Karlsruhe Institute of Technology

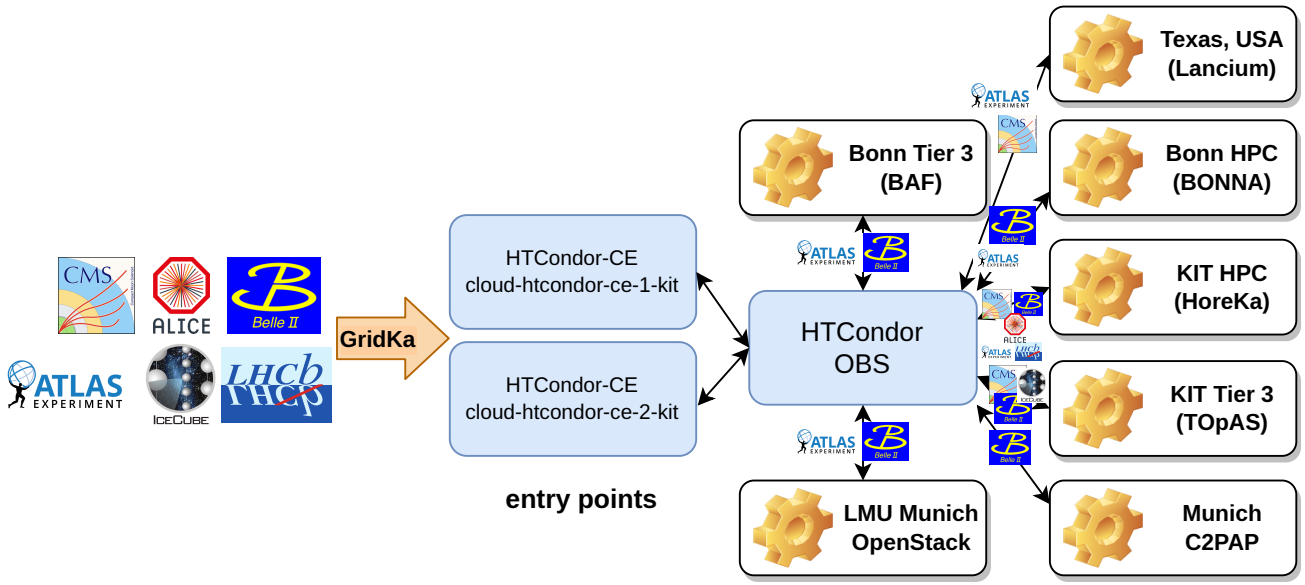
Institute of Experimental Particle Physics

Data-Aware Scheduling with Hash-Based Data Distribution

M. Giffels, A. Gottmann, **R. Hofsaess**, M. Horzela, G. Quast, M. Schnepf | 30.03.2023

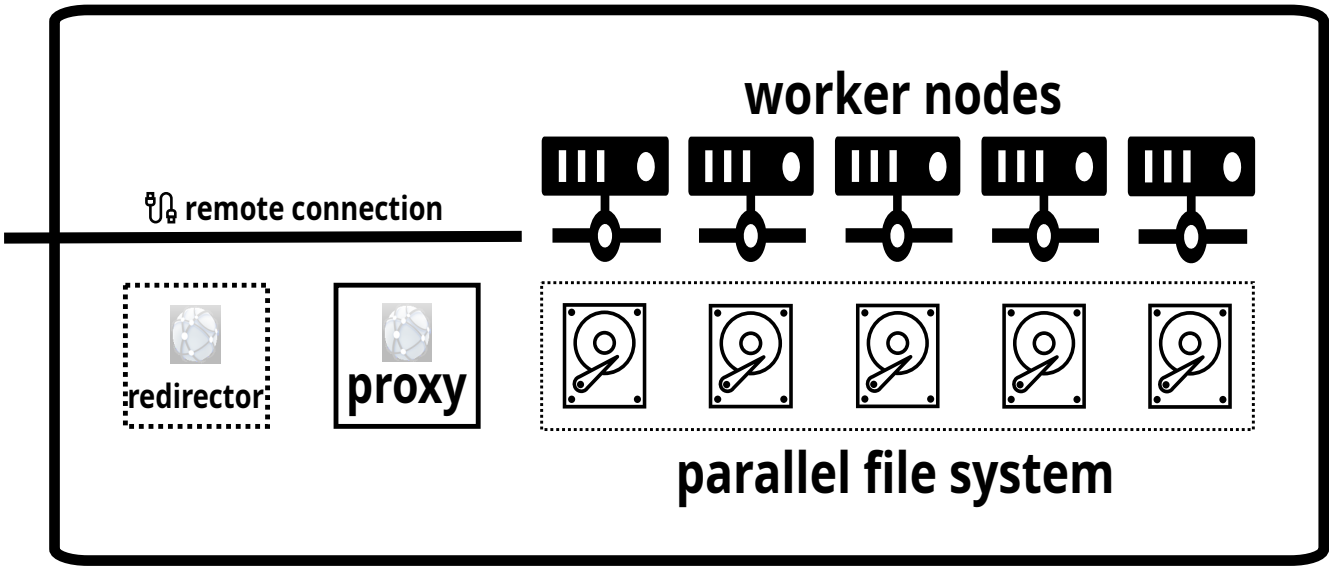
FTS and XRootD Workshop @ JSI (27-31 March 2023)

Integration of Opportunistic Resources



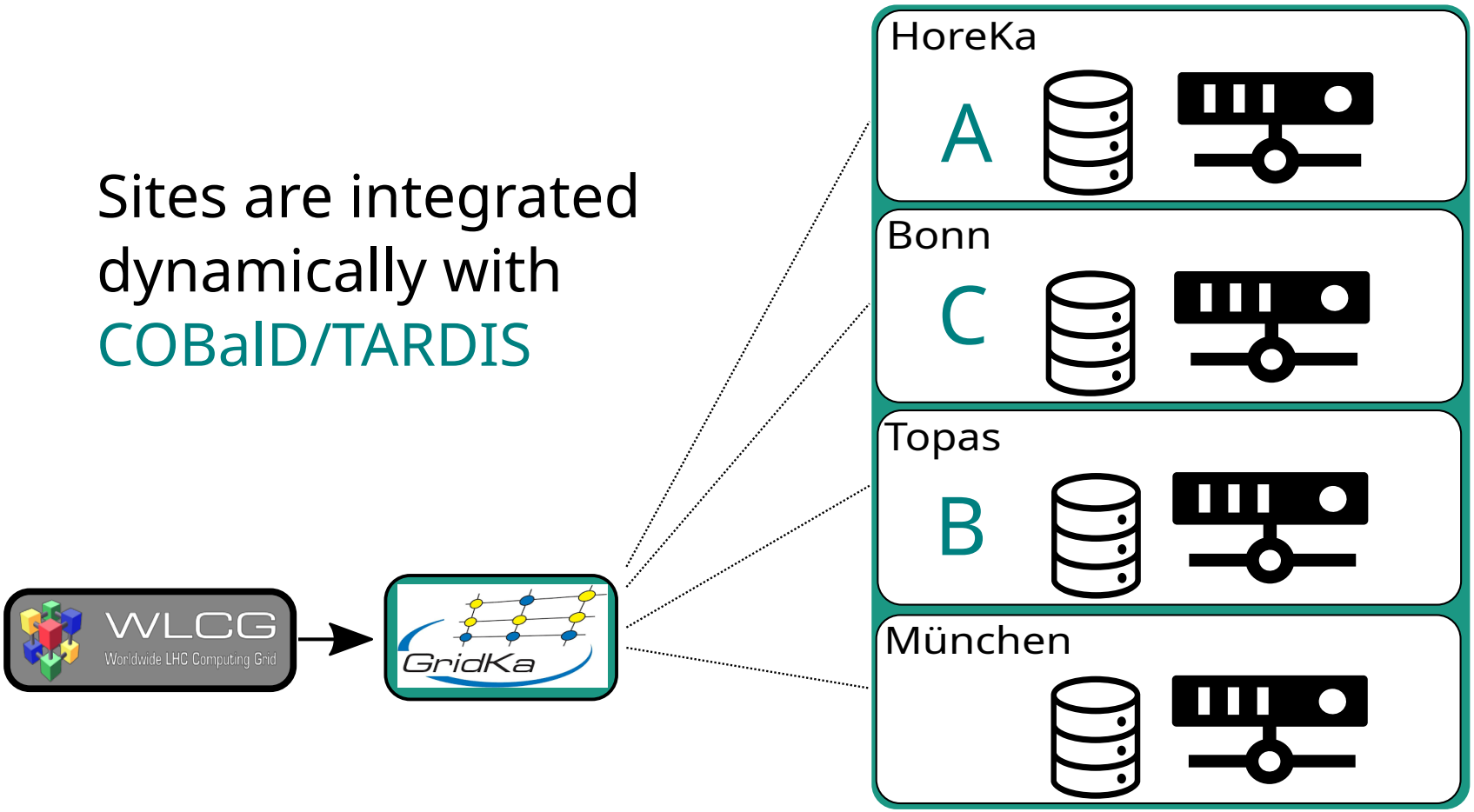
- Opportunistic resources are integrated dynamically with COBaID/TARDIS
- Very heterogeneous infrastructure
- No permanent, managed storage at integrated sites – only caches

