

# WG 4

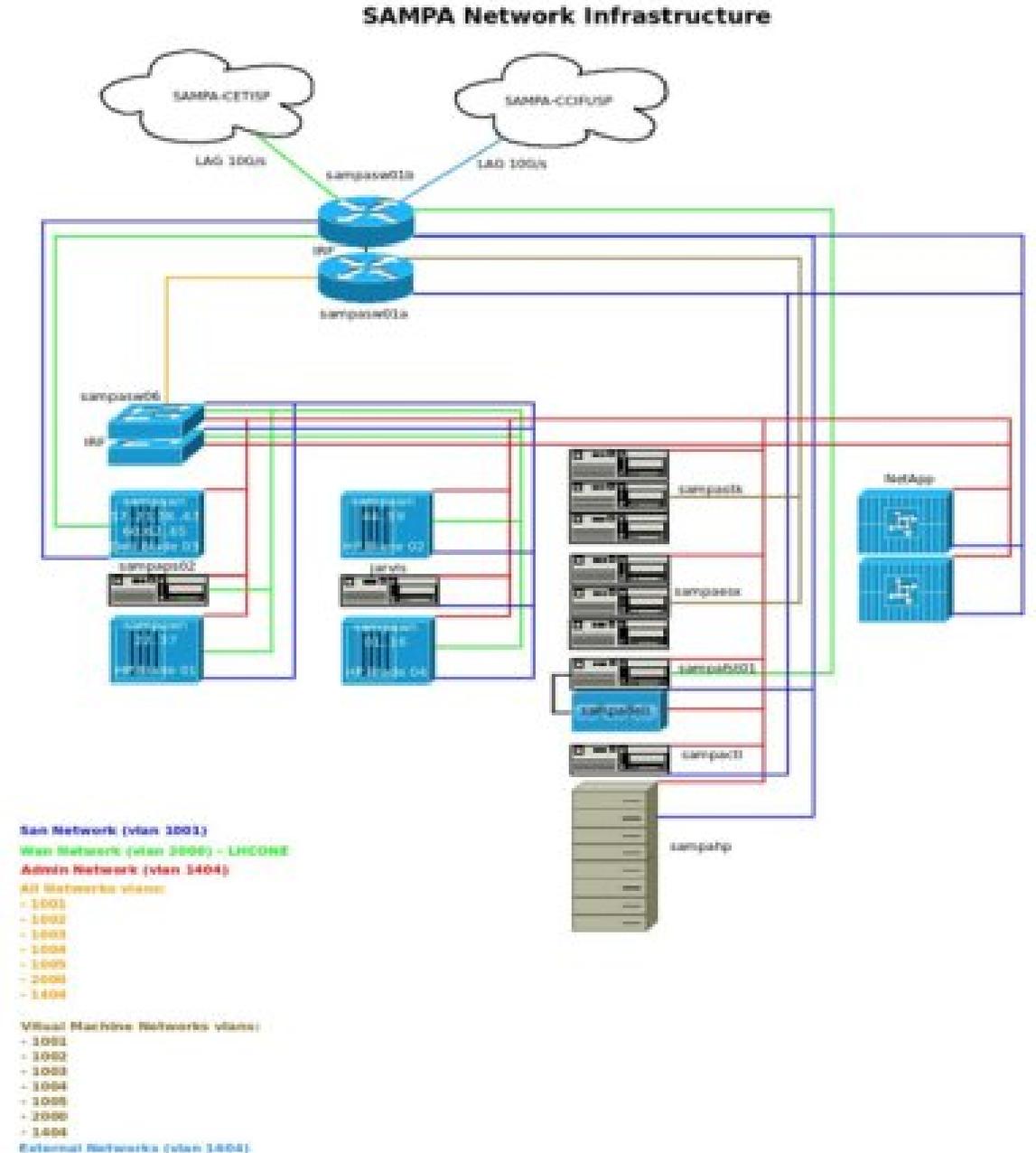
## status do sistema computacional do HEPIC

Alexandre Suaide

fevereiro/2024

# Overview

- Upgrades do cluster
  - Storage
  - Conectividade
- Processamento de dados
  - GRID/LHC
- Atualizações e upgrades
  - Federação de clusters do IFUSP
  - Em andamento
  - Alguma urgência imediata



