CCTVal/UTFSM Data Center: 10+ years of experience

Yuri Ivanov

Universidad Técnica Federico Santa María

- ## CCTVal/UTFSM Cluster
  - ### Research projects
  - ### HPC and Grid Computing
  - ### Brief History
  - ### Statistics over years
- ## Current state
  - ### Cluster layout
  - ### Network connectivity
  - ### Cluster 2020
- ## Resource demands
  - ### Mandatory maintenance
  - ### Minimal extension
 - ### Desired extension
- ## Conclusions

Research projects: UTFSM and other Chilean universities (U. of Chile, U. La Frontera, U. Santiago de Chile):

- Computations in **High Energy** and **Particle physics,** including **ATLAS analysis**
- **Biomedical** image processing (**Digital Pathology** on breast cancer etc.)
- **Satellite** image processing for **Environment protection**
- Modeling of **mechanical constructions** (turbulent flow around bridge piers etc.)
- Numerical simulation for **Biophysics**
- **Tsunami project**: modeling of **tsunami hydrodynamics**
- **Nanotechnology,** functionalization of s**emiconductor materials**
- **Machine Learning, Pattern Recognition, Deep neural networks**
- Computational Chemistry, **Quantum Chemistry**
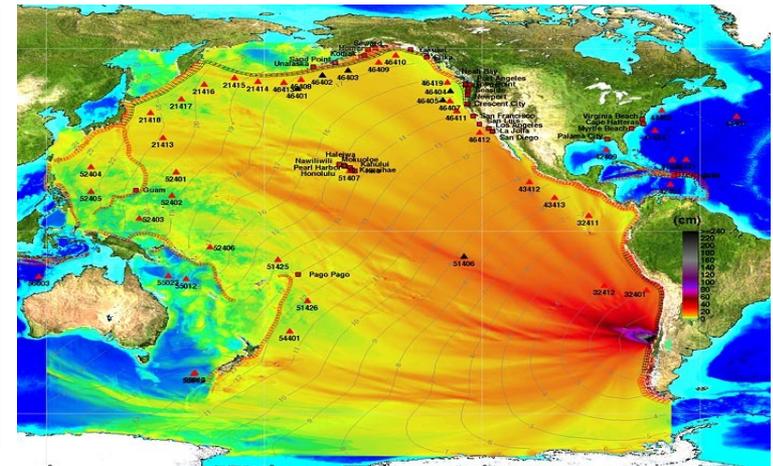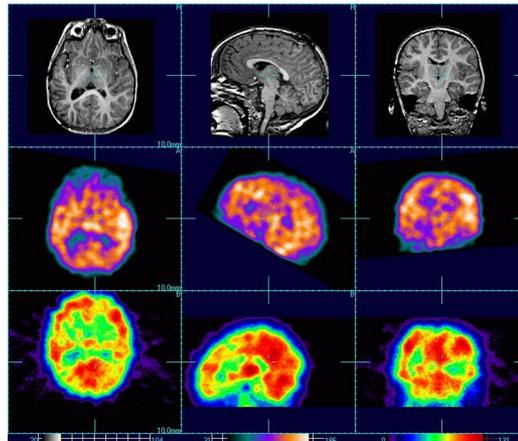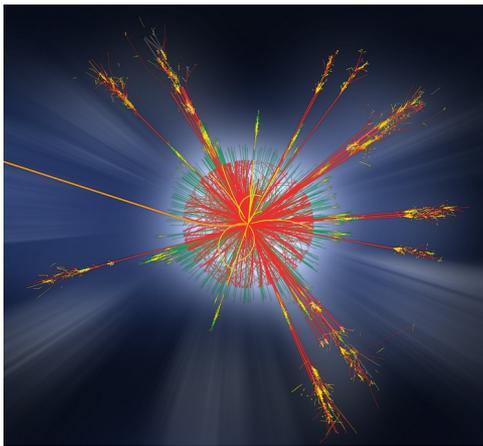
HPC (High-Performance Computing), HTC (High-Throughput Computing), MTC (Many-Task Computing). Batch processing is the running of jobs that can run without end user interaction, or can be scheduled to run as resources permit.
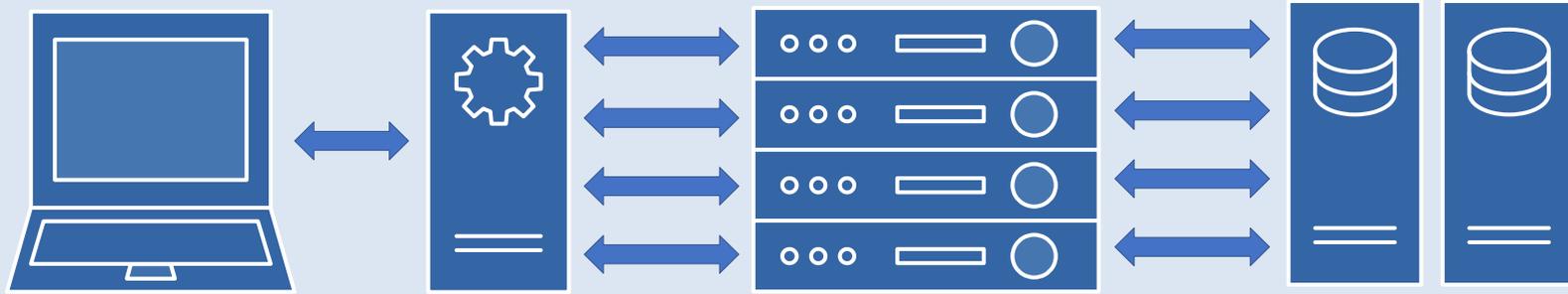