Opportunistic Resource

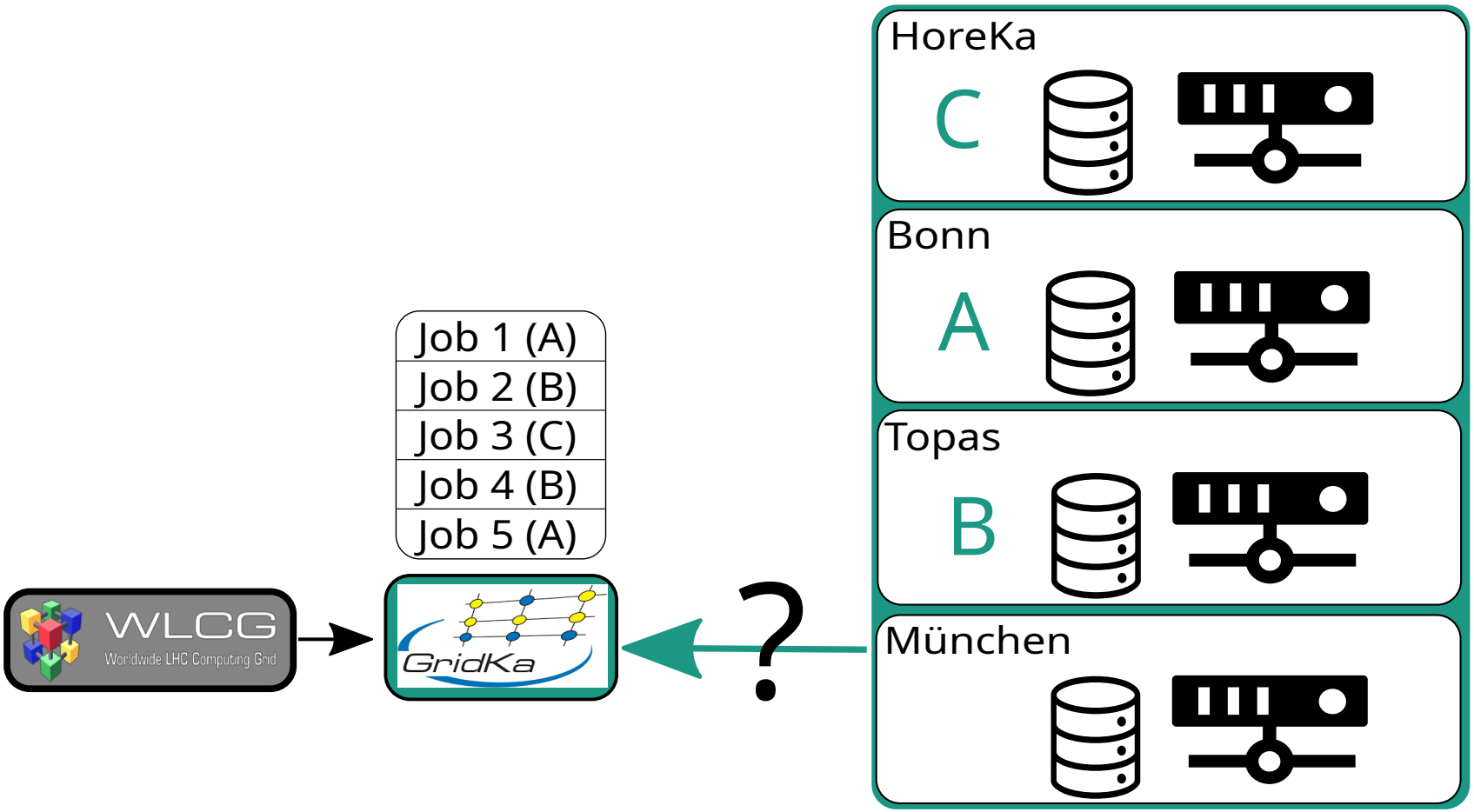


Data-Aware Scheduling for ORs

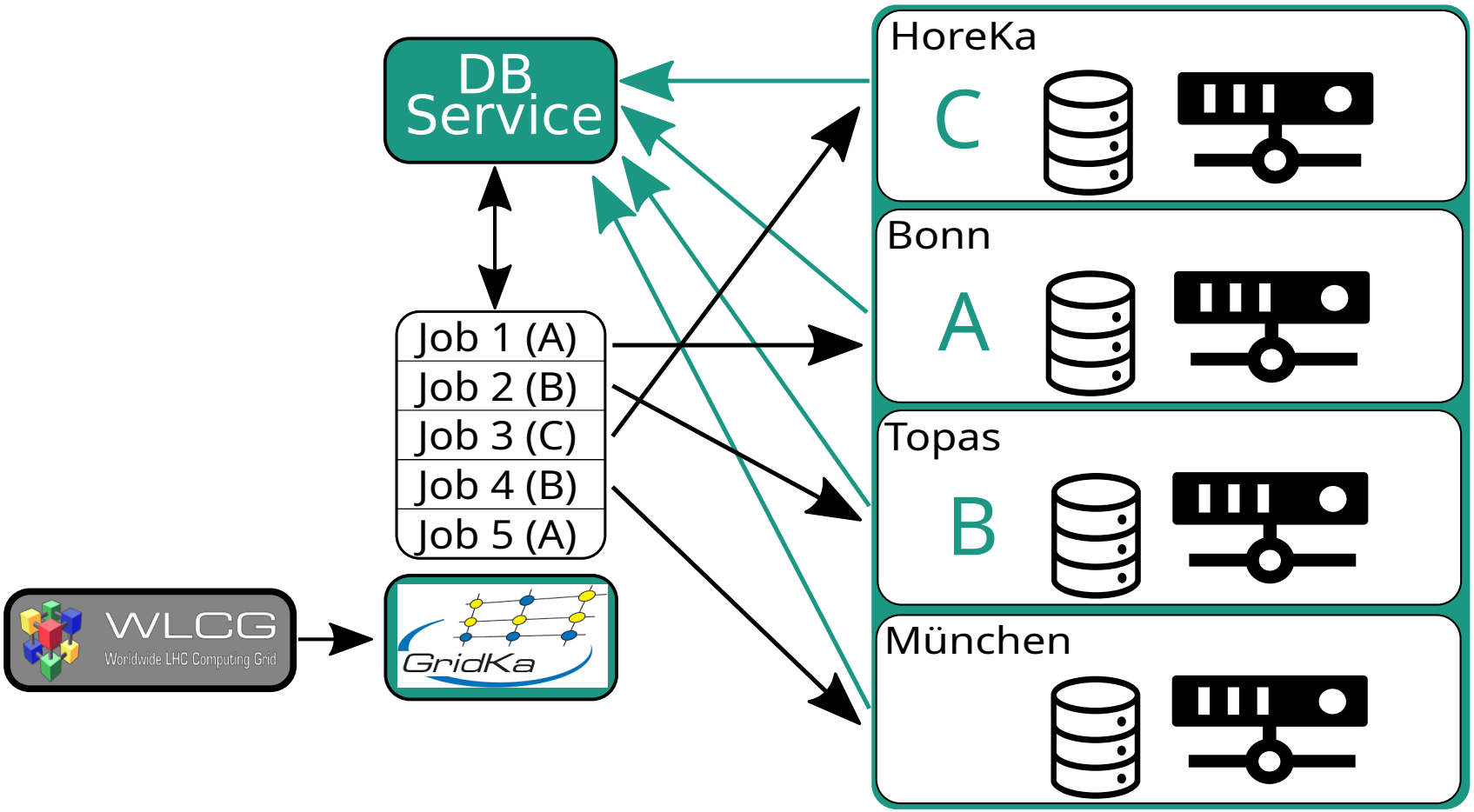
Sites are integrated dynamically with COBalD/TARDIS



