HEP DOMA Activities in Germany

A. Petzold

Disclaimer

- non-exhaustive list
- there a probably other projects I'm not aware of
- there are most likely other related activities outside of HEP computing

05.06.18   Andreas Petzold – WLCG DOMA Kickoff Meeting June 2018

SCC Steinbuch Centre for Computing

LMU

- ATLAS DE cloud
- Interest in using WN and site XCache, to compliment directio (Rod)
- LMU is part of a proposal for development activities focusing on data caching

SCC Steinbuch Centre for Computing

DESY

- DESY: multidisciplinary research center
  - photon, astro, HEP
  - important bridge to other communities

- dCache development center
  - great support for many Tier-1 and Tier-2 centers
  - history of engaging with communities outside of HEP

SCC Steinbuch Centre for Computing

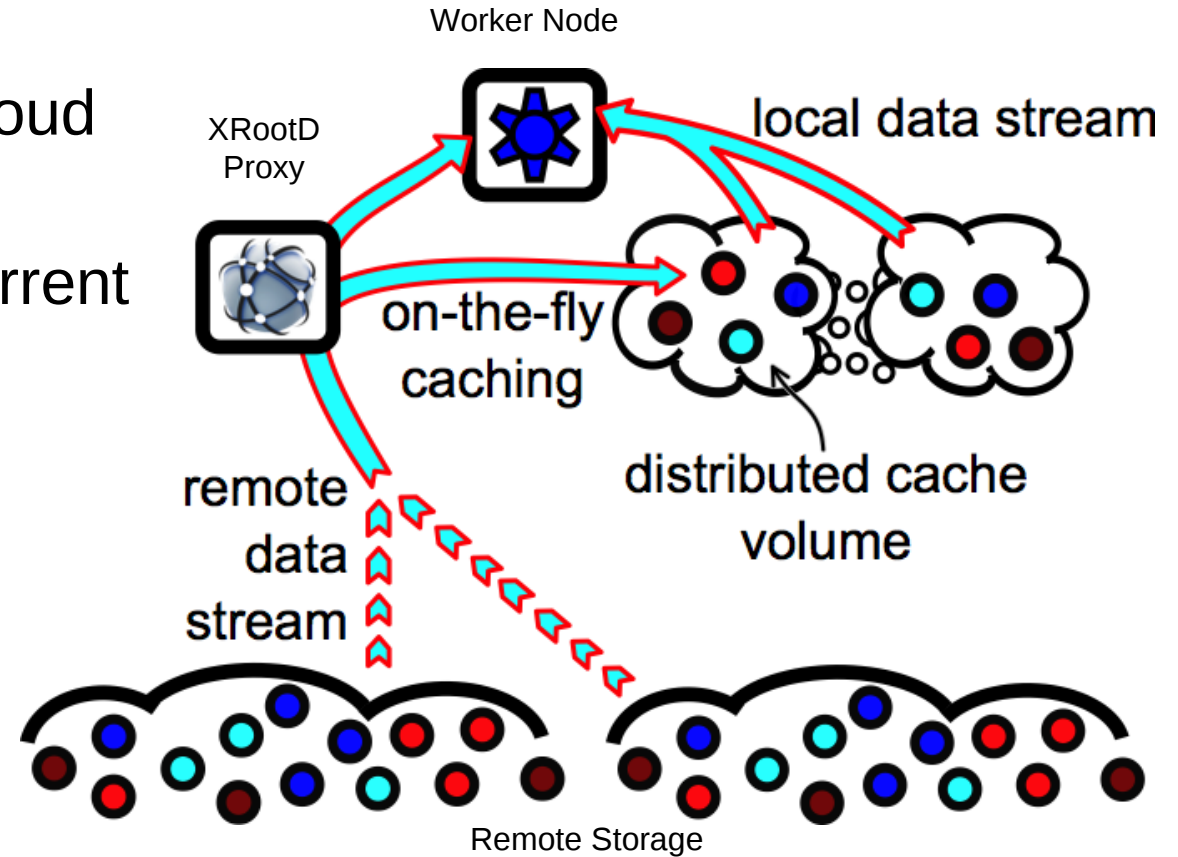# KIT – High Performance through Coordinated Caching

