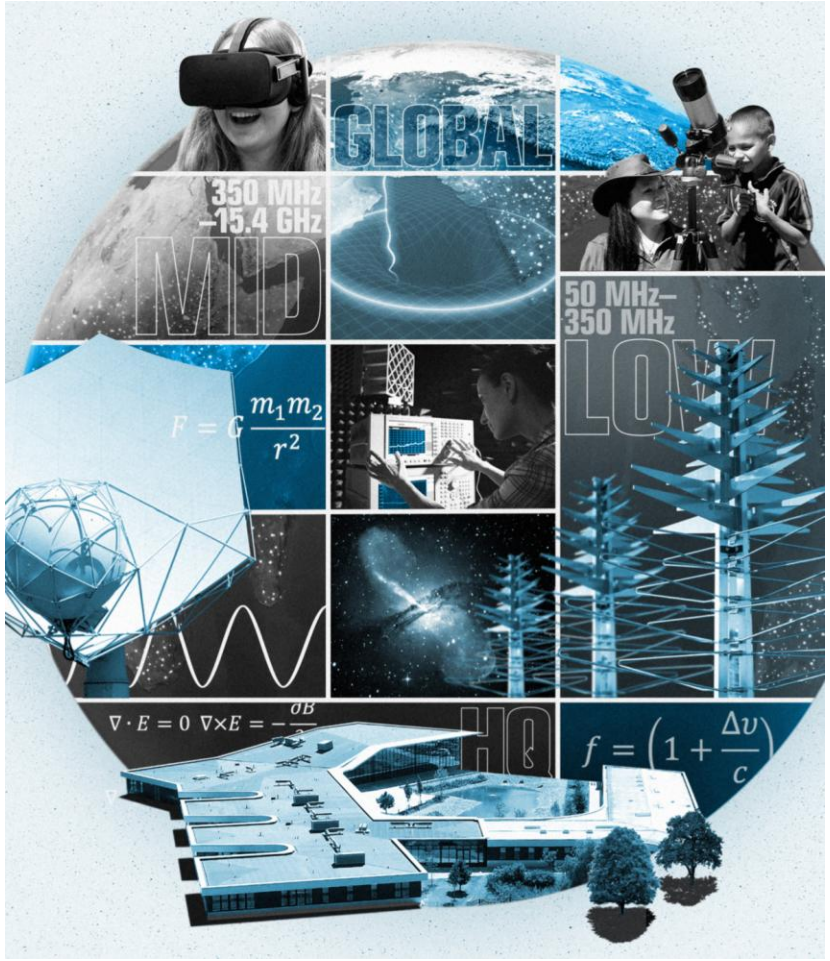
ENGAGE SKA RI: towards a PTSRC

Domingos Barbosa



SKAO Mission

"The SKAO's mission is to build and operate cutting-edge radio telescopes to transform our understanding of the Universe, and deliver benefits to society through global collaboration and innovation."



ENAbling Green E-science for SKA - Capacitation and Sustainability of Portuguese participation in the SKA with radioastronomy as an Innovation Open Living Lab – ENGAGE SKA

(2017-2022 – 1st period, 3.85M€)

(2023* - 2nd period)



ENGAGE SKA Goals:

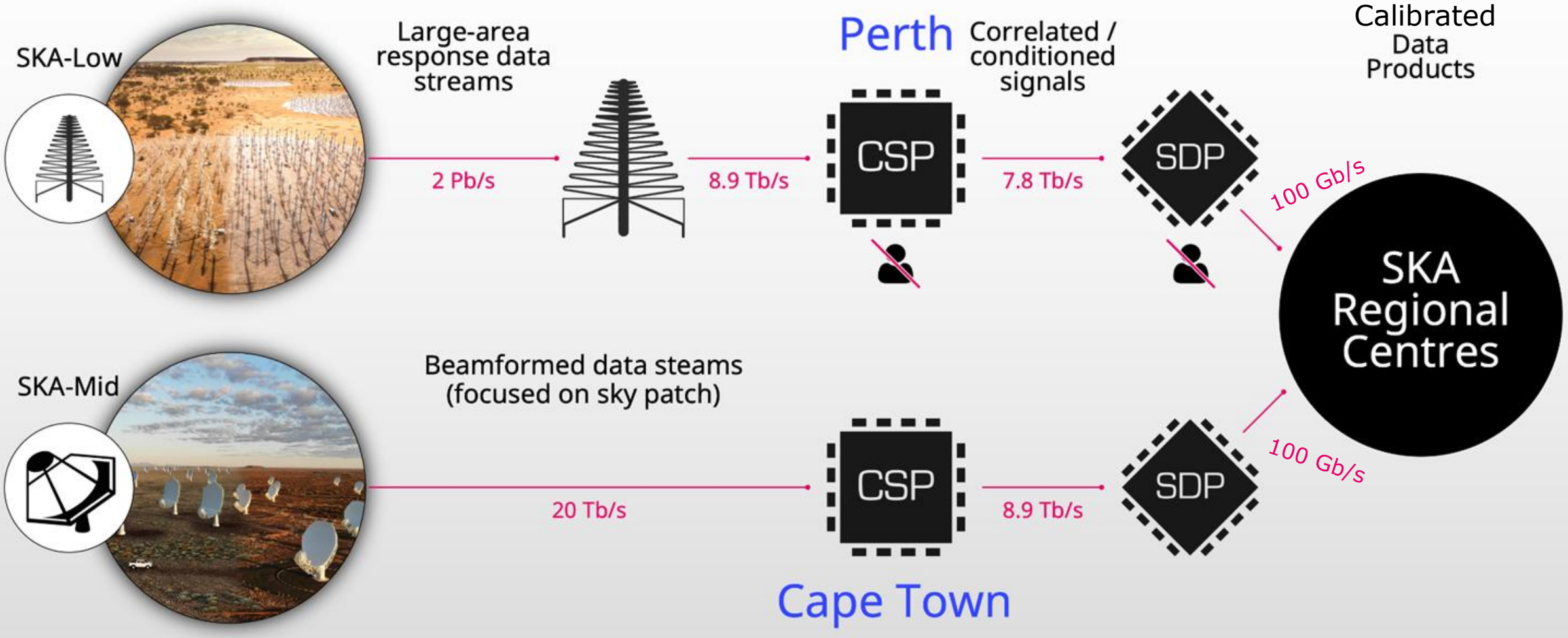
- Continue to **support National and International Agencies** in the Pre-Construction Phase of the SKA
- Foster **Science Presence and Human Capacitation**, presence in SKA Working Groups
- Maximize the non-astronomy industry opportunities from SKA, including **Big Data, Energy Efficiency** etc.
- Foster regional **research/industry connections** to broader Global linkages – **establish long-term tech Demonstrators in PT**
- Provide an **industry-led catalyst** for informing **national strategy** regarding Big Data, Energy and Environmental sensors
- Develop **SKA industry consortia**; develop example for future industry consortia-building (Research - Innovation - Market)

- 2012. ENGAGE SKA -> develops PPP
- 2015 ENGAGE SKA selected for “Roteiro”
  - The only Astronomy RI at the National Roadmap
- 2017 ENGAGE SKA awarded 3.8M€ (2017-2021)
  - Radiastronomy training
  - RNCA compute node -> Tier 2
  - Engage industry in Pre-construction
  - Strong DevOPS culture
  - Aligned with national/regional roadmaps
- 2019- PT SKA White Book
  - Mapping SKA Book contributions
  - Science updates / Industry update
  - Growing Participation
    - The biggest participation in SWGs of current PT astronomy projects.
- 2020 – SKA ratification (PT Parliament 3rd Nov)

2019; PT – Founding Signatory Member of IGO



SKA Regional Centres: SKAO data processing stages



SRC WGs

SKA Regional Centres Network (SRCNet)
Software Architecture Document

SRC-0000001	Revision 01
Classification:	UNRESTRICTED
Document type:	DDR
Date:	2023-06-01
Status:	RELEASED
Authors:	Salgado, Jesús; Wicenec, Andreas; Goliath, Sharon; Joshi, Rohini; Swinbank, John; Bolton, Rosie; Webster, Brendan; Oonk, Johannes; Grainge, Keith; Sánchez, Susana; Parra, Manuel; Dack, Thomas; Hardcastle, Martin; Barbosa, Domingos; Llopis, Pablo; Fabbro, Sébastien; Beswick, Robert; Villote, Jean-Pierre; Breen, Shari; Yates, Jeremy; Grange, Yan; Gaudet, Séverin; An, Tao; Possenti, Andrea; Darriba, Laura; Holanda, Victor; Mendoza, Mariangeles; Galluzzi, Vincenzo; Svedberg, Thomas; Lee-Waddell, Karen; Vitlacil, Dejan; Pandey, Vishambhar Nath; Akahori, Takuya; Chisholm, Louise; Horton, Maya; Watson, Robert;