# Upgrade do cluster - storage

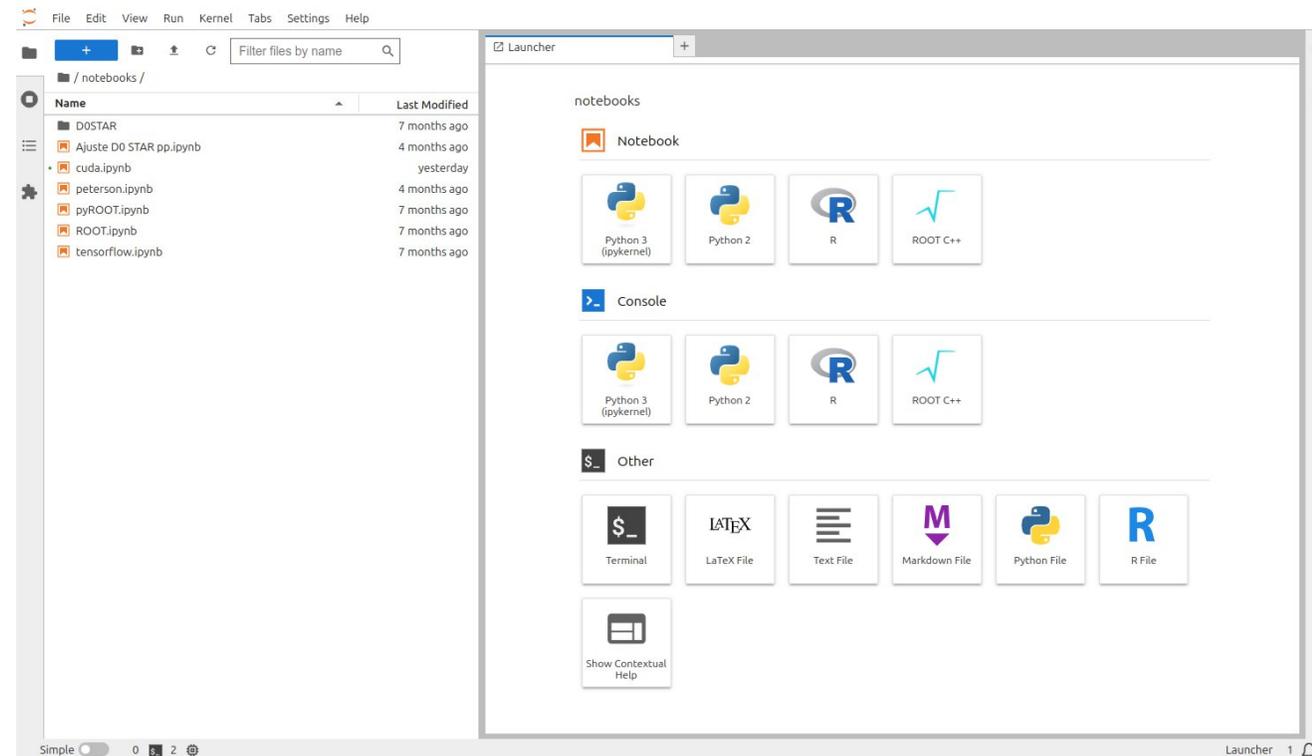
- Compra de 1.4 PB de STORAGE em 2023
  - Dificuldades de instalação de sistema de gerenciamento
- Espaço já disponível para uso
  - Em fase de migração de dados dos arquivos do GRID para nova arquitetura



SaoPaulo												
AliEn SE			Catalogue statistics						Storage-provided inform			
SE Name	AliEn name	Tier	Size	Used	Free	Usage	No. of files	Type	Size	Used	Free	Usage
1. SaoPaulo - EOS	ALICE::SaoPaulo::EOS	2	892 TB	289.2 TB	602.8 TB	32.42%	5,208,363	FILE	892.4 TB	665.9 TB	226.5 TB	74.62%
2. SaoPaulo - EOS_TEST	ALICE::SaoPaulo::EOS_TEST	2	892 TB	10 B	892 TB	-	1	FILE	881.4 TB	60.56 GB	881.4 TB	0.007%
<b>Total</b>			<b>1.742 PB</b>	<b>289.2 TB</b>	<b>1.46 PB</b>		<b>5,208,364</b>		<b>1.732 PB</b>	<b>666 TB</b>	<b>1.082 PB</b>	

