

Manchester/IRIS report

A. Forti

SKA

10 September 2019



IRIS

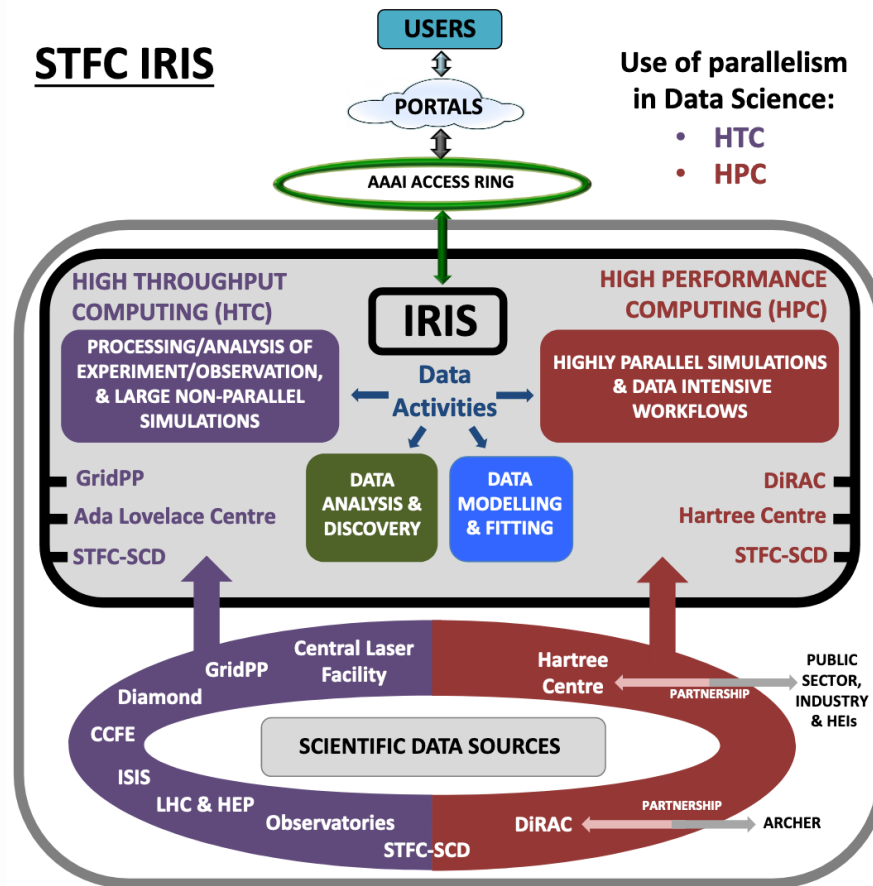
- Any STFC science activity needing HTC/Storage/Tape compute hardware resources
- IRIS
 - is a coordinating body
 - It does not have any resource for staff for operations, software support or user support
 - Though there is some money to develop some digital assets
 - It doesn't own the resources
 - Though provides the necessary coordination for various sciences to access them
 - Development is done by the science groups according to what their computing models
- GridPP (part of WLCG) is an IRIS partner



IRIS



STFC IRIS



Use of parallelism in Data Science:

- HTC
- HPC

All Data Activities contain elements of:

- Collection/generation
- Simulation
- Analysis/discovery
- Modelling/fitting

Processing/Analysis of Experiment/Observation, & Large Non-parallel Simulations
Bespoke systems that are directly connected to experiments, observatories & National Facilities

Highly Parallel Simulations & Data Intensive Workflows
Systems which generate data from a set of starting assumptions/equations

Data Analysis & Discovery
Systems analyze data to discover relationships, structures and meaning within, and between, datasets

Data Modelling & Fitting
Systems use a combination of simulations, experimental data and statistical techniques to test theories and estimate parameters from data

Manchester resources

- Manchester is a medium-large EGI/WLCG T2
 - ~7500 job slots
 - 3000 (IRIS)
- 7.3 PB of storage
 - 1 PB (IRIS)
- Mainly dedicated to LHC experiments (ATLAS/LHCb)
 - But as an EGI site has always supported also smaller groups
 - ~28 other groups have accessed the resources
 - Groups increasing in size and requirements
 - DUNE, SKA and LSST among the bigger new customers

