T2K and HK Computing

12th December 2019

Sophie King

JENNIFER 2
EU grant n.822070

Intro

This is a very brief overview of computing/software in the T2K and HK experiments

It is probably worth noting/explaining…

Currently **T2K and HK are seperate experiments**
→ being a member of one, does not grant you access to the other

Resources are treated **seperately**
Authorisation / access treated **seperately**

But worth noting that **T2K resources will (roughly speaking) transition into HK** resources when HK begins operation

… so the seperation is probably a bit blured in this talk

Computing resources

CPU

- GRID
- Kamioka cluster (Far detector site in Japan)
- KEK cluster (Near detector site in Japan)
- Canada compute (Cedar cluster)

Storage

- GRID
- Kamioka cluster
- KEK cluser
- iRODS hosted in Japan
- iRODS hosted in UK (soon to be Nextcloud or ownCloud)

Computing resources

Grid for T2K/ND280

→ Rely heavily on GridPP (UK)

→ Additional Grid storage at SFU (Canada)

→ Introducing storage and CPU at IN2P3 (France)

Grid for HK

→ GridPP (Need to look into IRIS funding and other options)

→ Plan to ask for more international Grid resources from participating countries
(To be discussed by HK International baord in January)

When HK begins operation, T2K Grid resources will essentially become HK resources

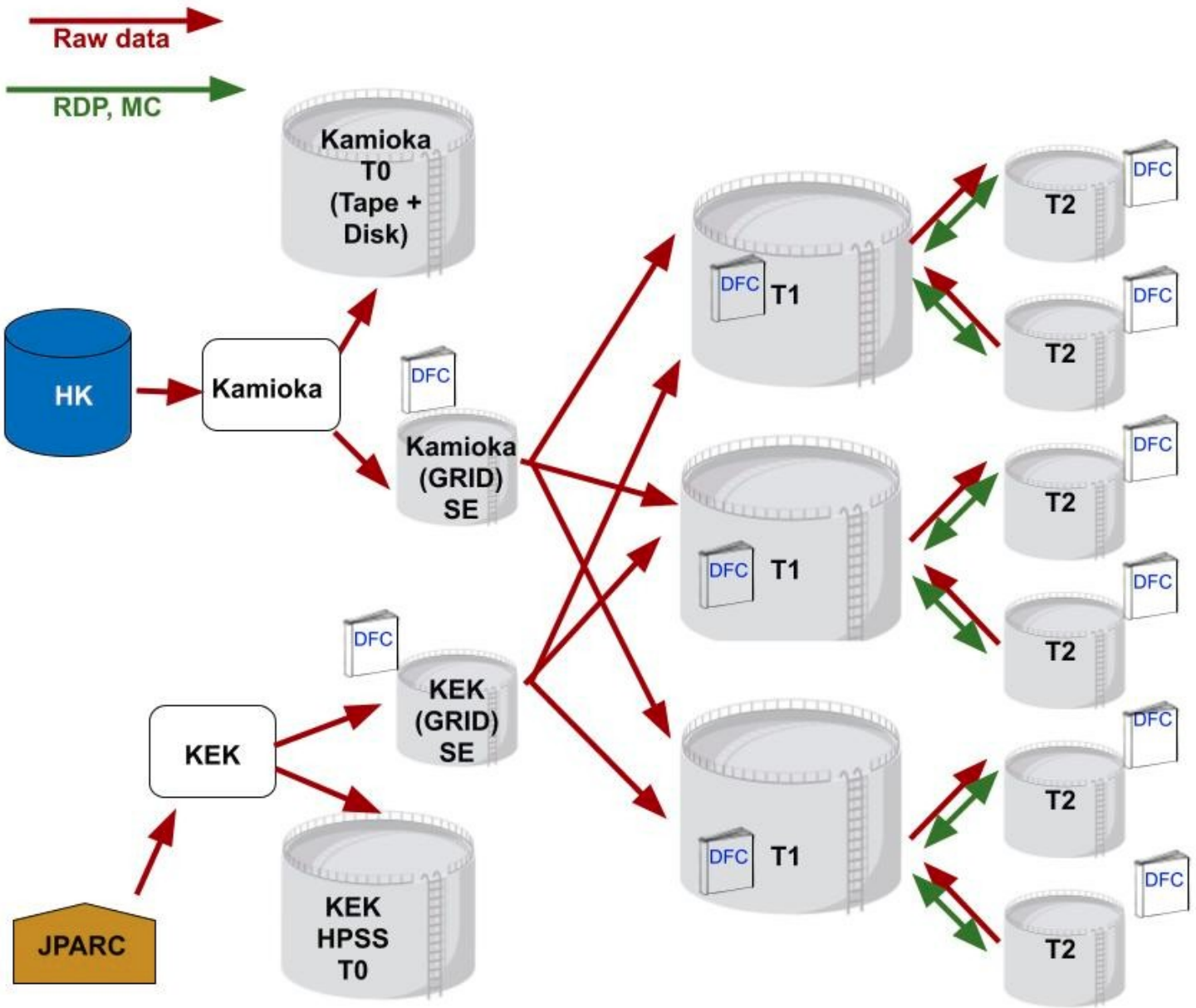
HK will be too large to live off LHC computing resource scraps!