Batch systems: HTCondor, LSF, PBS (Torque+Maui), SLURM, …

Storage systems: AFS, Ceph, EOS, GlusterFS, Luster, OpenAFS, …

Infrastructure services: DNS, LDAP, Kerberos, …

Grid computing is the use of widely distributed computer resources to reach a common goal. Grid computers also tend to be more heterogeneous and geographically dispersed (thus not physically coupled) than cluster computers.
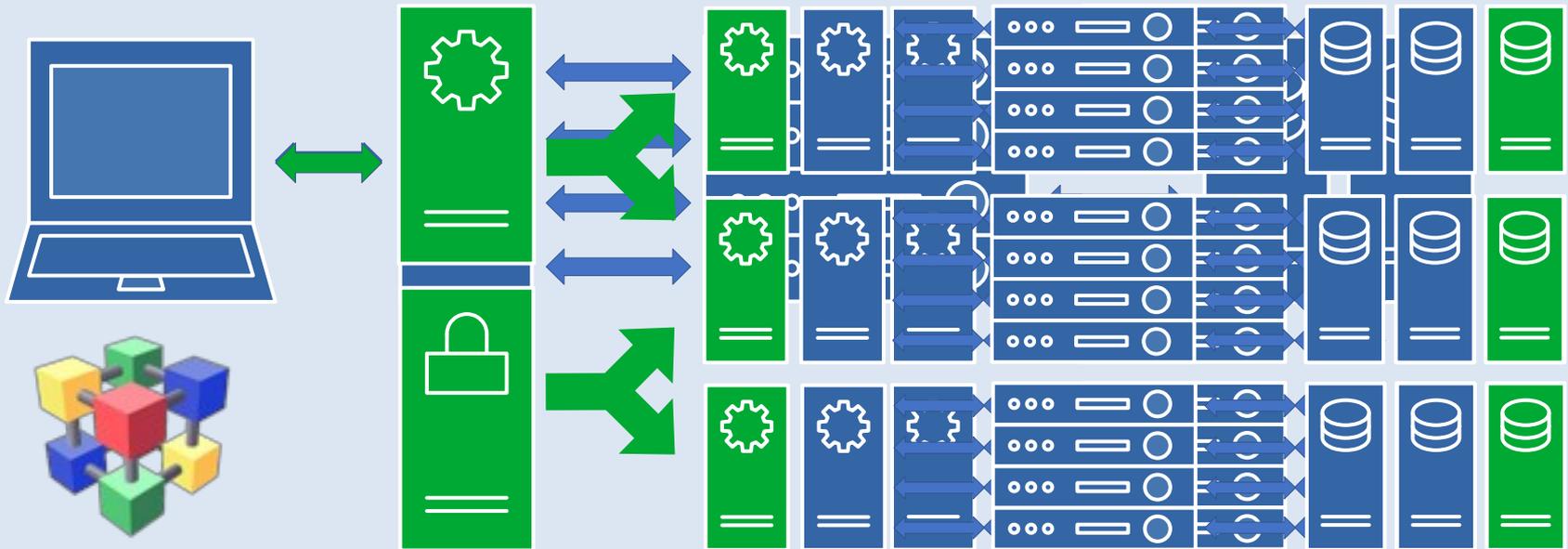
CE (Computing Element): ARC-CE, CREAM CE, HTCondor-G, ...

