

# Analysis Facilities and Use Cases Session Summary

Oliver Gutsche, Eduardo Rodrigues



Joint WLCG & HSF Workshop 2018, Naples, 29 March 2018

# Session overview

- Examples of new and/or innovative analysis flows techniques and facilities being used/investigated

16:00 → 18:00

## Analysis Facilities and Use Cases


Scarlattì (Hotel Vesuvio)

Conveners: Eduardo Rodrigues (University of Cincinnati (US)), Oliver Gutsche (Fermi National Accelerator Lab. (US))

16:00

### Overview of CWP paper from WG Data Analysis and Interpretation


Speaker: Mark Neubauer (Univ. Illinois at Urbana Champaign (US))

 Naples HSF Analy...

16:20

### Spark-like and query-like analysis systems and tools


Speaker: Jim Pivarski (Princeton University)

 pivarski-querylike.pdf

16:40

### HPC analyses

Speaker: Viktor Khristenko (CERN)

 deepest\_hpc4hep.pdf

17:00

### Real-time analyses - the LHCb case


Speaker: Rosen Matev (CERN)


 slides

17:20

### Analyses with SWAN and docker

Speaker: Pere Mato Vila (CERN)

 swan-wlcg-hsf-work...

 swan-wlcg-hsf-work...

17:40

### Open discussion on future avenues and collaboration

Speakers: Eduardo Rodrigues (University of Cincinnati (US)), Oliver Gutsche (Fermi National Accelerator Lab. (US))

# Data Analysis & Interpretation WG Paper

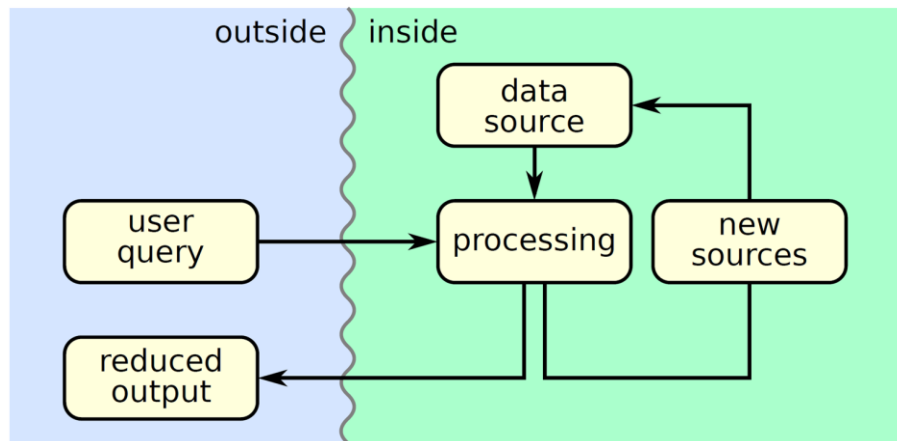
- In support of Community White Paper
- Cuts across many other WG areas
  - Data preservation, Machine Learning, ...
- About to be available on the arXiv
  
- Time-to-insight (data) is a crucial metric
- Scope for new ideas of analysis pipelines, collaborating with non-HEP communities

## Contents

- 1 Introduction**
- 2 HEP Analysis Software Ecosystem**
- 3 Analysis Languages**
  - 3.1 Python
  - 3.2 Declarative Languages
- 4 Analyzing Data**
  - 4.1 File format and compression algorithms
  - 4.2 Analysis Facilities
  - 4.3 Non-event data handling
- 5 Analysis Models and Future Systems**
  - 5.1 Sequential Ntuple Reduction
  - 5.2 “Spark”-like Analysis
  - 5.3 Query-based Analysis
- 6 Analysis Preservation**
- 7 Analysis Interpretation**
- 8 Analysis Roadmap**
  - 8.1 1-year Time Frame
  - 8.2 3-year Time Frame
  - 8.3 5-year Time Frame

# Explorations around data querying in analysis flows

- Or: what works, what doesn't, what's needed, what we're building
- Exploratory work around the following grand picture:
- Several proof-of-concept Python packages developed along the way

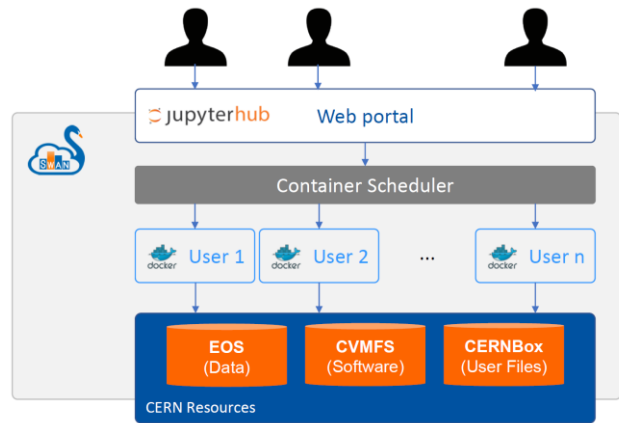


# HEP analysis pipelines on HPC facilities

- DEEP-EST: EU R&D project to use HPC facility toolkit for HEP computing
- Optimising analysis workflows for novel compute/memory/storage capabilities
- Utilising industry tools like Apache Spark: enabled to read ROOT files natively

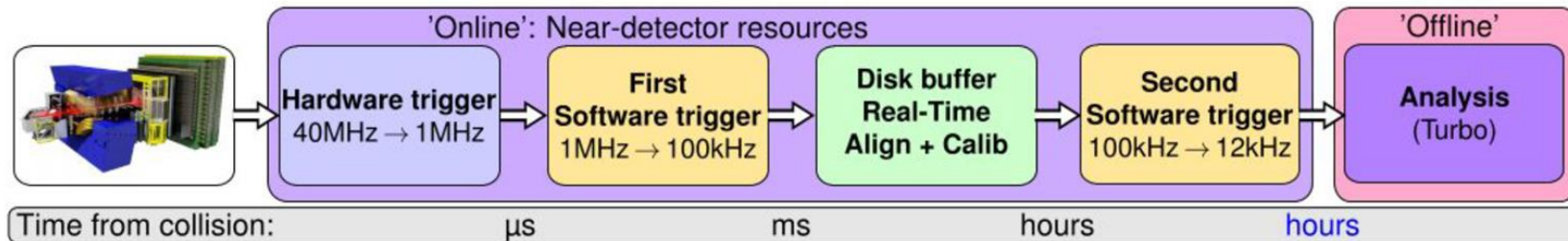
# SWAN - service for web-based analysis

- Users can do analysis only with their browser
  - Cloud-based analysis model
- Jupyter Notebooks on demand
  - Use personalised docker container instance and scale to larger resources like Spark clusters
- SWAN is first step towards a truly scalable and interactive distributed data analysis environment



# LHCb online & high-throughput analysis (flow)

- A reality, in production !



- Boosts up physics reach and allows for otherwise non-doable analyses
- Only saving reconstructed objects (no raw event) in the trigger demands much on online calibration and alignment, hence on a robust and efficient online monitoring of the data as well

# Thinking forward

- We want to foster collaboration and common projects on analysis-related topics.
- Questions:
  - Do we have analysis efforts that were not covered, that should be mentioned here?
  - Is more contact information needed for projects already discussed to start more collaboration?
- Final remarks:
  - Lots of activity post-CWP but no organisational structure / specific timelines to tackle roadmap
  - ... can the HSF help with information exchange?
  - Examples of funded efforts exist (e.g., DIANA/HEP), but need more
  - Links with communities outside HEP are emerging through the tools explored/used - great