Data-Aware Scheduling for ORs



Data-Aware Scheduling for ORs



Hash-Based Approach

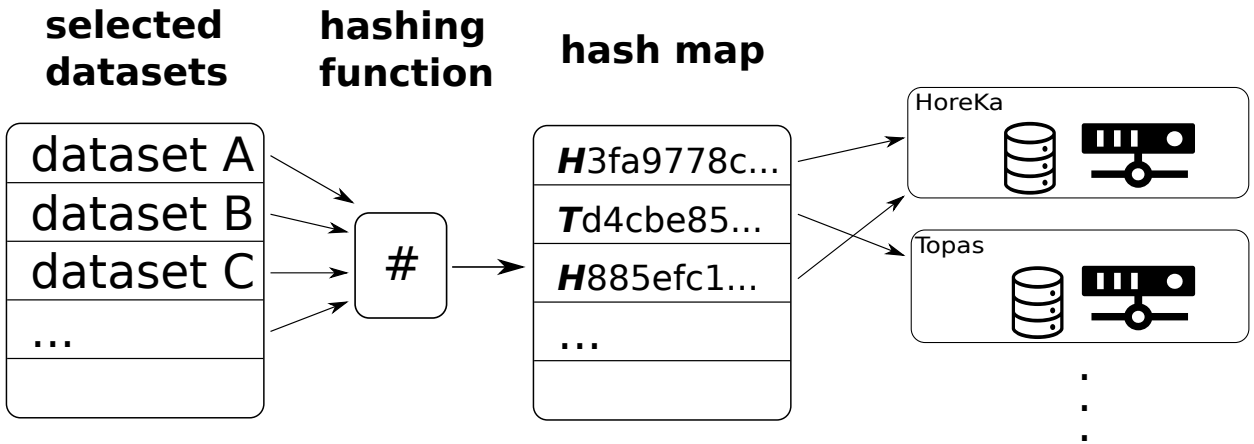
- Hash-based distribution of files inspired by Ceph/CRUSH

→ based on the file name, a hashing function calculates the data placement

Hash-Based Approach

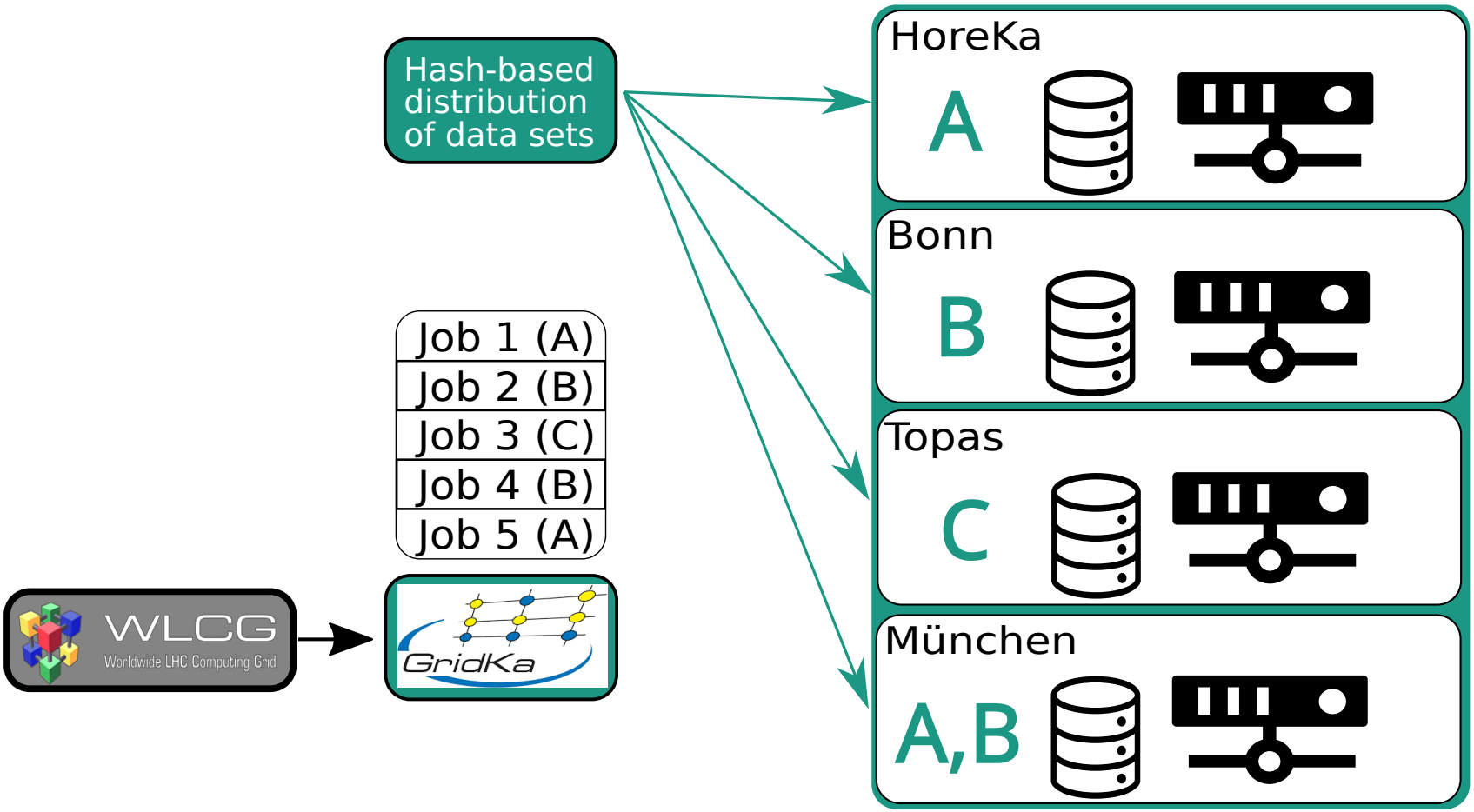
- Hash-based distribution of files inspired by Ceph/CRUSH

