates of the Curie and the Cray XE guide, which was renamed into Cray XE/XC. PRACE-4IP Task 7.3.B added new guides about Knights Landing and Haswell/Broadwell and provided updates of the Intel® Xeon Phi™ and the GPGPU Best Practice Guides. Finally, PRACE-5IP Task 7.3B added six new Best Practice Guides about the processors AMD EPYC and ARM64, Deep Learning in HPC, HPC for Data Science on the Cray Urika, Modern Interconnects and Parallel I/O.

Topics for these Best Practice Guides include: a short description of the processor architecture, optimal porting of applications (e.g., choice of numerical libraries and compiler options); architecture specific optimisation and scaling techniques; optimal system environment (e.g., tuneable system parameters, job placement and optimised system libraries); debugging tools, performance analysis tools and programming environment.

This report describes the process which led to the Best Practice Guides and the structure of the guides. For the Best Practice Guides itself we refer to the online versions on the PRACE RI web site [1] (cf. [2], [3], [4], [5], [6], [7]).

1 Introduction

Efficient use of PRACE systems requires detailed knowledge of architecture specific factors influencing performance, compilers, tools and libraries. The main goal of this task is to investigate such issues, collect best practices on how to achieve good performance on the systems, and disseminate this knowledge to users.

The purpose of this report is to give a description of the process which led to the Best Practice Guides and present the structure of the guides.

In Section 2 we describe the selection of the topics of the Best Practice Guides, the subtasks, the technology used for creating the Best Practice Guides, and finally, the generic table of contents.

According to the DoA, this deliverable D7.4 “Best Practice Guides for New and Emerging Architectures” should report on the final versions of Best Practice Guides, covering process and structure. As in the past, because of the total size of the Best Practice Guides, we decided not to include them as separate chapters in this report but to refer to the online versions on the PRACE RI web site [1] instead (cf. [2], [3], [4], [5], [6], [7]). Section 3 thus only gives an executive summary of the contents of the guides.

The target audience are users and support staff who are developing and enabling applications.

2 Approach to Best Practice Guides

2.1 Selection of Topics

In the DoA we announced to maintain and extend the successful series of Best Practice Guides to new architectures/systems and mentioned likely topics of technical interest: new processors (Intel Skylake, ARM64), new accelerators (NVIDIA Pascal) and new interconnect technology (OmniPath).

When PRACE-5IP started in February 2017 we considered recent trends in HPC/AI technology to select the topics of the Best Practice Guides to be written. Since in PRACE-4IP the major focus was on Intel processors (Knights Corner, Knights Landing, Haswell/Broadwell), in PRACE-5IP we wanted to concentrate on new technologies from different vendors like ARM and AMD. Instead of updating the GPGPU Best Practice Guide once again, we decided to cover Deep Learning in HPC as a hot new topic and include also information about recent GPU technologies in this guide.

Finally, considering also experiences in other tasks of the Work Package, the following topics have been selected:

- AMD EPYC
- ARM64
- Deep Learning in HPC
- HPC for Data Science on the Cray Urika
- Modern Interconnects
- Parallel I/O

2.2 Editors

The whole Best Practice Guide activity 7.3.B was led by Volker Weinberg (BADW-LRZ); the respective chief editors/sub-activity leaders were:

- Ole W. Saastad for the AMD EPYC - Best Practice Guide
- Wim Rijks for the ARM64 - Best Practice Guide
- Caspar van Leeuwen for the Deep Learning in HPC - Best Practice Guide
- Terence Sloan for the HPC for Data Science on the Cray Urika - Best Practice Guide
- Sebastian Lührs for the Modern Interconnects - Best Practice Guide
- Dominic Sloan-Murphy and Andrew Turner (in the beginning of the project) for the Parallel I/O - Best Practice Guide

2.3 Technology

We built on the long experience obtained during the corresponding PRACE-1IP, PRACE-2IP, PRACE-3IP and PRACE-4IP tasks (cf. [8], [9], [10], [11]). It was decided already in PRACE-1IP that high quality HTML versions as well as high quality, fully featured PDF versions should be created and published. To reach this goal, we use DocBook. DocBook (cf. [12]) is employed by a lot of open source projects amongst others by the Linux Documentation Project. The key feature is having single (XML) source and multiple fully cross-referenced output formats: HTML, PDF and others. To keep track of the source code versions we used the LRZ GitLab server, as the PRACE GitLab service was not available yet at the beginning of the project and the PRACE SVN server caused many problems in previous PRACE projects.

