

Estudo interno sobre o estado da arte do HTC
no contexto do
aviso de abertura de *calls* para uso da RNCA

Abril de 2020

FCT-FCCN

Índice

1	Distinção clássica entre HTC e HPC	1
2	Definições adicionais de HPC	2
3	Projetos HTC.....	3
4	GRID Computing e HTC.....	4
5	Utilização de HTC na RNCA.....	4
6	Anexo A	6
7	Anexo B – datacenters do EGI.....	7
8	Anexo C – aprovisionamento HPC em self-service web.....	10

1 Distinção clássica entre HTC e HPC

HTC = <i>High Throughput Computing</i>	HPC = <i>High Performance Computing</i>
Jobs correm em paralelo, em vários computadores ao mesmo tempo, mas com fraca necessidade de comunicação entre os computadores intervenientes	Jobs correm em paralelo com elevadas necessidades de comunicação entre os computadores intervenientes. No caso limite, a execução de um job HPC pode ser tão rápido quanto o seu componente mais lento, ou seja, todo o sistema pode ficar à espera de um resultado do componente mais lento. No caso no HPC é, portanto, crítica a latência e débito da rede de interligação dos computadores.
A computação é geograficamente distribuída, p. ex., em vários países ou mesmo continentes	A computação é geograficamente concentrada, p. ex. numa sala com cabos de interligação até 100 metros de comprimento
Existe um software de controlo ou <i>middleware</i> sofisticado, multi- <i>site</i> . Utilizadores podem ter que ser acreditados por uma entidade certificadora nacional.	O software de controlo de <i>jobs</i> é tipicamente um mero <i>batch-scheduler</i> local

Tabela 1- HTC vs HPC

O HTC não precisa de uma rede de baixa latência, mas pode tê-la e os seus processos de software estão prontos a incorporar modelos de recursos híbridos. Portanto a fronteira entre uma coisa e outra pode ser difícil de estabelecer em alguns casos concretos.

Por exemplo o EGI oferece um serviço chamado MPI *High-Throughput Compute* [1] que esbate as fronteiras entre HTC e HPC. No limite este serviço poderá ser prestado em *compute-nodes* intra-cluster, na mesma sala, deixando de haver componente de computação distribuída. No Anexo A apresenta-se informação sobre latências de rede de várias tecnologias populares. Para aplicações com frequentes e elevadas interdependências de cálculos paralelos é um fator crítico a latência das comunicações. Não se consegue fazer esse tipo de cálculo de forma geograficamente distribuída pois, no limite, a velocidade da luz passa a ser o fator limitante.

2 Definições adicionais de HPC

É útil neste estudo incluir os seguintes elementos adicionais definidores do modelo HTC:

<p>EGI [2]</p>	<p>With High-Throughput Compute you can run computational jobs at scale.... execute thousands of parallel computing tasks.</p> <p>... is provided by a distributed network of computing centres, accessible via a standard interface and membership of a virtual organisation</p> <p>... Large amounts of processing capacity over long periods of time</p> <p>... Shared resources among users, enabling collaborative research</p> <p>.... a computing paradigm that focuses on the efficient execution of a large number of loosely-coupled tasks</p>
<p>Wikipedia [3]</p>	<p>... use of many computing resources over long periods of time to accomplish a computational task.</p> <p>... run tasks on opportunistic resources</p>
<p>wisc.edu [4]</p>	<p>... effective management and exploitation of all available computing resources</p>

A definição de HTC aparece ligada ao conceito de **computação distribuída** e de partilha de recursos.

Neste contexto o serviço do EGI chamado MPI *High-Throughput Compute* se for providenciado de forma intra-*cluster*, numa mesma sala, então não seria computação distribuída e não se

¹ <https://marketplace.egi.eu/high-throughput-compute/44-mpi-high-throughput-compute.html>

² <https://www.egi.eu/services/high-throughput-compute/>

³ https://en.wikipedia.org/wiki/High-throughput_computing

⁴ <https://research.cs.wisc.edu/htcondor/htc.html>

poderia chamar de HTC. Seria portanto uma contradição nos seus termos, pois o MPI pressupõem latências de comunicação baixas.

