



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



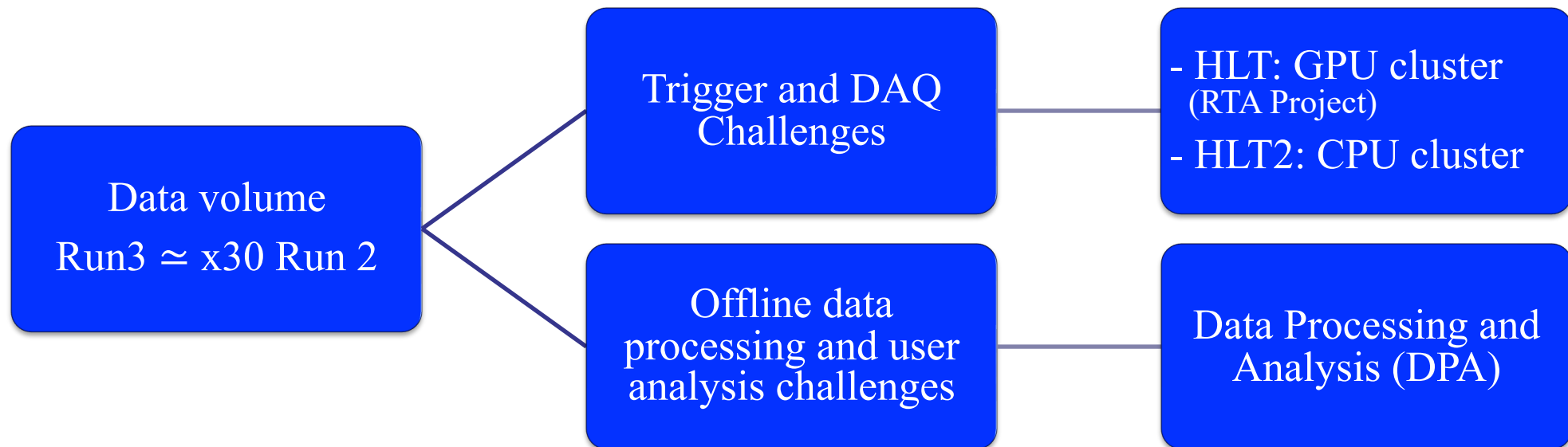
Analysis Facilities Forum Kick-off Meeting

Donatella Lucchesi
University of Padua and INFN
on behalf of LHCb Data Processing and Analysis Project



AF Activities in LHCb

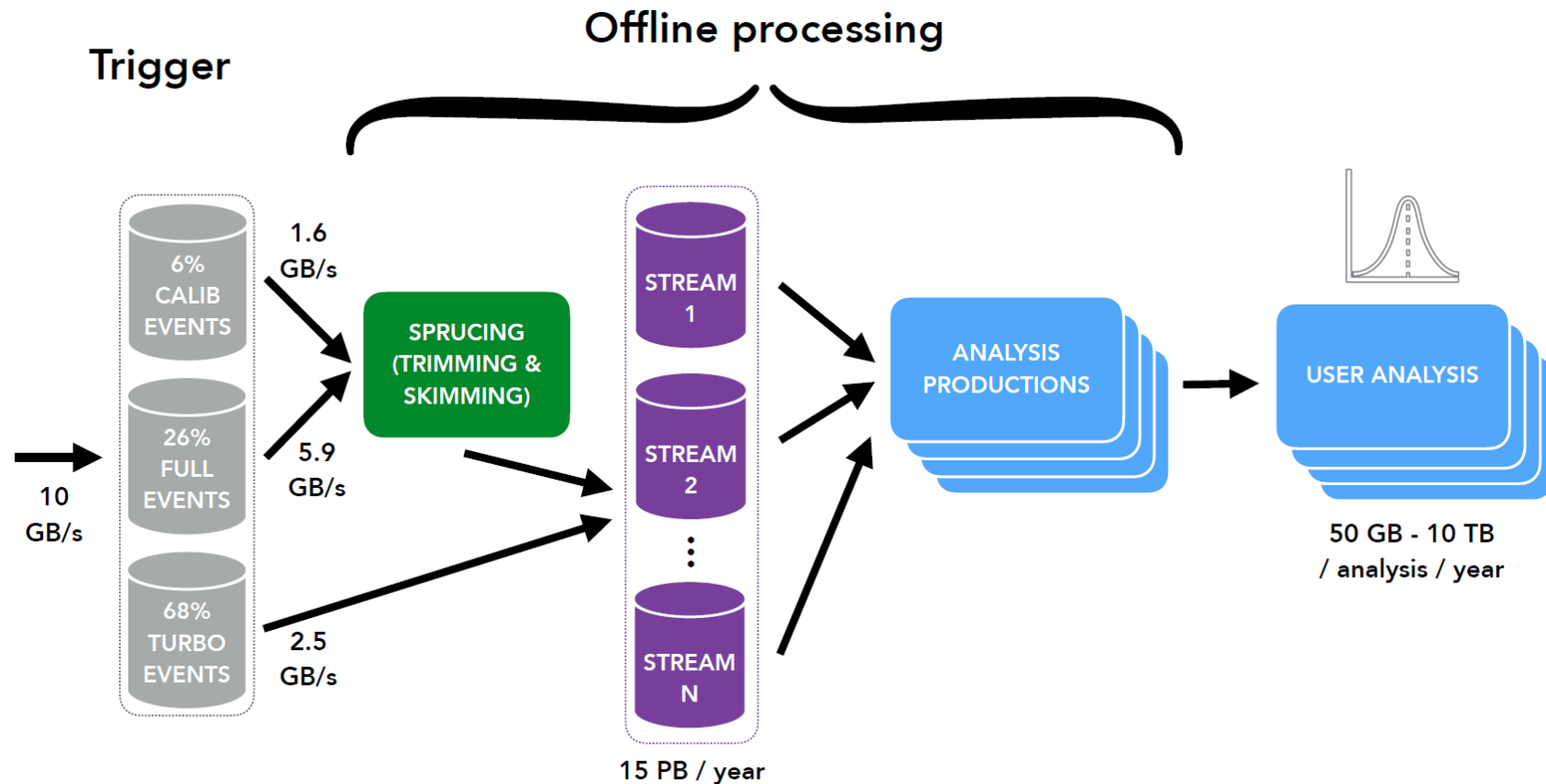
Large increase in data volume Run III vs. Run II \longrightarrow challenges to trigger, DAQ and to offline data processing and analysis