To keep track of the status of the Best Practice Guides, Wiki pages for the whole activity and subpages for the individual guides have been set up and maintained on the PRACE Wiki, including information about status of the individual chapters of the guides, DocBook XML, GitLab server access, timetables, generic table of contents, assignment of collaborators to topics and previous work within PRACE related to Best Practice Guides.

Figure 1 shows the Best Practice Guides available on the PRACE RI web site [1] as of January 2019. Though the milestone for the publication of the Best Practice Guides was in February 2019, initial versions of four guides have already been published in December 2018 and January 2019.

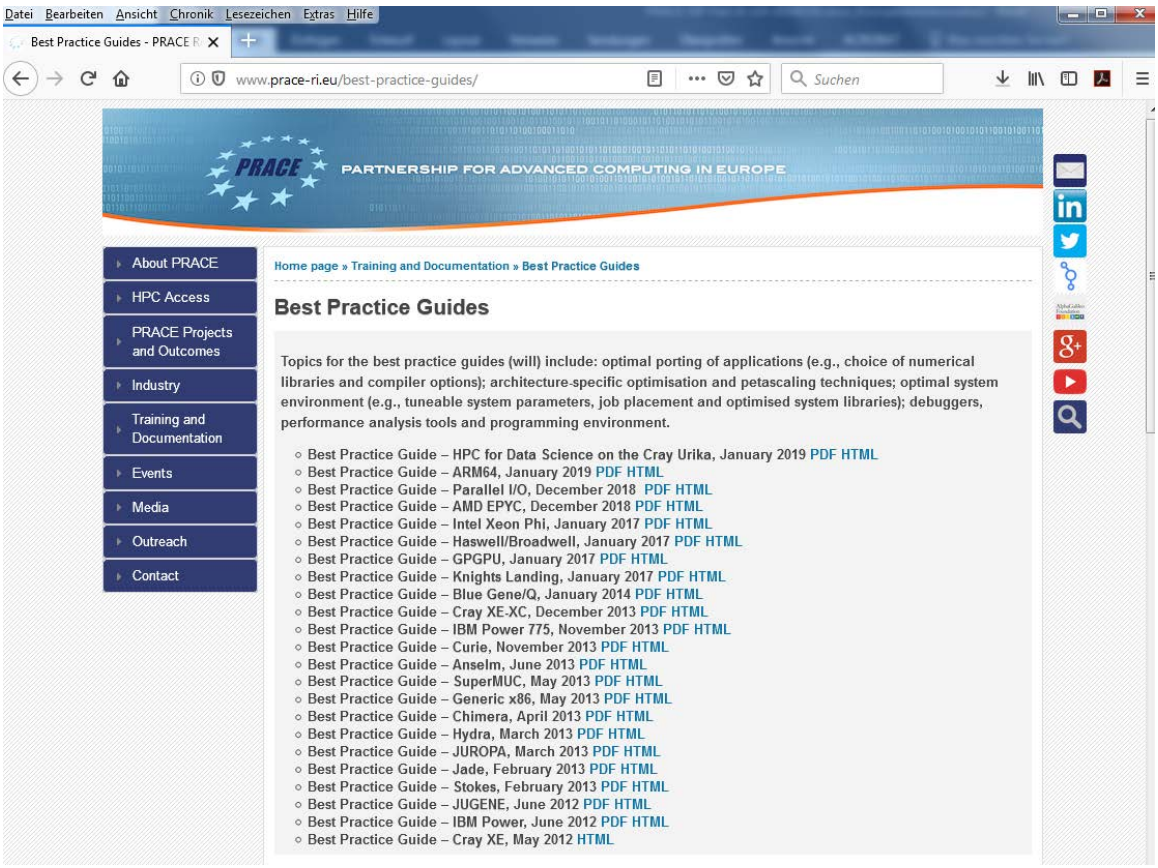

Figure 1: Best Practice Guides on the PRACE RI web site (as of January 2019).