Para haver computação distribuída é necessário que exista algum tipo de software multi-site. Elenca-se a seguir algum desse software:

- Unified Middleware Distribution (UMD)
 - Pacote de software disponibilizado no âmbito do EGI
 - Última *release*: Março de 2020
 - Módulo de distribuição da computação [⁵]: *Advanced Resource Connector (ARC) middleware is an Open Source software solution to enable distributed computing ... provides an abstraction layer over computational resources, complete with input and output data movement functionalities. The security model of ARC is identical to that of Grid solutions, relying on delegation of user credentials and the concept of Virtual Organisations.*
 - <https://wiki.egi.eu/wiki/Middleware>
- HTCondor
 - Pacote de software disponibilizado no âmbito da universidade UW-Madison
 - *HTCondor is a specialized workload management system for compute-intensive jobs. Like other full-featured batch systems, HTCondor provides a job queueing mechanism, scheduling policy, priority scheme, resource monitoring, and resource management. Users submit their serial or parallel jobs to HTCondor, HTCondor places them into a queue, chooses when and where to run the jobs based upon a policy, carefully monitors their progress, and ultimately informs the user upon completion.*
 - Última *release*: Fevereiro de 2020
 - <https://research.cs.wisc.edu/htcondor/>
- BOINC
 - Pacote de software disponibilizado no âmbito da universidade de Berkeley
 - *BOINC is a platform for high-throughput computing on a large scale (thousands or millions of computers). It can be used for volunteer computing (using consumer devices) or grid computing (using organizational resources)*
 - Última *release*: Novembro 2018
 - <https://boinc.berkeley.edu/>

3 Projetos HTC

Mantendo a componente de computação distribuída na definição de HTC, então os projetos de HTC, tendem a ser iniciativas globais, que atravessam diferentes domínios administrativos e de propriedade dos equipamentos, introduzindo-se processos como a definição de confiança nos participantes e o conceito de VO- *Virtual Organization*.

Pode ainda distinguir-se duas variantes de HTC com computação distribuída que são a computação voluntária ou computação em grelha (o antigo *GRID computing*).

⁵ Para além deste módulo o UMD tem muitos outros módulos para partilha de storage, autenticação, etc.

A lista seguinte foi extraída de [6] de projetos ou iniciativas operacionais HTC ativas em 2020:

Project	Active processing units	Research focus
Folding@home	4,630,510 (com GPUs)	Understand protein folding, misfolding, and related diseases, with a minor emphasis in protein structure prediction. Computing power also harnessed the power of PlayStation 3s.
Einstein@Home	12,013	Search for pulsars using radio signals and gravitational wave data[39]
NFS@Home	7,719	Performs parts of the Number Field Sieve in the factorization of large integers
Cleanmobility.now	5,135	Find safer and greener materials for e-vehicle batteries
SETI@home Beta	954	Test project of SETI@home
LHC@home Sixtrack	4,840	Improve the design of the Large Hadron Collider and its detectors

Tabela 2- Projetos HPC típicos

Para além destes projetos da variante de computação voluntária, há um exemplo importante de computação distribuída institucionalizada via EGI. O EGI tem mais de 150 moradas de *datacenters* participantes e quase duas dezenas de prestadores de serviço. A listagem completa está disponível no Anexo B. Esta vasta infraestrutura presta serviços principalmente à comunidade científica.