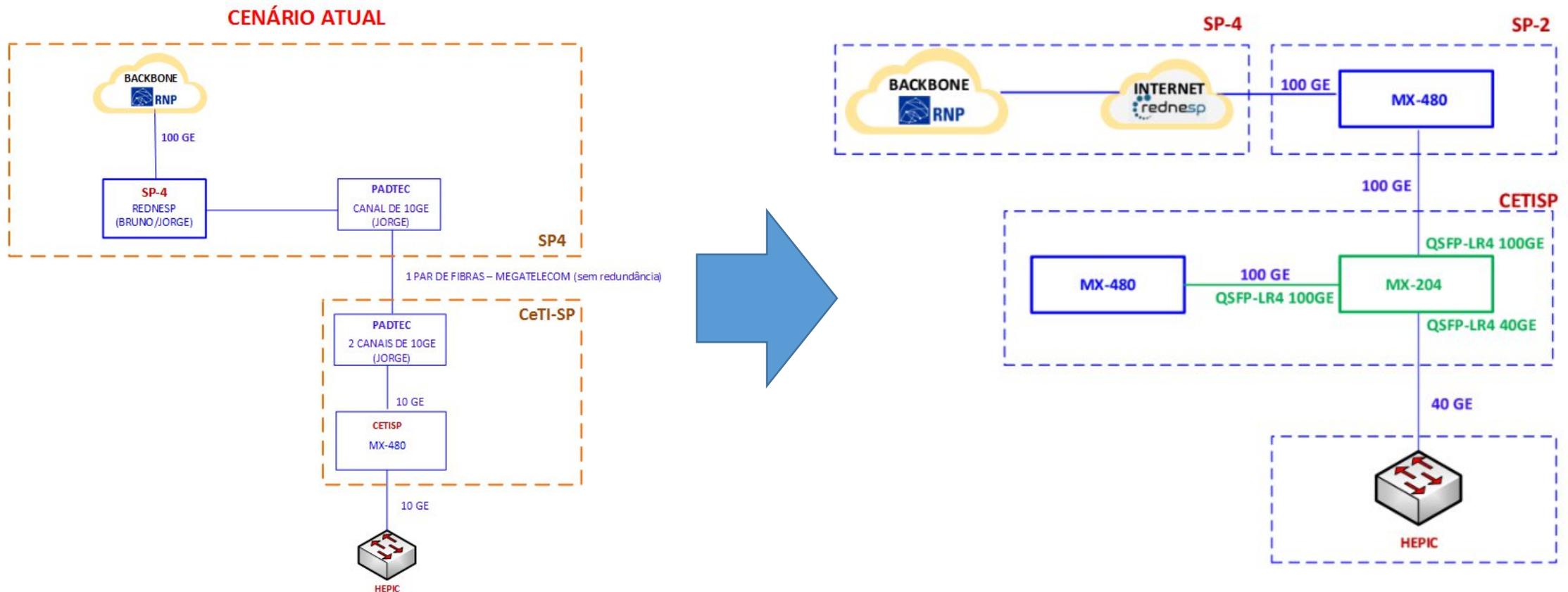
# Upgrade do cluster - JARVIS

- <http://jarvis.if.usp.br:8000>
  - Notebook com kernels diversos
  - Acesso via shell
- Pequeno cluster com GPUs
  - 4 x NVIDIA RTX 3060
    - CUDA cores: 1792
    - Threads: 43008
    - CPU clock: 1777 MHz
    - Memória: 12 GB
  - Pouco uso
- Basta ter conta no SAMPA



