E-Infrastructures
H2020-EINFRA-2014-2015

EINFRA-4-2014: Pan-European High Performance Computing Infrastructure and Services

PRACE-4IP

PRACE Fourth Implementation Phase Project

Grant Agreement Number: EINFRA-653838

D6.2
Final Report on the PRACE Operational Services

Final

Version: 1.0
Author(s): Giovanni Erbacci, CINECA, Miroslaw Kupczyk, PSNC
Date: 20.04.2017

Project and Deliverable Information Sheet

PRACE Project	**Project Ref. №: EINFRA-653838**	
	Project Title: PRACE Fourth Implementation Phase Project	
	Project Web Site: http://www.prace-project.eu	
	Deliverable ID: < **D6.2**>	
	Deliverable Nature: < Report >	
	Dissemination Level: PU *	**Contractual Date of Delivery:** 30 / April / 2017
		Actual Date of Delivery: 30 / April / 2017
	EC Project Officer: Leonardo Flores Añover	

* - 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 2991/844/EC.

Document Control Sheet

Document	**Title: Final Report on the PRACE Operational Services**	
	ID: D6.2	
	Version: <1.0>	**Status:** ***Final***
	Available at: http://www.prace-project.eu	
	Software Tool: Microsoft Word 2010	
	File(s): D6.2.docx	
Authorship	**Written by:**	Giovanni Erbacci, CINECA, Miroslaw Kupczyk, PSNC
	Contributors:	Martin Bidner, HLRS Thomas Boenisch, HLRS Xavier Delaruelle, CEA Lukasz Flis, CYFRONET Zoltan Kiss, NIIF Felip Moll, BSC Ralph Niederberger, FZJ Cristiano Padrin, CINECA Huub Stoffers, SURFSara Oscar Yerpes, BSC
	Reviewed by:	Walter Lioen, SURFsara Florian Berberich, FZJ
	Approved by:	MB/TB

Document Status Sheet

Version	**Date**	**Status**	**Comments**
0.1	01/February/2017	Draft	Structure and first elements of the document

0.2	15/March/2017	Draft	First draft of operational services
0.3	17/March/2017	Draft	Agreed the final structure of the document
0.4	28/March/2017	Draft	Added draft on user services
0.5	31/March/2017	Draft	Draft final version
0.6	01/April/2017	Draft	Final polishing and editing
0.7	03/April/2017	Draft	Draft version ready for internal review process
1.0	20/April/2017	Final version	Final version ready for EC submission

Document Keywords

Keywords:	PRACE, HPC, Research Infrastructure, Operational Services, Tier-0 Systems, KPI

Table of Contents

List of Figures

List of Tables

References and Applicable Documents

[1] http://www.prace-project.eu

[2] PRACE-1IP Deliverable D6.3, Second Annual Report on the Technical Operation and Evolution: http://prace-ri.eu/IMG/pdf/d6.3_1ip.pdf

[3] PRACE-3IP Deliverable D6.1.1, First Annual Operation Report, PRACE, June 2013

[4] Beskow System - https://www.pdc.kth.se/resources/computers/beskow

[5] Hazel Hen System - https://wickie.hlrs.de/platforms/index.php/Cray_XC40

[6] ARCHER - http://www.archer.ac.uk/about-archer/news-events/news/2014-10-31-Phase2_upgrade.php

[7] Salomon System - https://docs.it4i.cz/salomon

[8] Leo System - https://wiki.niif.hu/HPC_en#NIIF_HPC_resources

[9] Galileo System - http://www.hpc.cineca.it/hardware/galileo

[10] Minotauro BSC - https://www.bsc.es/innovation-and-services/supercomputers-and-facilities/minotauro

[11] Phitagoras NIIF - https://wiki.niif.hu/index.php?title=Budapest2_klaszter

[12] Marenostrum 4 BSC - https://www.bsc.es/news/bsc-news/marenostrum-4-supercomputer-be-12-times-more-powerful-marenostrum-3

[13] Aris GRNET - http://doc.aris.grnet.gr/

[14] Prometheus Cyfronet - http://www.cyfronet.krakow.pl/komputery/15207,artykul,prometheus.html

[15] Eagle PSNC - https://wiki.man.poznan.pl/kdm/index.php?title=Eagle

[16] Marconi CINECA - https://www.cineca.it/en/content/marconi

[17] BEM WCSS - http://kdm.wcss.wroc.pl/wiki/Bem

[18] MOAB IUCC - https://www.iucc.ac.il/services/computing-services/

[19] RFC 5280, Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List (CRL) Profile: https://tools.ietf.org/html/rfc5280

[20] EUGridPMA: https://www.eugridpma.org/

[21] IGTF: http://www.gridpma.org/

[22] Globus GSI-OpenSSH: http://www.globus.org/toolkit/security/gsiopenssh

[23] GSISSH-Term: https://www.lrz.de/services/compute/grid_en/software_en/gsisshterm_en/

[24] DART: http://prace-ri.eu/Accounting-Report-Tool

[25] A Trust Framework for Security Collaboration among Infrastructures http://pos.sissa.it/archive/conferences/179/011/ISGC%202013_011.pdf

[26] PRACE Helpdesk: https://tts.prace-ri.eu/

[27] Best Practical Request Tracker: https://www.bestpractical.com/rt-and-rtir

[28] PRACE-RI Website: http://www.prace-ri.eu/

[29] PRACE SVN repository: https://prace.osd.surfsara.nl/svn/

[30] MarkDown: https://en.wikipedia.org/wiki/Markdown

[31] Modules-Tcl: http://modules.sourceforge.net/

[32] Pandoc: http://pandoc.org/

[33] PRACE Software Stack document.

[34] INCA Monitoring system, http://inca.sdsc.edu/

[35] ARGO Monitoring System, http://argoeu.github.io/

[36] Nagios, https://www.nagios.com/

[37] Zabbix, www.zabbix.org

[38] ICINGA, https://www.icinga.org/

[39] Opsview, https://www.opsview.com/

[40] PRACE-4IP Deliverable D6.1, First Annual Report on the PRACE Operational Services, PRACE, April 2016

[41] Gtransfer, https://github.com/fr4nk5ch31n3r/gtransfer

List of Acronyms and Abbreviations

AAA	Authorization, Authentication, Accounting
aisbl	Association International Sans But Lucratif (legal form of the PRACE-RI)
APGridPMA	The Asian Pacific Grid Policy Management Authority
BCO	Benchmark Code Owner
BGP	Border Gateway Protocol
BSS	Batch Scheduling System
CA	Certificate Authority
CLI	Command Line Interfaces
CoE	Center of Excellence
CP/CPS	Certificate Policy/Certification Practice Statement
CPU	Central Processing Unit
CSIRT	Computer Security Incident Response Team
CUDA	Compute Unified Device Architecture (NVIDIA)
DARPA	Defense Advanced Research Projects Agency
DART	Distributed Accounting Reporting Tool
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
EUDAT	European Data Infrastructure
EUGridPMA	European Grid Policy Management Authority
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 =109 periods or clock cycles per second
GPU	Graphic Processing Unit
GSI	Grid Security Infrastructure
GT	Globus Toolkit
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
HTML	HyperText Markup Language
IGTF	Interoperable Global Trust Federation
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
KPI	Key Performance Indicator
LDAP	Lightweight Directory Access Protocol
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
MD-VPN	Multi Domain Virtual Private Network
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
PCPE	PRACE Common Production Environment
PKI	Public Key Infrastructure
PMA	Policy Management Authority
PRACE	Partnership for Advanced Computing in Europe; Project Acronym
PRACE 2	The next phase of the PRACE Research Infrastructure following the initial five year period
PRIDE	Project Information and Dissemination Event
RI	Research Infrastructure
RT	Request Tracker
SCI	Security for Collaborating Infrastructures
SDSC	San Diego Supercomputer Center
SSH	Secure Shell
SVN	SubVersioN: software versioning and revision system
TAGPMA	The Americas Grid PMA
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
TTS	Trouble Ticket System
UNICORE	Uniform Interface to Computing Resources. Grid software for seamless access to distributed resources.
VPN	Virtual Private Network
WISE	Wise Information Security for collaborating E-infrastructures

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 (3 rd 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)
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.
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 (3 rd Party to GENCI)
IST	Instituto Superior Técnico, Portugal (3rd Party to UC-LCA)
IT4Innovations	VYSOKA SKOLA BANSKA - TECHNICKA UNIVERZITA OSTRAVA, IT4Innovations National Supercomputing Center, Czech Republic
IUCC	INTER UNIVERSITY COMPUTATION CENTRE, Israel
JKU	Institut fuer Graphische und Parallele Datenverarbeitung der Johannes Kepler Universitaet Linz, Austria
JUELICH	Forschungszentrum Juelich GmbH, Germany
KIFÜ	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
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 (3 rd 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 Visualization 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 objective of this deliverable is to present the activity done to operate and coordinate the common PRACE Operational services, foreseen by Task 6.1 of WP6 in the PRACE-4IP project, with a special focus on the last reporting period (May 2016 - April 2017).

The operation of the PRACE distributed HPC infrastructure involves the coordination of a set of services which integrate the Tier-0 systems and a number of national Tier-1 systems, providing services for Tier-0, in a "single" pan-European HPC infrastructure.

This work is the continuation of the work done by Task 6.1 in the previous PRAC-IP projects to give continuity to the PRACE Operational services for the HPC eco-system. In turn, the activity presented here will continue and further progress in Task 6.1 of the PRACE-5IP project, just started.

Seven Tier-0 systems are operational in the second year of the PRACE-4IP project period:

- JUQUEEN at GCS@FZJ;
- CURIE at GENCI@CEA;
- HAZELHEN at GCS@HLRS;
- SuperMUC at GCS@LRZ;
- SuperMUC phase2 at GCS@LRZ
- MARCONI (BDW & KNL) at CINECA;
- MareNostrum 3 at BSC.

Furthermore, operational support has been provided to 28 national Tier-1 systems, which provide services for Tier-0 (i.e. used from SMEs for the SHAPE activity, or as stepping stone towards Tier-0 systems, or to prototype and asses new operational services). These Tier-1 systems are distributed among 16 different countries, ensuring a wide distribution of the European HPC eco-system.

The version of the PRACE Service Catalogue, approved from the PRACE Board of Directors (BoD) in March 2015, has been revised during the second year of PRACE-4IP. In the process towards establishing a PRACE Quality of Service and quality control, the work on PRACE Operational Key Performance Indicators has been addressed. During this second period of the project, for some services, an activity started to undertake the measurement of the KPIs and their evaluation.

Based on the procedures for incident and change management the complete set of PRACE common services as defined in the Service Catalogue (Networking, Data, Compute, AAA and Security, User, Monitoring and Generic) have been operated and monitored on a day-by-day basis to assure continuity and integrity of the services.

The current PRACE network (based on a 10 Gb/s star-topology designed more than 10 years ago) has been up-graded in the last months to a new MD-VPN (Multi Domain Virtual Private Network) architecture, which allow much for more flexibility in configuration, faster setup and cheaper connectivity costs per year.

The Security Forum, coordinated by Task 6.1, is responsible for all security related activities and, by means of periodic teleconferences, constantly assures the monitoring of the whole HPC infrastructure and prevents possible incidents, which could cause vulnerability on the PRACE RI.

The activity done in PRACE-4IP Task 6.1 will smoothly continue and progress in PRACE-5IP WP6, where the coordinated operation of the common PRACE operational services will involve an additional Tier-0 site (CSCS Zurich) which joins the group of the Hosting Members partners in the PRACE 2 framework.

1 Introduction

This deliverable describes the activities done in Task 6.1"Operation and coordination of the comprehensive common PRACE operational services" of WP6 "Operational Services for the HPC eco-system" in PRACE-4IP. This task is responsible for the operations of the set of common services, which presents the PRACE Tier-0 systems as an integrated pan-European HPC eco-system [1]. The operational services are extended also to national Tier-1 systems, essential as a stepping-stone towards Tier-0 systems (Tier-1 for Tier-0). Examples of Tier-1 for Tier-0 activities are the SHAPE activity toward SMEs, the prototyping and assessment of new operational services investigated in Tasks 6.2 and 6.3 of PRACE-4IP WP6, the testing and utilization of specific architectures and technologies, which are only available in specific countries.

The operation and coordination of the "common PRACE operational services provided in Task 6.1 have well established management procedures and organisation as set up already since PRACE-1IP, [1], [3]. The task further continues the implementation of the roadmap to a professional service level of sustainable services with a defined quality of service.

This report mainly focuses on the activities done in the second reporting period (May 2016-April 2017) and complements the activity undertaken during the first year of PRACE-4IP and documented in PRACE-4IP Deliverable 6.1 [40].

In this second reporting period, the operation of the common PRACE operational services has been coordinated and monitored constantly, by means of biweekly teleconferences, attended by all the partners involved in the Task 6.1 activity. Furthermore, two successful WP6 Face to Face meetings have been organised, in Ljubljana (Slovenia) on 5 - 6 September 2016 and in Athens (Greece) on 31 January – 1 February 2017. These meetings have been very useful to discuss the status of the operational activity, to plan the activity for the subsequent periods and to agree on the teams involved in the different activities. Both meetings had around 40 attendees and all the different sites involved in WP6 have been represented.

Section 2 describes the status of the Tier-0 systems and the Tier-1 systems involved in the Tier-1 for Tier-0 activity, composing the PRACE HPC eco-system. Section 3 gives a status overview of the common services for the different service areas:

- Network services (dedicated network provided by GEANT connecting Tier-0 and major Tier-1 centres);
- Data services;
- Compute services;
- Authorization, Authentication and Accounting;
- Operational security;
- User services;
- Monitoring services for operations;
- Generic services.