SRCs AGILE Teams (2022-*)

- ❑ Olive team – Convergence HPC & Cloud
- ❑ Federation of Openstack and requiremenst with JPSRC
- ❑ Collaboration with Coral team (SPSRC) on Cloud services
- ❑ *interrupted in late 2023.
- ❑ Moving FWD soon (HPC@UE lead)
- ❑ **PT : 1.5% SKA data; 10Pb /year (+2029)**



Oblivion Supercomputer
An RNCA node @UE (2022)









https://oblivion.hpc.uevora.pt/projects/

4 Design targets

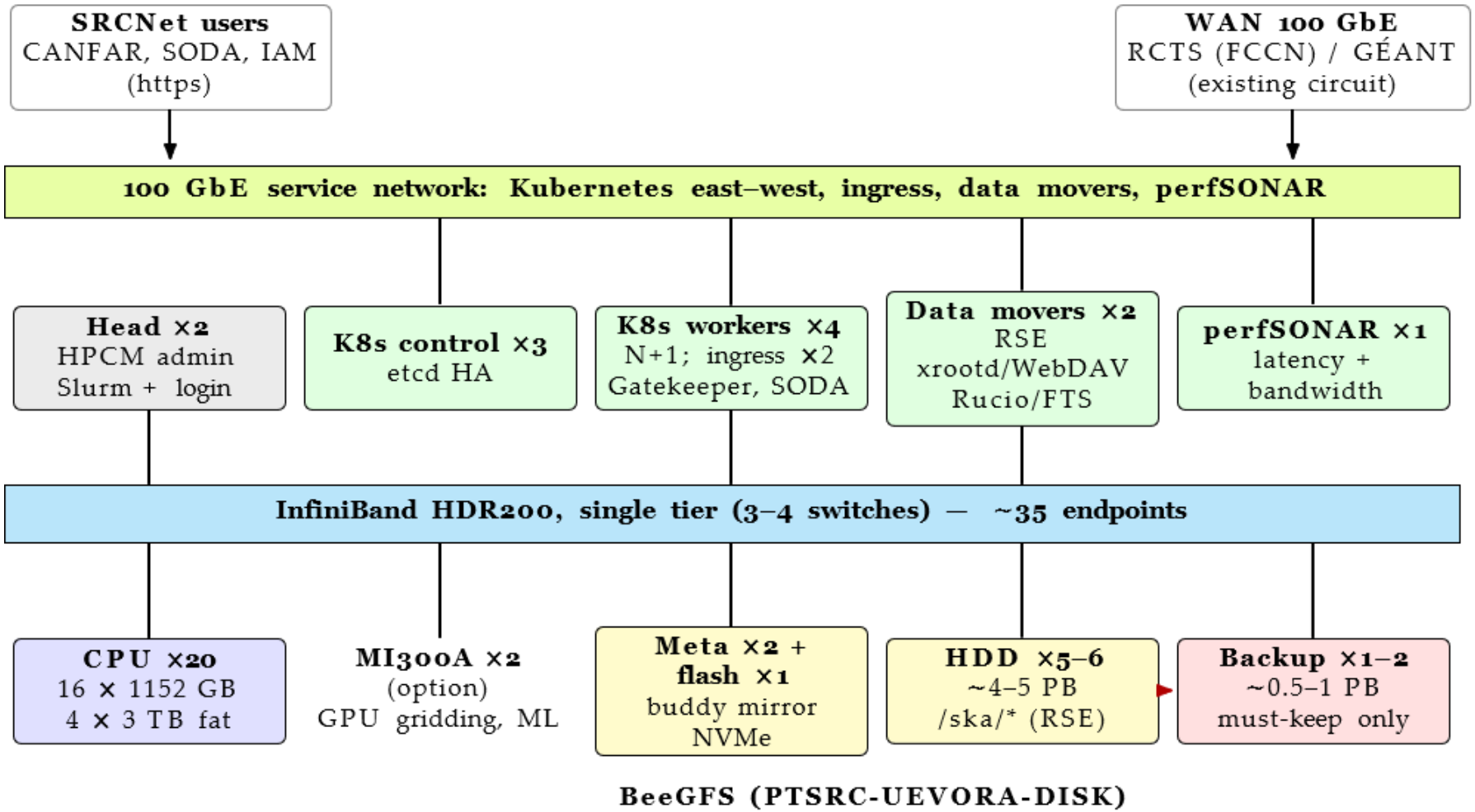
Item	Target
Role	PTSRC production node from end of 2028 **and** Chair/RNCA HPC
Management	HPE Performance Cluster Manager for every node (head, CPU compute, CPU+GPU, storage, Kubernetes)
SKA live data	15 PB ODP/ADP to hold and process (16 PB pool)
HPC user data	~3 PB HDD + ~0.6 PB NVMe flash
Live backup	~5 PB usable, separate servers, continuously synchronised
CPU	~4 × OBLIVION cores, **high memory** (1152 GB per node, 9 GB/core indicative; 3 TB fat subset)
CPU+GPU	AMD Instinct MI300A partition next to the data, GPU-aware MPI; GPUs as accelerators for SKA pipelines (priority) and Chair data processing / ML, not LLM training
Interconnect	InfiniBand HDR200, fat-tree at 1:1 or 1:2 (two options)
Kubernetes	Highly available control plane, N+2 workers, redundant ingress
Cost cap	~5 M€ hardware + fabric