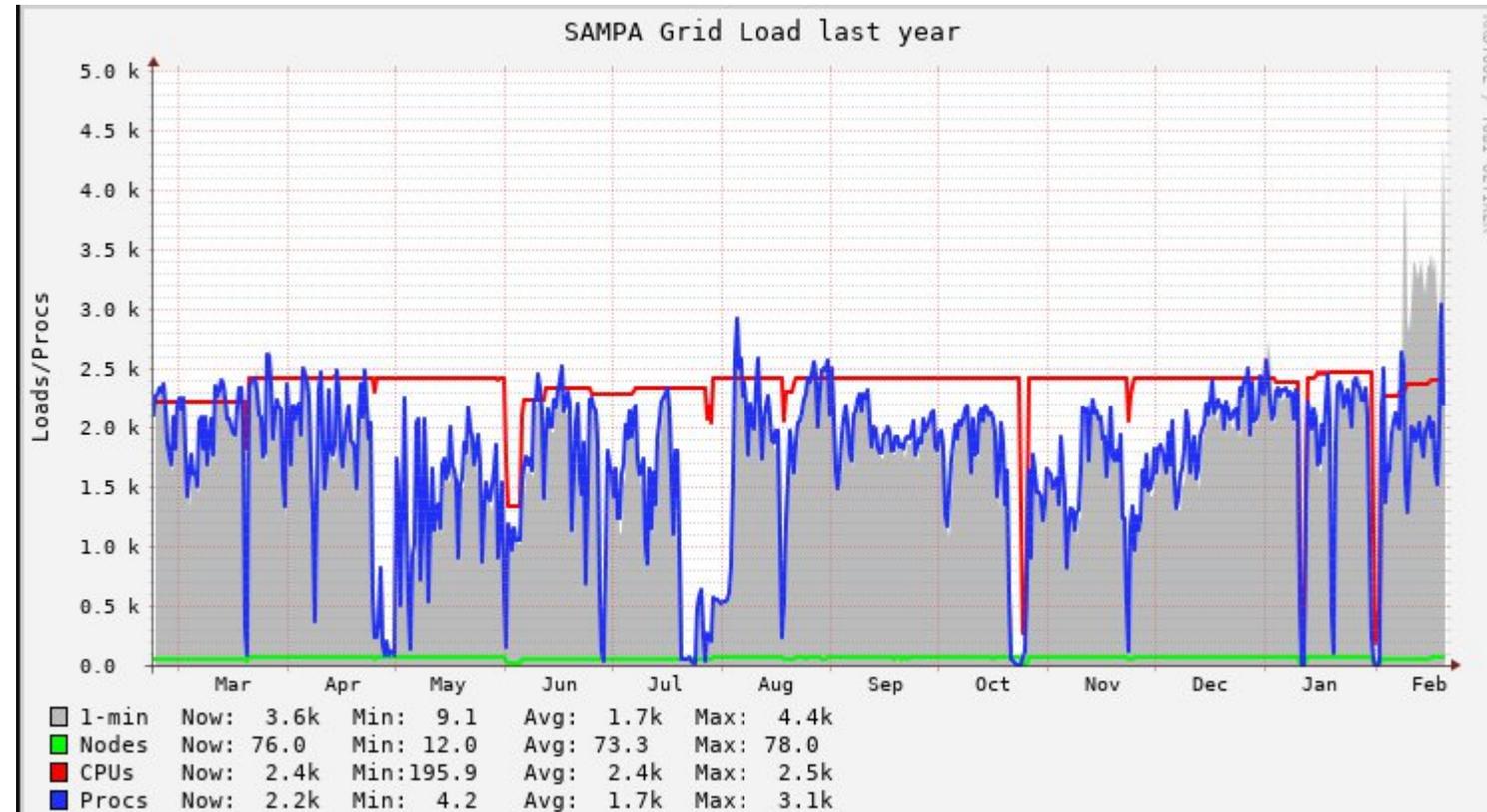
# Upgrade do cluster - conectividade

- Ativação do link de 40 Gbps até final de fevereiro após série de incidentes



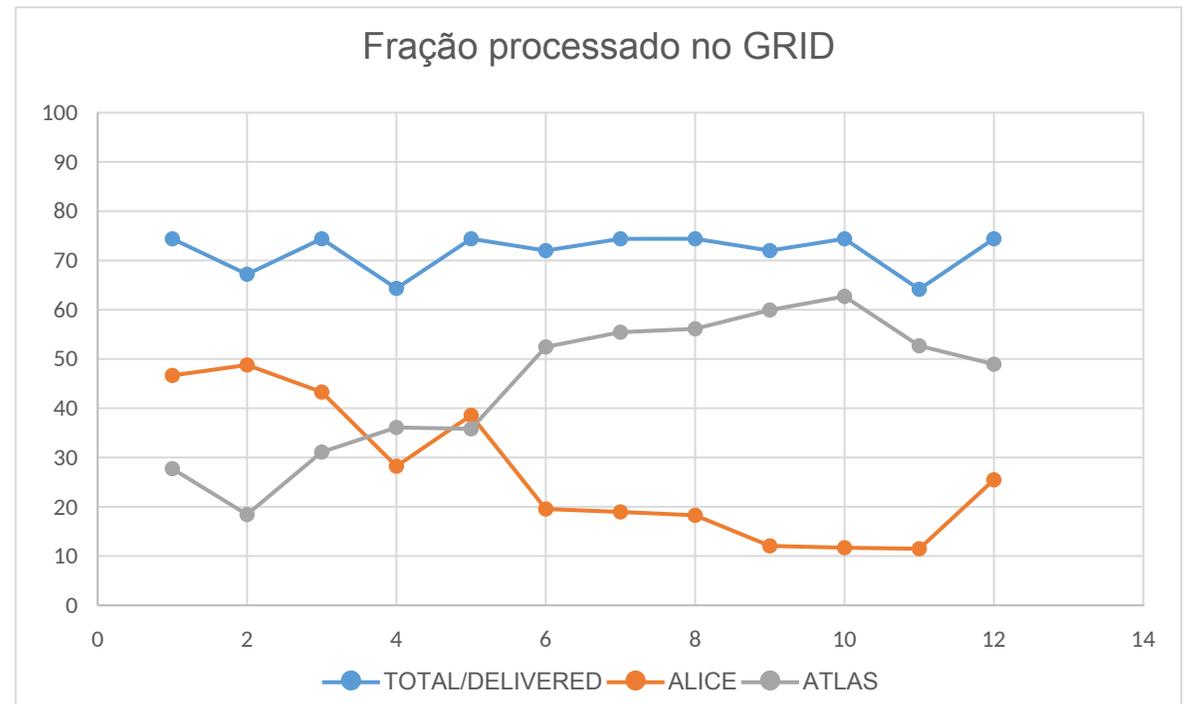
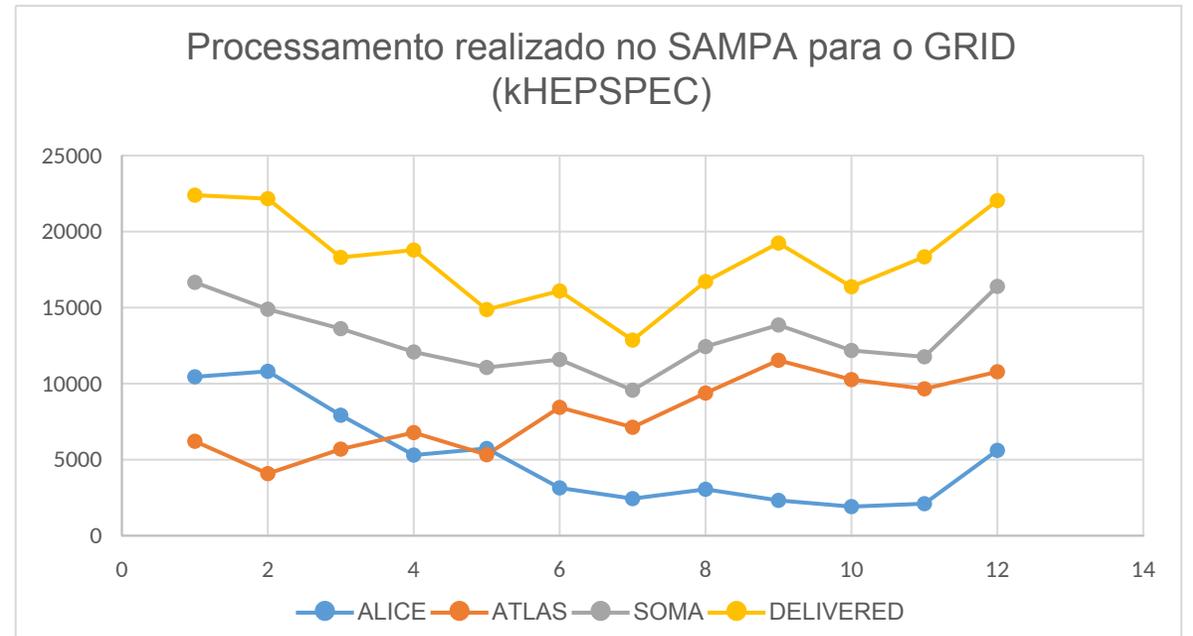
# Processamento de dados

