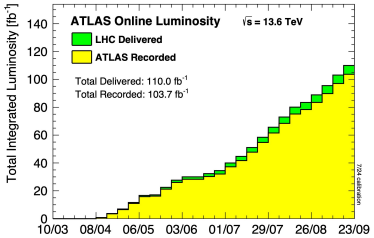
Technical Interchange Meeting Jan 21-23, 2025
TIM Goals

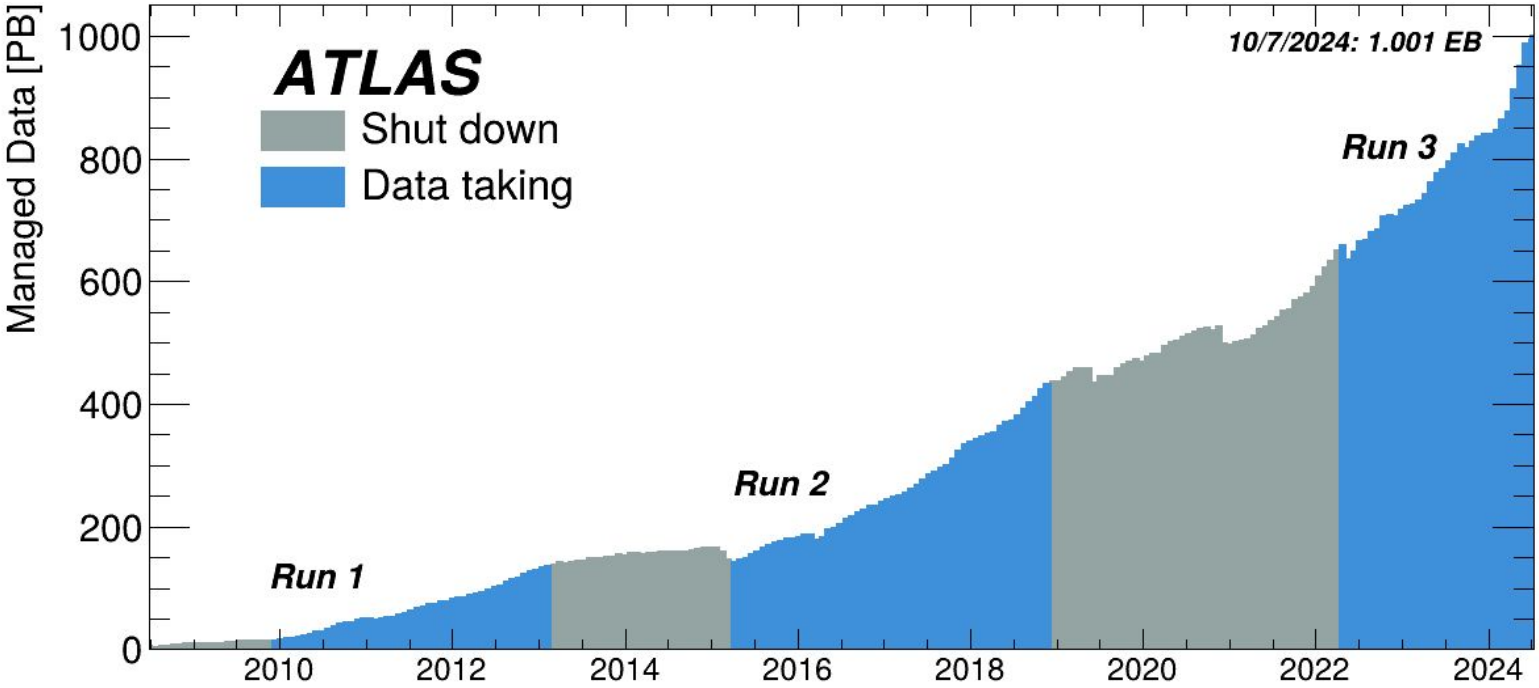
Alexei Klimentov (BNL) , Andreu Pacheco Pages (IFAE), Chris Bee (Stony Brook), Mario Lassnig (CERN), Shawn Mc Kee (University of Michigan), Torre Wenaus (BNL)

Where we are:
We made the right choices with the PanDA ecosystem and Rucio developed for ATLAS.