Finally, Section 4 presents some suggestions for the transition to the operational services in the PRACE-5IP project, just started, and draws some conclusions.

This deliverable aims to support the PRACE partners to better identify and integrate the PRACE pan-European HPC infrastructure, including possible new partners that will join the HPC infrastructure in the context of the PRACE 2 Framework. The intended audience of this deliverable is primarily the PRACE project partners where the cooperation and interactions between different WPs can require access to the operational services of the HPC Infrastructure. In addition to the previous audience, there are the members of the PRACE RI, the computational users and the members of the CoEs, which consider access to the PRACE HPC resources an important instrument to enhance their competitiveness in science and technology.

2 PRACE HPC eco-system: Tier-0 and Tier-1 sites, system upgrades and new systems

The first part of this section will present the changes implemented during the current reporting period (May 2016 – April 2017) that concern the status of Tier-0 sites, and the Tier-1 national sites providing Tier-1 for Tier-0 services. In the second part, the chronology and the status of the performed system upgrades is being detailed. Finally, the new systems that have been integrated into the PRACE eco-system are presented.

2.1 Maintaining the services

In the current reporting period, the main tasks of the operational procedures used to offer the PRACE services have been kept similar to what we had in the previous periods. Therefore, the work related to the managing of the operability procedures resulted more easily and has been well tested as was the continuation of the activity already in progress since the previous PRACE Projects.

Nevertheless, some minor changes to the operational procedures have been made and have been communicated to partners accordingly. Mainly these changes modified the way that data is organized in the PRACE Operations WIKI in order to improve productivity and reduce complexity. Furthermore, efforts have been kept to maintain the documentation up-to-date and feedback has been periodically inquired to the participant sites. In general terms, the procedure has been consolidated and now it is well established.

Consequently, all participants can introduce, find and update information of their PRACE HPC systems more easily and in standard way.

2.2 Status of Tier-0 & Tier-1 sites

The Tier-0 and Tier-1 systems constitute an HPC eco-system offering high level services to the European computational community. At present 7 Tier-0 systems and 28 Tier-1 systems are in production. Their status is monitored and operational quality is assured by employing a specific regular activity provided daily by the members of PRACE-4IP WP6 Task 6.1. This On-duty Activity is described below.

2.2.1 *On-duty Activity*

The On-duty Activity is supplied following a weekly schedule, and all partners who provide effort and/or systems in the WP6 guarantee it. The topics/incidents reported through this service are mainly related to operational issues and activities needed to maintain the distributed infrastructure in good shape. Specific requests from users are rare and normally are redirected to the local Help Desk of each individual site.

A report on the On-duty Activity is produced every week, reporting the status of the infrastructure, all occurred problems, the status and any other notification regarding the core services.

The related documentation on the WIKI site is constantly updated, using the report template agreed at the beginning of PRACE-4IP.

Starting from 1 April 2016, a schedule planning has been defined for the second year of activity engaging the 21 PRACE partners involve in the On-duty Activity. Each of them is in charge for monitoring the infrastructure, and reporting the related issues using the Trouble Ticket System (TTS). The 21 partners involved in the schedule are reported in the Table 1 below.

1 BSC	12 IDRIS
2 CASTORC	13 NCSA
3 CEA	14 NIIF
4 CINECA	15 PDC
5 CINES	16 PSNC
6 CSC	17 RZG
7 CYFRONET	18 SURFSARA
8 EPCC	19 VSB-TUO
9 JUELICH	20 UIO
10 HLRS	21 UHEM
11 ICHEC	

Table 1: PRACE partners involved in the on duty activity

A given partner is involved in the shift one week out of 21, i.e. every 4-5 months. The weekly reports of the On-duty Activity are stored on the PRACE Operation WIKI.

The TTS tool used by the On-duty Activity staff is the Best Practical RT 4.2.8, an enterprise-grade issue tracking system. It is freely available under the terms of Version 2 of the GNU License. It is hosted by CINECA on a Virtual Machine where it is maintained since its deployment during the PRACE-3IP project.

During the reporting period, the following KPI has been defined to better supervise the On-duty Activity (see Table 2):

Tickets resolving efficiency	
Description:	Number of tickets resolved by the Service Desk
Calculation:	O+R R/(O+R)*100
Inputs:	Number of open tickets reported (O) Number of resolved tickets reported (R)
Outputs:	Total number of worked tickets Percentage of resolved tickets with respect to total number of worked tickets

Tickets resolving efficiency	
	(%)
Time-interval:	Monthly every first working day (update during the first PRACE Operations meeting of the following month) Annual report
Threshold:	Percentage of resolved tickets respect worked tickets > 90% (Annual threshold)
Tools:	RT
ITIL Category:	Service Operation – Incident Management
'KPI Lead':	CINECA
Implementation plan:	Request Tracker (RT) provides all data necessary for computing this KPI. The necessary input data can be extracted using the Query Builder of RT.

Table 2: Tickets resolving efficiency KPI

In the last twelve months 101 tickets have been created and 94 of these have been resolved, with a percentage of resolved tickets with respect to the total number of worked tickets equal to 93,07%.

In PRACE-5IP, another KPI will be added to evaluate the service, as reported in Table 3.

Average Initial Response	
Description:	Average time taken between the time a user reports a ticket and the time that the Service Desk responds to that ticket
Calculation:	for i=1,(O+R) { T(i) = ST(i)-CT(i) } maximum value of vector T [sum of T(i)] / (O+R)
Inputs:	Number of open tickets (O) Number of resolved tickets (R) Vector of Creation times (CT) Vector of Starting times (ST) Vector of Initial Response times (T)
Outputs:	Maximum Initial Response Average Initial Response
Time-interval:	Monthly every first working day (update during the first PRACE Operations meeting of the following month)
Threshold:	< 1 working day without a motivation - unlimited with motivation
Tools:	RT
ITIL Category:	Service Operation – Incident Management
'KPI Lead':	CINECA
Implementation plan:	Request Tracker (RT) provides all data necessary for computing this KPI. The necessary output data can be extracted using the Query Builder of RT and managing the chart.

Table 3: Average initial response KPI

It is important to underline that the activity is in principal related to the traffic on the General queue; however the operator is in charge to report if a ticket moved into a site queue is not updated for more than a week, and/or if the owner is missing.

2.2.2 *Production Systems*

At present (April 2017) the Tier-0 ecosystem is made up of seven systems (MARCONI BDW & KNL is considered a single system), distributed over six sites, operated by six different partners, in the four Hosting Members countries (France, Germany, Italy and Spain) as reported in Table 4.

The peak performance of these Tier-0 systems ranges from more than1 PFlop/s up to more than 10.8 PFlop/s for the new MARCONI system in Italy. Only one Tier-0 is accelerated: The MareNostrum system at BSC, accelerated with Intel Xeon Phi.

All the Tier-0 systems are ranked in the Top500 (November 2016) and five of them are in the first 50 positions. The dominant vendor is IBM/Lenovo with five systems.

Since March 2017, CSCS (Switzerland) is a new Hosting Member in PRACE 2, and its Tier-0 system Piz Daint (Cray XC50, Xeon E5-2690v3, accelerated with NVIDIA Tesla P100) is going to be integrated in the PRACE eco-system.

Partner	**Country**	**Tier-0**	**Architecture CPU**	**Rpeak (TFlop/s) Top500#(nov16)**
JUELICH	Germany	Juqueen	IBM BlueGene/Q IBM PowerPC A2; 16-core; 1.60GHz	5802.0 19
GCS-LRZ		SuperMUC	IBM iDataPlex Intel Sandy Bridge-EP; 8-core; 2.7GHz	3185.0 36
		SuperMUC phase 2	Lenovo NeXtScale Intel Haswell-EP 14 core; 2.6GHz	3580.0 37
GCS-HLRS		Hazelhen	Cray XC40 Intel Haswell E5 2680v3; 12-core; 2.5GHz	7420.0 14
GENCI	France	Curie (Thin)	Bull Bullx B510 Intel Sandy Bridge-EP; 8-core; 2.70GHz	1741.8 129
BSC	Spain	Mare Nostrum	IBM iDataPlex Intel Sandy Bridge-EP; 8-core; 2.7GHz Intel Xeon Phi	1017.0 93
CINECA	Italy	MARCONI (BDW)	Lenovo NextScale (Intel OmniPath Cluster) Intel Xeon E5-2697 v4 (Broadwell); 18-core; 2.30 GHz	2000.0 57
		MARCONI (KNL)	Lenovo Adam Pass (Intel OmniPath Cluster) Intel Xeon Phi 7250 (Knights	11000.0 12

			Landing) 68-core; 1.40 GHz	

Table 4: PRACE Tier-0 systems

As far as the Tier-1 eco-system is concerned, 28 systems are operating as Tier-1 for Tier-0 services (the four Cartesius partitions have been considered as different systems). These Tier-1 systems are distributed among 19 different PRACE sites, operated by 16 partners, in 16 different European Countries. Table 5 presents the list of the Tier-1 systems.

The peak performance ranges from very small system partitions (<10 TFlop/s) up to large systems in excess of 2.5 PFlop/s. Nine systems deliver more than 1 PFlop/s.

About half of the Tier-1 systems are accelerated, either with Intel Xeon Phi or Nvidia cards. Several different vendors and architectures are present and this is a real advantage for the PRACE HPC infrastructure: HP, SGI, Bull, Cray, IBM and Lenovo are nearly equally represented; BlueGene, iDataPlex/NeXtScale, Bullx are the most popular architectures.

Partner	Country	Tier-1	Architecture CPU	Rpeak (TFlop/s)
GCS / RZG	Germany	Hydra	IBM iDataPlex Intel Sandy Bridge-EP; 8core; 2.6GHz + Intel Ivy Bridge; 10-core; 2.8 GHz, Nvidia GPU, Intel Xeon Phi	1501.7
GENCI	France	Turing	IBM BlueGene/Q IBM PowerPC A2; 16-core; 1.60 GHz	1258.0
EPCC	UK	Archer BlueJoule	Cray XC30 Intel Ivy Bridge; 12-core; 2.7GHz IBM BlueGene/Q IBM PowerPC A2; 16-core; 1.60 GHz	2550.5 1258.3
BSC	Spain	Minotauro	Bull Bullx B505 Intel Westmere-EP; 6-core; 2.53 GHz, Nvidia/M2090	185.8
CSC	Finland	Sisu	Cray XC40 Intel Haswell; 12-core; 2.60 GHz	1688.0
SURF sara	Netherlands	Cartesius phase 1 + 2	Bull Bullx B720/B710 Intel Haswell; 12-core; 2.6 GHz + Intel Ivy Bridge; 12-core; 2.4GHz+ Intel Sandy Bridge; 8-core; 2.7 GHz	1349.0
		Cartesius GPU extension	Bull Bullx B515 Intel Ivy Bridge 8 core; 2.5 GHz Nvidia K40	210.0
		Cartesius Sequana extension 1	Bull Sequana X1110 Intel Broadwell 16 core; 2.6 GHz	236.0
		Cartesius Sequana extension 2	Bull Sequana X1210 Intel Knights Landing 64 core; 1.3 GHz	48.0
SNIC	Sweden	Beskow	Cray XC40, Intel Haswell; 16-core; 2.3 GHz	1973.0
CINECA	Italy	Galileo	IBM NeXtScale, Intel Haswell; 8-core; 2.40 GHz Nvidia K80, Intel Xeon Phi	1000.0

Partner	Country	Tier-1	Architecture CPU	Rpeak (TFlop/s)
PSNC	Poland	Zeus BigMem	HP BL685c G7 AMD Interlagos; 16-core; 2.3 GHz	61.2
		Zeus GPGPU	HP SL390s Intel Westmere-EP; 6-core; 2.45 GHz, Nvidia/M2090	136.8
		Cane	SGI Rackable C1103-G15, AMD Interlagos; 12-core; 2.40 GHz, Nvidia M2050	224.3
		Chimera	SGI UV 1000 Intel Westmere-EX; 8-core; 2.67 GHz	21.8
		Supernova	HP Cluster Platform 3000 BL2x220 Intel Westmere-EP; 6-core; 2.67 GHz	30.0
SIGMA	Norway	Abel	MEGWARE MiriQuid Intel Sandy Bridge-EP; 8-core; 2.6 GHz	178.6
NUI Galway	Ireland	Fionn	SGI ICE X Intel Ivy Bridge; 12-core; 2.4 GHz	147.5
UHEM	Turkey	Karadeniz	HP Proliant BL460 Intel Nehalem-EP; 4-core; 2.67 GHz	2.5
CaSToRC	Cyprus	CyTera	IBM IDataPlex Intel Westmere; 12-core; 2.67 GHz, Nvidia M2070	35.0
NCSA	Bulgaria	EaEcnis	IBM BlueGene/P IBM PowerPC 450; 4-core; 850 MHz	27.8
IT4I/VSB-TUO	Czech Rep	Anselm	Bull Bullx B510/B515, Intel Sandy Bridge-EP; 8-core; 2.4 GHz	66.0
		Solomon	SGI ICE-X Intel Xeon E5-2680v3; 12-core; 2.5 GHz, Intel Xeon Phi	2000.0
NIIF	Hungary	NIIFI SC	HP Cluster Platform 4000SL AMD Magny-Cours; 12-core; 2.2 GHz	5.48
		Seged	HP Cluster Platform 4000SL AMD Magny-Cours; 12-core; 2.2 GHz, Nvidia M2070	14.0
		Leo	HP SL250s Intel Xeon E5-2650 v2; 2.60GHz, Nvidia K20, K 40	254.0
		PHItagoras	HP SL250s Intel Xeon E5-2680 v2; 2.80 GHz, Intel/Xeon Phi 7120	27.0

Table 5: PRACE Tier-1 systems

2.3 System Upgrades

In this section, we describe the activity related to the system upgrades and integration in the PRACE eco-system, being Tier-0 or Tier-1 systems involved in the Tier-1 for Tier-0 services.

