

SWIFT-HEP

logo placeholder

[SWIFT-HEP/ExcaliburHep
workshop](#)

2-3rd November 2021

WP5: Analysis

Analysis Grand Challenges and
collaboration venues with IRIS-HEP

Eduardo Rodrigues & Luke Kreczko



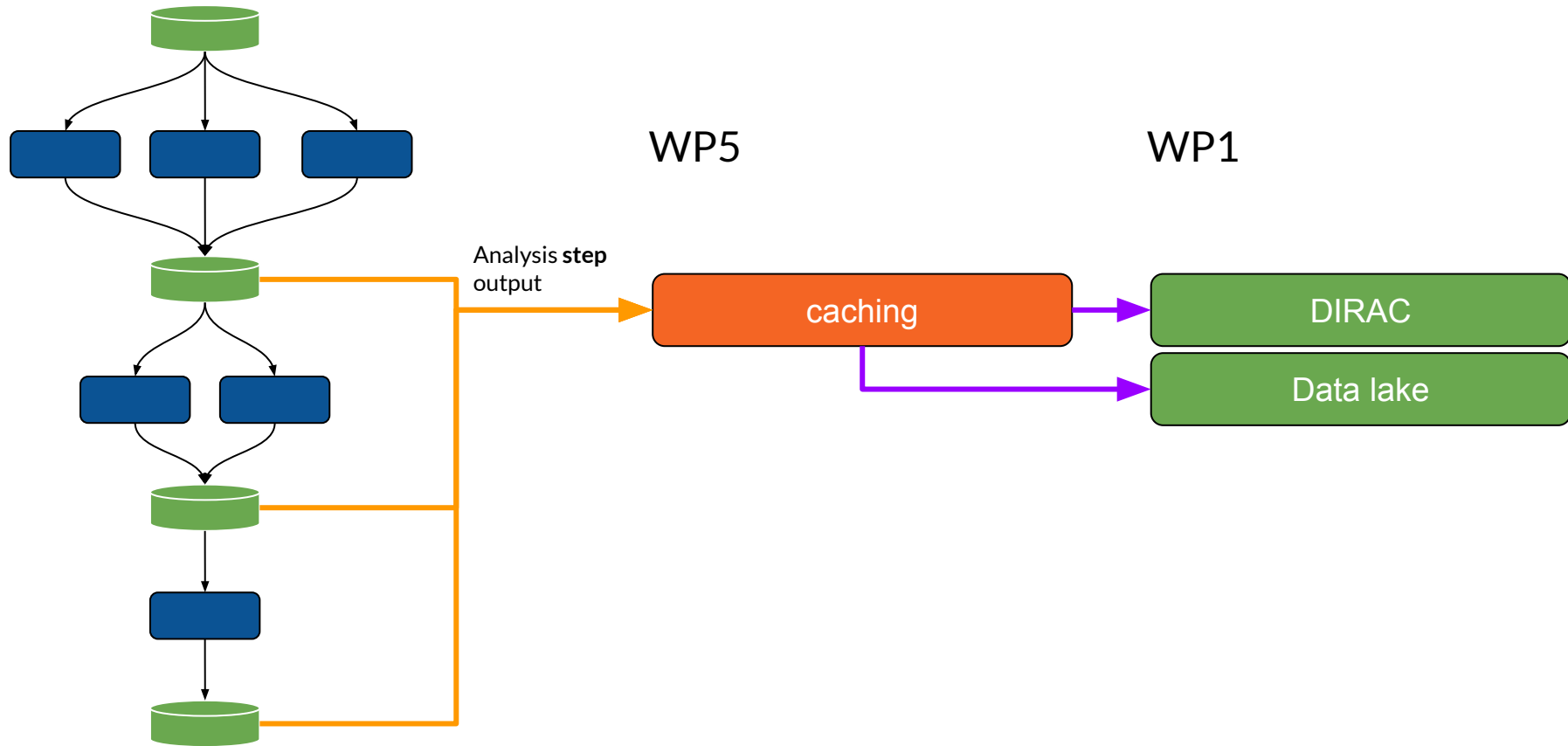