Major Rucio milestone - July 2024

- ATLAS has reached 1 Exabyte of Rucio-managed data
  - LHC delivered more than full 2024 **pp** target of 110/fb^{-1} early
  - Data-taking efficiency at 94.3% !
  - Frequent applications of Lifetime model and Catmore-rule necessary





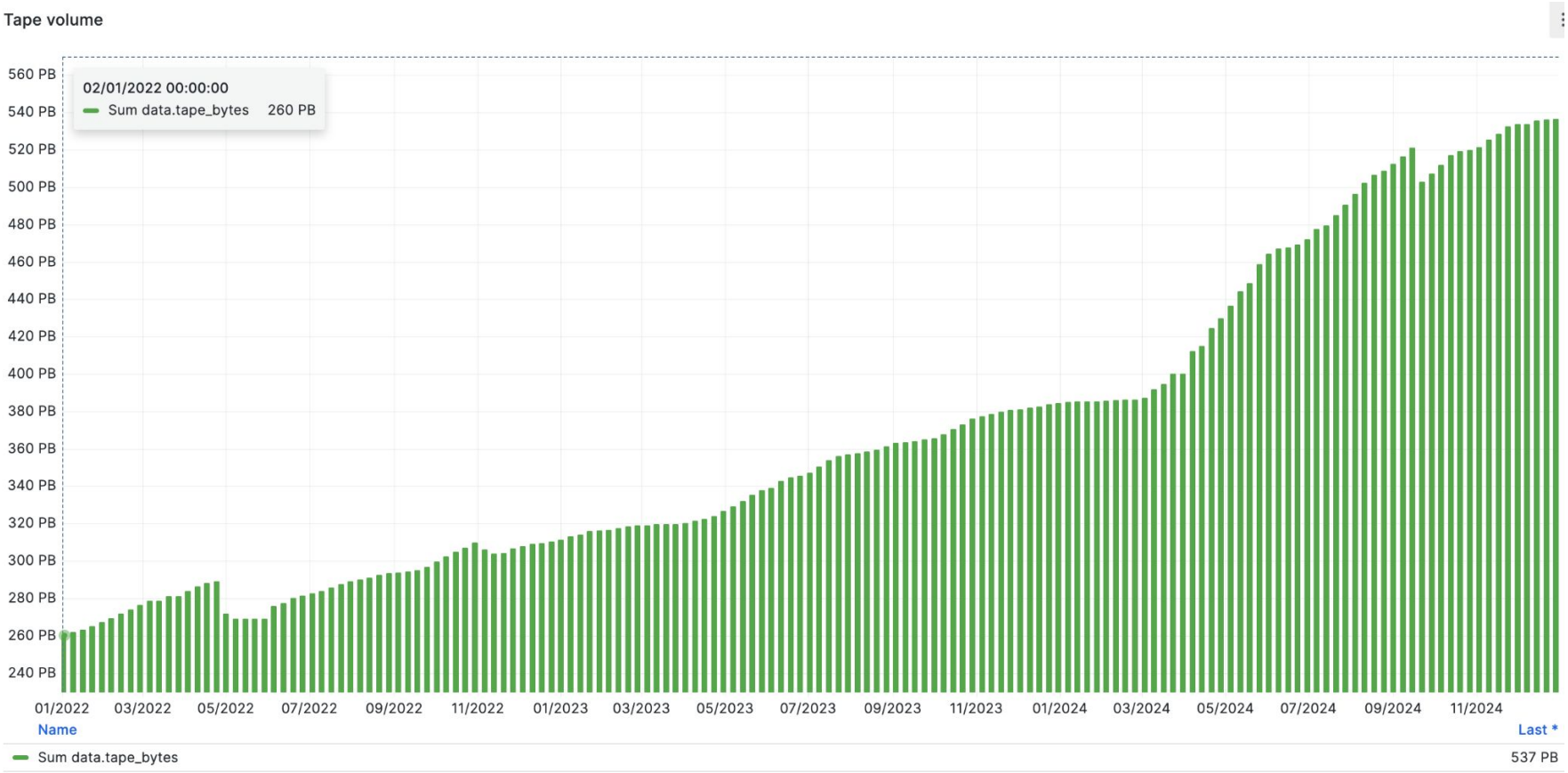
Where we are: Data in Rucio - 17 Jan 2024 (source)