2.4 Generic Table of Contents

As in the past, the following generic table of contents was used as a template if applicable for the topic of the Best Practice Guide:

1 Introduction
2 System Architecture / Configuration
    2.1 Processor Architecture / MCM Architecture (including caches)
    2.2 Building Block Architecture (node cards, nodes, drawers, supernodes, racks)
    2.3 Memory Architecture (including NUMA effects)
    2.4 (Node) Interconnect (including topology, system specific)
    2.5 I/O Subsystem Architecture (being system specific and not architecture specific!)
    2.6 Available File Systems
        2.6.1 Home, Scratch, Long Time Storage
        2.6.2 Performance of File Systems
3 System Access
    3.1 How to Reach the System (ssh, portals, file transfer, ...)
4 Production Environment
    4.1 Module Environment
    4.2 Batch System
    4.3 Accounting

- 5 Programming Environment / Basic Porting
  - 5.1 Available Compilers
    - 5.1.1 Compiler Flags
  - 5.2 Available (Vendor Optimised) Numerical Libraries
  - 5.3 Available MPI Implementations
  - 5.4 OpenMP
    - 5.4.1 Compiler Flags
  - 5.5 Batch System / Job Command Language
- 6 Performance Analysis
  - 6.1 Available Performance Analysis Tools
  - 6.2 Hints for Interpreting Results.
- 7 Tuning
  - 7.1 Advanced / Aggressive Compiler Flags
  - 7.2 Single Core Optimisation
  - 7.3 Advanced MPI usage
    - 7.3.1 Tuning / Environment Variables
    - 7.3.2 Mapping Tasks on Node Topology
    - 7.3.3 Task Affinity
    - 7.3.4 Adapter Affinity
  - 7.4 Advanced OpenMP Usage
    - 7.4.1 Tuning / Environment Variables
    - 7.4.2 Thread Affinity
  - 7.5 Hybrid Programming
    - 7.5.1 Optimal Tasks / Threads Strategy
  - 7.6 Memory Optimisation
    - 7.6.1 Memory Affinity (MPI/OpenMP/Hybrid)
    - 7.6.2 Memory Allocation (malloc) Tuning
    - 7.6.3 Using Huge Pages
  - 7.7 I/O Optimisation (Tuning / Scaling of Application I/O)
  - 7.8 Advanced Job Command Language (includes defining task topology, affinity, etc.)
  - 7.9 Possible Kernel Parameter Tuning (probably less relevant to the ‘average’ user but possibly relevant for large production runs)
- 8 Debugging
  - 8.1 Available Debuggers
  - 8.2 Compiler flags

2.5 Content

For all guides an inventory of the existing documentation was made that could be used as base material for some of the topics mentioned above. Many topics had to be complemented or written from scratch. Apart from this, experiences learned during the various enabling or benchmark activities in other tasks were added. As an internal quality assurance, T7.3B-internal cross-reviews were performed by the editors.

3 Best Practice Guides

The Best Practice Guides itself can be found online on the PRACE RI web site [1] (cf. [2], [3], [4], [5], [6], [7]). The following subsections give a short description of the contents of the guides.

3.1 Best Practice Guide – AMD EPYC (cf. [2])

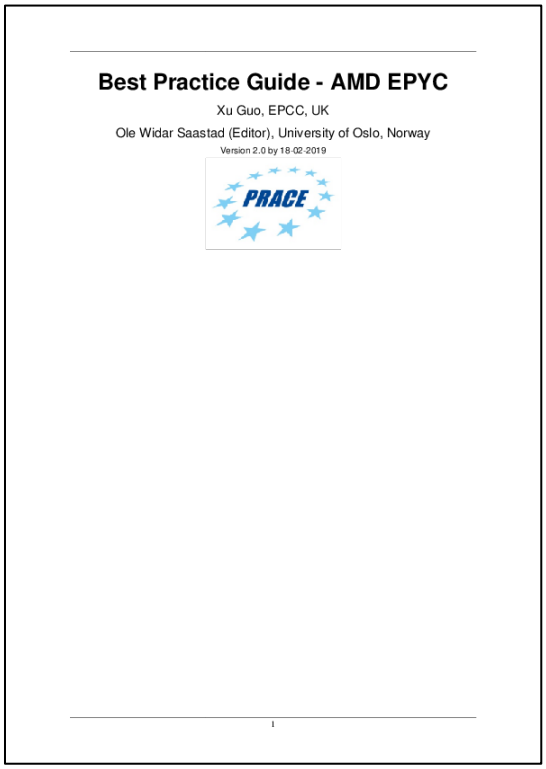

Figure 2: The AMD EPYC - Best Practice Guide.