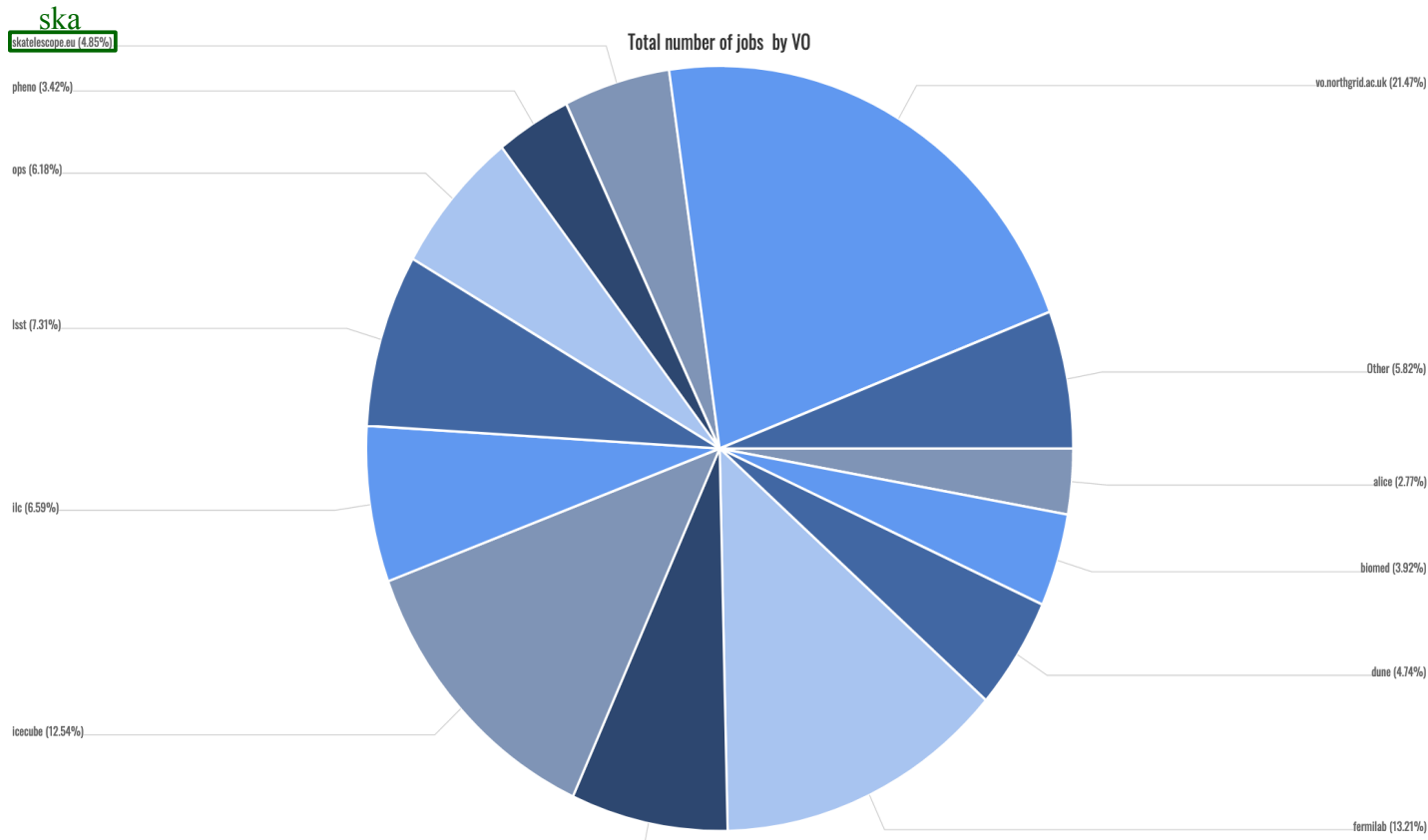


Distributed computing

- Most resources accessible via grid middleware
 - UK has a community Dirac WMS to support UK resources
 - LSST-UK, SKA jobs submitted via Dirac
- Openstack was requested by other IRIS partners
 - We now have a small openstack testbed
 - We also have 10% of resources used by VAC/Vcycle
 - VAC is a daemon that spawns Vms on a local hypervisor
 - Vcycle is a small agent that can instantiate VMs on various cloud resources
 - OpenStack, EC2, Google Cloud...
 - Dirac can also submit to VAC and Vcycle



Manchester usage



VO
None
alice
atlas
bes
biomed
cernatschool.org
comet.j-parc.jp
dune
fermilab
gridpp
hyperk.org
icecube
ilc
iris
iris.ac.uk
lhcb
lsst
lz
magrid
mice
na62.vo.gridpp.ac.uk
ops
pheno
skatelescope.eu
snoplus.snolab.ca
solidexperiment.org
t2k.org
vo.moedal.org
vo.northgrid.ac.uk
vo.scotgrid.ac.uk



SKA in Manchester

- SKA has different workloads some of which are suited for HTC
 - We also host some of their 1.5TB machines
 - Brokering from community Dirac still in development
- All jobs to the IRIS/GridPP resources are containerized
 - Singularity runtime
- Deployment model very similar to ATLAS
 - Containers
 - Multicore jobs
 - Setup also a CVMFS repository at RAL



SKA in Manchester

- Most jobs submitted to SARA (NL) until few months ago because of large memory
- Recently more optimized workflows run well in Manchester/UK
 - Ramp up in July and August