- Aproximadamente 70% de carga ao longo do último ano



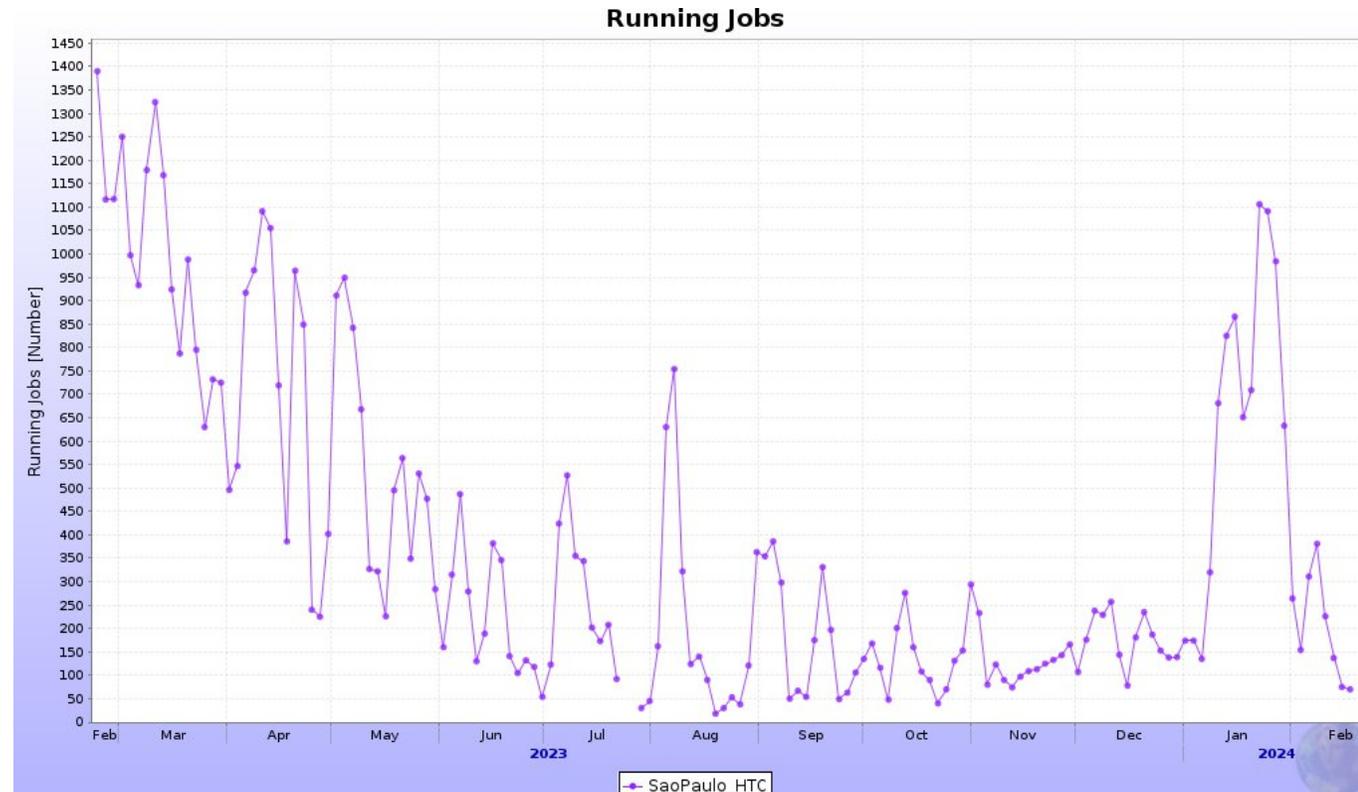
# Processamento de dados

- Balanço entre processamento do ATLAS e do ALICE foram alterando ao longo do ano



# Processamento de dados

- Número de jobs do ALICE diminui ao longo do ano
  - Perfil do job mudando de single thread para multithread
    - Tipicamente um job do ALICE aloca 8 cores
  - Número de jobs não é uma boa métrica para monitorar



# Processamento de dados

Select site: SaoPaulo\_HTC

MonALISA information Version: 22.11.05 (JDK 12.0.1) Service health NTP: SYNC, offset: 0.011s  
Running on: sampacvo.if.usp.br  
Administrator: <Douglas Vieira <dvieira@gmail.com> sampa-admin <sampa-admin@if.usp.br>>

**Services status**  
AliEn: ClusterMonitor: n/a  
PackMan: n/a  
CE: n/a  
CE info:  
Max running jobs: 320  
Max queued jobs: 10

**Proxies status**  
AliEn proxy: OK (1 day, 23:23)  
Delegated proxy: n/a (n/a)  
Proxy server: OK (57 days, 06:26)  
Proxy of the machine: OK (18:08)

**Current jobs status**  
Assigned: 0  
Running: 155  
Saving: 44

**Accounting**  
(last 24h)  
Success jobs: 390 (profile)  
Error jobs: 10788 + 43 expired  
kSI2k units: 0 / pledged

**Site averages**  
(last 24h)  
Active nodes: 1773  
Average kSI2k/core: 0

**Storages status**

Name	Status	Size	Used	Free	Usage	No of files	Type	ADD test
------	--------	------	------	------	-------	-------------	------	----------