UNIVERSITY OF
LIVERPOOL



University of
BRISTOL

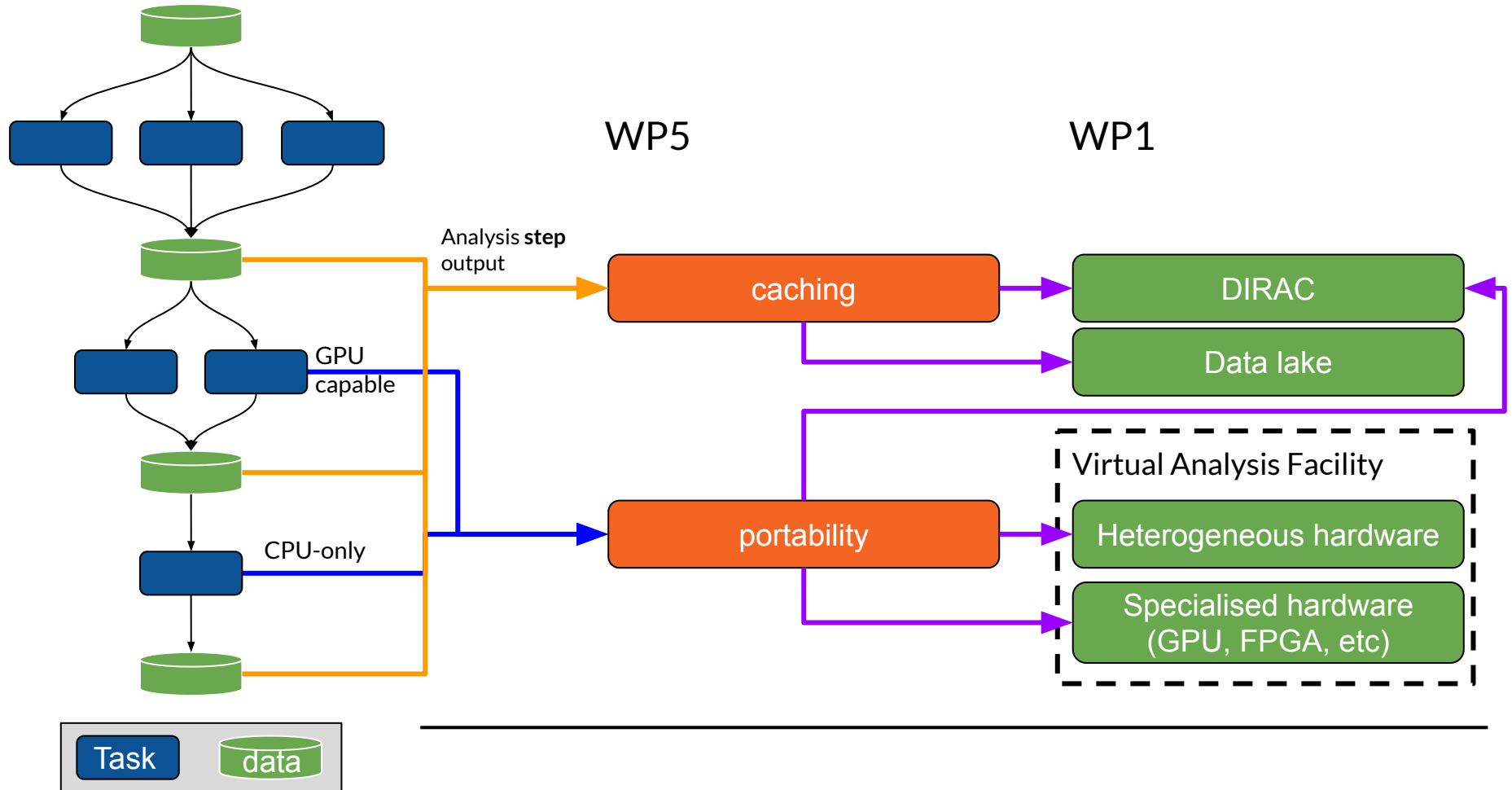
Analysis workflow

Reminder WP5: Analysis



Analysis workflow

Reminder WP5: Analysis



IRIS-HEP

Analysis Grand Challenges

[AGCs]

(incl. ATLAS, CMS and WLCG)

