



Integrating High Resolution Solar Physics

Big Data Storage

The KIS Science Data Centre (SDC) case

2nd SOLARNET Forum Meeting for Telescopes and Databases
26 Nov 2020

Peter Caligari, SDC Head
Petri Kehusmaa, SDC Manager and System Architect
Nazaret Bello González, WP5 Coordinator and SDC Project Scientist
and
the KIS SDC Team



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824135.



- Deliverable D2.22: Report on Big Data storage possibilities
- In 1st SOLARNET Forum Meeting: Report on our plans for development of a Science Data Centre at KIS -> **We got approved 5 years+ funds (starting 2021)**
- Today, SDC colleagues (Cale and Petri) will report on our concept to deal with storage and management of large data volumes from ground-based solar observatories

SDC Context



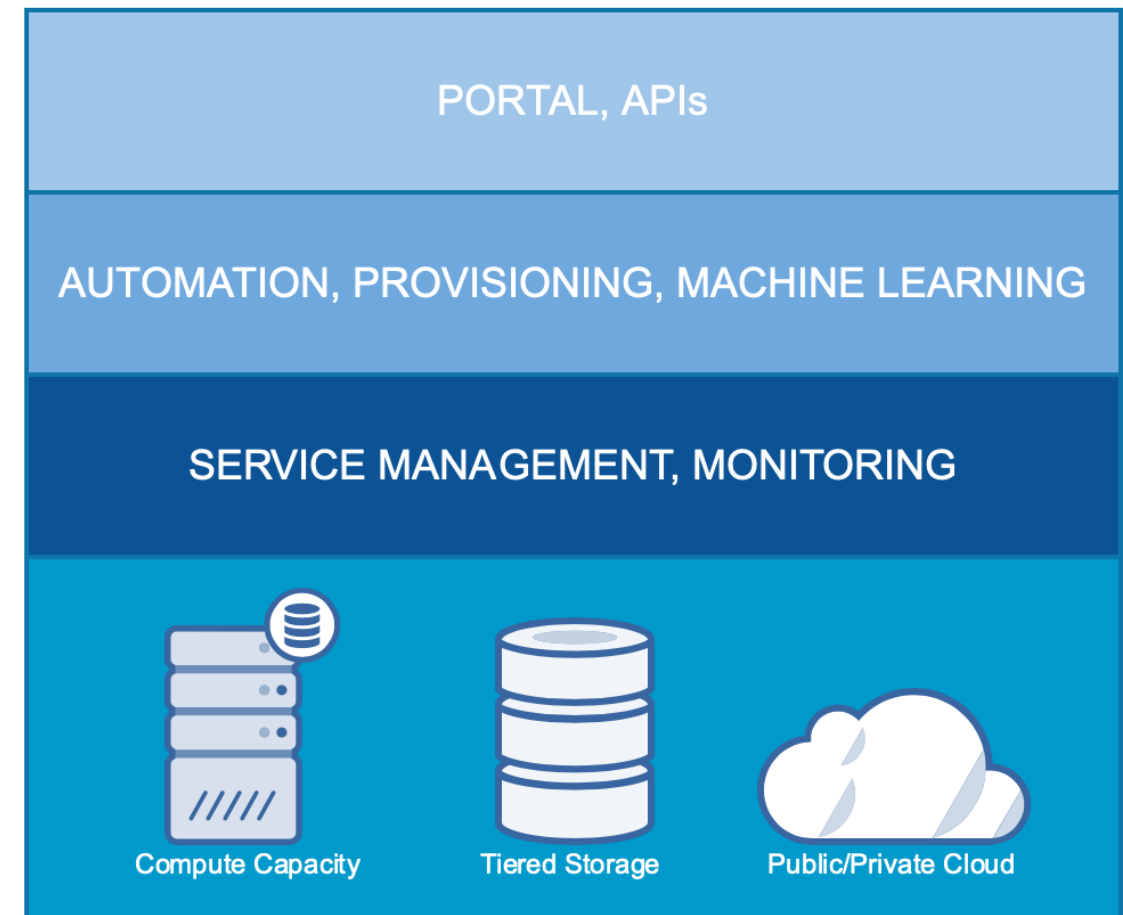
SDC @ KIS/KIT

- Tiered storage
 - Will start with disks only
 - Use cloud for scaling
- Main compute resources @ KIS/KIT

