SKA SDP PROT ISP Memo 002

HPC resources available within ISP: System architecture, documentation, access, application for compute time

Document number	SKA-TEL.PROT-ISP-MEMO-002
Authors	Willi Homberg George Beckett Jesus Lorenzana David Macleod Stephen Pickles Jeremy Yates
Version	0.1
Release Date	20 February 2014

Released by:

Name	Designation	Date	Signature
Willi Homberg	Lead Author	20FEB14	

Version	Date of Issue	Prepared by	Comments
0.1	20JAN14	Willi Homberg	First final.

Introduction

The compute platform for the SKA1 telescope Science Data Processor will be a large super-computer scale machine. In order to be able during the SDP design work to conduct tests on large-scale HPC systems the SDP consortium has a number of national supercomputing partners that can provide access to their large production machines.

The purpose of the ISP work package is to organize access to these supercomputing facilities and to deploy, run, interpret, and provide feedback on a range of experiments of them.

The following list comprises ISP partner sites which contribute HPC resources that potentially might be used for horizontal prototyping:

1) STFC Hartree Centre
2) DiRAC
3) Jülich Supercomputing Centre
4) CHPC
5) FCSCL
6) iVEC

For each partner site a list of HPC systems is provided including site web address, system documentation and contact address and a description of the application process for computing time.

HPC resources available at ISP partner sites

STFC Hartree Centre:

Site web address:

- STFC: http://www.stfc.ac.uk/
- Hartree Centre: http://www.stfc.ac.uk/Hartree/default.aspx

System/ Partner	Nodes	Processor Technology	Procs	CPU-Cores	Accel-Cores NVIDIA GPUs Intel Xeon Phi
BlueJoule STFC Hartree Centre[1]	7168[2]	64 bit, 1.60 GHz A2 PowerPC	7168	114688	
BlueWonder STFC Hartree Centre	512[3]	Intel Xeon E5-2670 2.6 GHz (SandyBridge)	1024	8192	nVidia M2090 GPUs[4]

[1] Additional systems forthcoming in 2014 include (a) Xeon Phi, (b) Maxeler FPGA, (c) IBM BGAS, (d) Hadoop appliance, (e) oil-cooled system, (f) ARM systems. Exact configurations and application procedures are not yet known (February 2014).
[2] Currently configured as 6+1 racks each of 1024 16-core processors. The limits for a single job are 98304 cores and 96 TB of distributed memory.
[3] Significant upgrade expected in 2014.
[4] A subset of 24 nodes each have 2 x nVidia M2090 GPUs.

System/ Partner	Main Memory (TB)	Local Disk (TB)	Network Technology	Global Filesystem/ Size/ Bandwidth
BlueJoule STFC Hartree Centre[5]	112 TB	N/A	5D Torus	GPFS[6] 5.7 PB
BlueWonder STFC Hartree Centre	40 TB[7]	256 TB	Infiniband	GPFS 5.7 PB

System web address:

- Hartree Centre wiki: http://community.hartree.stfc.ac.uk/

Application process for computing time:
A Hartree Centre project for SKA Science Data Processor has already been established, with a limited allocation of computing time up to the value of 510 k€. To gain access, send email to Stephen Pickles with a description of the intended work and an estimate of the computing time required.

Documentation:

- Documentation and support for Hartree Centre projects: http://community.hartree.stfc.ac.uk/wiki/site/admin/home.html

DiRAC:

Site web address: www.dirac.ac.uk

System/ Partner	Nodes	Processor Technology	Procs	CPU-Cores	Accel-Cores NVIDIA GPUs Intel Xeon Phi
DiRAC BlueGene Q	6144	64 bit, 1.60 GHz A2 PowerPC	6145	98304	
DiRAC Shared Memory System	58	Sandybridge E5-2670 2.6GHz	232	1856	31 Xeon Phi
DiRAC DataCentric	420	Sandybridge E5-2670 2.6GHz	840	6720	
DiRAC Data Analytic	600	Sandybridge E5-2670 2.6GHz	1200	9600	A separate system Wilkes is available see http://www.hpc.cam.ac.uk/services/wilkes.html
DiRAC Complexity	272	Sandybridge E5-2670 2.6GHz	544	4352	

[5] Additional systems forthcoming in 2014 include (a) Xeon Phi, (b) Maxeler FPGA, (c) IBM BGAS, (d) Hadoop appliance, (e) oil-cooled system, (f) ARM systems. Exact configurations and application procedures are not yet known (February 2014).
[6] Shared by BlueJoule and BlueWonder
[7] Not uniformly distributed across nodes. 4 nodes have 256 GB, 256 nodes have 128 GB, and the rest have 32 GB.