In the current period 2016-2017, this activity has been included under WP6 subtask 6.1.2 and is led by Barcelona Supercomputer Centre.

2.3.1 *Operational procedures for new systems and system upgrades*

Current operational procedures are well documented in the PRACE WIKI and BSCW and available at the following links:

- Integration of new Tier-0/Tier-1 sites
  - Procedure/Template:
    - WIKI template that new sites must fulfil in order to complete their integration into PRACE infrastructure (https://prace-wiki.fz-juelich.de/bin/view/Prace2IP/WP6/TemplateNewTier1Site)
  - Information for new Tier-1 sites
    - BSCW presentation about basic concepts that new sites should know before starting their integration into PRACE infrastructure (https://bscw.zam.kfa-juelich.de/bscw/bscw.cgi/d622792/Information for new sites.ppt)
- Upgrade of systems (Tier-0/Tier-1)
  - Procedure for upgrade of systems:
    - WIKI guide with information to upgrade systems in the PRACE infrastructure (https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/SystemUpgradeProcedure)
  - Report template for upgrades
    - WIKI template for the completion of the upgrade procedure (https://prace-wiki.fz-juelich.de/bin/view/PRACE/Operations/SystemUpgradeTemplate)

This documentation ensures that all systems follow equal procedures and are in line with the Service Catalogue requirements.

2.3.2 *Overview of System Upgrades in PRACE-4IP*

Eight sites started the upgrade of their respective systems, and six of them successfully completed the integration while one is still on-going and one did not succeed in the process for site specific reasons. Two of these sites have upgraded their Tier-0 systems, and the others their Tier-1 systems. At the time of writing this report, the latest announcement comes from BSC, for the full upgrade procedure of the new Marenostrum 4 Tier-0 system.

Some services, considered optional, have not been included in the Service Catalogue offer, of some sites, due to a lack of physical resources (like the connection to GEANT network), to services not heavily used (like Unicore), or some other motivations. The mission of this subtask is to ensure that all the sites can offer all the core services, and guarantee that they are installed, and run correctly, following the PRACE adopted rules.

All the six sites have completed the upgrades following the standard procedures defined internally in the PRACE Operations Wiki, with no major issues, providing a good feedback of the process to the subtask leaders.

In the Table 6 below, we show a brief overview of the upgraded systems and the completion dates.

System	Site	Tier	Status	Last Check
Beskow	KTH-PDC	Tier-1	Completed (2015-03-01)	02/12/2015

System	Site	Tier	Status	Last Check
Leo	NIIF	Tier-1	Completed (2016-02-05)	08/02/2016
Hazel Hen	HLRS	Tier-0	Completed (2015-10-16)	19/02/2016
Occigen	CINES	Tier-1	Aborted (2016-02-04)	19/02/2016
Minotauro	BSC	Tier-1	Completed (2016-05-25)	25/05/2016
Cartesisus	SURFSara	Tier-1	Completed (22/12/2016)	22/12/2016
Phitagoras	NIIF	Tier-1	Completed (08/02/2017)	08/02/2017

Table 6: System upgrades in PRACE-4IP

For Marenostrum 4 (Tier-0 system in BSC), the upgrading activity started in March 2017 and its completion is foreseen for July 2017.

2.3.3 *Details of upgraded sites*

The full list of the upgrades sites in the PRACE-4IP time-frame is reported below.

1. **Beskow – KTH/PDC – Tier-1**

Beskow has replaced the old Lindgren system that was fully dismantled after successful installation of the new system. Important points for Beskow are:

- Cray XC40 system architecture, based on Intel Haswell and Cray Aries interconnect, [4];
- Intended lifetime of 4 years, until 31/12/2018;
- This system is not connected to the PRACE network, so the following services could not be provided:
    - PRACE Network
    - Data services (GridFTP)
    - Compute services (Unicore)
    - AAA Services (GSISSH).

2. **Hazel Hen – HLRS – Tier-0**

Hazel Hen is an upgrade/follow on project of the previous system known as Hornet. The Cray XC40 Hazel Hen is based on Intel Haswell Processor and the Cray Aries network. Important points for Hazel Hen are:

- Cray XC40 system architecture, based on Intel Haswell and Cray Aries interconnect; Info: see [5];
- This is a Tier-0 system;
- System provides all the PRACE core services, except for:
    - UNICORE: Available only upon user request
    - LDAP: Not connected to PRACE LDAP
    - Accounting data: Available through their web interface
    - GSISSH: Not available.

3. **Leo – NIIF – Tier-1**

Leo is the system located in Debrecen (Hungary) and is an extension of the previous system named "Debrecen", adding now nodes with GPU capabilities. The system upgrade has been performed with no issues and all the services have been implemented.

- The system is deployed using the Intel Xeon E5-2650 v2 architecture and InfiniBand FDR as interconnect. GPUs model are Nvidia K20/K40;
- Info: see [8];
- The system provides all the core services defined in the PRACE catalogue.

4. **Occigen – CINES – Tier-1**

CINES started its upgrade procedure on June 2014, installed the system on January 2015, and, in February 2016, decided to not provide PRACE services. The reasons are mainly technical, as they are missing some front end and data transfer nodes and they are not going to connect the system to the PRACE network.

5. **Minotauro – BSC – Tier-1**

Minotauro has upgraded its compute power and now it provides two node partitions. One provided with 61 nodes equipped with Tesla M2090 GPUs, and another with 39 nodes equipped with NVIDIA K80 GPUs.

- Intel Xeon E5649;
- This is a Tier-1 system;
- Nvidia K80 GPUs;
- SLURM 15.08.10 + Centos 6.5 (Bull);
- Info: see [10];
- The system provides all the PRACE core services.

6. **Phitagoras – NIIF – Tier-1**

This system, called Budapest2 at NIIF consists of 14 compute nodes model HP SL250s.

- 2x Intel Xeon E5-2680v2 @ 2.8Ghz;
- 64GiB RAM;
- 14x2xIntel(r) Xeon Phi(tm);
- MIC SE10/7120;
- 500TB dedicated storage;
- InfiniBand NB FDR;
- Slurm Scheduler;
- Total of 27 Tflop/s;
- Info: see [11].

7. **Marenostrum 4 – BSC – Tier-0**

The Marenostrum 3 computer system will be fully decommissioned and a new system will be installed. It will be seen as an upgrade because all PRACE services will continue as usual but the core of the supercomputer will be changed entirely.

- Lenovo + Intel + Fujitsu technologies;
- Features:
  - Performance capacity of 13.7 PFlop/s, 12x improvement from Marenostrum 3;
  - Energy: 1.3Mw;

    - >10 PB IBM Spectrum Protect (GPFS) Storage;
- 3 emerging technologies clusters:
    - IBM POWER9 + NVIDIA GPUs (1.5 Pflop/s, Lenovo);
    - Intel Knights Landing and Intel Knights Hill (0.5 PFlop/s, Fujitsu/Lenovo);
    - 64 bit ARM v8 (0.5 PFlop/s, Fujitsu);
- Info: see [12].

8. Cartesius – SURFsara – Tier-1

Cartesius has been extended with a Bull Sequana cell, in which two new node types are introduced:

- 177 sequana X1110 thin nodes, 2 × 16-core 2.6 GHz Intel Xeon E5-2697A v4 (Broadwell) CPUs/node, 64 GB/node;
- 18 sequana X1210 Xeon Phi nodes, 1 × 64-core 1.3 GHz Intel Xeon Phi 7230 (Knights Landing) CPU/node, 96 GB/node.

The nodes have been integrated into the existing environment. Though each new type is in a distinct partition, they are integrated into the Cartesius batch environment (SLURM) and they are attached to the same system interconnect – which is both a storage and an MPI interconnect. The Sequana cell is the last extension of Cartesius. The intended lifetime for these nodes is until 15/10/2018.

2.4 New Tier-0/Tier-1 sites and systems

In this section, we present the new sites and systems that are going to be integrated in the PRACE infrastructure.

2.4.1 *Overview of new Tier-0/Tier-1 sites & systems on PRACE-4IP*

In PRACE-4IP ten new systems have been planned to be included in the PRACE infrastructure and to start providing services to it, as shown in the table below. Six systems have been successfully integrated, one is in the integration phase (Cyfronet) and three more are scheduled in the next few months (IUCC in Israel; UL and FIS in Slovenia).

Site	System Name	Systm. Details URL	Tier	Integration Start Date	Status	Last Check
Cyfronet (Poland)	Prometheus	https://kdm.cyfronet.pl/portal/Prometheus	Tier-1	aprox. 09/2016	In progress	28/09/2016
PSNC (Poland)	Eagle	https://wiki.man.poznan.pl/hpc/index.php?title=Eagle	Tier-1	aprox. 09/2016	Completed	01/02/2017
CINECA (Italy)	Marconi	http://www.hpc.cineca.it/hardware/marconi	Tier-0	02/01/2017	Completed (02/2017)	21/02/2017
UL (Slovenia)	n/d	n/d	Tier-1	aprox. 06/2016	Planned	24/02/2016
FIS (Slovenia)	n/d	n/d	Tier-1	aprox. 06/2016	Planned	24/02/2016
GRNET (Greece)	Aris	http://doc.aris.grnet.gr	Tier-1	01/09/2015	Completed	21/03/2017
CINECA (Italy)	Galileo	http://www.hpc.cineca.it/hardware/galileo	Tier-1	01/11/2015	Completed (01/2016)	26/01/2016

Site	System Name	Systm. Details URL	Tier	Integration Start Date	Status	Last Check
WCSS (Poland)	Bem	http://kdm.wcss.wroc.pl/wiki/Bem_overview	Tier	31/01/2016	Completed (03/2016)	05/04/2016
IUCC (Israel)	Moab	n/d	Tier-1	1/2/2017	Planned	23/01/2017
IT4I (Czech R.)	Salomon	https://docs.it4i.cz/salomon/hardware-overview	Tier-1	29/01/2016	Completed (01/2016)	01/02/2017

Table 7: New Tier-0/1 sites & systems in PRACE 4-IP

2.4.2 *Details of new Tier-0/Tier-1 sites*

The full list of the new systems integrated in the PRACE-4IP time-frame is reported below.

1. **Galileo – CINECA - Tier-1**

The integration of this system has started in November 2015 and completed on January 2016. This new system is equipped with up-to date Intel accelerators (Intel Xeon Phi 7120p) and NVIDIA accelerators (NVIDIA Tesla K80). Processor architecture is Intel Haswell based with 128GiB RAM each node.

Galileo is mainly used to develop and run applications targeted at hybrid architectures, leveraging software applications in the fields of computational fluid dynamics, material and life science, and geophysics.

The computing system has been made fully available to European researchers as a Tier-1 system of the PRACE infrastructure, providing all the services of the PRACE catalogue. For more info see [9].

2. **Salomon – IT4/VSB-TUO – Tier-1**

This system initially was categorized under a system upgrade because it was taken as an extension of the computing power of the centre, but it is, in fact, a new system so we moved it from upgrades to a new system.

- The system consists of 1008 compute nodes equipped with Intel Xeon E5-2680v3 architecture and at least 128GB RAM. Interconnection is InfiniBand based. For more info see [7].
- The site provides all the core services defined in the PRACE catalogue.

3. **ARIS – GRNET – Tier-1**

This system is composed of four islands of nodes with different architecture, it has completed its integration in February 2017. Its technical characteristics are specified below:

- 532 nodes, 4 islands:
    - 426 thin nodes – 64G RAM
        - Ivy Bridge – Xeon E5-2680v2
        - IBM NeXtScale nx360
    - 44 GPU (2xNvidia Tesla K40)
        - Dell PowerEdge R730
        - Haswell – Xeon E5-2660v3
    - 18 Phi: (2xIntel Xeon Phi 7210P)

        - Dell PowerEdge R760
        - Haswell – Xeon E5-2660v3
    - 44 Fat nodes – 512G RAM
        - Dell PowerEdge R820
        - SandyBridge – Xeon E5-4650v2
- InfiniBand FDR 56Gbps
- IBM GPFS 750TB

For more info see [13].

4. Prometheus - Cyfronet – Tier-1

Prometheus, a new supercomputer with 2.4 PFlop/s of theoretical performance, has been listed at 59th position in the Top500 list (November 2016). Hewlett-Packard has built this top-efficient supercomputer located in Poland, according to requirements and partial design provided by Cyfronet. To date, Prometheus is one of the biggest installations of this type in the world and the first one in Europe, based on the latest technology of direct liquid cooling.

Prometheus consists of more than 2,200 servers based on the HP Apollo 8000 platform, combined with super-fast InfiniBand network with 56 Gb/s capacity. Its energy saving and high-performance Intel Haswell latest-generation processors offer more than 53,000 cores. These are accompanied by 279 TB RAM in total, and by two storage file systems of 10 PB total capacity, and 180 GB/s access speed. Prometheus has been also equipped with 144 NVidia Tesla GPGPUs. For more info see [14].

5. Eagle – PSNC – Tier-1

The Eagle cluster installed at PSNC combines 1,032 Huawei blade servers in the E9000 chassis with CoolIT Systems Rack DCLC liquid cooling solution to manage over 80% of the server heat.