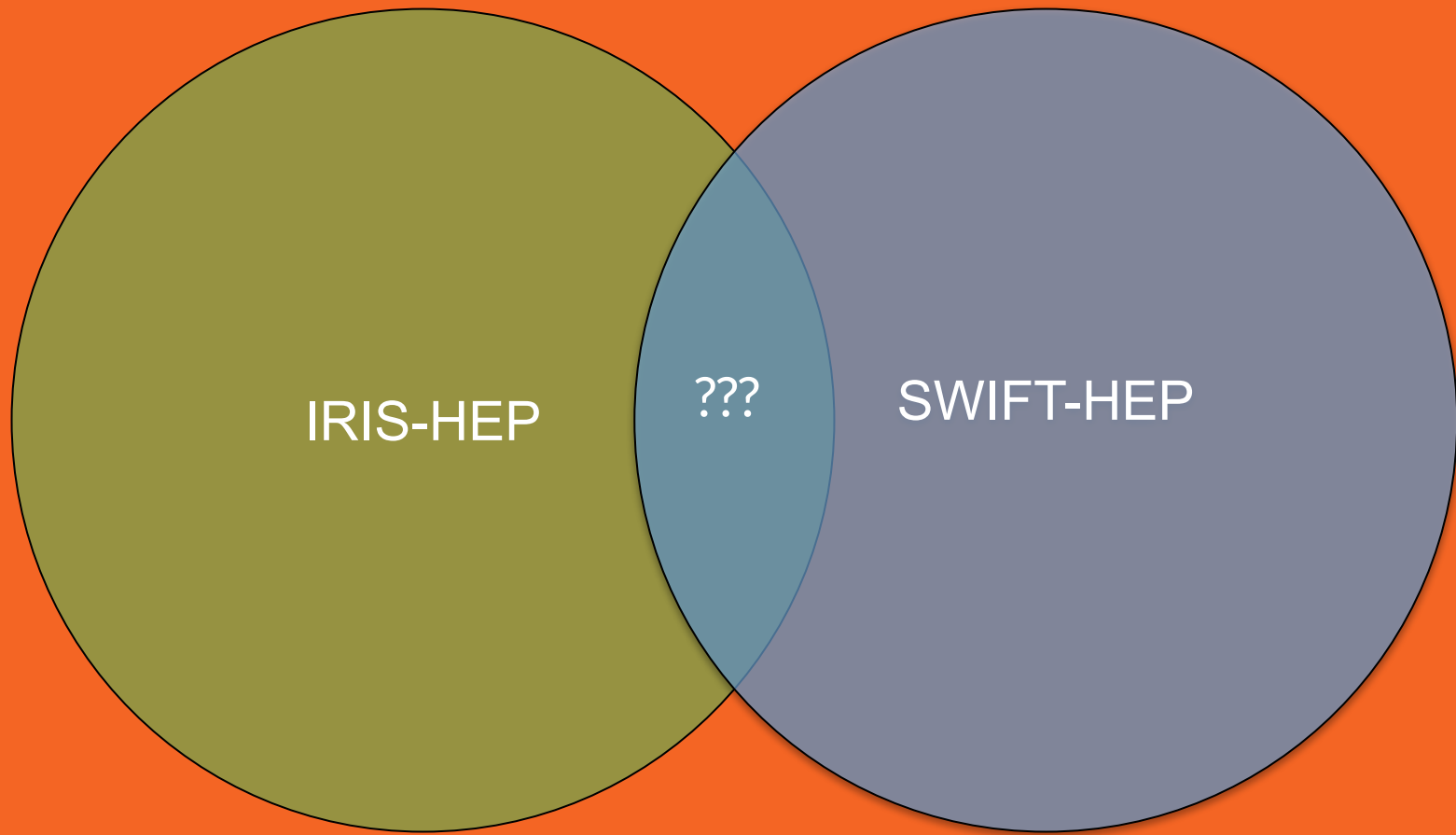
[Related IRIS-HEP workshop starting tomorrow](#)

Multiple challenges in the years
2021, 2023, 2025, 2027

Analysis: Demonstrate analysis
system can cope with increased data
volume while delivering enhanced
functionality**

Data volume: realistically sized
HL-LHC end-user analysis dataset (~
200 TB)

Reproducibility and
Reinterpretation



IRIS-HEP/SWIFT -HEP intersection

IRIS-HEP → SWIFT-HEP

Part of an international HEP
community

We have benefitted from [IRIS-HEP](#) +
[Scikit-HEP](#) developments

- we are users and contributors of awkward, uproot, hist, mplhep, etc
- Tested coffea (FermiLab + IRIS-HEP) as a backend

How can we benefit from AGCs?

- Do the proposed tests and benchmarks make sense for us?

IRIS-HEP/SWIFT -HEP intersection

SWIFT-HEP → IRIS-HEP

Are IRIS-HEP thinking about caching solutions?

- If yes, can we extend to our use-case?
- If no, would that be something of interest?

Adding DIRAC as backend

- Benefits more than IRIS-HEP

How can we contribute AGCs?