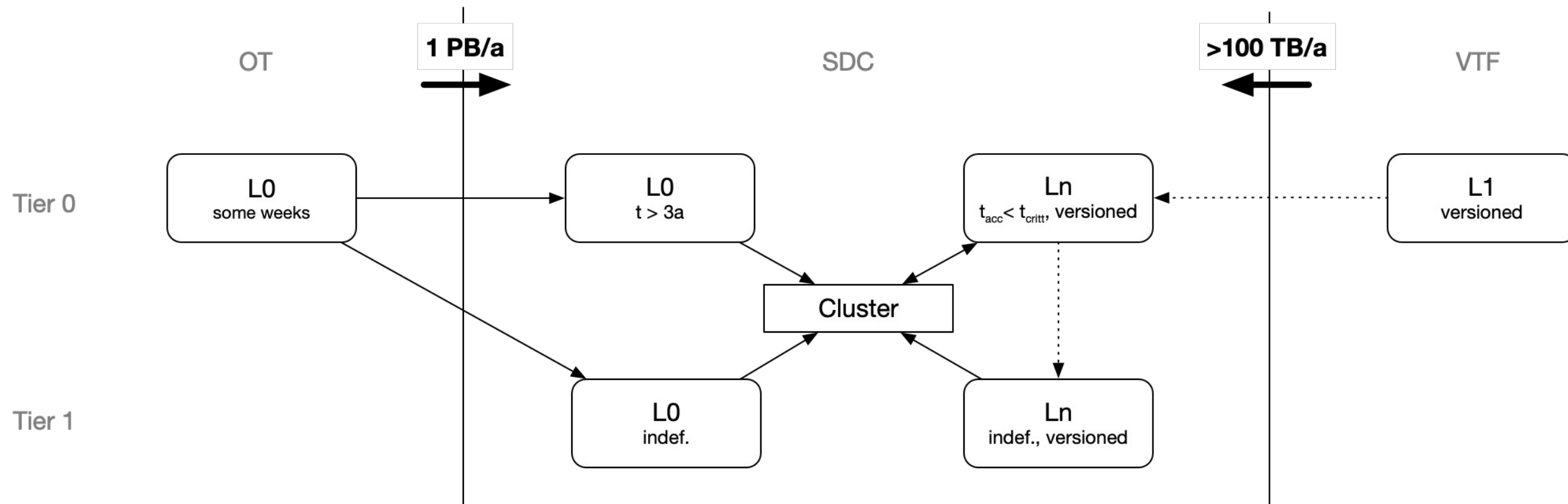
SDC IT-Framework



- Data governance (policies, standards etc.)
- Data lake (usage and cost optimization)
- Containerized applications
- User support (portal, FAQ, Docs)
- Public cloud for scalability (Amazon, Microsoft, Google)



Storage Structure @ SDC & Dataflow



Rucio

rule based data-distribution for SDC



- Currently:
 - mainly OT-data (GRIS, Chrotel, LARS, BBI) using an object-store in Mongo-DB
 - <https://www.sdc.leibniz-kis.de>
 - single-site, no tiers, proprietary, gateways almost all archives need to be developed
- Switch to Rucio (suggested by KIT)
 - + easily replicate between different sites with varying retention times
 - + keep track of data-usage (couple DOIs to collections, guarantee storage times > 10a)
 - + synergy effects within astronomy (SKA, ESCAPE, ...), easy to integrate with VTF
 - + abstracts storage layer, open source
 - no native embargoes
 - steep learning curve, no *HowTOs* & documentation not suitable for beginners (but very helpful developers)