SE (Storage Element): DPM, dCache, STORM, ...

Grid services: APEL, BDII, VOMS, LB, WMS, ...

2006

CPU: 40 Cores 1.6 GHz (DELL)
Disk: 2 TB
Net: 12 Mbps

2020

CPU: ~ 930 Cores 2.8-3.1 GHz (DELL,HP..)
Disk: ~ 570 TB (DELL, HP, Supermicro)
GPU: nVidia (2xM2050, 10xK20m)
Net: 1 Gbps

2006 40 CPU cores, 2TB
12 Mbps (shared)

2008 Mostly local users
Grid tests

2009 EELA-2
WLCG Tier-3

2011 250 cores, 15TB

2012 500 cores, 200TB
50 Mbps (REUNA)
ATLAS Production

2013 ATLAS Analysis,
WLCG Tier-2

2014 GPU (Tsunami)
NLHPC (+240 cores)
100 Mbps (REUNA)

2016 800 cores, 300TB
200 Mbps (REUNA)
10Gbps internal

2019 930 cores, 570TB
1Gbps (REUNA)
Disk replacements

Site Name	Logical CPU	HEP SPEC06	Disk (TB)
CBPF (ALICE, LHCb)	360	2,992	476
EELA-UTFSM (ATLAS)	384	4,440	445
ICN-UNAM (ALICE)	0..856	0..9,000	0..1,600
SAMPA (ALICE, LHCb)	2,400	28,800	939
UERJ (CMS)	0..600	0..4,800	> 1

The **MoU** between **Centro Latinoamericano de Física** (**CLAF**) and **WLCG** has been signed 10 September 2013. **CLAF** is an international organization aimed to promote and coordinate efforts in development of physics in Latin America (Argentina, Brazil, Colombia, Chile and many others). **CLAF** maintains special relations with many other international organizations (**CERN**, **JINR**, etc.). **CLAF** will work as an "umbrella" to all **ROC-LA** sites. The financing is coming from the hosting institutions, not from central sources.

- **Chilean researches participate in ATLAS and ALICE experiments**
  - 2007 : ATLAS - PUC (Santiago), UTFSM (Valparaiso)
  - 2011 : ALICE - UTALCA (Talca)

  PONTIFICA UNIVERSIDAD CATÓLICA DE CHILE

  UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

  TALCA UNIVERSIDAD CHILE

- **Grid sites in Chile**
  - EELA-UTFSM (Tier-2, ATLAS, UTFSM, 450 cores, 300TB)
  - ATLAND (Tier-3, **ATL**AS **And**ino group, PUC, 170 cores, in preparation)
  - AstrogridPUC (Tier-3, **Astro**nomy **Grid**, 4 cores, PUC, in preparation)
  - GrimaPUC (**Gr**upo de **I**ntelligencia de **Ma**quina, PUC, in preparation)