- Extend to non-LHC experiments (e.g. DUNE, LZ)

Analysis Grand Challenge Timeline

	Entity	Scope	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
Infrastructure	LHC	Global	Run-2	LS2			Run-3			LS3		Run-4	
	WLCG	Global	Global coordination of requirements, resources, policies, networking, security, etc.										
	GridPP	UK	GridPP5			GridPP6			?				
	IRIS-UK	UK	UKTO	IRIS 4yr x £4m				Support of non-LHC STFC communities?					
Experiments	ATLAS-CMS	Global	S&C Conceptual Design			S&C Technical Design			S&C deployment		Operation		
	LHCb	Global	S&C TDR	S&C deployment			Operation and Upgrade 2 preparation						
	DUNE	Global	Protodune		S&C CDR	ProtoDUNE Comp model	DUNE implementation and deployment			Operation			
	Others	Global	Experiments common software infrastructure design and development (neutrino, dark matter, etc)										
Software	HSF	Global	HEP Software Forum: White Paper --> Working Groups --> Community Meetings -->										
	IRIS-HEP	USA	S2I2	IRIS-HEP: 5yr x 5m USD					?				
	ECHEP	UK	£50k			ECHEP							
	Excalibur	UK	£240k		Excalibur		?						
	HSUK	UK					SWIFTHEP-1: 3 x £400k			SWIFTHEP-2: n x £2m?			

Fig. 2: In the proposal

The next challenge is planned for 2023

That's ~ two years in for WP1 and one year in for WP5 → perfect time to participate

Analysis Grand Challenge Timeline

	Entity	Scope	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028
Infrastructure	LHC	Global	Run-2	LS2			Run-3			LS3		Run-4	
	WLCG	Global	Global coordination of requirements, resources, policies, networking, security, etc.										
	GridPP	UK	GridPP5			GridPP6			?				
	IRIS-UK	UK	UKTO	IRIS 4yr x £4m				Support of non-LHC STFC communities?					
Experiments	ATLAS-CMS	Global	S&C Conceptual Design			S&C Technical Design			S&C deployment		Operation		
	LHCb	Global	S&C TDR	S&C deployment			Operation and Upgrade 2 preparation						
	DUNE	Global	Protodune		S&C CDR	ProtoDUNE Como model	DUNE implementation and deployment				Operation		

Fig. 2: In the proposal

Outcomes would naturally feed into
SWIFT-HEP extension proposal

	HSUK	UK						SWIFTHEP-1: 3 x £400k	SWIFTHEP-2: n x £2m?				
--	------	----	--	--	--	--	--	-----------------------	----------------------	--	--	--	--