This Best Practice Guide provides information about how to use the AMD EPYC processors in an HPC environment and describes some experiences with the use of some common tools for this processor. The EPYC processors are the latest generation of processors from AMD Inc. While not yet significant on the top-500 list, this is likely to change in the future. Information about the nature of the NUMA architecture is provided. In addition, some tuning and optimisation techniques as well as debugging are covered.

The guide covers the following tools: compilers, performance libraries, threading libraries (OpenMP), message passing libraries (MPI), memory access and allocation libraries, debuggers, performance profilers, etc.

Selected benchmarks comparing compilers and libraries have been performed. Furthermore, recommendations and hints about how to use the Intel tools with this processor are presented. In contrast to the Intel tools the GNU tools and tools from other independent vendors have full support for EPYC. A set of a compilers and development tools have been tested with satisfactory results.

3.2 Best Practice Guide – ARM64 (cf. [3])

Figure 3: The ARM64 - Best Practice Guide.

This Best Practice Guide provides information about the ARM64 architecture and the programming models that programmers can use in order to achieve good performance with their applications on this architecture.

The guide gives a description of the hardware of the ARM64 processor. It also provides information about the programming models and development environment as well as information about porting programs. Furthermore, it provides information about tools and strategies on how to analyse and improve the performance of applications.

Finally, the guide contains a description of test and production systems that already exist in Europe or are planned in near future.

3.3 Best Practice Guide – Deep Learning in HPC (cf. [4])

Figure 4: The Deep Learning in HPC - Best Practice Guide.

Deep learning is a class of machine learning algorithms that use multiple layers of nonlinear processing units for feature extraction and transformation. This allows models to represent multiple levels of abstraction from the data, which is a natural approach to many problems, such as image, sound, and text analysis.

Sparked by various inference and prediction challenges on publicly available large datasets and by having available open-source frameworks (such as TensorFlow, Caffe and PyTorch), the deep learning field has evolved rapidly over the past decade. With more complex neural networks and larger input data sets, scalability of deep learning algorithms is an increasingly important topic.

The main aim of this Best Practice Guide is to teach how to perform deep learning at large scale. Different algorithms, software frameworks and hardware platforms are discussed, with a focus on how these can be employed for large scale distributed deep learning. This should help the reader to pick the most suitable framework and hardware platform for a deep learning problem, and use it efficiently.

3.4 Best Practice Guide – HPC for Data Science on the Cray Urika (cf. [5])

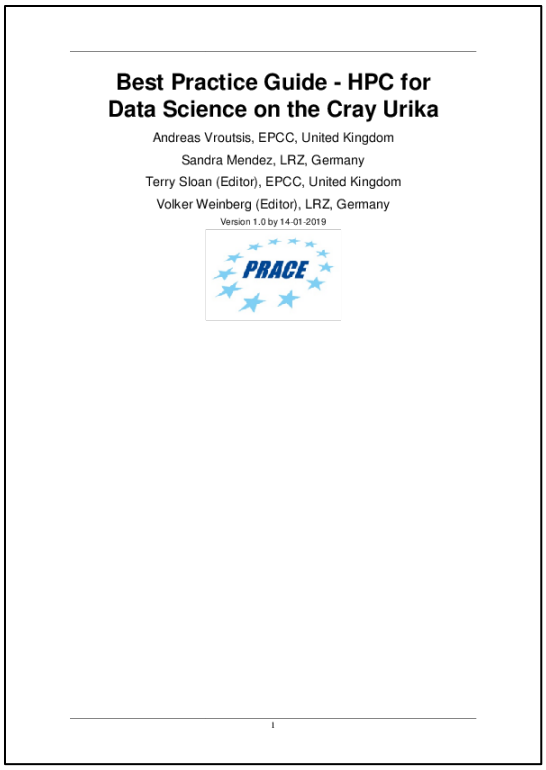

Figure 5: The HPC for Data Science on the Cray Urika - Best Practice Guide.

This Best Practice Guide provides information about exploiting HPC platforms and techniques for Data Science projects.

The first section covers Spark. This is an open source distributed data analytics platform. It provides access to many different data sources and enables parallel computations to be distributed across a cluster. This section of the guide describes Resilient Distributed Datasets, the concept at the heart of Spark. It also describes the architecture of Spark, its libraries and how to run Spark on a cluster.