System/ Partner	Main Memory (TB)	Local Disk (TB)	Network Technology	Global Filesystem/ Size/ Bandwidth
DiRAC BlueGene Q	96	N/A	5D Torus	GPFS 1.2PB
DiRAC Shared Memory System	14.4	N/A	NUMALink cache-coherent interconnect, all-to-all, MPI offload engine.	300TB, up to 6GB/s sustained I/O disk to/from memory
DiRAC DataCentric	53.760		Mellanox FDR10 Infiniband in a 2:1 blocking configuration	3PB (SD12K). 6 GPFS servers connected into the controllers over full FDR and using RDMA over the FDR10 network into the compute cluster.
DiRAC Data Analytic	38.4	0.376	Mellanox FDR ConnectX3 interconnect in a non blocking config	Lustre, 2PB, 7GB/s sustained
DiRAC Complexity	34.8	0.6	56Gb/s FDR non-blocking Infiniband network in a fat-tree topology	Panasas ActiveStor. 0.75TB

System web address:

DiRAC BlueGene Q http://www.epcc.ed.ac.uk/facilities/dirac
DiRAC Shared Memory System http://www.cosmos.damtp.cam.ac.uk/
DiRAC Data Centric http://icc.dur.ac.uk/index.php?content=Computing/Cosma
DiRAC Data Analytic http://www.hpc.cam.ac.uk/services/darwin.html#arch
DiRAC Complexity http://www2.le.ac.uk/offices/ithelp/services/hpc/dirac

Application process for computing time:
All systems are accessed via single application process, see https://www.dirac.ac.uk/access.html

All proposals are peer reviewed. If proposals are for a technical design study they must be accompanied by a letter for support – in this case it would be from Dr Rosie Bolton.

The deadline for the next call for Projects is 3/3/14. The application form is here https://www.dirac.ac.uk/DiRAC%20RAC%20Application%20Form.docx and the Application Guide is here https://www.dirac.ac.uk/DiRAC_RAC_Guidance_Notes_Dec_13.pdf

Small projects of order (<= 50k cpu-hours) can apply at any time)

Documentation:
DiRAC Shared Memory System http://www.cosmos.damtp.cam.ac.uk/user-guide
DiRAC BlueGene Q http://www.epcc.ed.ac.uk/facilities/dirac/userguide/index
DiRAC DataCentric http://icc.dur.ac.uk/index.php?content=Computing/Cosma_FAQ
DiRAC Data Analytics http://www.hpc.cam.ac.uk/user/
DiRAC Complexity http://www2.le.ac.uk/offices/ithelp/services/hpc/dirac

Jülich:

Site web address:
Jülich: http://www.fz-juelich.de/portal/DE/Home/home_node.html
JSC: http://www.fz-juelich.de/ias/jsc/DE/Home/home_node.html

System/ Partner	**Nodes**	**Processor Technology**	**Procs**	**CPU-Cores**	**Accel-Cores NVIDIA GPUs Intel Xeon Phi**
JUQUEEN Jülich	28672	IBM PowerPC A2	28672	458752	
JUROPA Jülich	3288	Intel Xeon X5570 (Nehalem-EP) quad-core	6576	26304	
JUDGE Jülich	206	Intel Xeon X5650	412	2472	NVIDIA Tesla M2070 /2050

System/ Partner	**Main Memory (TB)**	**Local Disk (TB)**	**Network Technology**	**Global Filesystem/ Size/ Bandwidth**
JUQUEEN Jülich	448		5D Torus	GPFS 7 PB 160 GB/s
JUROPA Jülich	79		InfiniBand QDR	Lustre 1.8 PB 50 GB/s
JUDGE Jülich	19.776		InfiniBand QDR	GPFS

System web address
JUQUEEN:
http://www.fz-uelich.de/ias/jsc/EN/Expertise/Supercomputers/JUQUEEN/JUQUEEN_node.html
JUROPA:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUROPA/JUROPA_node.html
An update of the JUROPA system is scheduled for the second half of 2014; exact configurations are not yet known.
JUDGE:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUDGE/JUDGE_node.html