→ based on the file name, a hashing function calculates the data placement

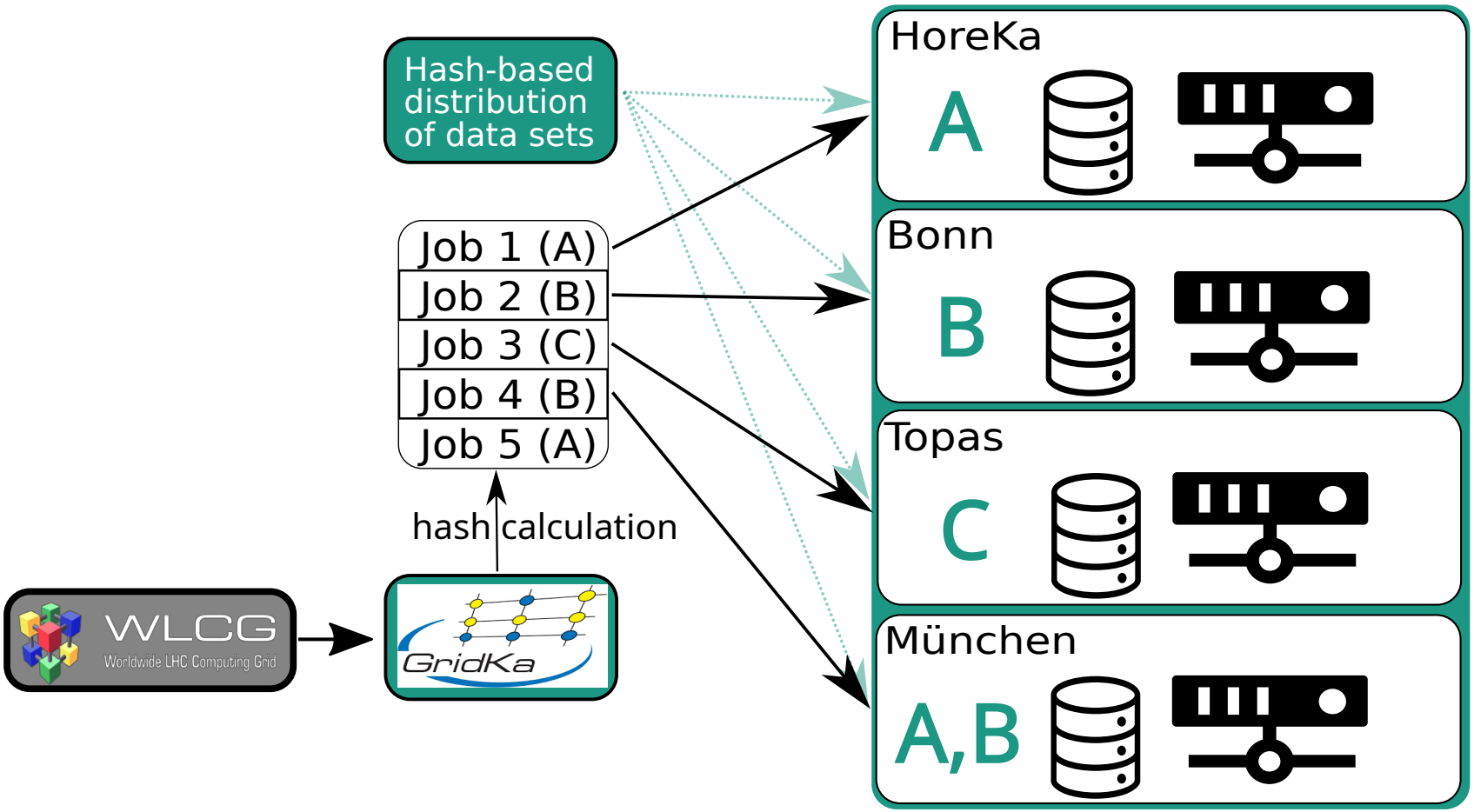


e.g.:
/SingleMuon/Run2022C-PromptNanoAODv10-v1/NANOAOD

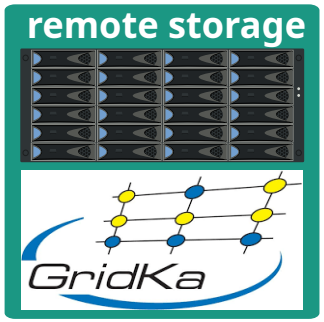
Hash-Based Distribution of Datasets



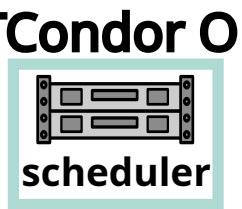
Hash-Based Distribution of Datasets

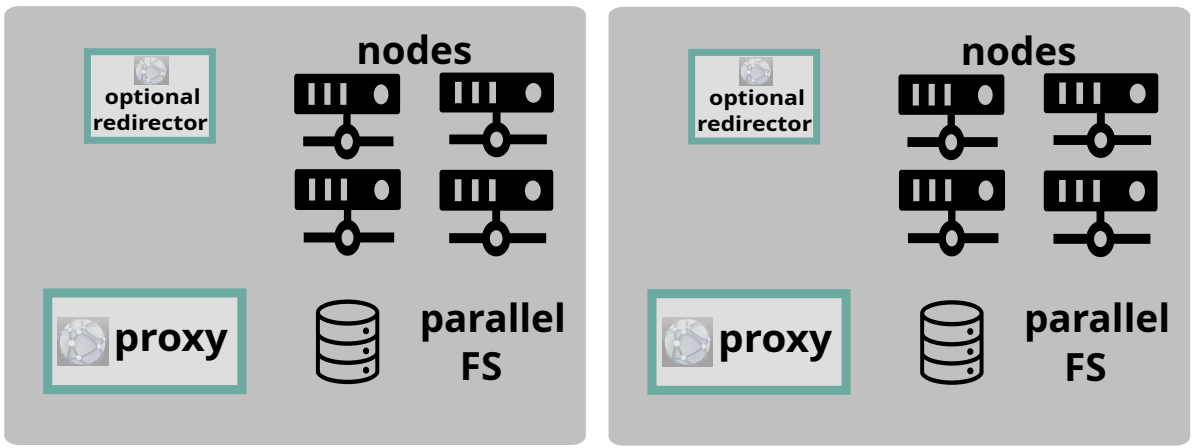


Concept and Implementation (I)