OBLIVION today: 3168 cores, 304 TFlops, 3 PB; 5 GB/core on the 58 Skylake nodes, 7 GB/core on the Ice Lake nodes. VISION (16 A100 today, new advanced node incoming) remains the Chair's large-scale AI / LLM training facility.

12 SRC-only minimum configuration

SKAO fixes no per-node hardware size: an SRC contributes what its country pledges to SRCNet. What is fixed is the service stack and its reliability (certification). The smallest machine that runs PTSRC alone, without the HPC Chair workload, is shown in Figure 7.

Block	Minimum	Why
Head	1 HPCM admin + 1 Slurm controller/login	users enter mainly via CANFAR and the Kubernetes services
Kubernetes	3 control-plane, 4 workers (N+1), 2 ingress, 2 data movers, 1 perfSONAR	mandatory SRCNet services (Rucio RSE, Gatekeeper, SODA, PrepareData, CANFAR, IAM, monitoring) with redundancy
BeeGFS	2 metadata (buddy mirror), 1 flash, 5–6 HDD servers, ~4–5 PB usable	SKA data lake and RSE PTSRC-UEVORA-DISK; grows in steps with the pledge
Backup	1–2 servers, ~0.5–1 PB	must-keep and PTSRC-only data; Rucio-managed data is replicated at other SRCs
CPU	16 × 1152 GB + 4 × 3 TB, ~2560 cores	ADP generation and user analysis; fat nodes for the largest imaging (§5.2.1)
GPU	0, or 2 MI300A nodes (option)	not required by SRCNet; GPU gridding and ML
Network	HDR200 single tier (3–4 switches) or 100 GbE only; management Ethernet; existing 100 GbE RCTS	~35 endpoints, no spine layer