- Established new approach for **user data analysis** via distributed, coordinated caching on local SSDs

- Introduced data locality to HTCondor batch system

- Performance gain of factor 3-4 on typical recurrent end user data analysis payloads (prototype system)

- Third party funding for production system was recently granted (hardware)
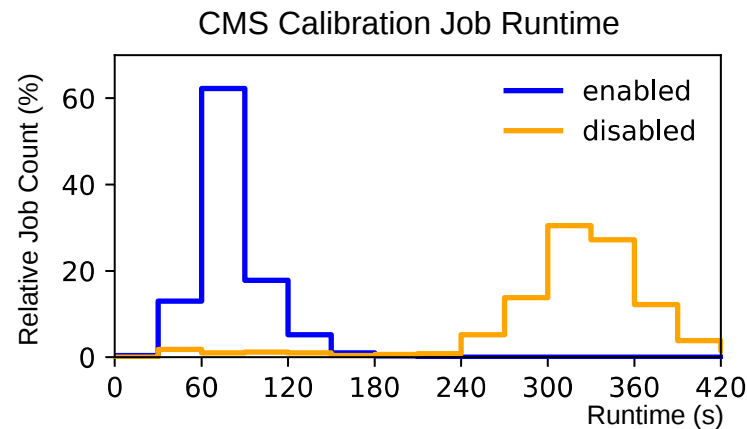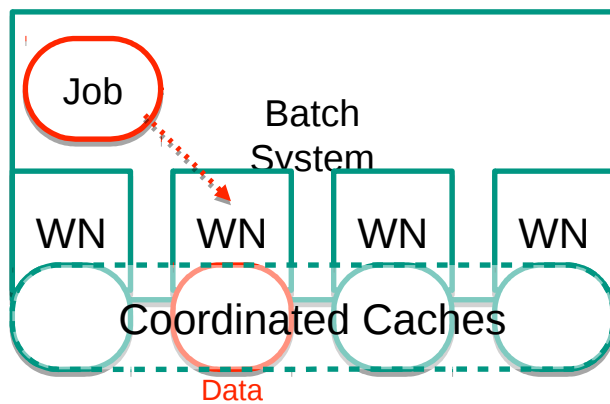
M. Fischer et al., Opportunistic Data Locality for End User Data Analysis, Journal of Physics **898**, 5 (2017)

SCC Steinbuch Centre for Computing

KIT – Caching Concepts on Opportunistic Sites

- Opportunistic resources usually well-suited for CPU-intense workflows

- Many opportunistic sites offering fast cloud storage or distributed storage

- Benefit from caching R&D to bring recurrent I/O-intense workflows to the cloud

Worker Node

XRootD Proxy

local data stream

on-the-fly caching

distributed cache volume

remote data stream

Remote Storage

ACAT 2017 C. Heidecker et al., Opportunistic Data Locality for HEP Analysis Workflows, to be published (2017)

SCC   Steinbuch Centre for Computing

KIT – Tier-1 GridKa

- no switch from dCache/ALICE xrootd planned
  - protocol zoo well covered
- Spectrum Scale serves as software defined storage layer
  - covers foreseeable scaling requirements w.r.t. capacity and performance
  - tiering with object storage might an interesting option
- fast WAN connections and corresponding strong LAN backbone
  - whatever will happen, network requirements will increase significantly!
  - ready to adapt available technology/bandwidth as soon as financially viable
  - participation in SDN trials
- Migration to HPSS Tape backend planned
  - fully exploit scalability and cost efficiency of tape

SCC Steinbuch Centre for Computing

'Data Lake' Questions

- Will we be able to 'plug in' with our existing storage technologies?

- Cost savings?
  - Do we expect to save on operations? Where? (Cost Modeling WG?)

- Storage at smaller sites?
  - NDGF model vs. cache-only vs. ???

SCC Steinbuch Centre for Computing