4 GRID Computing e HTC

O *GRID computing* [7] foi um termo popularizado em finais da década 1990 início da década 2000. Prometia enormes ganhos de eficiência e escala através da inclusão da capacidade computacional global numa arquitetura operacional por todos partilhada e por todos acedida. Porém esse objetivo só seria possível se os donos individuais da capacidade computacional estivessem motivados para incorrer nos custos marginais de instalar, manter e evoluir os agentes de software GRID, o que não aconteceu na prática em grande medida. O termo tem caído em desuso [8] face ao novo modelo de *cloud computing* que é mais polivalente e flexível. As iniciativas sobreviventes originalmente denominadas “GRID” tem evoluído o seu conceito original. O EGI, inicialmente **E**uropean **G**RID Initiative, surge agora livre da denominação GRID: *EGI is a federation of 19 cloud providers and hundreds of data centres, spread across Europe and worldwide.*

5 Utilização de HTC na RNCA

Há que distinguir duas alternativas para definir o HTC no âmbito da RNCA, que são

- Definição (A): Perspetiva clássica onde existe um software de controlo ou *middleware* multi-site.

⁶ https://en.wikipedia.org/wiki/List_of_distributed_computing_projects

⁷ https://en.wikipedia.org/wiki/Grid_computing

⁸ <https://storagemojo.com/2007/11/26/grid-is-dead/>

- O utilizador deve indicar que software multi-site pretende que esteja disponível para controlar os *jobs* que pretende correr.
- Nesta alternativa há uma marcada diferença entre HTC e HPC.
- Estas utilizações tendem a ser de longo prazo.
- Definição (B): Perspetiva alargada de HTC sem recorrer a um software de controlo ou *middleware* multi-site.

Nesta alternativa há pouca diferença entre HTC e HPC se os *jobs* HTC correrem *intra-cluster* no mesmo *datacenter*. De seguida apresentamos uma progressão até um cenário final onde o HTC se funde com o HPC:

- Menos performance:
 - As VM, lançadas via mecanismos *cloud*, correm MPI-over-TCP. Ligações *intra-cluster* são através de rede de *interconnect* de baixo custo.
- Mais performance:
 - As VM correm MPI com acesso facilitado à porta de *interconnect* (p. ex. Infiniband). Ligações *intra-cluster* são em *Interconnect* de baixa latência e alto débito. Há ainda uma penalização de performance devido à existência de camada de software virtualização e partilha de um servidor real por várias VM que lá correm ao mesmo tempo.
- Igual a HPC:
 - O sistema de *self service* WEB aprovisiona *compute-nodes* nativos sem camada de virtualização, dedicados temporariamente a um único utilizador. Exemplo de um sistema desse tipo: Ironic do Openstack. O Anexo C descreve esse tipo de software. Neste caso aprovisiona-se um sub-cluster HPC com acesso nativo ao *interconnect*.

No primeiro caso as utilizações do tipo (A) que se perspetivam são:

- Use-case#A/1 – registar a utilização histórica
 - Utilidade: registar os recursos afetos que vem do passado em modelos de GRID e/ou HTC, ou seja a participação em projetos HTC internacionais. Ex.: compromissos com o CERN ou o IberGRID.
- Use-case#A/2 – afetar capacidade da RNCA a um projeto HTC existente
 - A RNCA afeta uma capacidade computacional à participação em um projeto HTC existente, projeto esse que pode ter a sua sede em país estrangeiro.
 - Exemplo: um grupo de investigação nacional propõem-se participar num projeto computacional HTC em alguma área do conhecimento. Pode não ser simples de medir o ganho nacional dessa participação.
- Use-case#A/3 – afetar capacidade da RNCA a um projeto HTC emergente ou completamente novo, em projeto.
 - Um grupo de investigação nacional apresenta um problema computacional cuja solução assenta numa *framework* HTC colaborativa internacional. Apresenta também uma lista de interessados em arrancar como uma iniciativa HTC, ou seja, com uma VO- *Virtual Organization* nova. Pode reutilizar software

e processos existentes, ou então, pode propor instalar novo software de HTC, ou até mesmo fabricar software HTC novo.

No que diz respeito aos use-cases do tipo “A” o que parece mais provável de ocorrer é o A/1.