Application process for computing time:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/ComputingTime/computingTime_node.html
JUQUEEN:
http://www.gauss-centre.eu/gauss-centre/EN/HPCservices/HowToApply/LargeScaleProjects/largeScaleJUQUEEN.html
guidelines:
http://www.gauss-centre.eu/gauss-centre/EN/HPCservices/HowToApply/guidelinesJUQUEEN.html
A moderate amount of computing time (<50k core-hours) can be requested any time within an application period for scaling tests
JUROPA:
application process:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/ComputingTime/juropa- application.html

guidelines:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/ComputingTime/
juropa-guidelines.html
JUDGE:
meant for application/simulation projects at the Jülich site in collaboration with JSC; no formal application process

Documentation:
JUQUEEN:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUQUEEN/Documentation/
Documention_node.html
http://www.fz-juelich.de/ias/jsc/EN/Expertise/High-Q-Club/_node.html
JUROPA:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUROPA/Documentation/
Documentation_node.html
JUDGE:
http://www.fz-juelich.de/ias/jsc/EN/Expertise/Supercomputers/JUDGE/Documentation/
Documentation_node.html

CHPC:

Site web address: http://www.chpc.ac.za/

System/ Partner	**Nodes**	**Processor Technology**	**Procs**	**CPU-Cores**	**Accel-Cores NVIDIA GPUs Intel Xeon Phi**
Sun Nehalem	288	Intel Xeon X5570	576	2304	
Sun Westmere	96	Intel Xeon X5670	192	1152	
Dell Westmere	240	Intel Xeon X5670	480	2880	
Sun M9000	1	Sun SPARC64 VI+	64	256	-
GPU Cluster	5	Intel Xeon X5550	10	40	Nvidia Tesla C1060 and C2070
Accelerator Test-bed	12	Intel Xeon E5-2640	24	192	6x Nvidia Tesla K20 & 6x Intel Xeon Phi 5100
Research Test-bed	6	Intel Xeon E5-2690v2	12	120	2x Convey HC1-ex systems

System/ Partner	**Main Memory (TB)**	**Local Disk (TB)**	**Network Technology**	**Global Filesystem/ Size/ Bandwidth**
Sun Nehalem	3.4	-	IB QDR	Lustre 480TB
Sun Westmere	2.3	-	IB QDR	Lustre 480TB
Dell Westmere	8.4	-	IB QDR	Lustre 480TB
Sun M9000	2	XFS 5.3	-	-
GPU Cluster	0.02	-	IB QDR	Lustre 14TB
Accelerator Test-bed	0.75	-	IB QDR	Lustre 480TB
Research Test-bed	1.5	-	IB FDR	NFS 10TB 1GB/s

Application process for computing time:
http://www.chpc.ac.za/index.php/support-resources/apply-for-resources

For assistance with applications contact David, dmacleod@csir.co.za

Documentation:
http://www.chpc.ac.za/index.php/support-resources/logging-in
http://www.chpc.ac.za/index.php/support-resources/compiling
http://www.chpc.ac.za/index.php/support-resources/scaling
http://www.chpc.ac.za/index.php/support-resources/user-guide-documentation
http://wiki.chpc.ac.za/

FCSCL:

Site web address: www.fcsc.es

System/ Partner	**Nodes**	**Processor Technology**	**Procs**	**CPU-Cores**	**Accel-Cores NVIDIA GPUs Intel Xeon Phi**
Calendula/FCSCL	288	Intel Xeon E5450 - 3.00GHz	576	2304	
No name yet /FCSCL	6	Intel Xeon E5 2670v2	12	120	36 Intel Xeon PHI 5110p (6 each node)

System/ Partner	**Main Memory (TB)**	**Local Disk (TB)**	**Network Technology**	**Global Filesystem/ Size/ Bandwidth**
Calendula/FCSCL	4,5	33,75	Infiniband DDR	NFS NetApp FAS3100 cabinet / 100 TB
No name yet /FCSCL	0,7	24	Infiniband FDR / 10Gb Ethernet	Lustre installation in process FhGFS evaluation

System web address: www.fcsc.es

Application process for computing time:
(1)
Documentation:
(2)

(1) (2) At present we are changing our web site. Next month we will have updated technical and user information.
(3) More systems Phi coprocessors based and new storage devices are coming in 2014. Technical details are not yet known.
(4) No name system is now being installed and tested.

iVEC:

Site web address:

- Non-technical description available at http://www.ivec.org/services/supercomputing/.