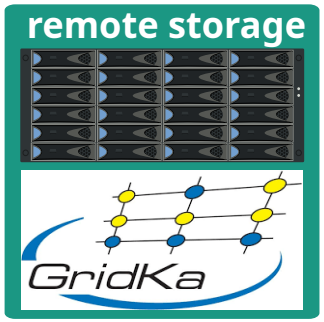


Remote storage at GridKa initially provides the data

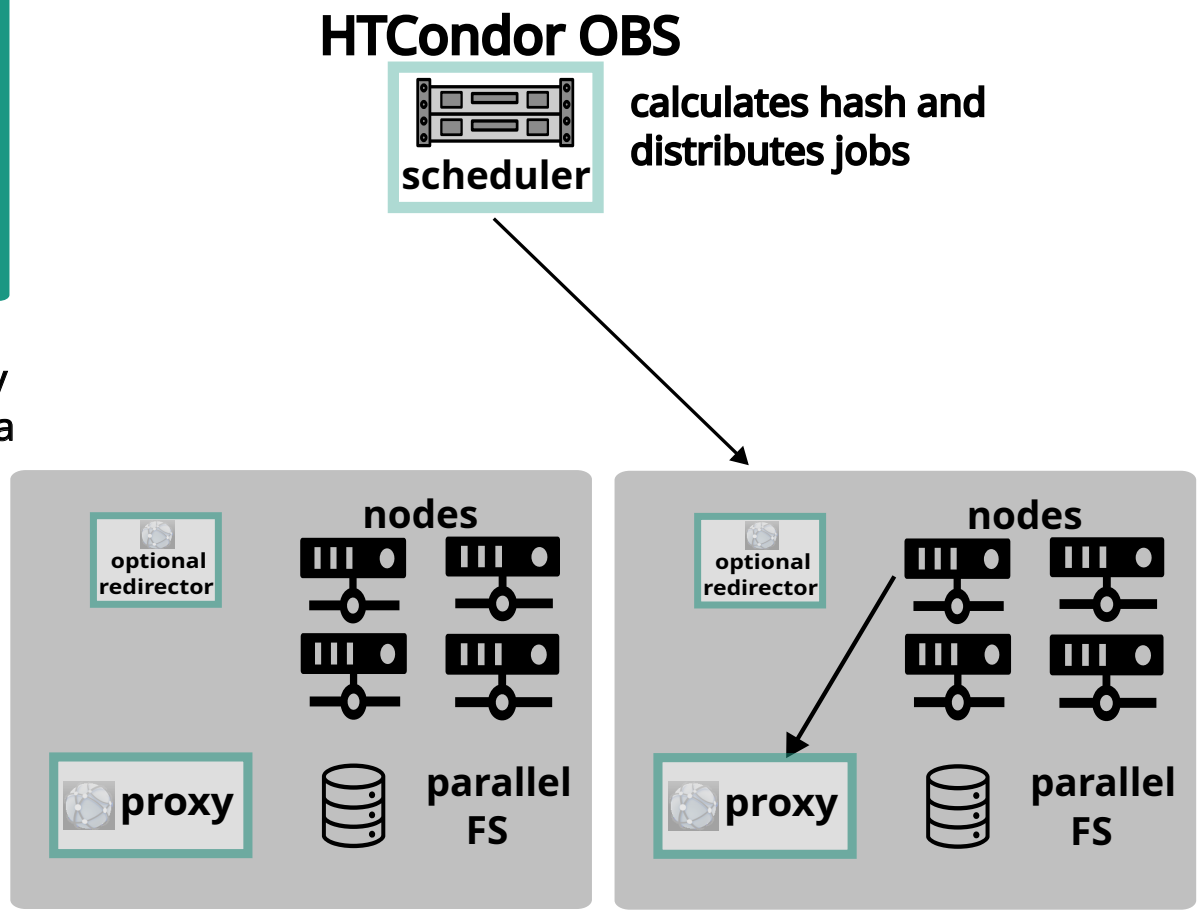




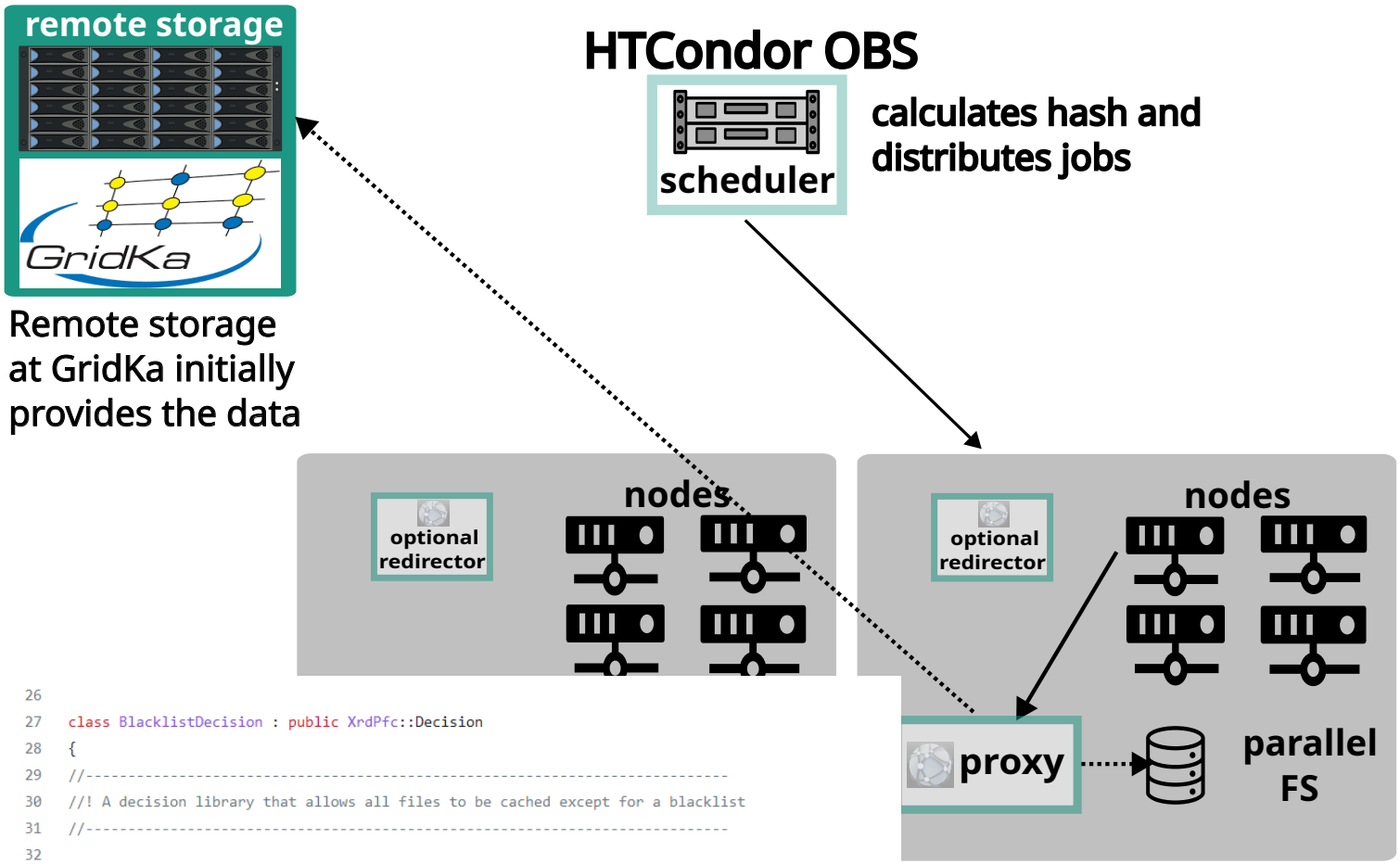
Concept and Implementation (I)



Remote storage at GridKa initially provides the data

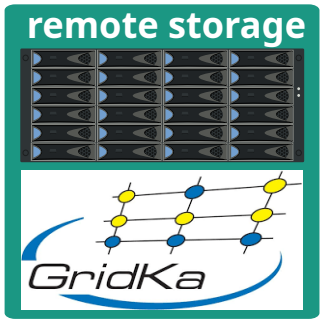


Concept and Implementation (I)