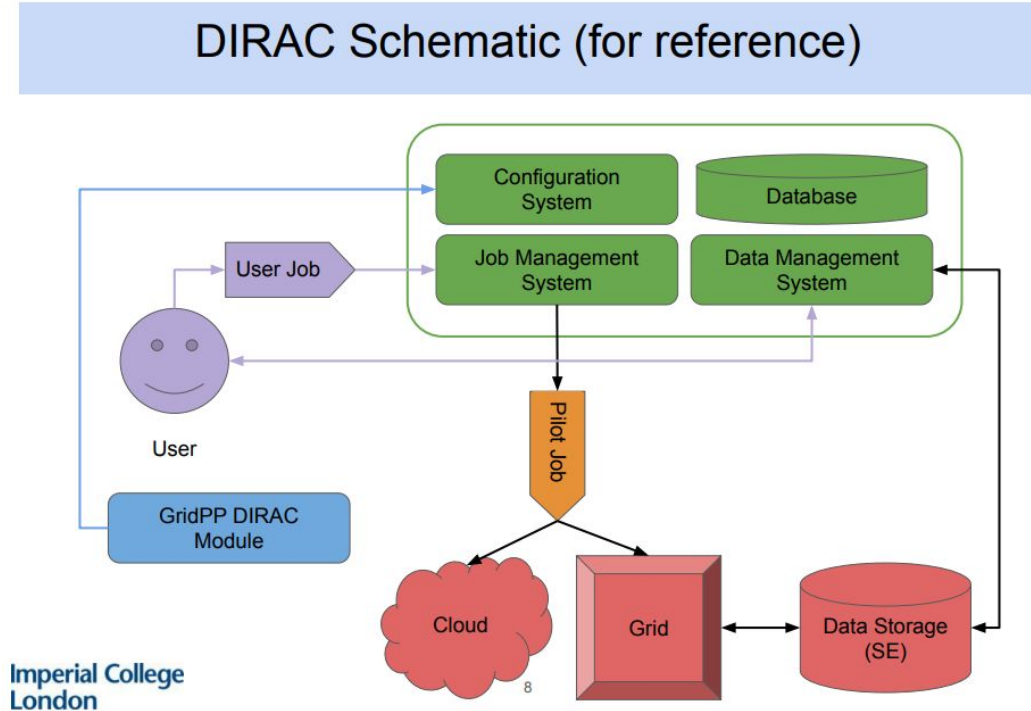
The next challenge is
planned for 2023

That's ~ two years in for
WP1 and one year in for
WP5 → perfect to
participate

Scheduling with DIRAC

In a nutshell: scheduling across job management systems

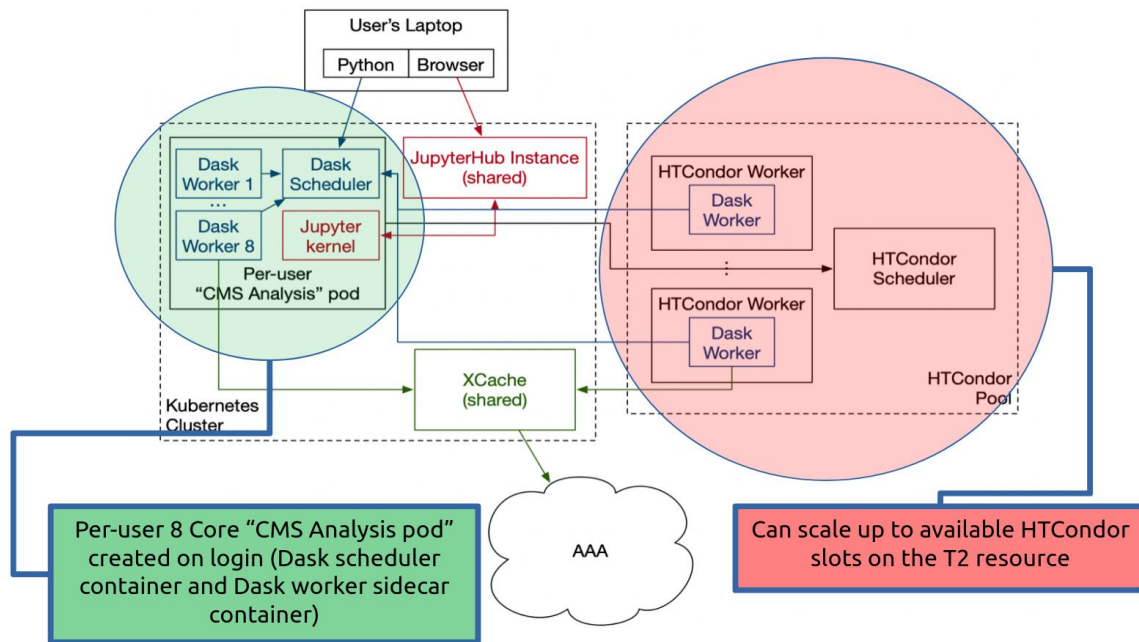
Data management system for access to data lake (here caching)



Slide stolen from Janusz's presentation at the [SWIFT-HEP May meeting](#)

Scheduling with coffea-casa

Uses Dask and [dask-jobqueue](#)