System/ Partner	Nodes	Processor Technology	Procs	CPU-Cores	Accel-Cores NVIDIA GPUs Intel Xeon Phi
Magnus/ iVEC (Cray XC30) [8] -2014Q2	208	Intel Xeon E5-2670 "Sandy Bridge" 8-core	416	3,328	None
Magnus/ iVEC (Cray XC30) 2014Q3-	1488	Intel Xeon "Haswell" – chip details covered by NDA	2976	NDA	None
Fornax/ iVEC (SGI cluster)[9]	96	Intel Xeon X5650 "Nehalem" 2×6-core	192	1,152	NVIDIA Tesla M2070 "Fermi" (1 per node)
Epic/ iVEC (HP Proliant) [10]	800	Intel Xeon X5650 "Nehalem" 2×6-core	1,600	9,600	None
Zythos/ iVEC (SGI UV 2000)	1	Intel Xeon E5-4610 "Sandy Bridge" 6-core	44	264	4 × NVIDIA Tesla K20 "Kepler"
Galaxy/ iVEC (Cray XC30)	472	Intel Xeon E5-2690v2 "Ivy Bridge" 10-core	944	9,440	None

System/ Partner	Main Memory (TB)	Local Disk (TB)	Network Technology	Global Filesystem/ Size/ Bandwidth
Magnus/ iVEC (Cray XC30) -2014Q2	13 TB (64GB /node)	-	Cray Aries Dragonfly Topology (⅓ populated)	3PB Sonexion 1600/ Lustre 70 GBytes/sec
Magnus/ iVEC (Cray XC30) 2014Q3-	93 TB (64GB /node)	-	Cray Aries w/ Dragonfly Topology (⅓+ populated)	3PB Sonexion 1600/ Lustre 70 GBytes/sec
Fornax/ iVEC (SGI cluster)	7 TB (72GB/ node)	7TB / node	2 × QDR Infiniband (MPI, I/O). Fat Tree	DDN 300 TB / Lustre ~1 GBytes/sec
Epic/ iVEC (HP Proliant)	19 TB (24GB/ node)	-	QDR Infiniband Fat Tree	HP 300 TB / Lustre ~1 GBytes/sec
Zythos/ iVEC (SGI UV 2000)	6 TB	8 TB	SGI NUMAlink	Sonexion storage from Magnus (see above)
Galaxy/ iVEC (Cray XC30)	29.5 TB (64GB/ node)	-	Cray Aries Dragonfly Topology (⅓ populated)	1PB Sonnexion 1600/ Lustre 50 GBytes/sec

System web address:

- Documentation and technical information at http://portal.ivec.org/

[8] Magnus to be upgraded to a Petaflops+ system in July 2014, and is expected to be accessible from late Q3/ early Q4

[9] Fornax is available until end of June 2015, at which point it will be decommissioned – to be replaced by a new machine, details to be confirmed.

[10] Epic is only available until end of December 2014, at which point it will be decommissioned – being replaced by Magnus.

Application process for computing time:
There are two routes through which compute time can be secured on iVEC resources (except for Galaxy):

- Significant amounts of compute time (100,000 core-hours or more) require a successful application to the iVEC Partner Merit Allocation Scheme, which runs annually. The call is open during mid-October—mid-November, for compute time in the subsequent calendar year. Application form is available at https://portal.ivec.org/ivecallocation/.
- Modest amounts of compute time (up to 100,000 core-hours) can be requested at any time from the iVEC Director's Share Scheme. The application process is responsive mode, with applications typically reviewed in 2—4 weeks. The application form is available at https://portal.ivec.org/ivecallocation/.

In addition, Magnus (Phase 2 machine) will be available during October—December 2014, in an early-adopter mode. During this time, SDP may have unmetered access to the system for running simulations and benchmarks.

Compute time on Galaxy is only available subject to special request – email george.beckett@ivec.org, in the first instance.

Documentation:
- Documentation and technical information at http://portal.ivec.org/.

Risks

Dependencies on other work packages in terms of software prototypes, use cases and HPC architecture requirements are not yet resolved and may cause delays.

References

None