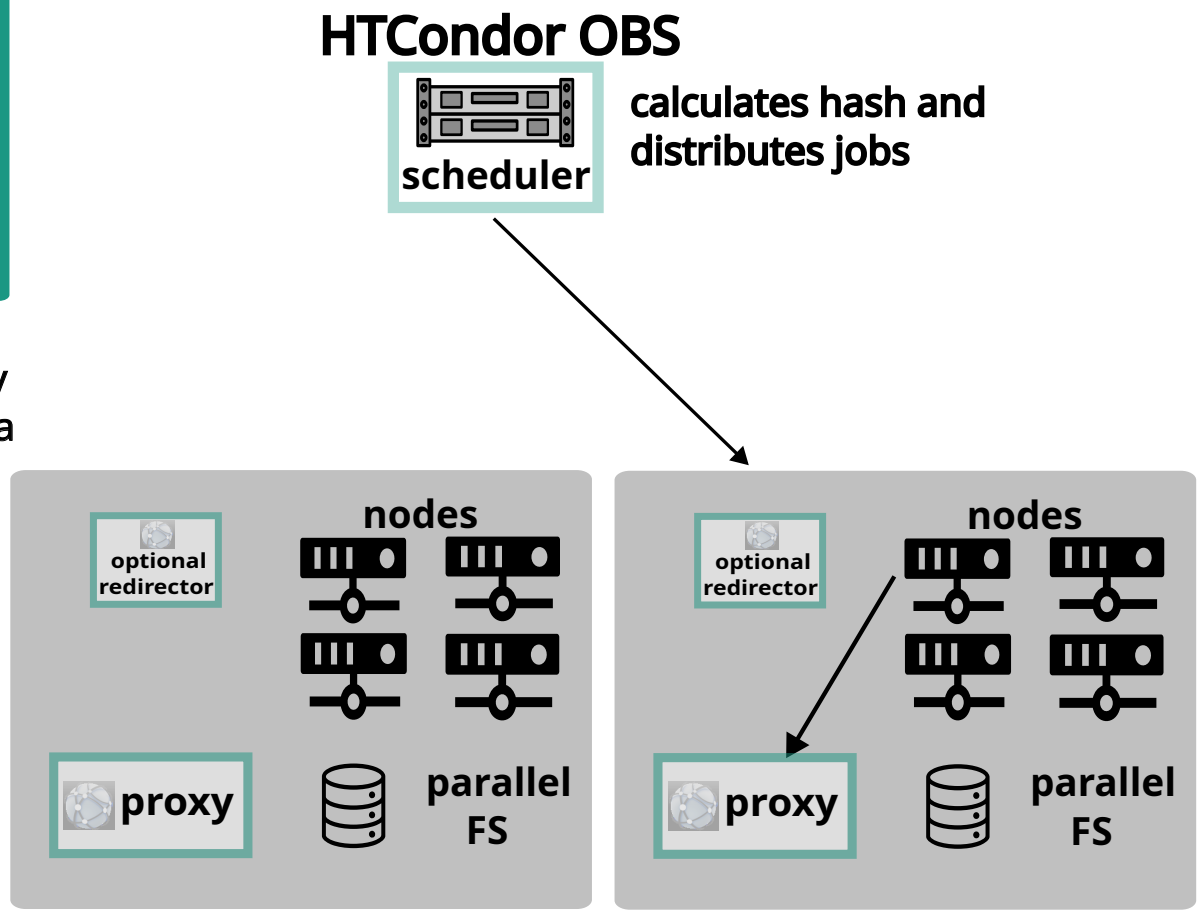



```
26
27  class BlacklistDecision : public XrdPfc::Decision
28  {
29  //------------------------------------------------------------------------
30  //! A decision library that allows all files to be cached except for a blacklist
31  //------------------------------------------------------------------------
32
```

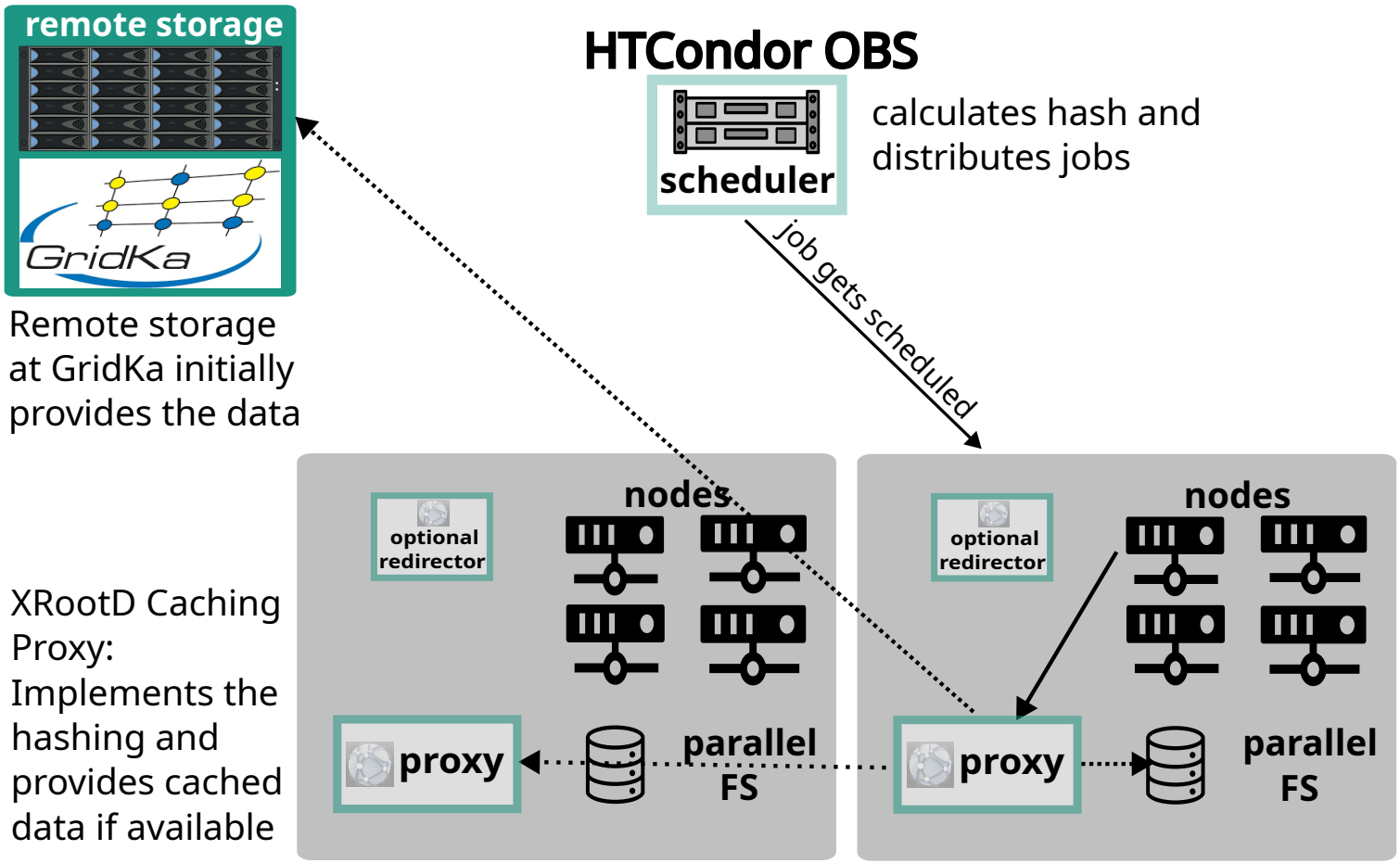
Concept and Implementation (II)