- Architecture: Intel Xeon E5-2697v3;
- Number of nodes: 1233;
- Network interconnections: InfiniBand FDR;
- Number of CPU cores: 32984;
- Total computing power: 1372,13 TFlop/s;
- Size of system memory: 120,6 TB.

For more info see [15].

6. Marconi – CINECA – Tier-0

MARCONI is the CINECA's Tier-0 system, co-designed by CINECA and Lenovo, and based on the Lenovo NeXtScale platform.

MARCONI, based on the next-generation of the Intel Xeon Phi™ product family alongside with Intel Xeon processor E5-2600v4 product family.

The new MARCONI system is composed of three different partitions and the installation has been scheduled in three different phases, between April 2016 and July 2017, according to the following plan:

- A1 MARCONI-BDW: 1512 nodes (2x18 cores) Intel Xeon E5-2697 v4 (Broadwell) at 2.3 GHz (128GB RAM), Internal Network Intel OmniPath. The system has a computational power of 2 PFlop/s. This partition has been put in production in July 2016;

- A2 MARCONI-KNL: 3600 nodes Intel Knights Landing (Intel Xeon Phi7250 at 1.4 GHz). The system, based on many-core architecture (68 cores per node), enables an overall configuration of 244800 cores with a peak performance of 11 Pflop/s. The system has been installed in October 2016 and has been ranked at number 12 in the Top 500 list (November 2016);
- A3 MARCONI-SKL: A partition based on the future generation Intel Xeon processors (Sky Lakes) will be integrated in July 2017.

When the third partition will be in production, the whole MARCONI system will reach a total computational power in excess of 20 Pflop/s.

A high-performance Lenovo GSS storage subsystem, that integrates the IBM Spectrum Scale™ (GPFS) file system, is connected to the new Intel Omni-Path Fabric and provides the data storage capacity (20 PB raw local storage plus 20 PB data repository). For more info see [16].

7. Bem – WCSS– Tier-1

This cluster was installed in WCSS (Poland) in March 2015. The computer architecture is based on Intel Xeon processor E5-2670v3 (Haswell microarchitecture):

- 720 nodes of 24-core computing (Intel Xeon E5-2670 v3 2.3 GHz, Haswell);
- 192 nodes computational 28-core (Intel Xeon E5-2697 v3 2.6 GHz, Haswell);
- 22,656 computational cores;
- 74.6 TB of RAM (64, 128 or 512 GB / node);
- Space for home directories: 11 TB NFS / home;
- Temporary space: 1.1 PB Lustre;
- Computational network: InfiniBand FDR, 56 Gb/s (fat-tree topology, pruning factor 3 : 1);
- Management network: Ethernet 1G / 10G;
- Operating system: CentOS 6.8;
- Computing power: 860 TFlop/s.

For more info see [17].

8. Moab – IUCC – Tier-1

IUCC (Israel) announced recently to WP6 the intention to integrate their Tier-1 system in the PRACE HPC infrastructure. At the time of this report, we do not still have the specifications of the system. For more info see [18].

9. UL and FIS – Slovenia – Tier-1

These two sites still need to define the specifications of the systems to integrate in PRACE.

3 Operational Services

Since the previous PRACE IP Projects we defined a process of selection and deployment of a common set of services aiming at presenting all Tier-0 centres as a single distributed infrastructure, instead of a set of individual systems/computing facilities.

Common services are divided into thematic categories: Network, Data, Compute, AAA, User, Monitoring and Generic. Each service category has a responsible person who is in charge of managing all the information and decisions related to a specific service area.

Selection of common services is published in the PRACE service catalogue and once chosen, the responsibility for a service is taken by the respective service area.

The following sections provide an up-date of the status of each service category and the main achievements within the current reporting period.

3.1 Network Services

In deliverable D6.1 [40] we have described in detail the history of the PRACE network and the evolution since the early beginnings. We also discussed in that document the pros and cons of the chosen physical setup and the implications on going „virtual“.

The PRACE network setup in the past can be seen in the Figure 1 below:

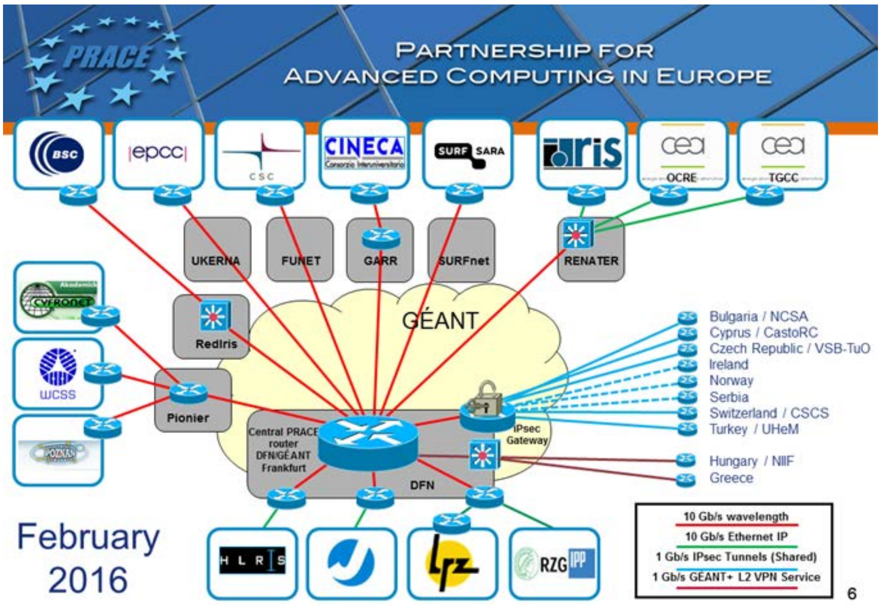

Figure 1: Old PRACE network topology

3.1.1 *The new PRACE network*

Since each site further agreed to connect no other systems to the PRACE backbone and not to route any other traffic via this infrastructure the “Net of Trust” idea will not change by migrating to a virtual environment. This idea, and, more in general, the design of a new network architecture for the PRACE infrastructure, was discussed for long internally in WP6, and with GEANT.

The idea was to move from a 10 Gb/s star-topology (designed in2006) to a MD-VPN (Multi Domain Virtual Private Network) solution which allows connecting systems or small networks with a security level comparable to the current dedicated links. Setting up a PRACE-VPN with this technology would replace the existing dedicated wavelengths between the different PRACE partners by a VPN infrastructure, i.e. separate VLANs on the existing PRACE partner’s NREN connections.

The proposed project was submitted to the PRACE Management Board that, in November 2016, approved the migration to the new infrastructure. From then on, the migration has started. Setting up processes took time so that, so far, only seven sites have connected to the PRACE-VPN. Actually, several others partners have started the negotiations with their local NRENs to be connected to the new infrastructure.

In the meantime technical pre-requisites have been setup and operational procedures have been implemented to monitor the new infrastructure and to provide a smooth migration phase. For this, a physical connection between the old and the new infrastructure has been implemented which allows partners having migrated to reach the other partners still connected to the old network. This current setup is sketched in Figure 2.

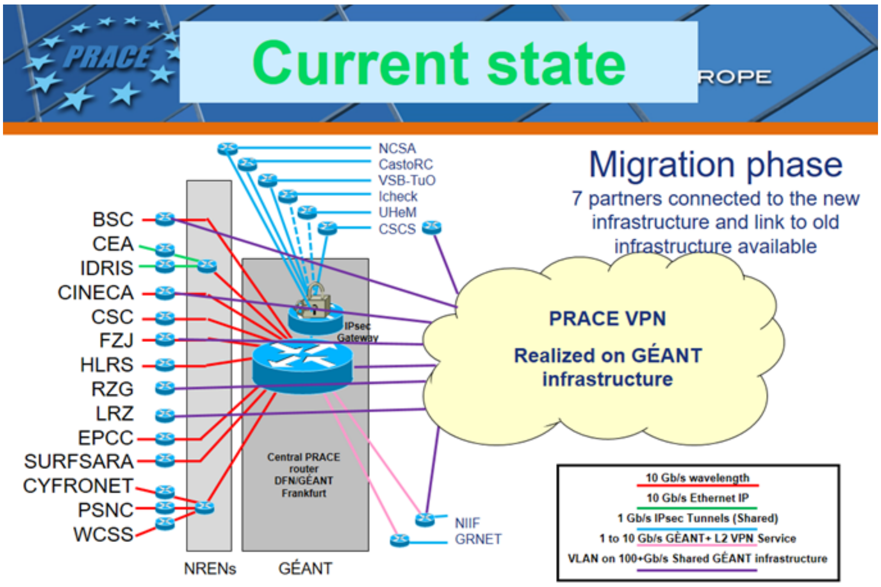

Figure 2: PRACE network topology

It is planned to complete the full migration at the end of April 2017, so that the central router and IPSEC tunnel router in Frankfurt can be decommissioned hopefully in May this year. The future PRACE-VPN will not have a central router anymore, but instead provide route information by the BGP protocol in the PRACE-VPN network. Figure 3 below presents a logical view of the final layout.

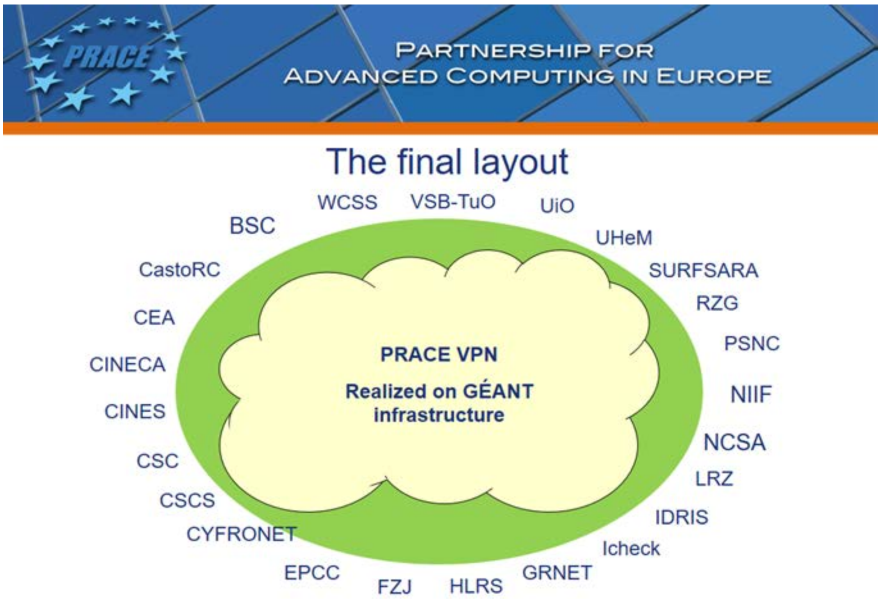

Figure 3: New PRACE network topology layout

Besides the migration activities, the main tasks performed have been the continuous optimization, operation and monitoring of the network infrastructure, attaching the new and detaching the old HPC systems to/from the network infrastructure.

The new PRACE network infrastructure is now progressing toward the completion phase, resulting in a stable, performant and less expensive solution than the old dedicated PRACE network.

3.2 Data Services

Data services within PRACE-4IP are delivered by GridFTP as a core service, gtransfer as an additional service and MC-GPFS as an optional service.

3.2.1 *GridFTP*

GridFTP is the core service for data transfer and it is therefore mandatory to be installed on every PRACE site attached to the PRACE network. GridFTP is a widely-used and state-of-the-art protocol for high performance file transfers. It can utilize high bandwidths among PRACE computing centres; therefore, it has been chosen as the core service for data transfer within PRACE. The GridFTP protocol enables high performance transfers by parallel streaming of data through the network.

PRACE internal documentation for GridFTP including an installation and tuning description as well as a special init script is available via the PRACE WIKI.

Status of deployment

Table 8 gives the status of GridFTP deployment on Tier-0 systems in April 2017. All the Tier-0 systems have deployed GridFTP and use the split configuration as recommended in the previous time-frame of PRACE-4IP [40].

Site / Tier-0 System	Version (GridFTP / Globus Toolkit)
BSC / MareNostrum	6.19 / GT 6.0.14
CEA / CURIE	Frontend: 8.7/GT 6.0 Backend: 7.25 / GT 6.0.0
CINECA / MARCONI	8.7 / GT 6.0
FZJ / JUQUEEN	6.38 / GT 6.0
HLRS / Hazel Hen	7.26 / GT 6.0
LRZ / SuperMUC	9.4 / GT 6.0

Table 8: GridFTP deployment status on Tier-0 systems in March 2017

Deployment status of GridFTP servers on Tier-1 systems in March 2017 is shown in Table 9. So far, 18 Tier-1 sites successfully deploy GridFTP.

Site / Tier-1 System	Version (GridFTP / Globus Toolkit)
BSC / MinoTauro	6.19 / GT 6.0.14
CaSToRC / CY-TERA	9.1 / GT 6.0
CINECA / GALILEO	7.26 / GT 6.0

Site / Tier-1 System	Version (GridFTP / Globus Toolkit)
CSC / SISU	9.3 / GT 6.0
Cyfronet / ZEUS	7.26 / GT 6.0
EPCC	GT 6.0
GRNET	11.1 /GT 6.0
ICHEC / Fionn	8.9 / GT 6.0
IDRIS / Turing	11.1 / GT 6.0
NIIF / NIIFI SC	11.8 / GT 6.0
PSNC	7.11 / GT 6.0
RZG / Hydra	9.17 / GT 6.0
SURFsara / Cartesius	11.8 / GT 6.0
UHeM / Karadeniz	7.25 / GT 6.0
UiO / Abel	9.4 / GT 6.0
IT4I/VSB-TUO / Anselm	10.6 / GT 6.0
IT4I/VSB-TUO / Salomon	10.6 / GT 6.0
WCSS / Bem	9.4 / GT 6.0

Table 9: GridFTP deployment status on Tier-1 systems in March 2017

Security & Bugs

Globus Toolkit 5.2.5 reached end of life effective 1 November 2015. Therefore, all sites have upgraded to Globus Toolkit 6.0 within the last year of the PRACE-4IP project.