- **Academic institutions in Chile use REUNA (Chilean NREN)**
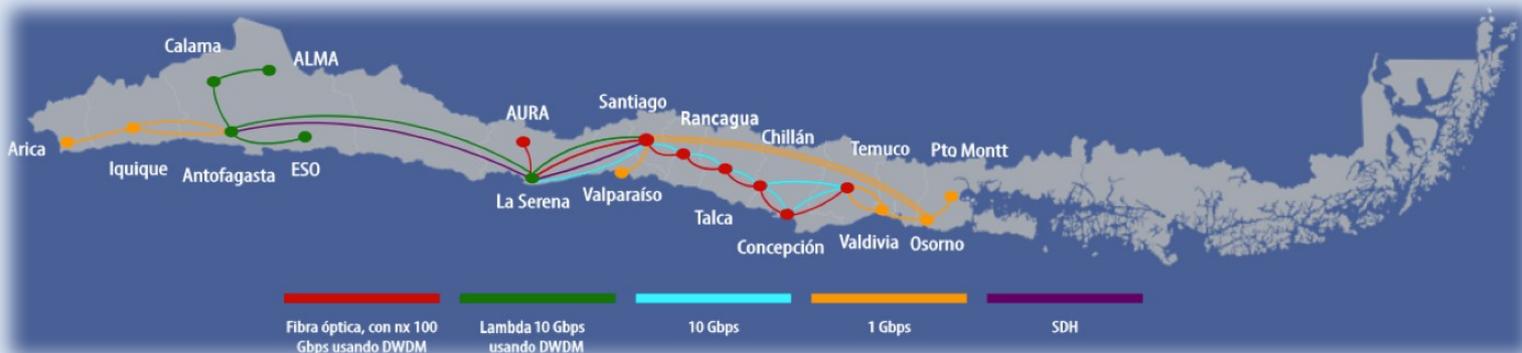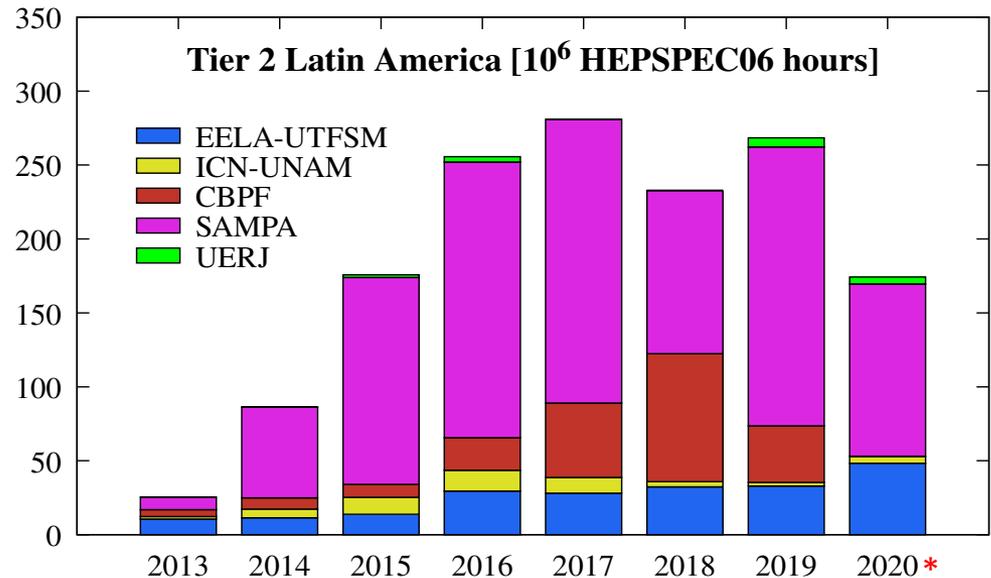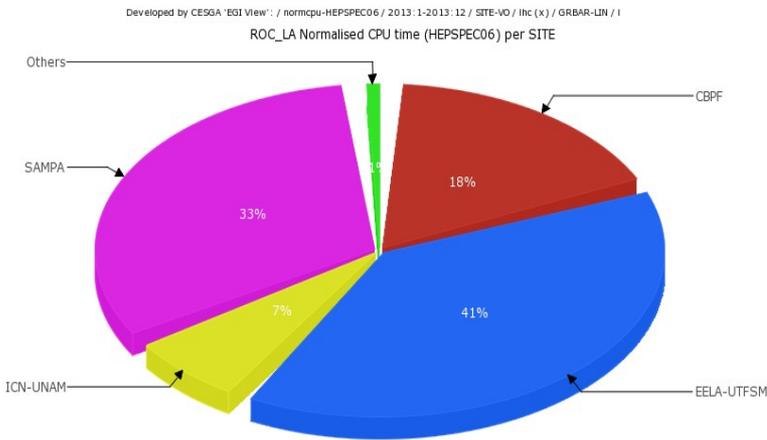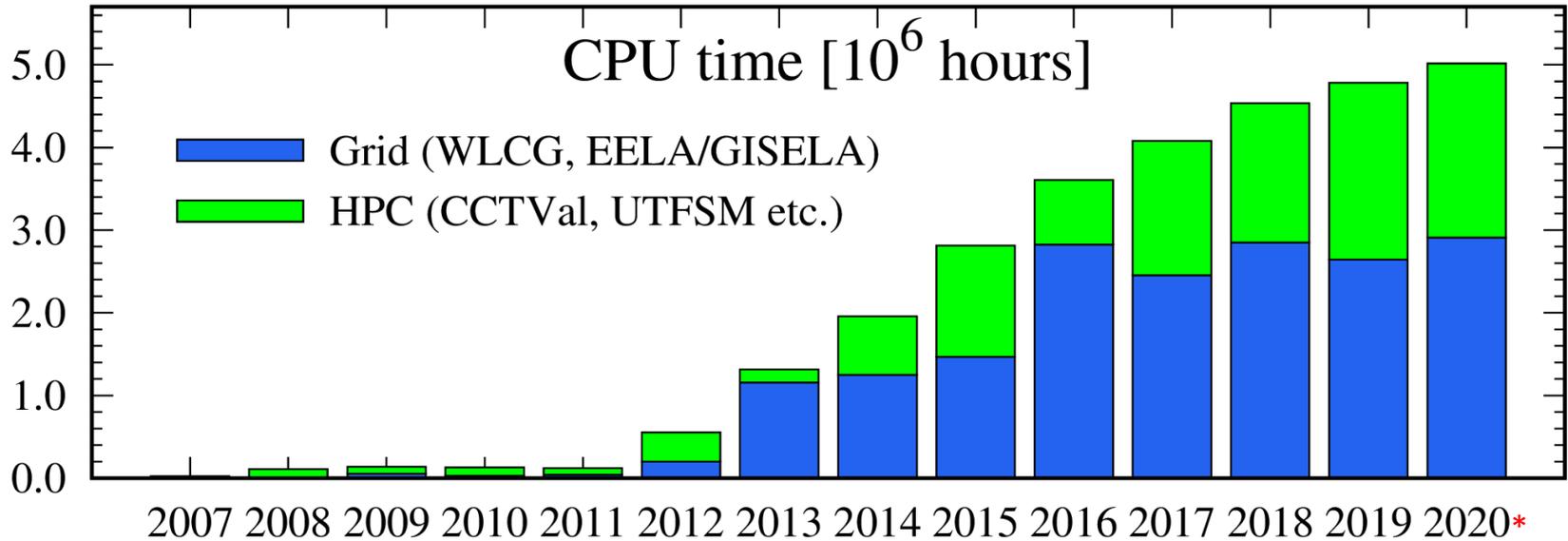
ATLAS Computing Acknowledgements
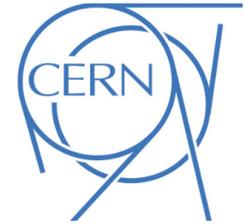
We acknowledge below the major contributors of ATLAS computing resources: the host laboratory CERN, the ATLAS Tier-1 and Tier-2 centres and the facilities contributing more than 5 Millions wall-clock processing hours in the period from 1st of January 2018 to the 31st of December 2019, normalised at 10 HS06 per core slot and showing a CPU to wall-clock time ratio of at least 40%. Numbers are extracted from the ATLAS accounting dashboard. We also acknowledge the contribution of volunteer computing, in the form of the ATLAS@HOME project.