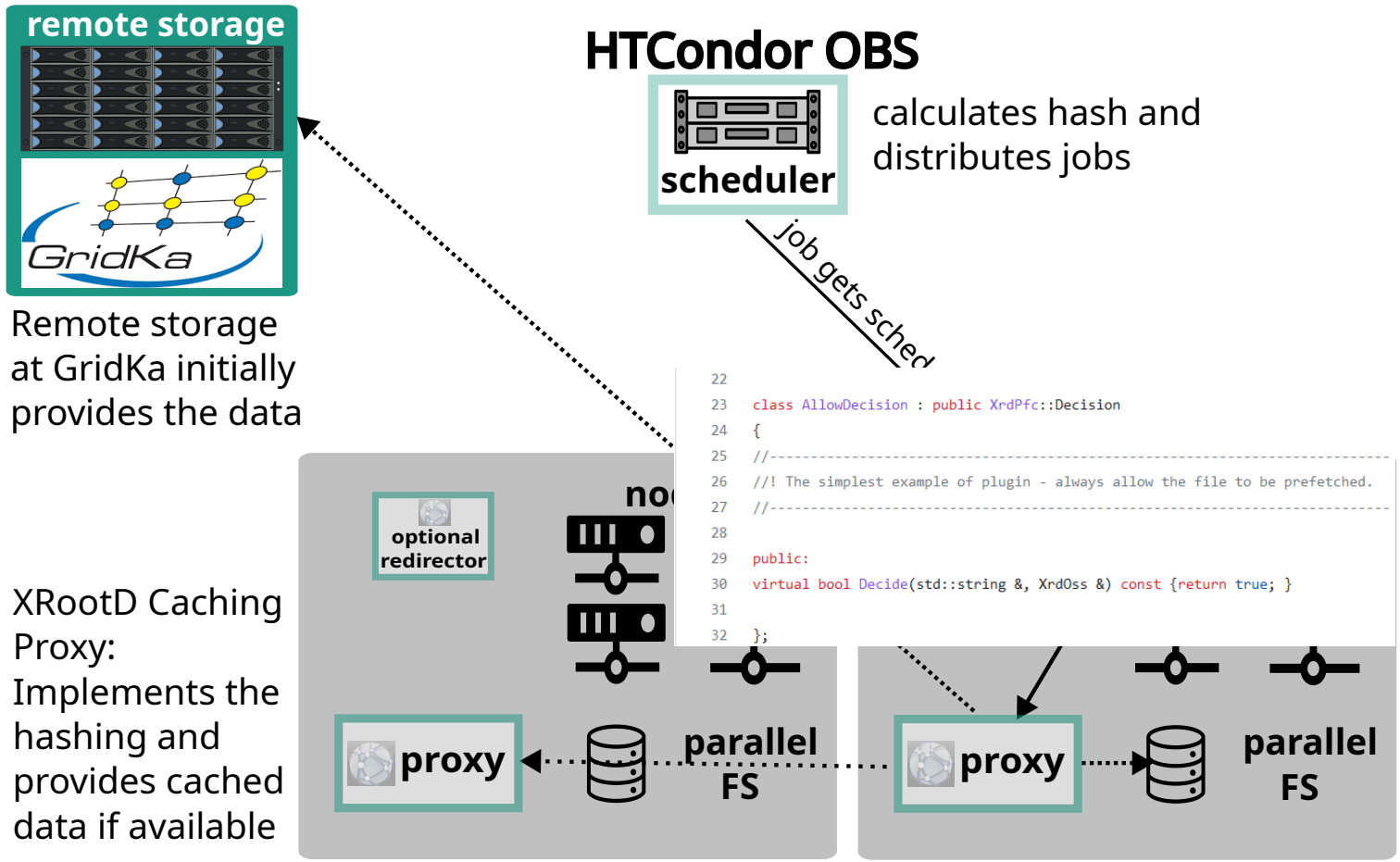
Remote storage at GridKa initially provides the data



Concept and Implementation (II)



Concept and Implementation (II)



Next Steps

- Proof of Concept:
  - Setup a test system at our institute
  - There, we also integrate some external resources (NEMO@Freiburg and our own T3) into our HTCondor batch system
  - First: fixed file lists.
  - Later: hash-based data distribution
- Finally: adapt system to the opportunistic resources

Summary

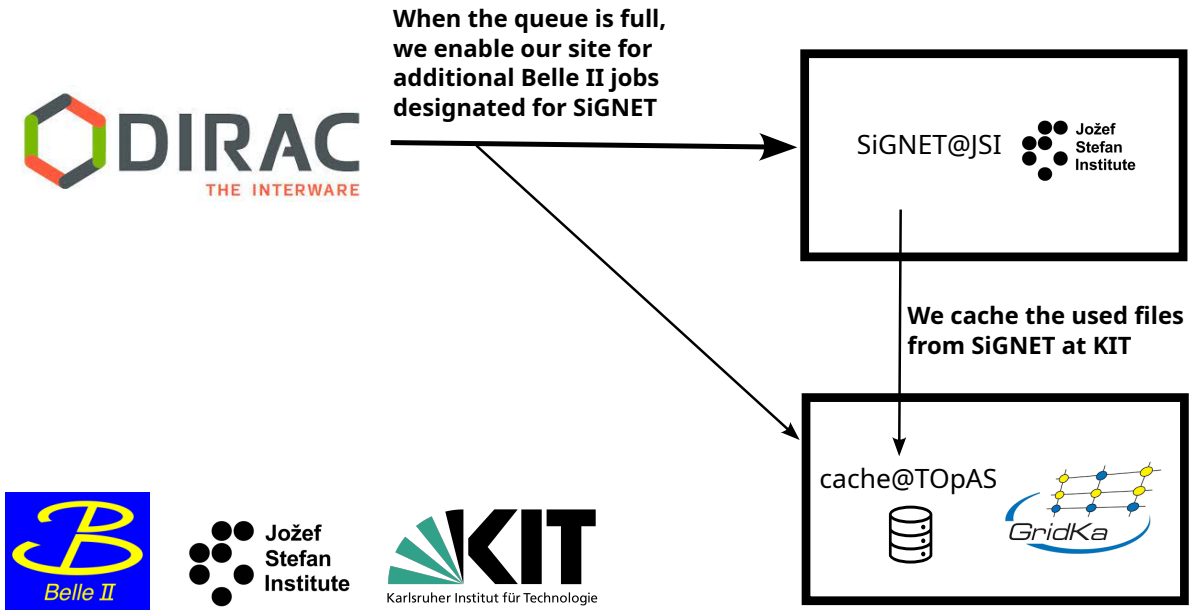
With our hash-based approach:

- We avoid actually keeping track of thousands of files on volatile caches
- The caching decision is determined locally (or centrally)
  $\rightarrow$ **horizontally scalable**
- With local on-the-fly caching, no prefetching is necessary (but possible)
- No database service needed (or at least more lightweight, tbd)
- Highly automatizable $\rightarrow$ **no active data management necessary**
- Allows inclusion of HPC sites without remote connection of WNs

$\rightarrow$ reduction of data transfers and more efficient utilization of resources achievable