**VoBox health**  
CPUs: x MHz  
Mem usage: % of  
Processes: 220  
Sockets: 71 TCP / 28 UDP  
Uptime:

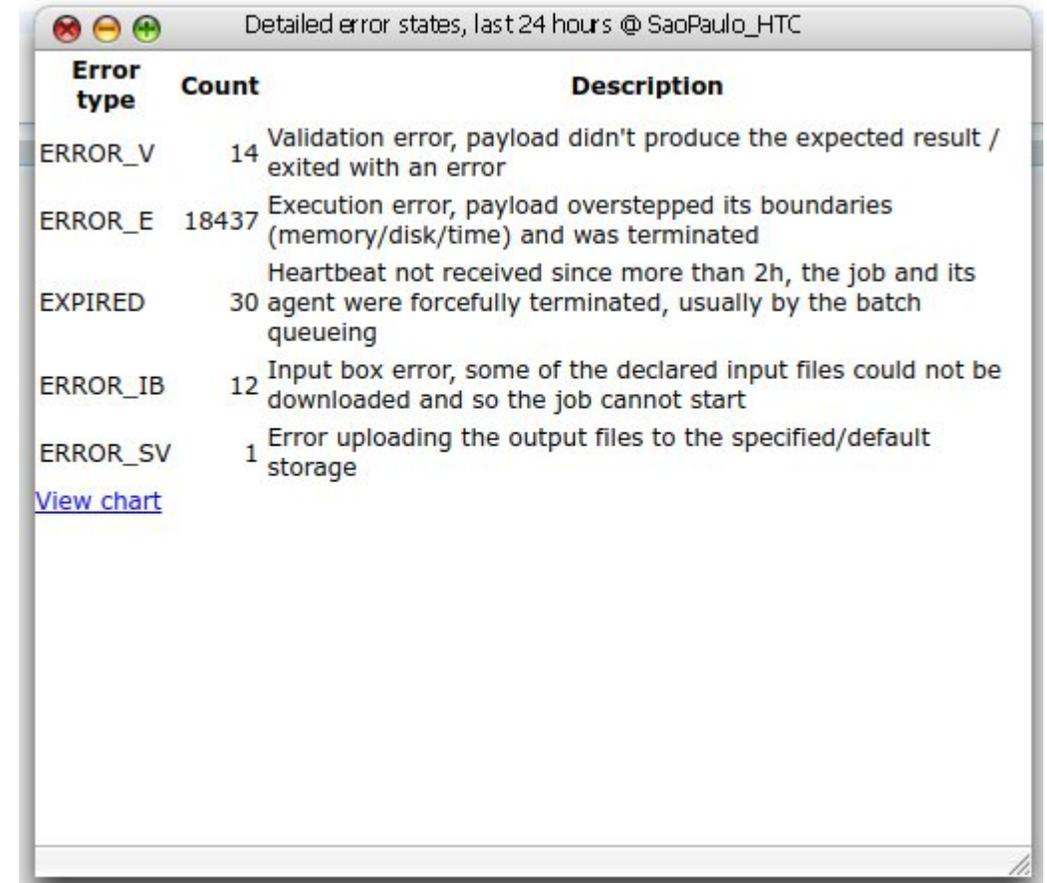
**CPU usage**  
(last 1h avg)  
Load:  
User: 1.959%  
System: 0.19%  
IOWait: 0%  
Idle: 97.84%

Int: 0%  
Soft int: 0.006%  
Nice: 0%  
Steal: 0.001%

AliEn LDAP var	VoBox path	Size	Used	Free	Use%
TMP		n/a	n/a	n/a	n/a
LOG		n/a	n/a	n/a	n/a
CACHE		n/a	n/a	n/a	n/a

# Processamento de dados

- Muitos erros por conta de falta de recursos em disco/memória/tempo
  - Vamos discutir isso mais a frente. Necessidade de ação urgente



Error type	Count	Description
ERROR_V	14	Validation error, payload didn't produce the expected result / exited with an error
ERROR_E	18437	Execution error, payload overstepped its boundaries (memory/disk/time) and was terminated
EXPIRED	30	Heartbeat not received since more than 2h, the job and its agent were forcefully terminated, usually by the batch queueing
ERROR_IB	12	Input box error, some of the declared input files could not be downloaded and so the job cannot start
ERROR_SV	1	Error uploading the output files to the specified/default storage