GridFTP Installation Screencast

To ease the GridFTP installation procedure for new sites in PRACE, as a Tier-0 or a Tier-1 site, a screen-cast about GridFTP installation was produced. The recording was performed with the Linux tool ttyrec and it can now be replayed with the tool ttyplay. This way, the whole installation and configuration procedure can be followed stepwise. The screen-cast is available for all PRACE sites on the internal GridFTP installation WIKI page.

GridFTP Logfile Analyzer Tool

Logging of the data transfers provides meaningful information, for example on network usage and GridFTP status. For this, a GridFTP logfile analyzer tool has been developed at the end of PRACE-3IP.

The tool uses the GridFTP log files to extract the required information. Automated analysis of these log files seems to be easy, but the used split configuration adds complexity as parts of the information is separated. However, each transfer is uniquely identified by the keys NBYTES and FILE. So the data is independently read from the frontend and from the backend(s) and stored in distinct DB tables. These tables are then combined to one single table containing all the necessary information. Due to privacy reasons, all transfer data is anonymized.

The implementation is via shell scripts and can be used with MySQL and sqlite databases and is available on the PRACE SVN.

The GridFTP logfile analyzer tool has been extensively and successfully tested in a distributed effort on several PRACE sites.

3.2.2 *Gtransfer*

Gtransfer is an advanced toolkit for performing data transfers. It is a CLI toolkit based on GridFTP that makes use of other tools (like tgftp, globus-url-copy (guc) and uberftp) to provide an easy to use service for users. Gtransfer has been developed at HLRS and is supported by Frank Scheiner (HLRS).

Gtransfer is an additional service in PRACE, however, many sites, such as CINECA, CINES, CSC, HLRS, ICHEC, IT4I/VSB-TUO and SURFsara provide the service and are using it in production.

With the recent update to version 0.8.0, wrappers for uberftp have been included into the gtransfer toolkit, which provide shorthand functionality to users for common operations on remote GridFTP servers, like, for example, gls, to list remote directory contents, gmkdir, to create directories on remote GridFTP servers, or grm, to remove remote files. Other new additional functionality can be followed on the gtransfer releases page on GitHub, [41].

3.2.3 *Gsatellite*

Gsatellite is a personal job scheduler for the command line which executes jobs in the background without the need for user intervention. Development and support is also done by Frank Scheiner (HLRS). In principle gsatellite can schedule and run any job that can be described in a script just like any ordinary batch scheduling system (BSS). But together with gtransfer it can be used to submit and schedule long running data transfers as jobs. Data transfer jobs can be interrupted and continued later on and are monitored and restarted automatically in case of recoverable errors during the data transfer. This is accomplished by gsatellite's programmable per job type ability to react differently on specific job exit codes.

Gsatellite was successfully used in WP6 Task 6.2 for a prototype project as part of two workflows which in general accomplished the task of feeding input data to a computational job, executing the computational job and finally transferring the results back to the user in a fault tolerant way. This provides an easy to use service to users as they only need to provide the input data and get the results back without further user intervention. More details can be found in the upcoming white paper “Using Gsatellite for data intensive procedures between large-scale scientific instruments and HPC”

It is proposed that gsatellite is added to the service catalogue. It then would undergo the service certification process in PRACE-5IP.

3.2.4 *MC-GPFS*

MC-GPFS is an optional service but actually, no site does provide it anymore. For this reason the service is currently removed from the service catalogue.

3.3 Compute Services

Compute services provide interfaces between users and computing capabilities. The target of this activity is to find out what services can be commonly provided in PRACE and try to maintain the uniformity among all sites providing computational hours.

In the PRACE-4IP timeframe the work consisted on maintaining the different components that are provided in the service catalogue, mainly the batch scheduling systems and Unicore. An update of the activity done in the reporting period for these services is provided in the following.

3.3.1 *Local batch scheduling*

Local batch scheduling systems identify the software tools responsible for managing user jobs. The service consists on maintaining a list of different software components that current supercomputers deploy now, thus providing a statistic of which are the most used ones. In Table 10 it is possible to see the relationship between supercomputer and batch system. One of the most interesting systems is SLURM, that is step by step overtaking others. SLURM is an Open Source software maintained by SchedMD and one of its strong points is its modularity and flexibility, making relatively easy to adapt source code to site needs. These characteristics ensures a rapid growth and adaption to new systems, platforms and architectures and we think this is the reason that many sites are using it. The code of this software is open so it is possible to see that it has a very high quality. Sites are getting support from the community but also directly from the developers and, moreover, there is the possibility of having a contract with SchedMD for premium support. All these features and other ones make SLURM the preferred solution among different sites. Another responsibility of Compute Services in PRACE-4IP has been the maintenance of a set of scripts to wrap the different inputs that different systems could take as parameters, motivated by the differences in interaction with batch systems for different products. A scripted file must be edited with a full description of a job, e.g. resources like core/cpu-hours, number of cores/cpu, I/O staging, service level, and all needed actions for preparing the execution environment.

These differences made it interesting to develop a set of wrappers with a common syntax for describing a job around the different batch systems and platforms. Such wrappers should make it easy for users to migrate their work from one system to another. Nowadays, given the complexity of systems and architectures, with different features, the work should be enhanced and reviewed in PRACE-5IP to better enable the wrappers and adapt to different types of systems.

3.3.2 *Unicore*

Another way to create and operate on a common layer is through an even higher abstraction level. The UNICORE software, which is part of the PRACE software portfolio since the beginning, and which was originally adopted in the DEISA projects, accomplishes this. It allowed a user to manage single jobs as well as a workflow of jobs remotely through a Java-based graphical interface (URC) and/or a command line client (UCC). Access to end-systems relies on account information stored in the PRACE central LDAP, while resources are published on a central directory service by sites or resource providers.

Partner	**Site**	**Machine Name**	**Tier**	**Batch System**
1. FZJ (de)	FZJ	Juqueen	Tier-0	LoadLeveler
2. GCS (de)	LRZ	SuperMUC phase 1	Tier-0	LoadLeveler
2. GCS (de)	LRZ	SuperMUC phase 2	Tier-0	LoadLeveler
2. GCS (de)	HLRS	Hazelhen	Tier-0	Torque/Moab
2. GCS (de)	RZG	Hydra	Tier-1	LoadLeveler
3. GENCI (fr)	CEA	Curie (Fat)	Tier-0	SLURM
3. GENCI (fr)	CEA	Curie (Hybrid)	Tier-0	SLURM
3. GENCI (fr)	CEA	Curie (Thin)	Tier-0	SLURM

3. GENCI (fr)	IDRIS	Turing	Tier-1	LoadLeveler
4. EPCC (uk)	EPCC	ARCHER	Tier-1	PBSPro/APLS
4. EPCC (uk)	STFC	Blue Joule	Tier-1	LoadLeveler
5. BSC (es)	BSC	Marenostrum III	Tier-0	SLURM
5. BSC (es)	BSC	Minotauro	Tier-1	SLURM
6. CSC (fi)	CSC	Sisu	Tier-1	SLURM
7. ETH (ch)	CSCS	Rosa	Tier-1	SLURM
8. SURFSARA (nl)	SURFsara	Cartesius	Tier-1	SLURM
8. SURFSARA (nl)	SURFsara	Cartesius (GPU)	Tier-1	SLURM
10. SNIC (se)	PDC	Beskow	Tier-1	SLURM
11. CINECA (it)	CINECA	Marconi	Tier-0	PBS Pro
11. CINECA (it)	CINECA	Galileo	Tier-1	PBS Pro
12. PSNC (pl)	CYFRONET	Zeus (BigMem)	Tier-1	PBS Pro
12. PSNC (pl)	CYFRONET	Zeus (GPGPU)	Tier-1	PBS Pro
12. PSNC (pl)	PSNC	Eagle	Tier-1	SLURM
12. PSNC (pl)	PSNC	Cane	Tier-1	SLURM
12. PSNC (pl)	PSNC	Chimera	Tier-1	SLURM
12. PSNC (pl)	WCSS	Supernova	Tier-1	PBS Pro
12. PSNC (pl)	WCSS	Bem	Tier-1	PBS Pro
13. Sigma (no)	UIO	Abel	Tier-1	SLURM
16. NUI (ie)	ICHEC	Fionn	Tier-1	Torque/Moab
17. UHEM (tr)	UHEM	Karadeniz	Tier-1	IBM LSF®
18. CaSToRC (cy)	CaSToRC	CY-TERA	Tier-1	SLURM
19. NCSA (bg)	NCSA	EA-ECNIS	Tier-1	LoadLeveler
20. IT4I (cz)	IT4I/VSB-TUO	Anselm	Tier-1	PBS Pro
20. IT4I (cz)	IT4I/VSB-TUO	Salomon	Tier-1	PBS Pro
21. IPB (rs)	IPB	PARADOX	Tier-1	PBS Pro
22. NIIF (hu)	NIIF	NIIFI SC	Tier-1	SLURM
22. NIIF (hu)	NIIF	Seged	Tier-1	SLURM
22. NIIF (hu)	NIIF	Leo	Tier-1	SLURM
22. NIIF (hu)	NIIF	Phitagoras	Tier-1	SLURM
24. GRNET(gr)	GRNET	Aris	Tier-1	SLURM

Table 10: Local batch systems in PRACE-4IP

Table 11 shows the deployed components on Tier-0 Systems that allow having UNICORE working in PRACE.

Despite of the interesting features that UNICORE provided in the past to the infrastructure, in WP6, asking to the different Tier-0 and Tier-1 sites, we have noticed that there is not much interest from the PRACE community to use this kind of software. As commented before, different architectures and systems make it complex to group all the possible system features under a unique set of wrappers or under a single piece of software. This led us to evaluate the convenience of keeping this software as a core service in the catalogue. After a wide discussion and inquiry on the different PRACE sites, WP6 decided to maintain this software as a core service for the Tier-0 systems and changed it into an optional service for the Tier-1 sites.

In the future new approaches and an analysis of real user needs must be assessed in order to better adapt and invest efforts in the correct direction.

Component	Description	Deployment
REGISTRY	Directory service publishing Tier-0 resources.	JUELICH

Component	Description	Deployment
		(Primary), CINECA (Backup)
UNICORE/X	Translate abstract jobs into concrete jobs for a specific target system. Job submission and monitoring	Tier-0 System
GATEWAY	Main entrance to each Tier-0 system. Client connections go through the gateway, which forwards them to internal components, and vice versa.	Tier-0 System
XUUDB	User Database for authentication and authorization. It is synchronised with the PRACE LDAP.	Tier-0 System
TSI	Interface with the local batch system and storage resources	Tier-0 System
URC	Graphical user client (based on Eclipse Java)	Client-side
UCC	Command-line user client	Client-side

Table 11: UNICORE Software components

3.4 AAA Services and operational security

The AAA activity is responsible for services, which provide Authentication, Authorization and Accounting facilities on the infrastructure. The services are fundamental for providing access to many other PRACE services and resources. E.g. services that provide interactive access to the PRACE systems must rely on the public key infrastructure PKI provided by the PRACE AAA services and the PRACE LDAP service.

The design and setup of the AAA services is stable and robust and, apart from a minor configuration update of the LDAP schema, to facilitate user administration information exchange with EUDAT, did not undergo any recent changes. The primary focus of operations is on keeping the information content provided by these services correct and up to date: new digital certificates are issued, and existing ones expire, consequently distributions of relevant certificates must be regularly compiled and distributed, mutations in the user administration must be processed at least on a daily basis by the central PRACE LDAP and be processed by the local user administrations of PRACE execution systems and “door node” services (see below). Occasionally, when a new partner site joins PRACE, a new branch to the LDAP is added. For example, in the course of September and October 2015, shortly after another, two branches for two new Spanish partner sites, CESGA and UPM, were added. Conversely, when a partner site leaves PRACE, its LDAP branch is closed – i.e.: frozen, as for reasons related to accounting and security, it is not allowed to immediately remove expired user administration items.

The processing of such changes, including the AAA-wise accommodation of new partner sites, is considered to be routine operation of stable services. Services for which currently no drastic changes are foreseen in the near future. However, in a sessions to be characterized as “strategic brainstorm”, PRACE and GEANT have been discussing the feasibility of a more general federated authentication and authorization infrastructure.

Though there is no doubt about the quality of the PRACE AAA services and the relevant servers are included in the PRACE monitoring infrastructure, developing a key performance indicator (KPI) for AAA services as such has been put on the operations agenda discussed. A good KPI – preferably an easily understood and measurable single quantity – would make the services more visible and would demonstrate their relevance and performance, but as shown below, there is no obvious ideal single measure.

3.4.1 *Public Key Infrastructure - PKI*

Several PRACE services rely on X.509 certificates [19] for authentication and authorization. These certificates must be issued by entities trusted by the service providers. PRACE relies on the Certificate Authorities (CA) accredited as a member by the EUGridPMA, the European Policy Management Authority (PMA) [20], or by one of the two other organizations, the TAGPMA and the APGridPMA, all three federated in the IGTF [21]. These PMAs require a minimum set of requirements for the CP/CPS (Certificate Policy/Certification Practice Statement) of the member CAs, as published and maintained by the PMAs in profile documents. PRACE itself is a member of the EUGridPMA as Relying Partner and participates in its activities.