No segundo caso as utilizações do tipo (B) que se perspetivam são:

- Use-case#B/1 – cálculo que não precise de uma rede de baixa latência e alto débito
 - Pode correr em VMs intra-cluster através de rede de *interconnect* de baixo custo
 - Utilizadores das VMs podem recorrer a MPI-over-TCP, com tempos de *round trip time*, na ordem de 1 mili segundo ou sub-mili segundo. Pode haver *jobs* que possam correr nestas condições.
- Use-case#B/2 – cálculo que precise de uma rede de baixa latência e alto débito
 - Existindo uma de rede de *interconnect* de baixa latência e alto débito, então ela poderá ser usada de várias formas. No limite poderão ser provisionados sub-clusters HPC via software do tipo Ironic do Openstack

No que diz respeito aos use-cases do tipo “B” todas elas parecem prováveis de ocorrer. O futuro do HPC poderá mesmo passar pelo Use-case#B/2

6 Anexo A

A seguir encontram-se algumas referências de latência de rede, com valores abaixo de um micro segundo para HPC.

Network Solutions	End-to-End Latency (µsec)
GigE	29-100 (47.1 for Netpipe results in Fig. 1)
10GigE (TCP)	12.51 Mellanox ConnectX + 10GigE
DDR InfiniBand	1.72
QDR InfiniBand	1.67

Outras referências [9]:

⁹ <https://www.microway.com/knowledge-center-articles/performance-characteristics-of-common-network-fabrics/>

Data Rate	Theoretical Bandwidth (unidirectional)	End-to-End Latency	Technology
Gigabit Ethernet	125 MB/s	25 ~ 65 microseconds	
10G Ethernet	1.25 GB/s	1.3 microseconds (RDMA application) 4 microseconds (sockets application)	Mellanox ConnectX-3 VPI
40G Ethernet	5 GB/s	1.3 microseconds (RDMA application) 4 microseconds (sockets application)	Mellanox ConnectX-3 VPI

Data Rate	MPI Bandwidth (unidirectional)	End-to-End Latency	Generation
10Gb/s SDR	1 GB/s	2.6 microseconds	Mellanox InfiniHost III
20Gb/s DDR	2 GB/s	2.6 microseconds	Mellanox InfiniHost III
40Gb/s QDR	4 GB/s	1.07 microseconds	Mellanox ConnectX-3
40Gb/s FDR-10	5.16 GB/s	1.07 microseconds	Mellanox ConnectX-3
56Gb/s FDR	6.82 GB/s	1.07 microseconds	Mellanox ConnectX-3
100Gb/s EDR	12.08 GB/s	1.01 microseconds	Mellanox ConnectX-4
100Gb/s Omni-Path	12.36 GB/s	1.04 microseconds	Intel 100G Omni-Path

Data Rate	Hop Latency	Generation
40Gb/s QDR	0.10 microseconds	Mellanox InfiniScale IV
56Gb/s FDR	0.20 microseconds	Mellanox SwitchX-2
100Gb/s EDR	0.09 microseconds	Mellanox Switch-IB
100Gb/s Omni-Path	0.10 microseconds	Intel 100G Omni-Path
200Gb/s HDR	<0.09 microseconds	Mellanox Quantum

7 Anexo B – datacenters do EGI

<https://www.egi.eu/federation/data-centres/>

Where are the EGI data centres?