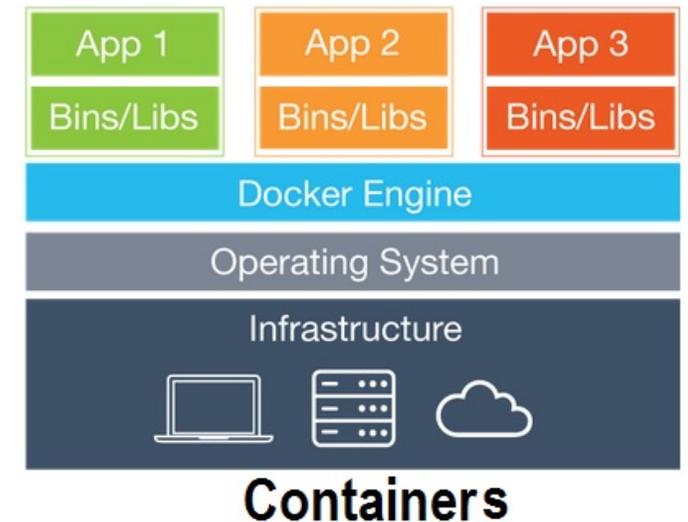
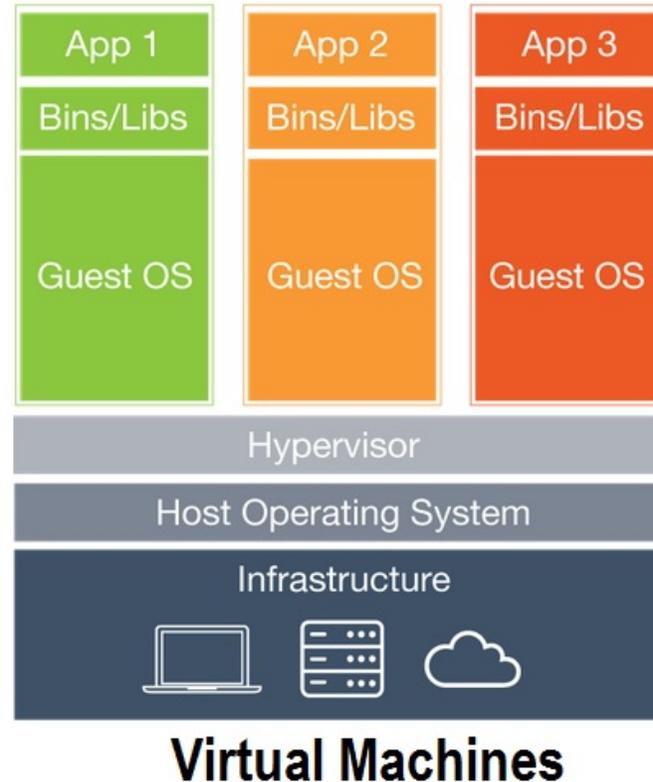
[View chart](#)

# Upgrade de hardware urgente

- Muitos desses jobs não rodam por falta de espaço temporário em disco local no WN
  - Hoje: 600 GB/WN de espaço no WN (bem menos disponível para job)
    - Quase não é suficiente para 1 Job do ATLAS ou ALICE
  - Ação:
    - Atualização para 2 TB/WN (2 x SSD 1TB)
    - Custo
      - Discos de prateleira: R\$ 500,00/disco x 2 x 48 WN = R\$ 48.000,00
        - Precisamos testar essa solução. Aguardando compras da USP fazer sua magia
      - Discos HP: R\$ 1.700,00/disco x 2 x 48 WN = R\$ 163.000,00

# Atualizações e upgrades

- Atualização dos sistemas do cluster de RedHat para ALMA Linux em preparação e andamento
  - Esperada conclusão nesse semestre
  - Mudança de arquitetura de uso diretamente sobre o hardware para containers



# Federação de clusters do IFUSP

- Agregar recursos computacionais oportunistas do IFUSP e outros sites para uso coletivo
  - Canalizar recursos de HPC para beneficiar os pesquisadores
  - Flexibilização
  - Descentralização
  - Independência de administração
  - Tornar acessível o financiamento e a participação de interessados
  - Alocação racional do dinheiro público
- Projeto EMU submetido à FAPESP
  - R\$ 2,5 M