→ **Need to look at different options available for computing resources**

- understand how we can best utlise these to bring our various resources together (e.g. through DIRAC)
- how to manage them
- how to fund them

This project can aid our decisions/direction

Computing resources



Standard LHC style tiered Grid usage

For T2K, only ND280 uses the GRID
→ **KEK T0**

For HK, plan to sync HK data from Kamioka to our T1s
→ **Kamioka T0**

Storage

	Construction (MC, 7y)	
Detector	CPU hrs	Storage (TB) (One copy of each file)
Hyper-K (FD)	20M	500
INGrid	0.13M	7
ND280	19.2M	2250
IWCD	97M	52
Flux	KEKCC	15
Total	133.43M	2,824 (~2.8 PB)

Predicted resouces at the end of the 7year HK construction period

	Data Taking (Data, 10y)	Data Taking (MC, 10y)	
Detector	Data Storage (TB) (One copy of each file)	MC CPU hrs	MC Storage (TB) (One copy of each file)
Hyper-K (FD)	18,300 + 140 reduced	25M	500
INGrid	226	0.51M	26
ND280	669	42.2M	4,950
IWCD	620	684M	367
Flux	N/A	KEKCC	15
Total	19,815 (~19.9 PB)	751.71M	5,858 (~5.9 PB)

Predicted resouces after 10 years of HK operation

Note: This does not account for replication. Need to multiple by ~2.5
Note: To normalise CPU to HEPSPEC06 times by ~10

Software management

CVS and CMT (previouly the standard for T2K)

GIT and CMAKE (large parts of T2K recently transitioning) – A. Finch
→ T2K GITLAB hosted in Poland

CVMFS
- hosted by RAL
- basic use for Grid production jobs
- potential to be used for container images

Containers
- ‘unofficial’ use of Docker and Singularity used in both T2K and HK
- work underway to officialise/standardise the use of singularity - M. Guigue

Continuous integration
- some T2K and HK packages have had CI in the past
  - but no official/standard method
- work ongoing to introduce continuous integration with GITLAB - M. Guigue

DIRAC

T2K and HK both currently use the DIRAC service hosted at Imperial

→ https://dirac.gridpp.ac.uk:8443/DIRAC/

→ **Multi VO DIRAC service**

→ Not managed by T2K/HK, but Imperial staff (D. Bauer, S. Fayer) always very helpful

→ Currently use standard DIRAC client software (No T2K/HK modifications)

→ Cloud resources linked to this DIRAC service (though not T2K/HK resources)

→ Should be fine to integrate some T2K/HK cloud resources into this service (at least for purpose of demonstrator)

DIRAC usage pretty basic

→ Job submission (T2K and HK software to automate JDL writing and job submission)

→ DFC (hosted at imperial)

Cloud

Neither T2K nor HK currently have experience using cloud resources or access to cloud resources

Options

* Cloud at Imperial, but need to access via IRIS funding
  (I will investigate this posibility, but will not be a quick turn around)

* Pay-to-use e.g. Amazon

* Other options for shortish term? (i.e. 2-3 years)

* See Silvio's slides in the final session for more options

Quick fix temporary options to get things started
- sneakily use some of the cloud resources at Imperial for tests
- see if we can use a small amount of Belle II cloud resources

Cloud

Neither T2K nor HK currently have experience using cloud resources or access to cloud resources

For this demonstrator we need to decide what to do about the T2K / HK split

a) work on something accessilble to both
- → multi VO setup
  - → as long as we are only sharing CPU and not storage, this shouldn't (?) cause political problems
- → I would guess this wouldn't be much more effort (??)
  (I guess this is more of a question to the experts)

b) chose to limit to T2K only or HK only
(Note: only for the cloud test, this project in general is both T2K and HK)
- → T2K more in need of resources in the short term.. might be more helpful
- → though depends on the size of the test, and if this line of work is continued as to whether this would actually benefit T2K

Summary

T2K computing relies mainly on GridPP and follows standard Tierd Grid Model
Non-grid resources are independent and files are copied to the grid

T2K software management being revamped
- GITLAB
- CMAKE
- Containers

HK needs to look to the future
- bigger storage needs
- new technology

Need to find cloud resources for demonstrator

Consider other options of collaboration (sharing tools, ideas etc.)