EGI Council participants

1. BEgrid-ULB-VUB – Belgium	82. NIKHEF-ELPROD – Netherlands
2. BEgrid-BELNET – Belgium	83. SARA-MATRIX – Netherlands
3. BG01-IPP – Bulgaria	84. GRIDOPS-SAM – Netherlands
4. egee.srce.hr – Croatia	85. EGI-DATAHUB – Netherlands
5. egee.irb.hr – Croatia	86. EGI-NOTEBOOKS – Netherlands
6. GRIDOPS-MON – Croatia	87. TASK – Poland
7. pragueicg2 – Czech Republic	88. CYFRONET-LCG2 – Poland
	89. PSNC – Poland

8.	prague_cesnet_lcg2 – Czech Republic	90.	ICM – Poland
9.	GRIDOPS-CTOOLS – Czech Republic	91.	WCSS64 – Poland
10.	GRIDOPS-SECTOOLS – Czech Republic	92.	NGI_PL_SERVICES – Poland
11.	GRIDOPS-PERUN – Czech Republic	93.	CAMK – Poland
12.	CESNET-MCC – Czech Republic	94.	NCBJ-CIS – Poland
13.	T2_Estonia – Estonia	95.	WUT – Poland
14.	CSC – Finland	96.	CYFRONET-CLOUD – Poland
15.	FI_Helsinki – Finland	97.	CYFRONET-PROMETHEUS – Poland
16.	FI_Aalto – Finland	98.	GRIDOPS-DIRAC4EGI – Poland
17.	FI_UTU – Finland	99.	EGI-MARKETPLACE – Poland
18.	FI_HIP_T2 – Finland	100.	NCG-INGRID-PT – Portugal
19.	FI_JYU – Finland	101.	UPorto – Portugal
20.	FI_LUT – Finland	102.	RO-14-ITIM – Romania
21.	FI_TUT – Finland	103.	RO-11-NIPNE – Romania
22.	FI_UEF – Finland	104.	NIHAM – Romania
23.	FI_AA – Finland	105.	RO-03-UPB – Romania
24.	FI_Oulu – Finland	106.	RO-07-NIPNE – Romania
25.	AUVERGRID – France	107.	RO-13-ISS – Romania
26.	IN2P3-CPPM – France	108.	RO-02-NIPNE – Romania
27.	OBSPM – France	109.	RO-16-UAIC – Romania
28.	IN2P3-CC-T2 – France	110.	GRIDIFIN – Romania
29.	GRIF – France	2.	CLOUDIFIN – Romania
30.	IN2P3-LAPP – France	111.	TU-Kosice – Slovakia
31.	IN2P3-IPNL – France	112.	IISAS-Bratislava – Slovakia
32.	UNIV-LILLE – France	113.	IEPSAS-Kosice – Slovakia
33.	IN2P3-LPC – France	114.	IISAS-FedCloud – Slovakia
34.	IN2P3-LPSC – France	115.	FMPHi-UNIBA – Slovakia
35.	IN2P3-SUBATECH – France	116.	UMB-BB – Slovakia
36.	IN2P3-IRES – France	117.	IISAS-GPUCloud – Slovakia
37.	IN2P3-CC – France	118.	IISAS-Nebula – Slovakia
38.	CREATIS-INSALYON – France	119.	SiGNET – Slovenia
39.	GRIDOPS-OPSPORTAL – France	120.	ARNES – Slovenia
40.	HG-02-IASA – Greece	121.	CESGA – Spain
41.	HG-06-EKT – Greece	122.	IFCA-LCG2 – Spain
42.	GRIDOPS-MSG – Greece	123.	UAM-LCG2 – Spain
43.	GRIDOPS-APPDB – Greece	124.	pic – Spain
44.	HG-08-Okeanos – Greece	125.	ifae – Spain
45.	GR-07-UOI-HEPLAB – Greece	126.	UPV-GRyCAP – Spain
46.	HG-09-Okeanos-Cloud – Greece	127.	USC-LCG2 – Spain
47.	GRIDOPS-UMDREPO – Greece	128.	IFIC-LCG2 – Spain
48.	GRIDOPS-CATCHALL – Greece	129.	CIEMAT-LCG2 – Spain
49.	GR-12-TEIKAV – Greece	130.	GRIDOPS-ACCPORTAL – Spain
50.	GRIDOPS-CheckIn – Greece	131.	GRIDOPS-EC3 – Spain
51.	INFN-GENOVA – Italy	132.	SE-SNIC-T2 – Sweden
52.	INFN-MILANO-ATLASC – Italy	133.	NDGF-T1 – Sweden
53.	INFN-ROMA1 – Italy	134.	CSCS-LCG2 – Switzerland
54.	CNR-ILC-PISA – Italy	135.	CERN-PROD – Switzerland
55.	RECAS-NAPOLI – Italy	136.	T3_CH_PSI – Switzerland
1.	INFN-T1 – Italy	137.	UNIGE-DPNC – Switzerland
56.	INFN-BARI – Italy	138.	UNIBE-ID – Switzerland
		139.	UNIBE-LHEP – Switzerland
		140.	TR-03-METU – Turkey