Institute of Physics of Azerbaijan National Academy of Sciences (Azerbaijan)

CoEPP, Melbourne (Australia)

Leopold-Franzens-Universität, Innsbruck (Austria)

McGill University (Canada)
Simon Fraser University (Canada)
TRIUMF (Canada)
Universite de Sherbrooke, (Canada)
University of Toronto (Canada)
University of Victoria (Canada)

Universidad Tecnica Federico Santa Maria (Chile)

Institute of High Energy Physics, Chinese Academy of Sciences (China)

Institute of Physics, Prague (Czech Republic)

Nordic e-Infrastructure Collaboration (Denmark, Finland, Norway, Sweden)

Cluster Layout

HPC Users

UI: ui01,ui02 ...

Batch Computing

SLURM server   **Worker Nodes**

CPU (wn01..wn38)

GPU (gp03..gp09)

Grid Services

APEL

CE (ce01,ce02)

SE (dCache)

Site BDII

Storage

FS: fs01..fs05 (EOS, NFS)

SE: se + sp01..sp07 (dCache)

Infrastructure Services

DNS, Kerberos, LDAP, Mail, MedeA, Proxy, Web, perfSONAR, ...

Network Connectivity

Cluster 2020

Batch system: Migration from PBS (Torque + Maui) to SLURM

Grid CE: Migration from CREAM-CE to ARC-CE
   CREAM (Computing Resource Execution And Management)
   ARC (Advanced Resource Connector)

Grid SE: dCache upgrade 5.2 to 6.2
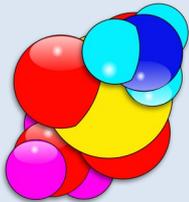
Covid@ATLAS - ATLAS, CERN, DEZY, … : contribution to study of diseases via the volunteer computing initiative Folding@Home

Cluster A/C systems: reparation and reconstruction (rats!)

Year	Computational nodes (CPU)	Status
2010	16 x 16 cores 2.8 GHz, 48 GB, 1 TB	Replaced: 23 HD, 2 RAM
2012	8 x 24 cores 3.1 GHz, 48 GB, 1 TB	Replaced: 2 RAM; Bad: 4 RAM
2014	10 x 24 cores 2.9 GHz, 48 GB, 1 TB	Replaced: 6 M/B, 10 FAN
2015	2 x 32 cores 3.2 GHz, 132 GB, 8 x 3TB	Replaced: 16 HD (disk pools)
2018	2 x 48 cores 3.0 GHz, 256 GB, 1 TB	OK

Year	Computational nodes (GPU)	Status
2011	2 x 24 cores 2.7 GHz, 24 GB, 2 x M2050	Replaced: 2 FAN; Bad: 2 x GPU
2014	5 x 12 cores 3.5 GHz, 64 GB, 2 x K20m	OK

Year	Storage nodes	Status
2010	12 cores 2.3 GHz, 8 GB, 15 x 2 TB	Replaced: 2 HD
2012	2 x 24 cores 3.1 GHz, 48 GB, 6 x 2 TB	Replaced: 2 HD; Err: 1 RAM
2012	3 x 12 cores 2.8 GHz, 24 GB, 24 x 6 TB	Replaced: 90 HD; Err: 2 RAM
2014	4 cores 2.5 GHz, 64 GB, 8 x 8 TB	Replaced: 10 HD
2016	2 x 16 cores 2.4 GHz, 64 GB, 12 x 4 TB	OK
2018	2 x 40 cores 2.2 GHz, 256 GB, 8 x 4 TB	OK