For PRACE partners CA information is maintained at a central repository. The information is based on the IGTF distribution and updates are made available to the PRACE partners for download in several formats. To guarantee a high level of availability, the distributions are stored on a virtualized webserver that can migrate between two physical hosts at SURFsara. New IGTF distributions are, and have been, provided on an almost monthly basis.

3.4.2 *User Administration*

Information about users and their accounts is maintained in a Lightweight Directory Access Protocol (LDAP) based repository. This facility is used to share among PRACE partners the authorization information needed by PRACE services and is used to retrieve information about users and their projects. Authorization information is provided among others for data transfers with GridFTP, interactive access through GSI-SSH, job submission with UNICORE, accounting services and access to the helpdesk facilities.

A single LDAP server is used for PRACE Tier-0 accounts. For most Tier-1 accounts the same server is used, however some partners host a local LDAP server for the registration of Tier-1 users. Separate databases are used for Tier-0 and Tier-1 accounts. An overview of the LDAP namespace is shown in Figure 4. For Tier-0 accounts the “ou=ua,dc=prace-ri,dc=eu” part of the name space is used and for Tier-1 accounts “ou=ua,dc=deci,dc=org” is used as top part. The Tier-1 accounts registered by other LDAP servers can be accessed through referrals.

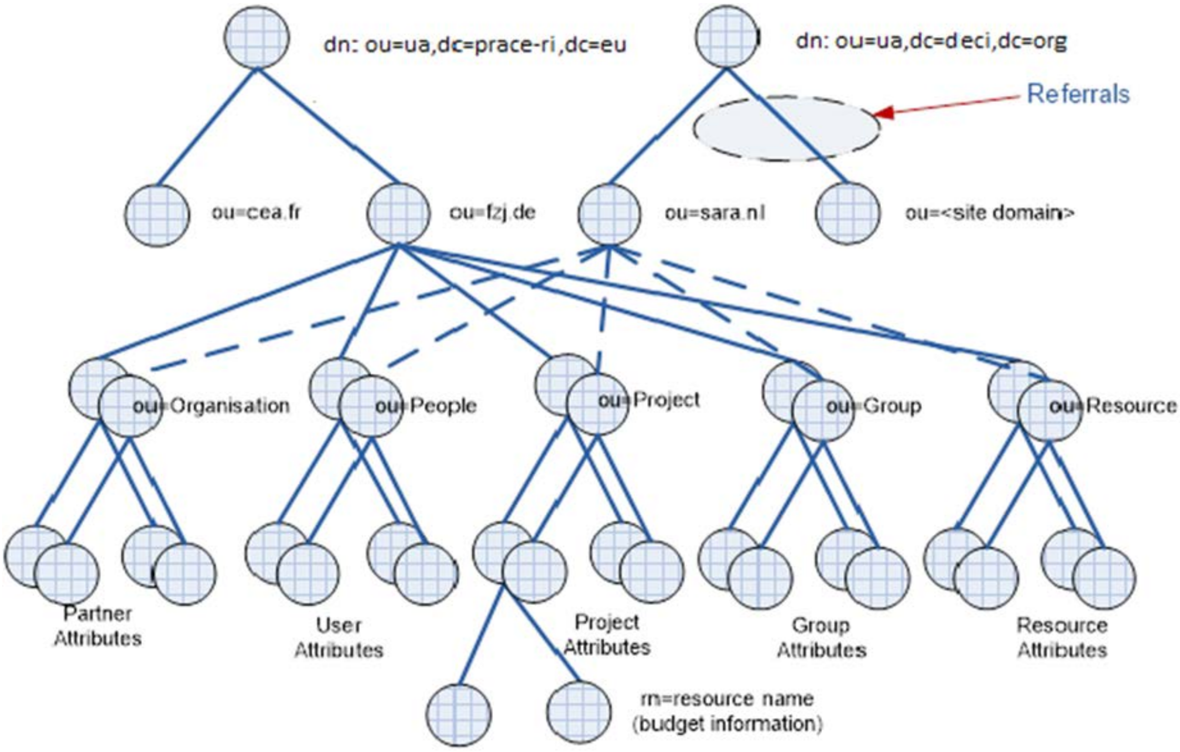

Figure 4: PRACE LDAP directory tree

All Tier-0 sites manage their own branch in the LDAP repository. They register the users who have access to their site in their LDAP branch, which starts at ou=<site domain>.

Each Tier-1 site also manages its own branch in the LDAP Directory Information Tree (DIT), which is everything below the ou=<site domain> attribute. Each user of Tier-1 resources is assigned to a Home site, which is one of the Tier-1 partners and it’s the Home site which creates and updates the account information of the user. The Home site in general is the site from the country where the user is working and not the site where Tier-1 resources have been granted to the user (the Exec site). The Home site mostly already knows the user and this makes the registration procedure easier. The Exec site retrieves the information from the LDAP server for the creation of a local account.

The main PRACE LDAP server and a replica server are operated by SURFsara. Both servers are virtualized and migratable. To avoid dependency on a single site for this core service, an additional remote replica server is operated by HLRS. The replica servers can be used if the primary server is unavailable for some reason. It is the distributed responsibility of the collective of Home sites to enter the correct information pertaining to their respective users in the central LDAP. It is the responsibility of each Exec site to retrieve the relevant information from the central LDAP and disperse it into their local user administration.

There is no specific prescription how to do this, merely that it has to be done and that the delay should not be more than 24 hours. The Exec site may choose to fully automate this process in a way that fits its particular local user administration, or it may care of it entirely by a manual operation. This abstention from any specific definition of a required interoperability interface between the PRACE LDAP and a site specific user administration in practice had made it easy to integrate new sites.

Like Exec sites, door nodes for interactive access, explained below, must process PRACE LDAP mutations in their local user administration. Unlike Exec sites, due to their function they cannot be selective and must always keep track of the complete active user community.

3.4.3 *Interactive Access*

Interactive access to the Tier-0 systems is provided either by the SSH (Secure Shell) facilities, provided by most distributions of operating systems, or X.509 based SSH facilities. The latter are mainly used for access from other PRACE sites using the dedicated PRACE network. On PRACE Tier-0 and Tier-1 sites the recommended X.509 based SSH facility is GSI-OpenSSH [22] (GSI-SSH for short), distributed by the Globus community. All sites that support GSI-SSH are now on GT version 6.

For access from their workplace users can use GSISSH-Term, a GSI-SSH Java based client, which is supported by the PRACE partner LRZ [23].

Not all sites support GSI-SSH based access directly from the internet. Therefore, two partners, CINECA (IT) and SURFsara (NL), provide GSI-SSH based access for all PRACE users who can use these sites as door nodes to other PRACE sites using the PRACE network.

3.4.4 *Accounting Services*

Users can get accounting information locally at the sites at which they consume resources. In addition users and PRACE staff can get accounting information from several sites in a uniform way with the DART tool [24]. With this tool users can display their accounting information for five Tier-0 systems and nine Tier-1 sites.

3.4.5 *Measuring the performance of AAA: towards a KPI*

A KPI for AAA should definitely focus on the LDAP service as it is fundamental for authentication and authorization and has a role in accounting as well. One point of view is to use the availability of the central LDAP service as the key performance indicator. It is measurable, as the availability of the servers involved is already monitored by the PRACE monitor infrastructure. It will lead to a very good track record, as the migratable virtualized service setup has proven to be very robust. Even in the period where the hosting partner, SURFsara, was migrating to a new data centre, redundancy was temporarily decreased, but the primary server never was unavailable.

Measuring only availability of the service however implies that certain problems that do impact the performance of the service are masked. The mere availability of the service does not imply that the information it serves is correct and up to date, and if it is not, this does impact the possibility of users to access resources that they are entitled to.

It is possible, and occasionally does happen, that information is entered into the LDAP that is not compliant. If that happens, this is usually detected on the side of the execution system or by the user administration used by a door node. An impacted end user may complain and a ticket is generated to solve the issue. In practice, non-compliant information, such as a missing mandatory field or a certificate with an improper subject format, are usually caught before an end user is aware of any problem. As the central LDAP is hosted by the same site that runs a door node, which has to “know” all active users, such problems are detected early, when the mutations in the LDAP are dispersed into the user administration used by the door node. An internal ticket, from PRACE staff to PRACE staff, is generated that notifies the site responsible for the erroneous data entry to correct it. Another approach to a KPI for the AAA service would be in terms of the number of tickets and/or the time to solution of tickets that specifically pertain to issues with the user administrations service.

3.5 Security Forum

The PRACE Security Forum has three main tasks:

- Defining security related Policy and Procedures;
- The Risk Review of new services or service upgrades;
- The management of operational security.

All PRACE partners contributing services to the PRACE infrastructure are represented in the Forum and can vote on the decisions. In this section we report on the internal activities of the Security Forum. The security activities in collaboration with other organisations are described in PRACE-4IP Deliverable D6.6.

3.5.1 *Security Policies and Procedures*

The PRACE common services provide access among PRACE sites and because of this sites cannot be completely isolated from each other. So, sites are vulnerable for security incidents at one of the other sites. This means that sites must trust the security measures at other sites. For this reason, partners had been asked to provide documents describing the local security policies and procedures. There are still some sites which are not able to provide relevant information since their local security policies do not allow to publish this information. To get rid of this unlucky situation, the work done by the WISE (WISE Information Security for collaborating E-infrastructures) community, formally SCI, can help. PRACE security staff has been involved in WISE work since the beginning and will go on in the future as well. With the SCI V1 paper, it is possible for those organisations at least to do a self-assessment and provide the outcome to the other partners. This document provides a list of policies and procedures which should be implemented as a basis for trust among infrastructures, but equally can be used to build trust among partners within a collaboration.

3.5.2 *Risk Reviews*

The Security Forum performs a risk assessment of a new service or the update of an existing service if there are changes in the security set-up of the latter.

In this reporting period no new risk review had to be done.

3.5.3 *Operational Security*

All partners providing services to the PRACE service infrastructure are member of the PRACE CSIRT team. Incidents are reported and discussed using an e-mail list and video/telcon facilities. Emergency phone numbers are available for all sites.

There were no security incidents reported by sites in this period. Nevertheless information on several vulnerabilities in software have been shared among partners. Although information about vulnerabilities in general will be received by partners by other means too, it is regarded as helpful if specific information is also provided through internal channels. It will guarantee that attention will be given to those vulnerabilities.

3.6 User Services

The User Services subtask is composed of the following four activities: the PRACE Common Production Environment (PCPE), the Trouble Ticket System (TTS), Helper scripts and PRACE User Documentation.

3.6.1 *PRACE Common Production Environment*

The PRACE Common Production Environment (PCPE) presents a common application development environment interface across all PRACE execution sites regardless of the underlying architecture. It guarantees availability of a minimum set of software components to support the running of highly-scalable parallel simulation software. Thus, it allows automated monitoring of key software components across all PRACE execution sites.

PCPE allows users to gain many benefits, such as improved efficiency in porting and running applications due to commonality of interface across PRACE; confidence that the required software for their research is in place at the start of their project; automated access to optimization best practice for local compilers, irrespective of vendor and hardware architecture.

The specification of PCPE has not evolved during the period. The focus has been put on the “module” tool command, which is at the root of the PCPE to enable or disable parts of this software collection. The module command has different implementations and at the time of the DEISA project the implementation made by the Modules-Tcl project [31] was the one chosen as the default to handle the DEISA Common Production Environment (DCPE). Work has been achieved this year to publish a release of this module flavour that can easily be installed at sites from source or package distribution. This work has been integrated in the upstream Modules-Tcl open source project.

3.6.2 *Trouble Ticket System*

The centralised Helpdesk [26] was initially deployed as part of the PRACE-1IP project. It is an important tool for the PRACE project staff to communicate among sites about problems. The Helpdesk is also available for users. However, users are directly redirected to the specific queues of the Tier-0 and Tier-1 sites.

TTS is based on the Request Tracker (RT) issue tracking system [27]. Since its deployment in PRACE-1IP, this service is hosted and operated at CINECA.

An improvement was made this year to better cope with spam emails or emails coming from automatic remote systems. TTS has been adapted not to send an email notification to a ticket requestor when his ticket is rejected or deleted. By doing so we avoid replying to spam emails or to robot system which was previously leading sometimes to the creation of a new spam ticket on PRACE TTS.

3.6.3 *Helper Scripts*

User services subtask provides helper scripts for PRACE users and staff to easily access the PRACE services in operations. Among these scripts, the “prace_service” utility gives access to the directory of services helping users and staff to get information on server addresses and ports for the various services.

This year the prace_service configuration has been updated many times to include the new PRACE Tier-0 and Tier-1 systems or to remove the systems that have been decommissioned.

No change was made on the prace_service script itself, as no new need arise during this second year.

3.6.4 *PRACE User Documentation*

The User documentation for PRACE is available online on the PRACE Research Infrastructure website [28]. During the first year a specific focus was put on this activity. A migration in terms of document format occurred and the whole user documentation content was moved from HTML to MarkDown [30] format in the PRACE SVN repository [29].

After this migration of the user documents, the publication process was also clarified, as recalled on Figure 5.