57. INFN-NAPOLI-ATLAS – Italy	141. TR-10-ULAKBIM – Turkey
58. INFN-COSENZA – Italy	142. TR-FC1-ULAKBIM – Turkey
59. INFN-TRIESTE – Italy	143. UKI-LT2-UCL-HEP – United Kingdom
60. INFN-LECCE – Italy	144. UKI-SCOTGRID-ECDF – United Kingdom
61. CIRMMP – Italy	145. UKI-SOUTHGRID-BHAM-HEP – United Kingdom
62. GRISU-UNINA – Italy	146. UKI-NORTHGRID-LIV-HEP – United Kingdom
63. NGI_IT – Italy	147. UKI-NORTHGRID-MAN-HEP – United Kingdom
64. INFN-CNAF-LHCB – Italy	148. UKI-NORTHGRID-SHEF-HEP – United Kingdom
65. INFN-PISA – Italy	149. UKI-SCOTGRID-DURHAM – United Kingdom
66. TRIGRID-INFN-CATANIA – Italy	150. UKI-LT2-Brunel – United Kingdom
67. INFN-LNL-2 – Italy	151. GRIDOPS-APEL – United Kingdom
68. INFN-PADOVA – Italy	152. UKI-LT2-RHUL – United Kingdom
69. UNINA-EGEE – Italy	153. GRIDOPS-GOCDB – United Kingdom
70. RECAS-BARI – Italy	154. UKI-SOUTHGRID-CAM-HEP – United Kingdom
71. INFN-CATANIA – Italy	155. UKI-SOUTHGRID-BRIS-HEP – United Kingdom
72. INFN-FERRARA – Italy	156. 100IT – United Kingdom
73. INFN-TORINO – Italy	157. UKI-NORTHGRID-LANCS-HEP – United Kingdom
74. INFN-ROMA3 – Italy	158. UKI-LT2-QMUL – United Kingdom
75. GARR-01-DIR – Italy	159. RAL-LCG2 – United Kingdom
76. INFN-FRASCATI – Italy	160. UKI-SOUTHGRID-OX-HEP – United Kingdom
77. INFN-ROMA1-CMS – Italy	161. UKI-LT2-IC-HEP – United Kingdom
78. INFN-MIB – Italy	162. UKI-SOUTHGRID-SUSX – United Kingdom
79. INFN-CATANIA-STACK – Italy	163. UKI-SCOTGRID-GLASGOW – United Kingdom
80. INFN-PADOVA-STACK – Italy	164. UKI-SOUTHGRID-RALPP – United Kingdom
81. EGI-FGSG – Italy	165. GRIDOPS-CVMFS – United Kingdom