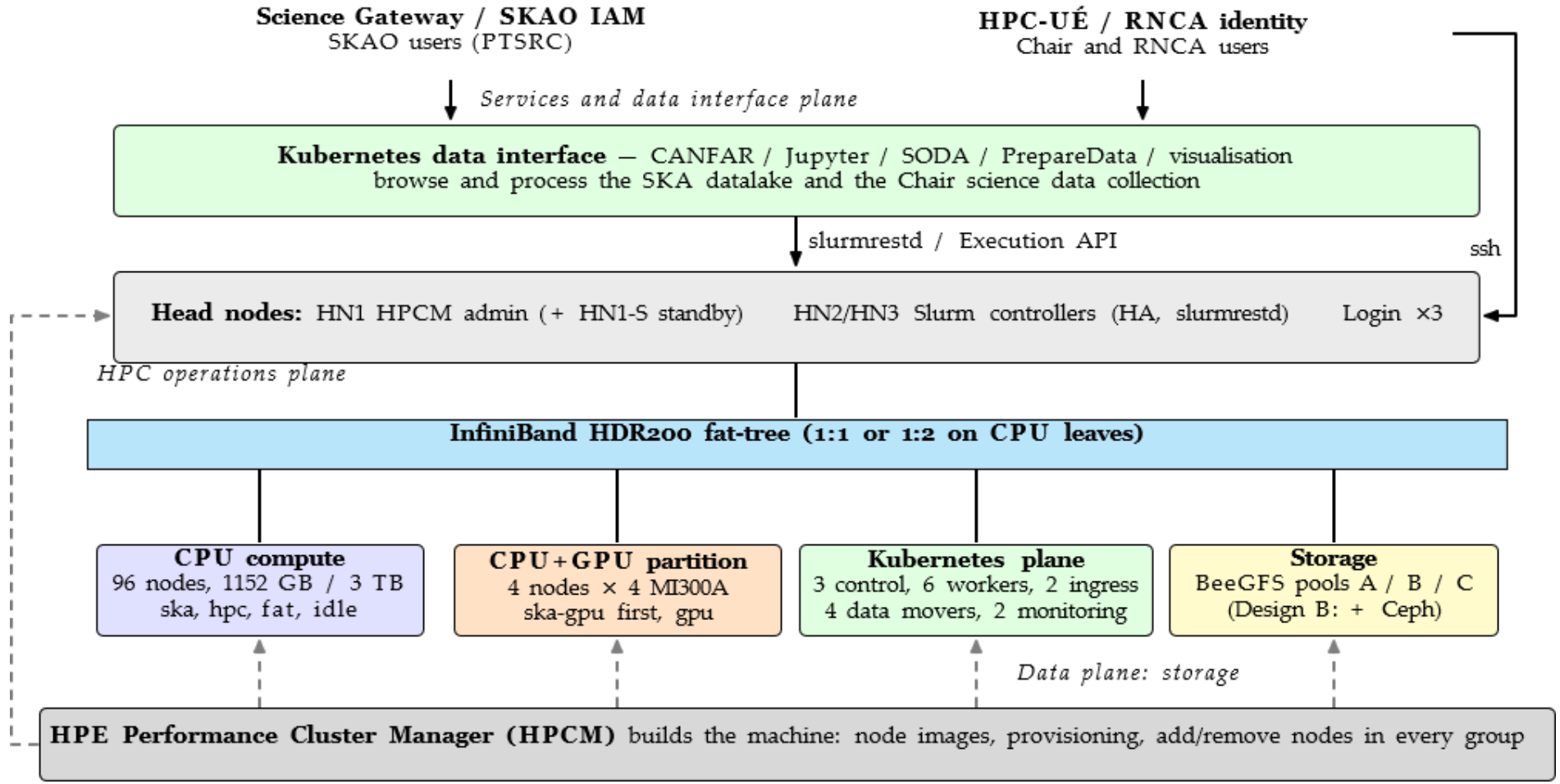


One HPCM-built machine, ~35 nodes, 4–5 racks, ~35–50 kW; ~1.2–1.6 M**e** (1.4–1.9 M**e** with MI300A).
All SRCNet mandatory services kept with redundancy; storage grows in HDD-server steps with the pledge.

Two identities, one plant

	SKA production (PTSRC)	Chair / RNCA HPC
Identity	SKAO IAM	HPC-UÉ / RNCA accounts
Entry	Science Gateway, CANFAR, SODA, Execution API	Login nodes, Slurm, CANFAR/Jupyter data interface
Data	Rucio RSE /ska/products, /ska/scratch, /ska/homes	/hpc/home, /hpc/scratch, /hpc/collections
CPU+GPU	ska-gpu partition, priority and reserved windows during ingest/ADP campaigns	gpu partition on the same nodes when SKA is idle: GPU-accelerated codes and data-processing ML. LLM / large AI training stays on VISION
CPU	ska partition, pledged SRC hours	hpc partition, RNCA / Chair queues
Idle policy	SRCNet allows back-fill of pledged compute when campaigns are idle	idle back-fill partition on CPU and CPU+GPU nodes, pre-emptible by ska / ska-gpu

Isolation is by **identity, Slurm partitions/QoS and filesystem namespaces/quotas**. /ska/products is writable only by the Rucio/PrepareData service identities and the ska / ska-gpu production accounts, never by an ordinary HPC job.