DPA built around 2 main ideas:

- Centralised skimming and trimming for physics cases where HLT2 output bandwidth is too high to be persisted straight to disk storage
- Centralised analysis productions for physics WGs and users

LHCb-FIGURE-2020-016





DPA Description of Activities (1)

The screenshot shows the DPA project website. The sidebar on the left contains a search bar and several menu categories: WORK PACKAGES, DOCUMENTATION, and MISCELLANEOUS. The main content area has a header 'Welcome to the Data Processing & Analysis (DPA) project' and a table listing the work packages. Two rows in the table, 'WP2 - Analysis Productions' and 'WP4 - Innovative Analysis Techniques', are highlighted with red boxes. Red arrows point from these boxes to the text 'Most relevant for AF discussion' below the table.

Docs » Welcome to the Data Processing & Analysis (DPA) project [View page source](#)

Welcome to the Data Processing & Analysis (DPA) project

The **Data Processing & Analysis, DPA, project** addresses the challenges for offline data processing and analysis due to the very large increase in data volume with respect to Run II. DPA is built around 2 main ideas:

- Centralised skimming and trimming (aka Sprucing) of a significant fraction of HLT2 outputs.
- Centralised analysis productions for physics WGs and users.

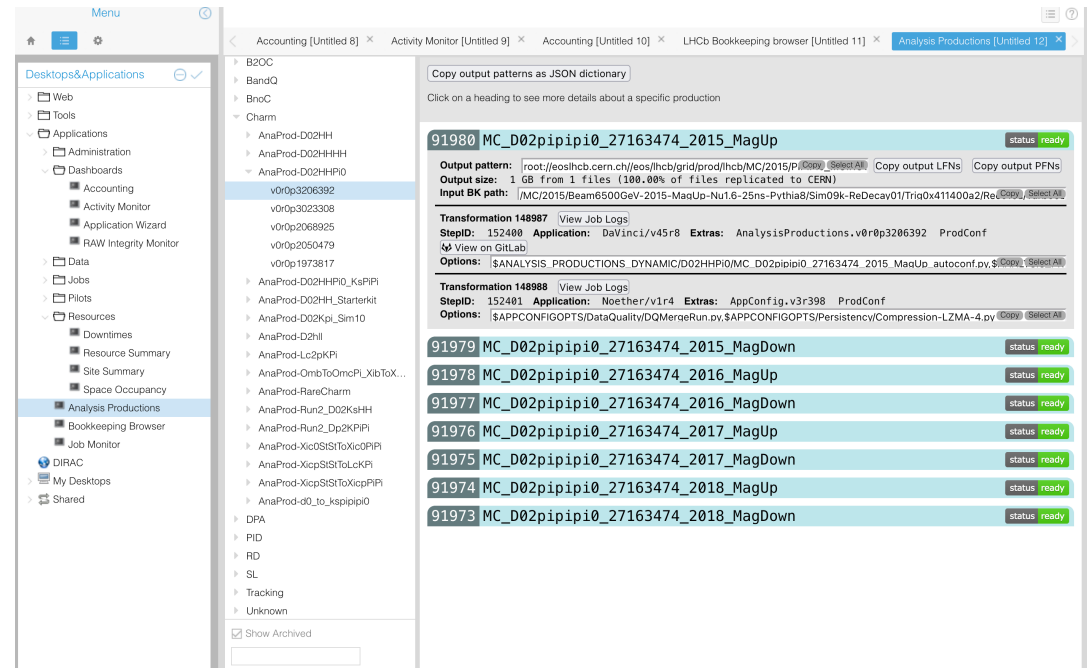
Overviews of the project Work Packages and offline processing flow are given below. The general project mailing list is [lhcb-dpa-general](#).