Integrated resource providers

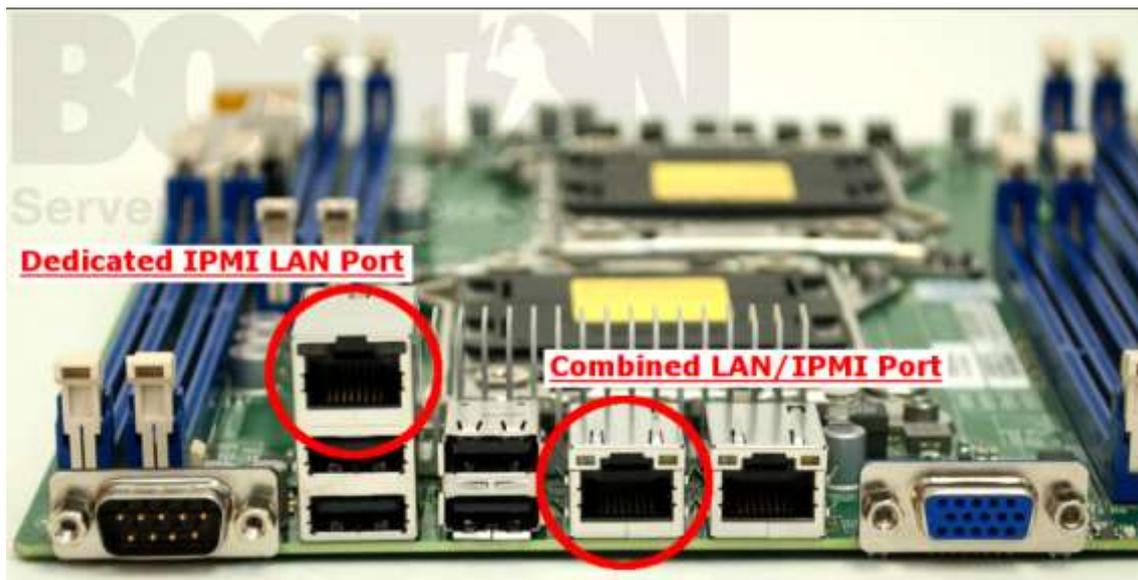
(including associate participants of the EGI Council)

1. DZ-01-ARN – Algeria	46. WEIZMANN-LCG2 – Israel
2. Australia-ATLAS – Australia	47. IL-TAU-HEP – Israel
3. Australia-T2 – Australia	48. TECHNION-HEP – Israel
4. HEPHY-UIBK – Austria	49. JP-KEK-CRC-02 – Japan
5. Hephy-Vienna – Austria	50. TOKYO-LCG2 – Japan
	51. ICN-UNAM – Mexico