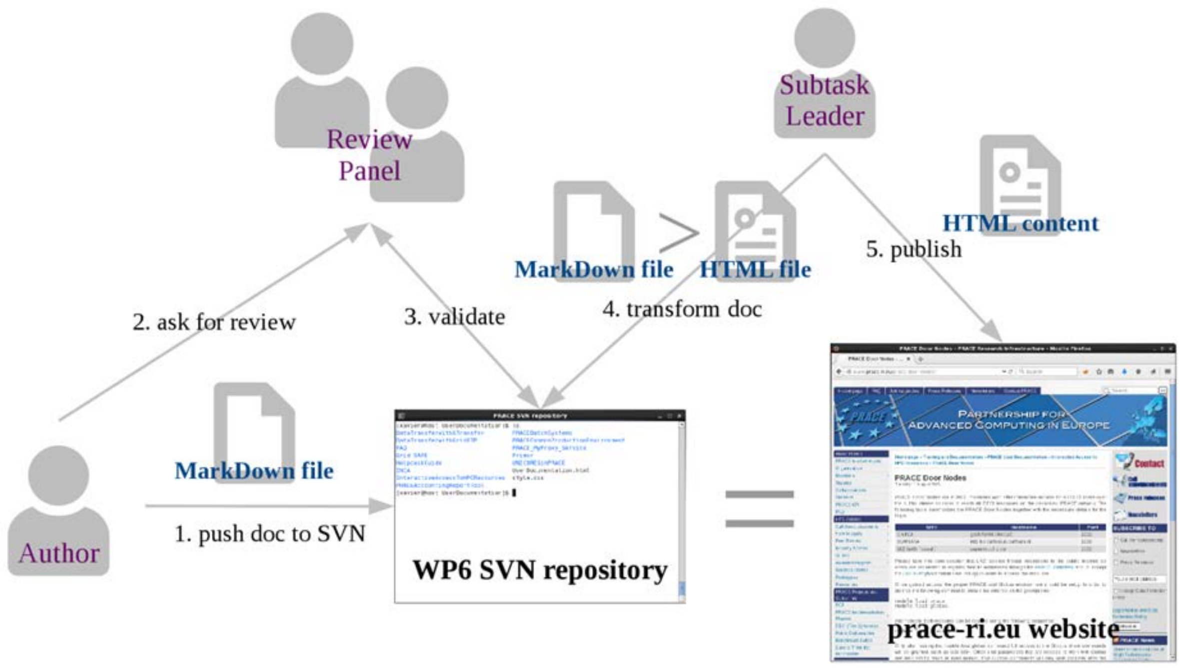

Figure 5: Publication process of the PRACE User Documentation

This second year of activity some updates of documents happened mainly to add references to the new PRACE Tier-0 and Tier-1 systems and to remove descriptions of systems that have been decommissioned.

3.7 Monitoring Services

Monitoring service is a crucial part of the PRACE-RI infrastructure. Its basic functionality is to gather the states of the infrastructure services, tools and machines among all the sites connected through PRACE network. The collected and processed data reflecting the infrastructure condition is made available with monitoring web interface for site administrators, managers and infrastructure operators. Information gathered by monitoring system provides foundation for infrastructure availability reporting and verification. Monitoring is also the complementary service for PRACE Network accessibility. The monitoring service features are defined in the PRACE service catalogue as “core service”.

The main work performed in the Monitoring Services sub-task during the current reporting period consists of maintenance and development and planning activities.

Maintenance of the monitoring services is an activity which is required to keep monitoring infrastructure operational and in a good shape. Its main parts include the following tasks:

- Service hosting: keeping service machines highly-available by maintaining underlying hardware and network infrastructure;
- Updates: Keeping operating systems, system software, middleware and service software up-to-date on all machines related to the services;
- Security: addressing security issues and advisories by applying necessary patches, updates or/and workarounds as soon as possible to reduce the risk of security incidents and breaches;
- Configuration: adjusting the monitoring service configuration to reflect current changes in the infrastructure, which include adding and removing computing resources and services provided by PRACE infrastructure. This activity covers also network configuration changes.

Development in the Monitoring Services sub-task is necessary to adapt the ICINGA2 monitoring software stack, adopted in the first year of PRACE-4IP (see [40]), to the specifics of PRACE infrastructure and involve the following activities:

- Implementation of new functionalities in the monitoring software to address project requirements requested by WP6 and other work packages;
- Creation of monitoring programs and scripts necessary to perform functionality/ availability checks for particular PRACE services;
- Development of automated configuration generator for translating service configuration database (prace_service.conf) to ICINGA2 configuration format.
- Integration with other services.

3.7.1 *Monitoring services - evolution and improvements*

The new ICINGA2 based monitoring service has successfully been integrated with almost all the sites in the PRACE infrastructure. Figure 6 shows an overview of the core services status, as reported by Icinga2. During the current reporting period, we have completed the integration of four new systems with the monitoring system:

- PSNC Eagle
- NIIF Phitagoras
- GRNET
- CINECA: Marconi

In addition to the new computing resources, also the monitoring of the new generic services has been integrated with the monitoring system. The change has added 10 web based PRACE services which include:

- BSCW - Document repository hosted by Juelich
- DPMDB - Accounting Database hosted in SURFsara
- DECI Peer review website (CINES)
- PRACE Event Website
- PRACE Peer review website
- PRACE Summer of HPC website
- PRACE Training websites
- PRACE Public website (www.prace-ri.eu)
- PRACE WIKI
- SVN code repository for PRACE

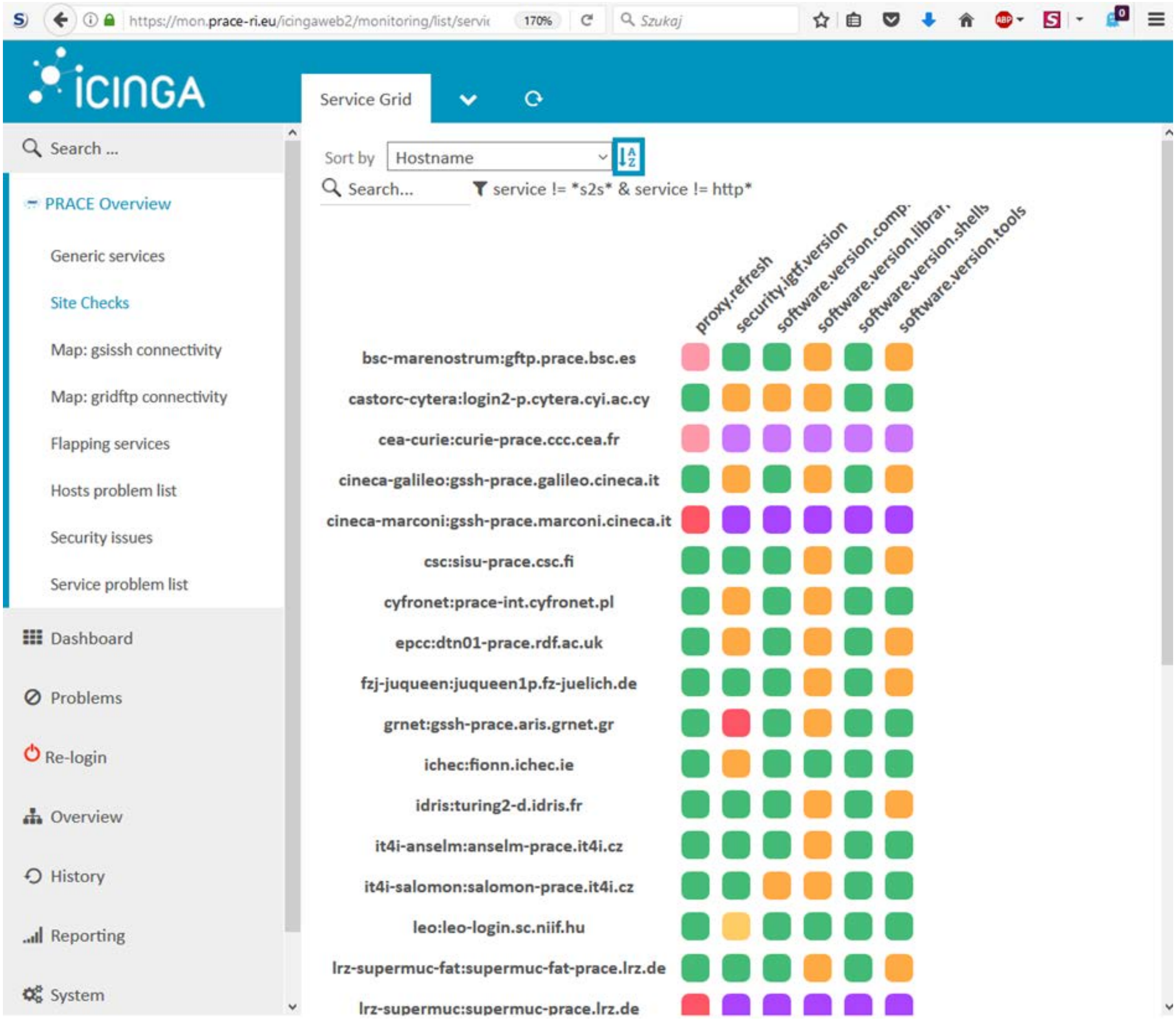

Figure 6: Screenshot of Icinga2: an overview of Core Services status

Currently the supported metrics include:

- generic services (both http/https protocols)
- gsissh service checks in all-to-all mode and monitoring to all mode.
- gridftp service checks in all-to-all mode, this metric has been redesigned to take advantage of 3rd party transfers
- software version - PCPE checks including:
  - Development libraries
  - Tools
  - Shells
- monitoring tools version check
- security related checks:
  - IGTF distribution version verification
- hosts availability checks

A set of new functionalities are being added to the existing monitoring eco-system. In order to allow site accessibility and service reliability statistics, a reporting method for ICINGA2 has been developed and is currently being tested and validated. As an example, Figure 7 presents the status of the availability of the Tier-1 system Salomon in IT4I site for the period 1st October – 30th November 2016.

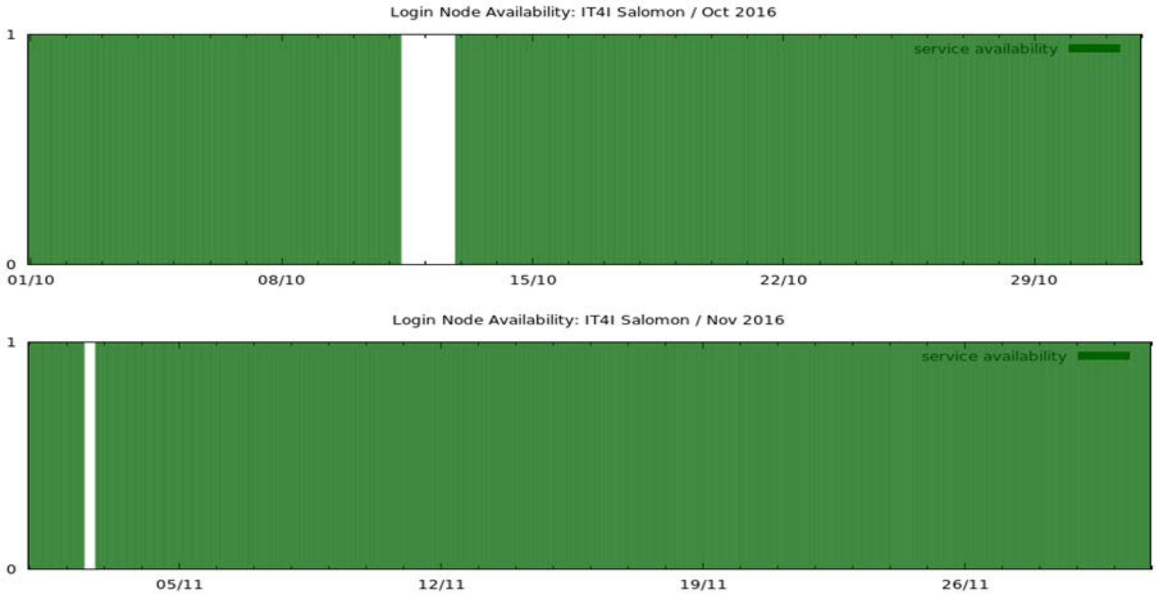

Figure 7: Example of availability graph (here Login Node Availability)

Based on a request from the resource providers, we have enhanced the monitoring tools package with logging functionality, allowing sites to keep track on what and when is being executed by monitoring systems on site owned resources. This change serves mainly for security auditing and debugging purposes.

In addition, there has been the change in the field of available monitoring protocols. During the monitoring deployment phase, we had the evidence that all the project resource providers preferred the gsissh-based protocol compared to NRPE. The gsissh-based collection protocol requires almost no maintenance by site administrators, compared to the NRPE-based solution. The gsissh-based solution proved to be a choice more flexible and safe, so the NRPE solution has been deprecated and will no longer be maintained.

The ICINGA 2 system, which has been chosen as the best possible INCA replacement in the PRACE infrastructure, still requires an effort for adaptation and maintenance as described in the previous section. Within the reporting period many issues have been addressed and resolved in order to keep the infrastructure in good condition.

At the time of writing, ongoing work includes testing of improved notification functionality and downtime handling in the monitoring software, as well as new service check development, in order to cover more aspects of infrastructure.

At the end of the reporting period (April 2017) all sites registered in the resource database (prace_service.config) are integrated within the monitoring services. Table 12 gives a picture of the monitored objects in the PRACE HPC infrastructure.