Work package	Coordinator(s)	Mailing list	Mattermost
Overall coordination	Eduardo Rodrigues		
WP1 - Sprucing	Nicole Skidmore	lhcb-dpa-wp1	link
WP2 - Analysis Productions	Chris Burr	lhcb-dpa-wp2	link
WP3 - Offline Analysis Tools	Patrick Koppenburg	lhcb-dpa-wp3	link
WP4 - Innovative Analysis Techniques	Donatella Lucchesi	lhcb-dpa-wp4	
WP5 - Legacy Software & Data	Federico Leo Redi	lhcb-dpa-wp5	Stripping, DaVinci
WP6 - Analysis Preservation & Open Data	Sebastian Neubert	lhcb-data-preservation	link

Most relevant for
AF discussion

Support user processing of data and simulation using the DIRAC transformation system.

- User save time, no GRID jobs monitoring
- Job details/configuration/logs automatically preserved in LHCb bookkeeping/EOS
- Automated error interpretation/advice
- Results displayed on webpage



Even if very flexible, not all user requests can be satisfied:

- Dedicated analysis workflow
- AI/ML training
- Hybrid resource usage



DPA Project: Innovative Analysis Techniques

Think tank for innovative analysis techniques and exploitation of new analysis facilities with heterogeneous computing resources (GPU/CPU/FPGA)

Two main topic so far:

➤ GPU resources usage in analyses:

- ML @HPC cluster “Marconi” Cineca Bologna: DNN for b- vs c- jet tagging
- Zfit and likelihood inference in Zurich
- DNN and ultra-fast simulation Florence
- Amplitude Analyses, Aix Marseille, CNRS/IN2P3
- Charm Analyses, Manchester
- Amplitude Analysis Λ_b Tsinghua
- Analysis of ϕ_s Santiago

➤ Quantum Computing application to HEP

- Quantum Machine Learning for b-jet identification. Use of GPU cluster for hardware simulation

Users access resources with custom code/methods



Marconi100@Cineca tests

Marconi 100 is an HPC cluster based at CINECA computing center in Bologna Italy
Agreement between INFN and CINECA to exploit resources for LHC experiments

- CPU part is PowerPC \Rightarrow LHCb software not available for this architecture
- Most of the power is provided by GPUs
- Computing experts need to be involved for proper data management and workload suitable implementation.
- In progress:
 - Changes to LHCbDirac middleware to submit jobs to this cluster
 - CVMFS on WN an outbound connection guarantee from/to CNAF Tier-1
 - CONDA available using lb-conda on CVMFS to manage virtual environments



Next Steps toward AF

- Several Tier-2, Tier-3, medium-large centers in US and in Europe are proposing AF prototypes developed and supported by ATLAS/CMS \Rightarrow LHCb may want to participate locally to these R&D
- LHCb will leverage the very powerful GPU HLT1 farm
- For the analysis, users want something easy to use, flexible and powerful enough... needless to say
- Given the progress in HSF, the papers submitted to Snowmass and the proposed prototypes, within LHCb starting:
 - Collect all the use cases, available and used resources, code developed, etc.
 - Identify the users needs for next round of analysis
 - Proceed with a structured activity that may led:
 - Different AF configurations depending on the site: CERN may need something dedicated to incorporate HLT1, exploit what available in country, design a dedicated configuration, ect.
 - Definition and identification of mandatory LHCb-specific requests