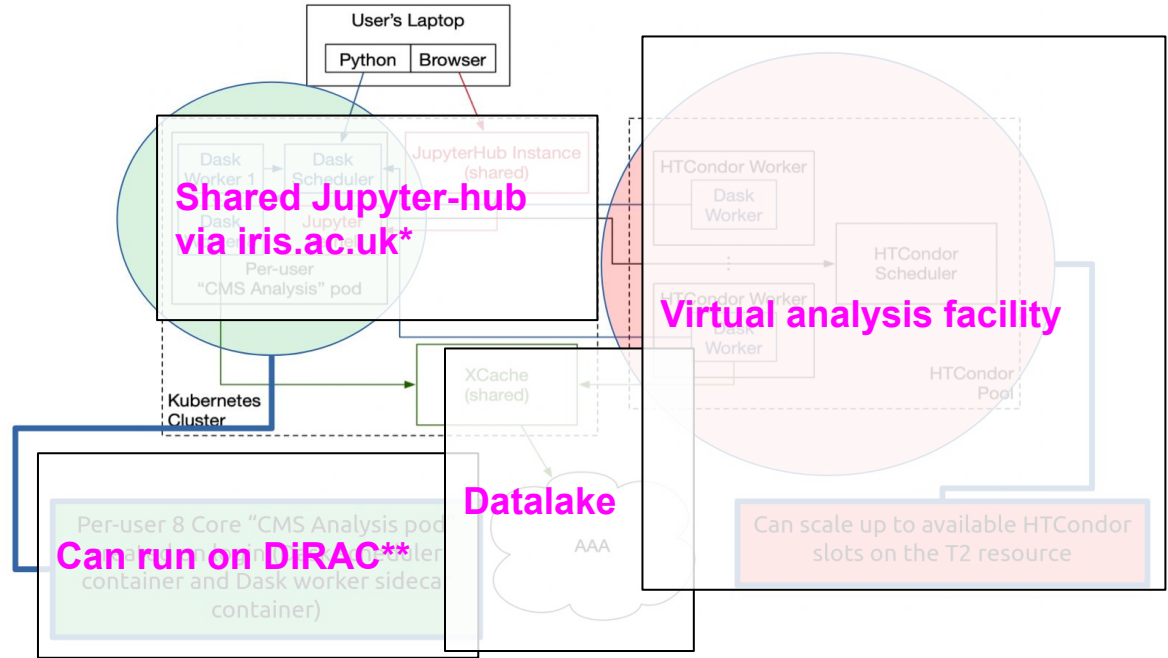


From [coffea-casa docs](#)

SWIFT-HEP

“Adaptation”

As simple as adding
DIRAC jobqueue to
[dask-jobqueue](#)?



*no relation to IRIS-HEP; **no relation to DIRAC (the

SWIFT-HEP WP5 requirements

Scheduling

1. Inspect analysis workflow
2. Parse extra requirements (CPU-only, GPU-capable, hashes for caching, etc) to DIRAC
3. Scheduling tradeoff: if no GPU resources available, try to schedule CPU instead? → ranking
4. All of the above with as little user input as possible → ease of use

Caching

1. Same data + same algorithm + same config = same hash
2. Cached analysis steps (CAS) might be stored anywhere (data lake)
3. Analysis groups: CAS need to be readable/writable by a set of users
4. CAS needs an expiry date: deletions are automated
5. CAS should produce little to no overhead for scheduling/running

IRIS* funding for FAST-HEP

Shameless plug

*no relation to IRIS-HEP;

Short-notice funding call from IRIS* for software development came over the summer

FAST-HEP received 1 FTE @ 6 months for

- Multi-tree and GPU support via awkward1/uproot4
- Faster merging via boost_histogram
- Benchmark analysis examples for CMS, DUNE, and LZ

Good to see dedicated funding for software development

Summary

Substantial groundwork has been done (IRIS-HEP, Scikit-HEP & Co)

Analysis Grand Challenges sound like a good opportunity for us

Valid intersection between IRIS-HEP and SWIFT-HEP → can we efficiently amplify respective efforts?

backup

Reminder WP5

Analysis Work Package (WP 5)
[starting April 2022] aims to

- Implement **caching and portability** for the average user
- **Integrate with WP1 developments** (data lake, workflow management, virtual analysis facility)
- Engage with the wider community as we see fit
 - e.g. analysis facilities R&D, Python ecosystem

But as always, the devil is in the detail

→ **can we learn from existing efforts?**

WP5.1 Caching of intermediate analysis results

- D5.1: Interoperability with UK data lake
 - Output of analysis **steps** will be cached in UK data-lake (D1.1)
- D5.2: Integrate caching mechanism with workflow management
 - Analysis steps that have not changed will be skipped

WP5.2 Portability of analysis code

- D5.3: Optimizations of workloads
 - Allow analysis steps to optionally request specialized resources (GPU, FPGA, NVMe storage, etc)
- D5.3: Integration with workflow management
 - Workload resource requests are ranked → e.g. run on GPU if available, otherwise fallback to CPU

IRIS-HEP Tools and SWIFT-HEP

[Related IRIS-HEP workshop starting tomorrow](#)

If FAST-HEP is taken as base-line

- we are user (and contributors) of awkward, uproot, hist, mplhep
- Scaling out with coffea has been tested

For SWIFT-HEP (taking coffea-casa as example)

- Would it be possible to schedule coffea via DIRAC?
- Are IRIS-HEP thinking of any caching solutions?

→ If the conditions are right, it would make sense to contribute (instead of new)