Monitored Sites	28
Monitored Generic Services	10
Monitored Host Objects	98
Monitored Service Objects	1612

Table 12: Number of the monitored objects in the PRACE-RI at MON.PRACE-RI.EU

3.7.2 *Monitoring service – KPI proposal*

For the Monitoring Service a KPI has been analysed and proposed. The KPI is based on the availability of testing of the remote endpoints located in the PRACE Infrastructure. This is the time, when ICINGA 2 server can send out the requests for checking, even not getting the response back. This KPI does not include the network availability nor the state of remote subsystems, responsible for proper assembling of answer and sending them back to monitoring server. This KPI strictly concerns the service installed at mon.prace-ri.eu. The lack of its availability in a certain time has no direct influence for other PRACE services, (see Table 13).

RI Monitoring availability	
Description	Availability of monitoring middleware.
Calculation	((A-B) / A) * 100
Inputs	Committed hours of availability (A) – the time when ICINGA is ready to send checks and to process the responses. Outage hours excluding scheduled maintenance (B)
Outputs	Availability (%)
Time-interval	Bi-weekly (during every PRACE Operations meeting). Annual report.
Threshold	99% (Annual threshold)
Tools	Icinga (mon.prace-ri.eu)
ITIL Category	Service Design – Availability Management
KPI Lead	PSNC, Cyfronet
Implementation plan	ICINGA2 provides all the data necessary to calculate this KPI. All the data related to the service, for a given time period, have to be extracted from ICINGA2.

Table 13: RI Monitoring availability KPI

3.8 Generic Services

This section describes the actions done during the reporting period within Generic Services subtask. In general, all services that need an operational basis and a centralized distribution for the PRACE project (or a part of it) could be assumed as Generic Services.

The goal of this task is the provisioning of these services and the supervision of their operation, as they are crucial for the day-to-day work of the project.

The Service leader of this task is Zoltan Kiss (NIIFI) since the beginning of PRACE-4IP.

3.8.1 *Roles and Actions*

The leader of generic services is acting as a liason / consultant with:

- PRACE aisbl;
- Other WPs;
- External providers.

The consultation is activated upon request when IT or operational issues happens or requests from the above entities rises. In the following we list the main activities arisen during the current reported period:

- Questions of another WP related to the operation of new or existing services:
    - Integration between Training, Events portals, CRM and PRACE Web site;
    - Negotiating data repository task with WP3, handover of detailed requirements;
    - Planning, negotiation and migration of CodeVault service from WP7 to WP6 involving WP4.
- Questions of PRACE aisbl when negotiating with service providers offering service for PRACE:
    - Consulting with service hosts, research on available providers;
    - Helping managing domain portfolio;
    - Help with managing certificate portfolio.
- Helping internal workflows of using a shared PRACE-owned resource
    - Handling *.prace-ri.eu wildcard certificate, general help, helping with new certificate requests, help with wildcard renewal, duplicate, SAN management and contact with providers;
    - Handling prace-ri.eu domain: subdomain requests, helping with resolving domain issues, complete migration to new domain provider, help with line-by-line transfer of complex zone file.

Another activity involved the reaction to the recent SSL related vulnerabilities. A big effort was spent to maintain a high-level status of security among generic services. An online SSL security analysis was made on the servers hosting Generic Services, and the security rating was improved in numerous cases.

Furthermore, a collaboration was issued with the Monitoring Service team in order to introduce the certificate validity check on top of HTTP/HTTPS availability. This way, the Generic Services leader and the service operators have been enabled to monitor certificates of services to get notification when a certificate expired or will expire soon. Figure 8 presents a picture of the monitored generic services related to certificates and webpages items.

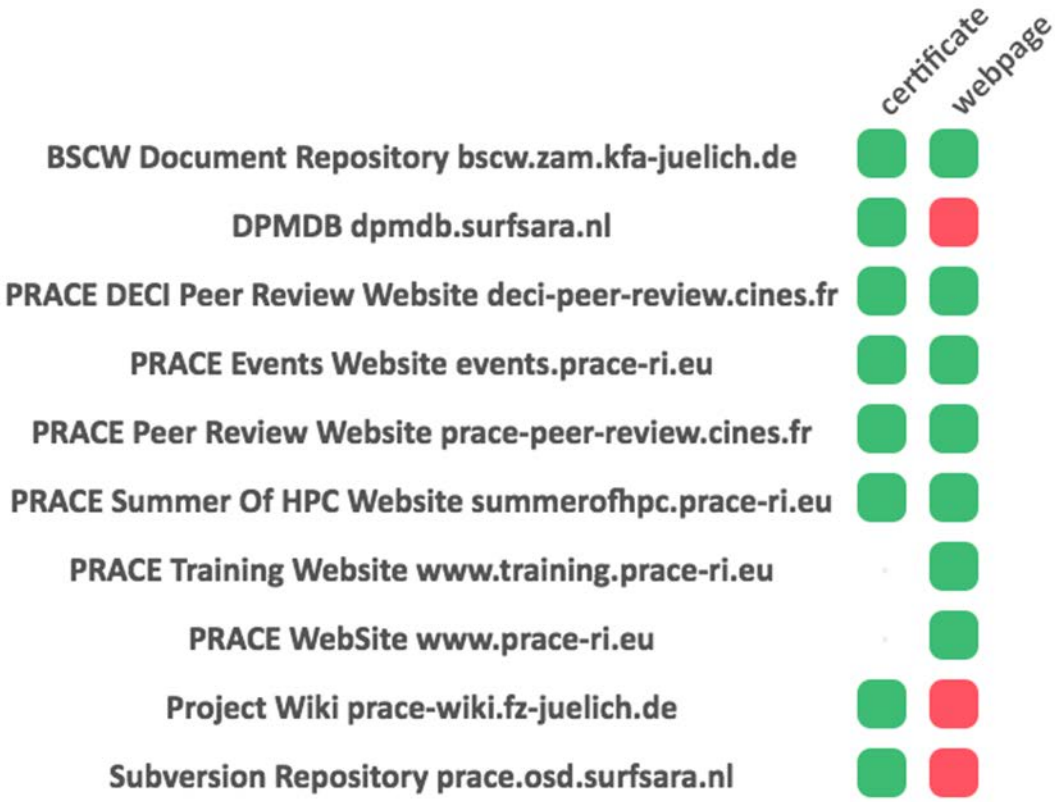

Figure 8: PRACE monitoring system: Generic Service entries

3.8.2 *Generic services exposed in the Service Catalogue*

During the reporting period a big effort has been undertaken to revise the service catalogue with respect to the Generic Services, analysing the services to maintain or to add to the service catalogue and identifying a host and a service operator for each service. Furthermore, some effort has been undertaken to identify some possible KPIs.

Table 14 presents a list of the services analysed and proposed to update the service catalogue.

Service name	Host	Service Operator	Service Catalogue	Suggestion
BSCW	FZJ	FZJ	core	core
DPMDB	SURFsara	SURFsara	core	core
SVN	SURFsara	SURFsara	core	core
WIKI	FZJ	FZJ	core	core
MailingLists	FZJ	FZJ	core	core
WEB	CINES	NIIF	core	core
Training Portal	CSCS	CaSToRC	core	core
PRACE EventsPortal	NIIF	NIIF	-	core
CRM	CaSToRC	CaSToRC	-	(External service)
PPR Tool	CINES	CINES	core	core

Table 14: Possible update of service catalogue

The proposed KPI, common to all the services investigated, consists in the measure of the availability of the service. As most of these services are available to the public, the new PRACE monitoring system seems to be an ideal way to measure service availability by introducing a new ICINGA hostgroup and new checks within this group.

The ICINGA monitoring system offers automated availability reports, therefore the KPIs can be easily produced. In addition, thanks to a built-in notification system able to notify the service maintainer in case of problems with the service, quicker reactions can be issued when the monitoring system is not available locally.

4 Conclusions

In this reporting period, Task 6.1 has continued the successful operation of the PRACE common services for the Tier-0 sites and the Tier-1 sites providing services for Tier-0. The operational procedures have continued to be successful in maintaining a reliable and available set of integrated services. The on-duty activity has a weekly schedule and continuously supervises the HPC infrastructure to assure a more concrete and punctual support on the day by day operation monitoring actions and support. This activity involves 21 PRACE partners with weekly shifts. Every week a report on the duty activity is produced, reporting the status of the infrastructure, of the core service and the problems occurred.

Seven Tier-0 and 28 Tier-1 systems are currently in production in the PRACE HPC Infrastructure. In the reporting period a total of seven sites have upgraded their systems (Beskow at KTK/PDC, Leo at NIIF, Hazel Hen at HLRS (Tier-0), Occigen at CINES, Minotauro at BSC, Cartesius at URFsara and Phitagoras at NIIF). Furthermore, in the PRACE-4IP time-frame, ten new systems have been installed and integrated in the PRACE infrastructure, or planned to be installed: Six system have been successfully integrated (Galileo and Marconi (Tier-0) at CINECA in Italy, Salomon at IT4I in Czech Republic, Aris at GRNET in Greece, Eagle at PSNC and Bem at WCSS in Poland); one is still in the integration phase (Prometheus at Cyfronet in Poland) and three more are scheduled in the coming months (IUCC in Israel, UL and FIS in Slovenia).

To ensure a seamless use of the PRACE distributed research infrastructure the PRACE common services are available on all PRACE Tier-0 systems and on the identified Tier-1 systems providing services for Tier-0. Through this common service provision, the Tier-0 and Tier-1 systems are presented as a single infrastructure, which enables a smooth interoperation of Tier-0 and Tier-1 systems.

The version of the PRACE Service Catalogue approved by PRACE BoD in March 2015 has guided the operational activity of the common services and a revision has been started in this reporting period to introduce new services and delete the deprecated ones. An analysis of the KPIs associated to the different services is in progress and will continue in PRACE-5IP.

Some of the major achievements in Task 6.1 are underlined here:

- The old dedicated wavelengths PRACE network (10 Gb/s star-topology) has been replaced with a Multi Domain Virtual Private Network (MD-VPN) which offers a much more flexible configuration, a faster setup, lower connectivity costs and a security level comparable to the dedicated links.
- The new version of the network monitoring tools (iperf V. 3) is fully operational and allows a stable .monitoring of the entire PRACE network.

- The cooperation between PRACE WP6 and GEANT is progressing well. A joint meeting has been held in 16 March 2017 to reinforce the agreement in place with the common MoU and strengthen the cooperation on the topics of AAI, performance network monitoring, security and Data transfer feasibility.
- The Data services have mainly focused on GridFTP and Gtransfer (file transfer). Currently, all the Tier-0 systems and 18 Tier-1 systems successfully deploy GridFTP.
- All the sites adopted Globus Toolkit version 6.0.
- Gtransfer, an advanced toolkit for performing data transfer is deployed on the Tier-0 systems and several Tier-1 sites. Different functionalities for common operations on remote GridFTP servers have been tested in different sites.
- The gsatellite job scheduler for executing jobs in the background, without any user intervention, has been fully evaluated and recently added to the service catalogue.
- In the Compute Services activity, different wrappers with a common syntax for describing a job around different batch systems and platforms have been analysed and fully tested to adopt a set of standards to implement from different sites in the context of the PRACE Common Production Environment.
- PRACE continues to maintain, in a central repository, the updated PRACE Partners information for the Certificate Authorities, as different PRACE services rely on X.509 certificates for AAA procedures. PRACE is a member of the EUGridPMA (the European Policy Management Authority) as Relying Partner and participates in its activities. The last EUGridPMA meeting was held in Florence in January 2017 and was attended by PRACE WP6 partners.
- The information related to the PRACE users and their accounts continues to be maintained and updated in the LDAP based repository hosted by SURFsara. This facility is used to share among PRACE partners the authorisation information needed by PRACE services (i.e. data transfer via GridFTP, interactive access through GSI-SSH, job submission via UNICORE, etc.).
- The PRACE Security Forum continues to coordinate all security related issues. Regular teleconferences, which see the participation of each PRACE contributing partner assure the continuity of this service and assure that the reported security vulnerabilities are managed efficiently.
- The PRACE Common Production Environment (PCPE) is stable. In the second year of PRACE-4IP the activity focused on the “module” tool command to allow users to enable ore disable part of the software tools from source or a package distribution.
- Improvements have been done to adapt the Trouble Ticket System to not send email notifications to a ticket requester when the corresponding ticket is rejected or deleted, to avoid replying to spam e-mails.
- A continuous improvement and revision process has been set up for the user documentation available on-line on the PRACE-RI Web site and the SVN repository. The revision process mainly concerned the references to the new PRACE Tier-0 and Tier-1 systems or the removal of documentation related to decommissioned systems.
- The full integration of the ICINGA Version 2 has been completed for all Tier-0 and Tier-1 systems to completely monitor the PRACE infrastructure. In addition to the PRACE computing resources also the generic services have been integrated and are constantly monitored by means of ICINGA 2.
- The generic services team strictly interact with the other WP (mainly WP3 and WP4) and the PRACE aisbl to provide and support them for all the services that need an

operational basis and a centralised distribution (i.e. BSCW, WIKI, SVN, mailing lists, Web portal, CRM, etc.). All these services are constantly monitored and managed.

- The Service Catalogue has been revised adding some new services and deleting some deprecated ones. For some services, specific KPIs have been defined and the measurement of the indicators and their evaluation has started; the process is still in progress and will be completed in PRACE-5IP to better assure the PRACE quality of services and quality control.

Thanks to the procedures for incident and change management, Task 6.1 operates and monitors on a day-by-day basis the complete set of PRACE common services, as defined in the Service Catalogue. This activity will continue and improve in PRACE-5IP in the context of the PRACE 2 programme, where new powerful Tier-0 and Tier-1 systems will further improve the infrastructure and enrich the whole European HPC eco-system at the service of Science.