The second section covers the Cray Urika GX system. The Cray Urika GX is an HPC platform dedicated to highly interactive and iterative data analytics that require supercomputer levels of computing performance. This chapter describes the production and programming environment on the platform. This environment includes Spark, Hadoop, R and graph databases. The chapter's contents are based on the Urika GX system hosted by EPCC and Cray on behalf of the Alan Turing Institute in the UK.

3.5 Best Practice Guide – Modern Interconnects (cf. [6])

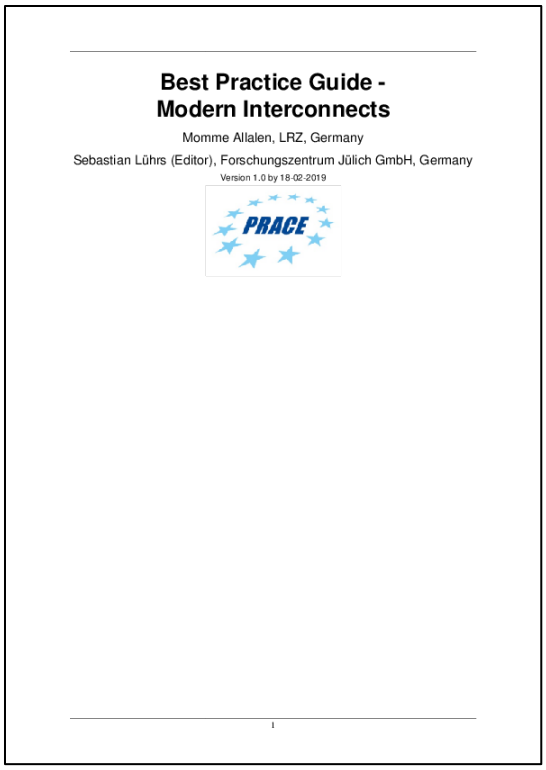

Figure 6: The Modern Interconnects - Best Practice Guide.

Having a high bandwidth and low latency interconnect usually makes the main difference between computers which are connected via a regular low bandwidth, high latency network and a so called supercomputer or HPC system. Different interconnect types are available, either by individual vendors for their specific HPC setup or in form of a separate product, to be used in a variety of different systems. For the user of such a system the interconnect is often used as a black box.

This Best Practice Guide gives an overview about the most common types of interconnects in the current generation of HPC systems: Omni-Path, InfiniBand, Aries, NVLink, Tofu and Ethernet. It describes the key features of each type and introduces the most common types of network topologies like fat tree, torus, hypercube and dragonfly.

The interconnect types within the current generation of PRACE Tier-0 systems are listed and the last section gives some hints concerning network benchmarking.

3.6 Best Practice Guide – Parallel I/O (cf. [7])

Figure 7: The Parallel I/O - Best Practice Guide.

This Best Practice Guide provides information about High Performance I/O systems, parallel I/O APIs and I/O optimisation techniques. It presents a description of the storage architecture and the I/O software stack.

The guide covers the basic concepts of parallel I/O in HPC, including: general parallel I/O strategies (e.g. shared files, file per process), parallel file systems, the parallel I/O software stack and potential performance bottlenecks. It also describes the major parallel file systems in use on HPC systems: Lustre and Spectrum Scale/GPFS, and touches on BeeGFS, a file system gaining popularity in HPC. Furthermore, a brief introduction to MPI-IO is given with links to further, detailed information along with tips on how to get best performance out of the MPI-IO library on parallel file systems. Another chapter describes the file per process model for parallel I/O along with performance considerations. The guide also covers the use of high-level libraries (HDF5, NetCDF, SIONlib) and considerations for getting best performance. Finally, information on how to gather performance data is given.

4 Conclusion

The successful series of Best Practice Guides has been initiated in PRACE-1IP and has been continuously extended since then.

Within PRACE-5IP the following six new Best Practice Guides have been published on the PRACE webserver:

- AMD EPYC
- ARM64
- Deep Learning in HPC
- HPC for Data Science on the Cray Urika
- Modern Interconnects
- Parallel I/O

PRACE-6IP will maintain and further extend the series of Best Practice Guides to new technologies and systems. Likely future topics of technical interest include new processors or GPUs, new memory technologies (MCDRAM, NVRAM), new interconnects, and workflows for HPC job processing and data management. A special focus will be put on how to exploit future PRACE and EuroHPC (Pre-) Exascale systems.