Science Gateway / SKAO IAM
SKAO users (PTSRC)
HPC-UÉ / RNCA identity
Chair and RNCA users
Services and data interface plane
Kubernetes data interface — CANFAR / Jupyter / SODA / PrepareData / visualisation
browse and process the SKA datalake and the Chair science data collection
slurmrestd / Execution API
ssh
Head nodes: HN1 HPCM admin (+ HN1-S standby) HN2/HN3 Slurm controllers (HA, slurmrestd) Login ×3
HPC operations plane
InfiniBand HDR200 fat-tree (1:1 or 1:2 on CPU leaves)
CPU compute
96 nodes, 1152 GB / 3 TB
ska, hpc, fat, idle
CPU+GPU partition
4 nodes × 4 MI300A
ska-gpu first, gpu
Kubernetes plane
3 control, 6 workers, 2 ingress
4 data movers, 2 monitoring
Storage
BeeGFS pools A / B / C
(Design B: + Ceph)
Data plane: storage
HPE Performance Cluster Manager (HPCM) builds the machine: node images, provisioning, add/remove nodes in every group

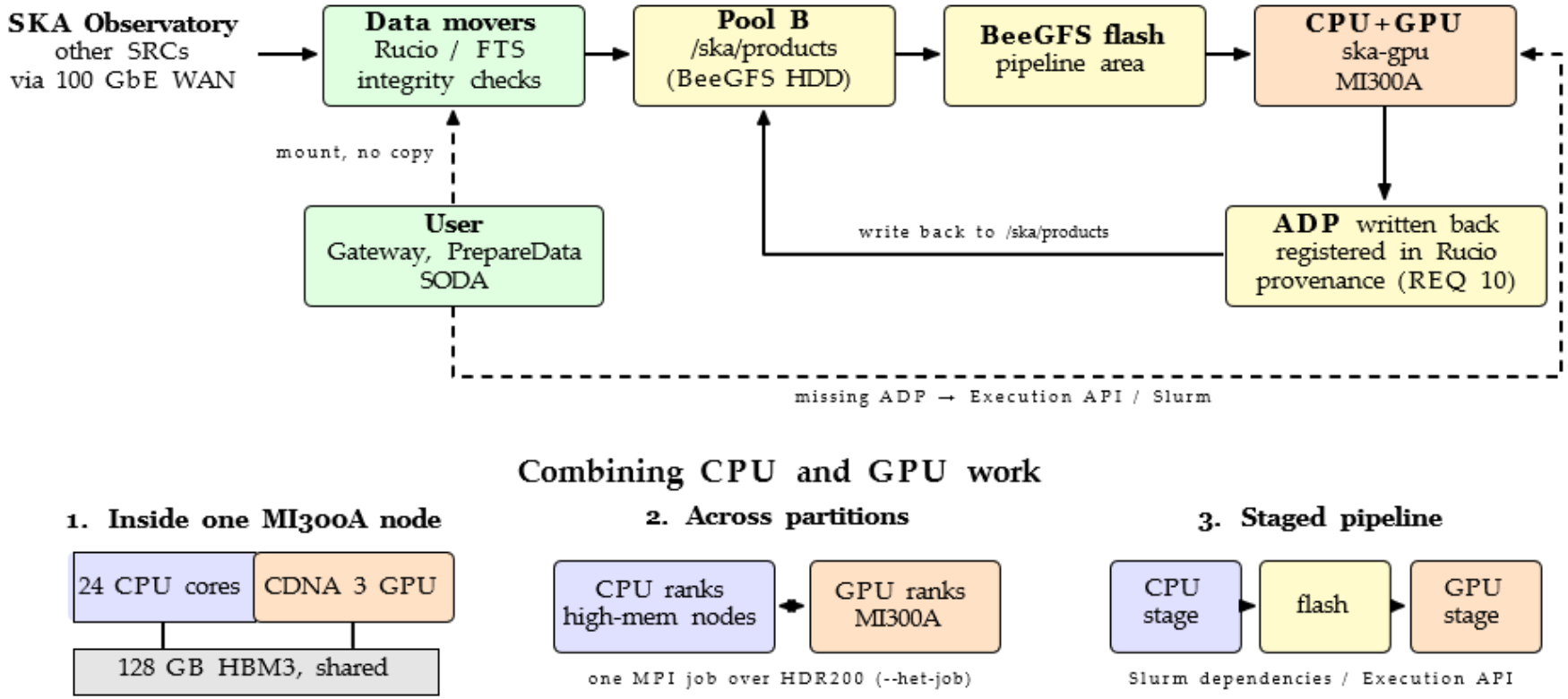


Figure 2: SKA data and processing pipeline on the CPU+GPU partition, and the three ways of combining CPU and GPU work.



Moitas Grazas, Muito Obrigado, Muchas Gracias