Belle II Caching@GridKa

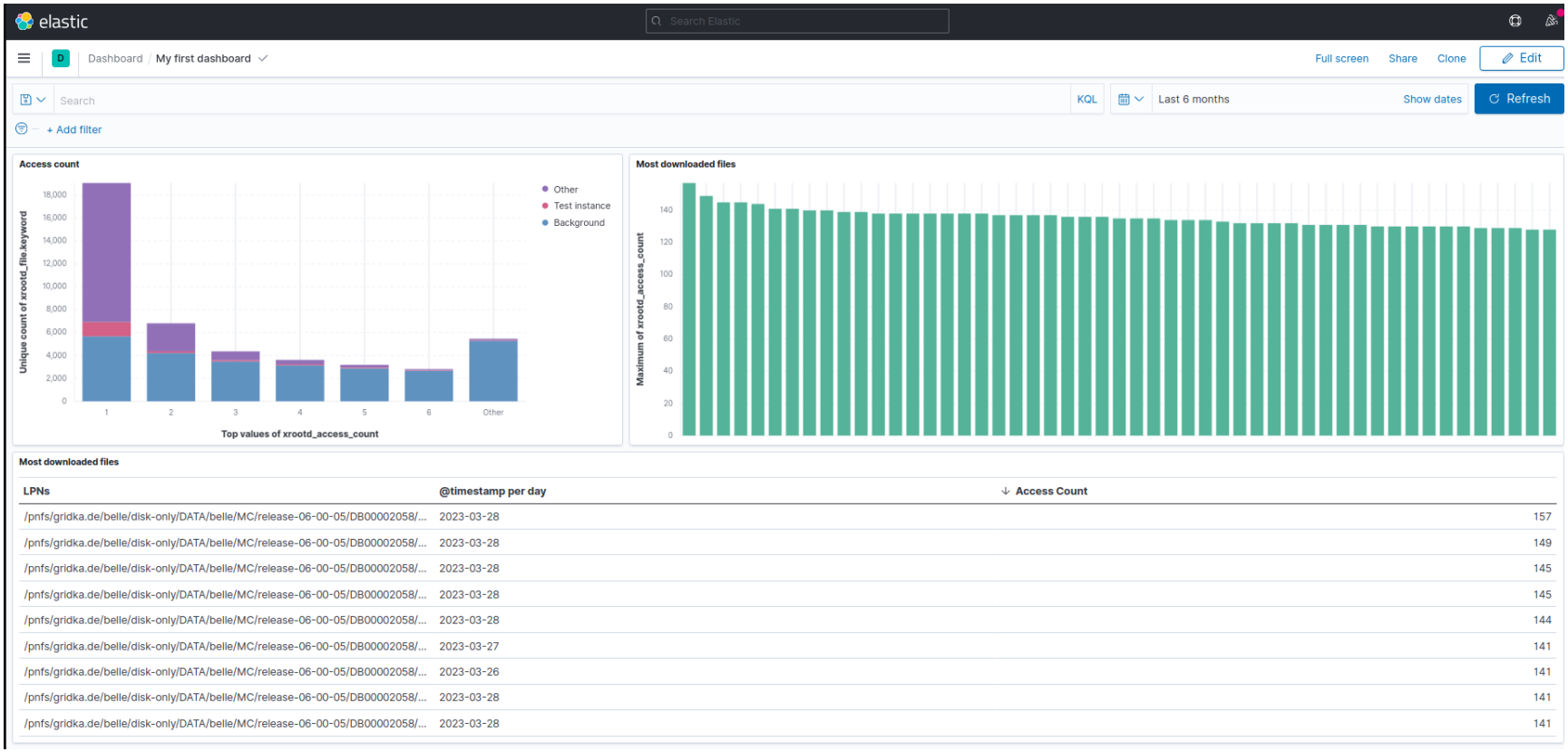
Currently, we have a XRootD caching project with SiGNET/JSI ongoing!
(I am not directly involved, credits go to: Moritz Bauer and Matthias Schnepf)



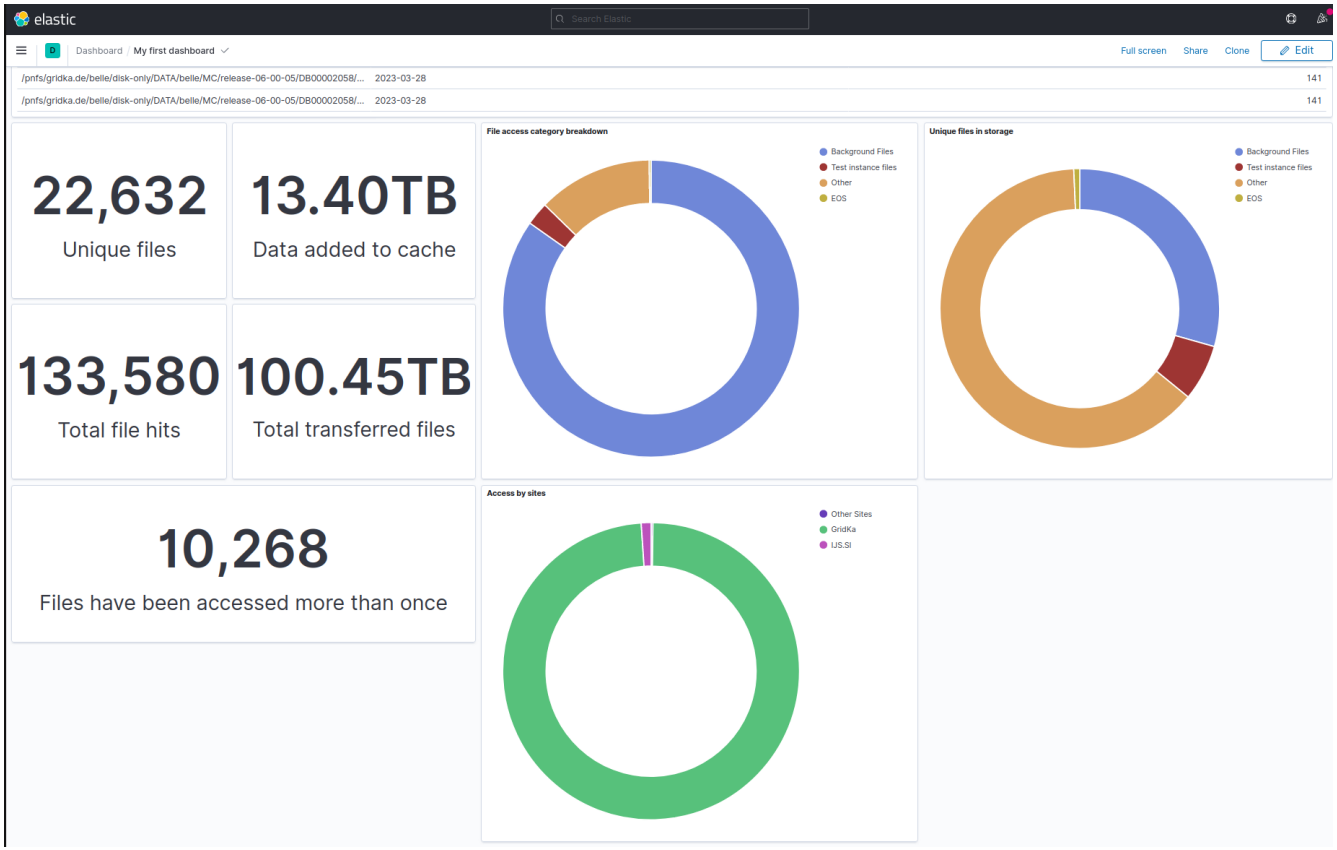
Belle II Caching@GridKa

```
xrd.port 1094
all.export /xroot:/ nostage
all.export /root:/ nostage
all.export /https:/ nostage
all.export /davs:/ nostage
all.export /http:/ nostage

xrootd.chksum max 4 adler32
ofs.osslib  libXrdPss.so
pss.cachelib default
ofs.ckslib * libXrdPss.so
pss.origin =
oss.localroot /ceph/belle2-xrootd-cache/
pfc.ram 32g
pfc.diskusage 0.9 0.95
pss.debug

xrootd.monitor all flush 10s window 5s fstat 5 lfn gbuff 8k dest fstat info user pfc 129.13.101.187:9921
xrootd.mongstream pfc use flush 10s

continue /etc/opt/belle2-xrootd/config.d/
~
~
~
~
~
~
~
~
"b2-xrootd-cache-server.cfg" 21L, 537C                                  1,1           All
```

Belle II Caching@GridKa



Belle II Caching@GridKa



XRootD Interactive

Karlsruhe Institute of Technology

Available at GitHub: xrd-interactive

- set of convenience functions to use xrootd interactively (as python questionary)
- Intended to make the usage more easy for non-powerusers (bachelor/master students etc)
- It is not really matured, nor generalized... (mainly adapted to GridKa)
- No plans on further development, just a small gimmick

XRootD Interactive