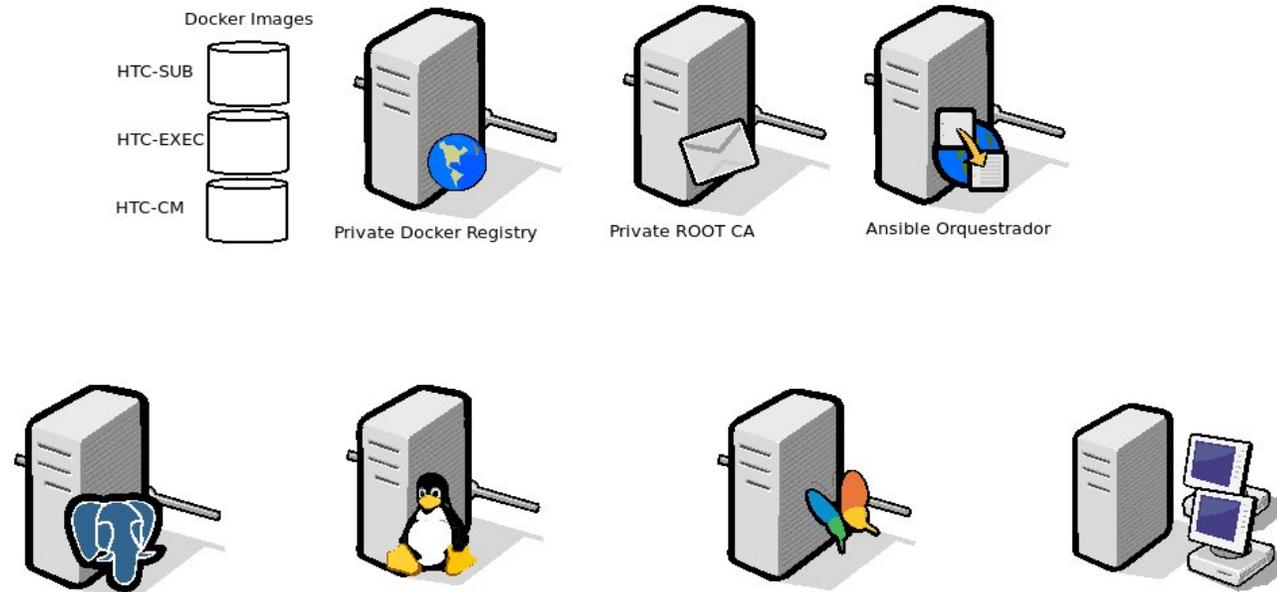
Responsável	Departamento/Grupo	Qtdd	CPU	Freq	RAM
Prof. Oscar Eboli	FMA	5	40 Cores	2.20GHz	251GB
Grupo HEPIC	FNC/SAMPA	48	32 Cores	1.70GHz	125GB
Grupo HEPIC	FNC/SAMPA	16	48 cores	1.80GHz	188GB
Grupo HEPIC	FNC/SAMPA	9	32 Cores		125GB
Prof. Helena	FMT	1	24 Cores	1.20GHz	65GB
Prof. Helena	FMT	2	32 Cores	3.30GHz	264GB
Prof. Eric	FMT	1	96 Cores	2.10GHz	512GB
Prof. Luis Gregório	FMT	4	32 Cores	2.13GHz	8 GB
Prof. Frederique	FMA	1	2x20 Cores	2.1GHz	128GB
Wanderley	FAP	1	12 Cores	1.9 GHz	125.6 GB
Wanderley	FAP	1	40 Cores	2.9 GHz	125.6 GB
Wanderley	FAP	1	224 Tensor cores • 3804 CUDA cores		
Wanderley	FAP	1	12 Cores	800 MHz	125.6 GB
Wanderley	FAP	1	16 Cores	1.7 GHz	94.28 GB
Wanderley	FAP	1	16 Cores	1.7 GHz	94.28 GB
Wanderley	FAP	1	16 Cores	1.7 GHz	94.28 GB
Wanderley	FAP	1	16 Cores	1.7 GHz	94.28 GB
Wanderley	FAP	1	16 Cores	1.7 GHz	94.28 GB
Wanderley	FAP	1	32 Cores	1107 MHz	62.89 GB
Prof. Airton Depmann	FEP/Grenac	1	2x10 Cores	2.20GHz	32 GB
Prof. Airton Depmann	FEP/Grenac	6	2x10 Cores	2.20GHz	32 GB
Profa. Kaline	FGE	7	Total 207 cores + 2GPUs		563 GB
Profa. Kaline	FGE	12	Total 304 cores		544 GB
Profa. Kaline	FGE	1	40 cores		125,42 GB
Profa. Kaline	FGE	1	8 cores		30 GB
Profa. Kaline	FGE	1	23 cores		94 GB
Profa. Kaline	FGE	1	48 cores + 2GPUs		125 GB
Prof. Henrique M.J. Barbosa	FAP	1	12 cores		64GB
Prof. Henrique M.J. Barbosa	FAP	4	12 cores		24Gb
Prof. Henrique M.J. Barbosa	FAP	4	20 cores		64GB
Prof. Caetano	FMT/SAMPA	31	4x64 cores	2.30GHz	128GB
Prof. Caetano	FMT/SAMPA	5	2x14 Cores	2.20GHz	96GB
Prof. Caetano	FMT/SAMPA	9	2x18Cores	2.20GHz	96GB
Prof. Caetano	FMT/SAMPA	7	4x12Cores	2.20GHz	192GB
Prof. Caetano	FMT/SAMPA	1	10 Cores		
Prof. Caetano	FMT/SAMPA	2	72Cores		
Prof. Caetano	FMT/SAMPA	2	GPUs		
Prof Airton	FEP	1	2x10 Cores	2.20GHz	32 GB
Prof Airton	FEP	2	2x10 Cores	2.20GHz	32Gb
Prof Airton	FEP	4	2x4 Cores	2.60GHz	32GB
Prof Airton	FEP	1	2x20 Cores	2.10GHz	96GB
Prof. Aléssio	FEP	1	2x24 Cores		192GB
Marco Leite	FNC	3	8 Cores	2.20GHz	32GB
Marco Leite	FNC	1	8 Cores	2.20GHz	32GB

# Federação de clusters do IFUSP

- Escalabilidade
  - Não existe um ponto centralizado gerente do sistema ou ui único
  - Todos os Clusters estão no mesmo nível hierárquico
  - Todos os Clusters podem sair a qualquer momento
  - Todos os Clusters podem otimizar suas filas conforme sua necessidade
  - Cada Cluster depende exclusivamente da disponibilidade da sua rede
  - Cada Cluster pode crescer independente
  - Cada Cluster pode oferecer recursos conforme seu potencial de processamento
  - Toda a Federação ganha conforme os membros crescem ou se associam
- Distribuição de imagens via DOCKER

# Federação de clusters do IFUSP

- O cluster SAMPA é protótipo nessa iniciativa
  - Já em operação
  - Outros clusters do IFUSP serão incorporados em breve



# Em resumo

- Muitas coisas acontecendo/para acontecer no SAMPA em breve
- É importante que usuários fiquem atentos às mensagens nesse período
  - Vamos tentar agendar algumas "oficinas" para bom uso dessas novas facilidades
- Obrigado ao Ricardo Romão e ao Fábio Pádoa que têm sido essenciais nesse processo.