6.	AZ-IFAN – Azerbaijan	52.	GRID-UNAM – Mexico
7.	BY-NCPHEP – Belarus	53.	SUPERCOMPUTO-UNAM – Mexico
8.	CBPF – Brazil	54.	MA-01-CNRST – Morocco
9.	SAMPA – Brazil	55.	MA-04-CNRST-ATLAS – Morocco
10.	TRIUMF-LCG2 – Canada	56.	NCP-LCG2 – Pakistan
11.	ROC_Canada_SERVICES – Canada	57.	RU-SPbSU – Russia
12.	CA-ALBERTA-WESTGRID-T2 – Canada	58.	RRC-KI-T1 – Russia
13.	CA-VICTORIA-WESTGRID-T2 – Canada	59.	ITEP – Russia
14.	CA-TRIUMF-T2K – Canada	60.	ru-PNPI – Russia
15.	CA-SFU-T2 – Canada	61.	RU-Protvino-IHEP – Russia
16.	CA-WATERLOO-T2 – Canada	62.	JINR-LCG2 – Russia
17.	EELA-UTFSM – Chile	63.	RRC-KI – Russia
18.	ATLAND – Chile	64.	Ru-Troitsk-INR-LCG2 – Russia
19.	AstrogridPUC – Chile	65.	JINR-T1 – Russia
20.	HK-HKU-CC-01 – China	66.	RU-SARFTI – Russia
21.	BEIJING-LCG2 – China	67.	AEGIS01-IPB-SCL – Serbia
22.	HK-LCG2 – China	68.	AEGIS03-ELEF-LEDA – Serbia
23.	UNICPH-NBI – Denmark	69.	AEGIS02-RCUB – Serbia
24.	FZK-LCG2 – Germany	70.	ZA-CHPC – South Africa
25.	wuppertalprod – Germany	71.	ZA-WITS-CORE – South Africa
26.	SCAI – Germany	72.	KR-KISTI-GSDC-01 – South Korea
27.	UNI-SIEGEN-HEP – Germany	73.	KR-KNU-T3 – South Korea
28.	UNI-BONN – Germany	74.	KR-KISTI-GSDC-02 – South Korea
29.	GRIDOPS-GGUS – Germany	75.	Taiwan-LCG2 – Taiwan
30.	GoeGrid – Germany	76.	TW-FTT – Taiwan
31.	DESY-HH – Germany	77.	TW-NTU-HEP – Taiwan
32.	LRZ-LMU – Germany	78.	TW-NCHC – Taiwan
33.	MPPMU – Germany	79.	T2-TH-SUT – Thailand
34.	DESY-ZN – Germany	80.	UA-KNU – Ukraine
35.	UNI-FREIBURG – Germany	81.	UA-BITP – Ukraine
36.	FZJ – Germany	82.	UA-ISMA – Ukraine
37.	RWTH-Aachen – Germany	83.	UA_ICYB_ARC – Ukraine
38.	GRIDOPS-XGUS – Germany	84.	UA-PIMEE – Ukraine
39.	mainz – Germany	85.	Kharkov-KIPT-LCG2 – Ukraine
40.	EOSchub-Helpdesk – Germany	86.	UA_BITP_ARC – Ukraine
41.	BUDAPEST – Hungary	87.	UA_ILTPE_ARC – Ukraine
42.	GRIDOPS-WS-PGRADE – Hungary	88.	UA-MHI – Ukraine
43.	IN-DAE-VECC-02 – India	89.	NGI_UA_SERVICES – Ukraine
44.	INDIACMS-TIFR – India	90.	UA-NSCMBR – Ukraine
45.	IR-IPM-HEP – Iran	91.	UA_ICMP_ARC – Ukraine
		92.	UA-IRE – Ukraine
		93.	UA-IFBG – Ukraine

8 Anexo C – aprovisionamento HPC em self-service web

Os servidores informáticos da classe de datacenter normalmente vem equipamentos com uma porta de rede do tipo *Intelligent Platform Management Interface* (IPMI) e um chip do tipo *baseboard management controller* (BMC) com funcionalidades variáveis, notoriamente a capacidade de fazer *reboots* ao servidor, aceder às configurações da BIOS, medir temperaturas internas do equipamento, etc.



Esse tipo de servidores informáticos tem também a capacidade de fazer boot-over-LAN, ou seja de arrancar com um Sistema Operativo que é descarregado a partir da rede local.

Através da porta IPMI e da funcionalidade boot-over-LAN é possível instalar automaticamente 1 (ou 1.000) servidores informáticos no datacenter, sendo os passos a dar, simplificadaamente, os seguintes:

- Forçar um “reboot” ao servidor através da porta de rede IPMI.
- O servidor arranca automaticamente através da funcionalidade boot-over-LAN (PXE) com uma imagem pré-feita pela equipa de administração informática.
- O servidor corre com essa imagem, que pode ser, por exemplo, ser a imagem de um compute-node HPC, com todo o software necessário pré-instalado.

Todos estes passos estão consolidados em software do tipo Ironic ^[10] do openstack ou outro similar e as respetivas funcionalidades podem ser acedidas comodamente por um utilizador não especializado, através do sistema de self-service WEB.

Um utilizador comum pode mandar arrancar um número arbitrário de compute-nodes através do sistema de self-service WEB, desde que para tal tenha quota de utilização e que toda a infraestrutura esteja preparada para isso. O utilizador pode mandar arrancar um cluster HPC pronto em poucas horas ou até menos tempo, pronto a usar, com todos os drivers infiniband testados, com todas as bibliotecas MPI testadas, com um Slurm a funcionar, etc.

¹⁰ <https://wiki.openstack.org/wiki/Ironic>