Mandatory maintenance

Just to keep the computational power at the level we have

	Mandatory maintenance	
1	Mantenimiento de los tres equipos de aire acondicionado (dos veces)	$6 \times 70{,}000$ CLP
2	Reemplazo de las baterías en UPS-es, tipo 12V 5AH	$48 \times 20{,}000$ CLP
3	Reemplazo de las baterías en UPS-es, tipo 12V 18 AH	$6 \times 40{,}000$ CLP
4	Reemplazo de los discos duros SATA-3 6TB	$16 \times 210{,}000$ CLP
5	Reemplazo de los discos duros SATA-3 8TB	$14 \times 320{,}000$ CLP
6	Reparación de fallas previstas (memoria, fuentes de alimentación)	3,000,000 CLP
7	Conexión a través de REUNA a Redes Internacionales	$12 \times 1{,}500{,}000$ CLP
		30,460,000 CLP
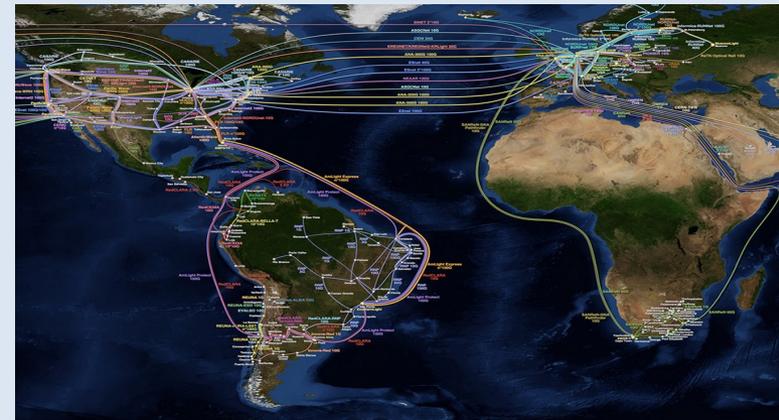
Increase the computing power up to 12% and the storage up to 10%

Minimal update		
1	Servidor computacional (2 Xeon Gold 6242, 256GB RAM, 10G SFP+)	2× 11,000,000 CLP
2	Servidor de almacenamiento (Xeon Silver 4210, 12 x 8TB, 10G SFP+)	1× 10,000,000 CLP
3	APC Smart-UPS RT 6000VA, 3U rackmountCable 10G SFP+, 5m	1,500,000 CLP
4	Actualización del enlace de red a REUNA (desde 1Gbps a 10Gbps)	3,000,000 CLP
		36,500,000 CLP

Increase the computing power up to 40% and the storage up to 20%

Desirable upgrade		
1	Supermicro SuperBlade Module Dual-Socket SBI-6429P-T3N, 6U	7×7,000,000 CLP
2	Supermicro SuperBlade Enclosure SBE-614E-422	5,300,000 CLP
3	APC Smart-UPS RT 6000VA, 3U rackmount	1,500,000 CLP
4	Servidor de almacenamiento (Xeon Silver 4210, 12 x 8TB, 10G SFP+)	2×10,000,000 CLP
5	Accesorios (cables, conectores, etc.)	1,200,000 CLP
6	Actualización del enlace de red a REUNA (desde 1Gbps a 10Gbps)	3,000,000 CLP
		80,000,000 CLP

Conclusions

Installation/configuration stage:

- ✓ Proper hardware: servers (computing, storage), network etc.
- ✓ Location, backup power lines, cooling systems etc.
- ✓ Software: OS (RHEL 3..7), packages, services etc.

Maintenance stage:

- ✓ Regular replacement/reparation (hard drives, UPS batteries, etc.)
- ✓ Regular hardware and software updates and new installations
- ✓ Reliable network channel: good ISP required!
- ✓ Users support: web, tutorials, online etc.

Human resources (always):

- ✓ Qualified personnel. Permanent increase of qualifications.
- ✓ Some strong internal motivation to keep everything running...

For details on cluster usage, please, visit www.hpc.utfsm.cl

Thank you for your attention