Disk Size	Primary Size	Disk Files	Primary Files
378 PB	254 PB	885014186	824445332

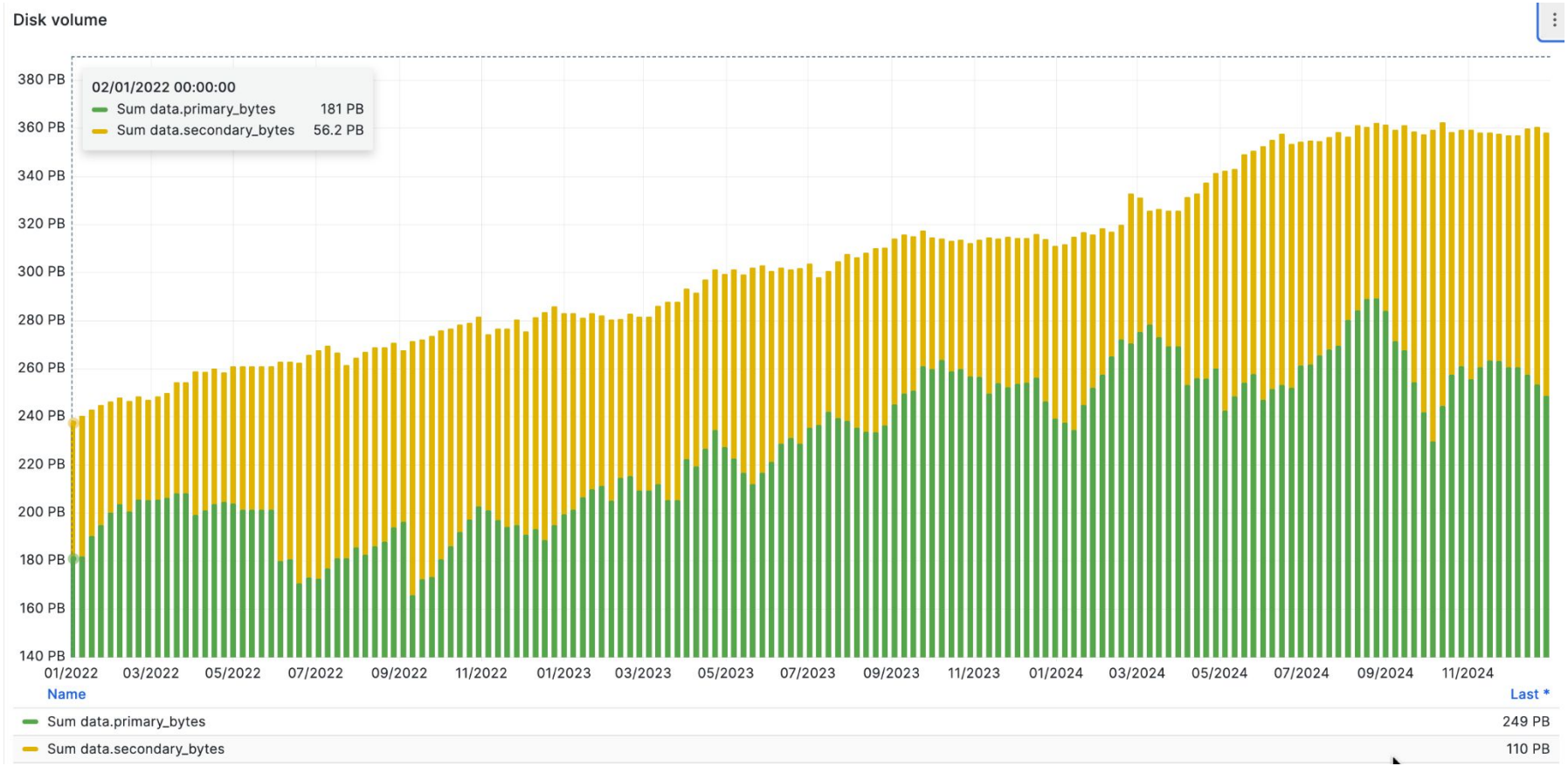
Tape Size	Secondary Size	Tape Files	Secondary Files
662 PB	124 PB	374358894	60568854

Total Size	Total Files
1.04 EB	1259373080

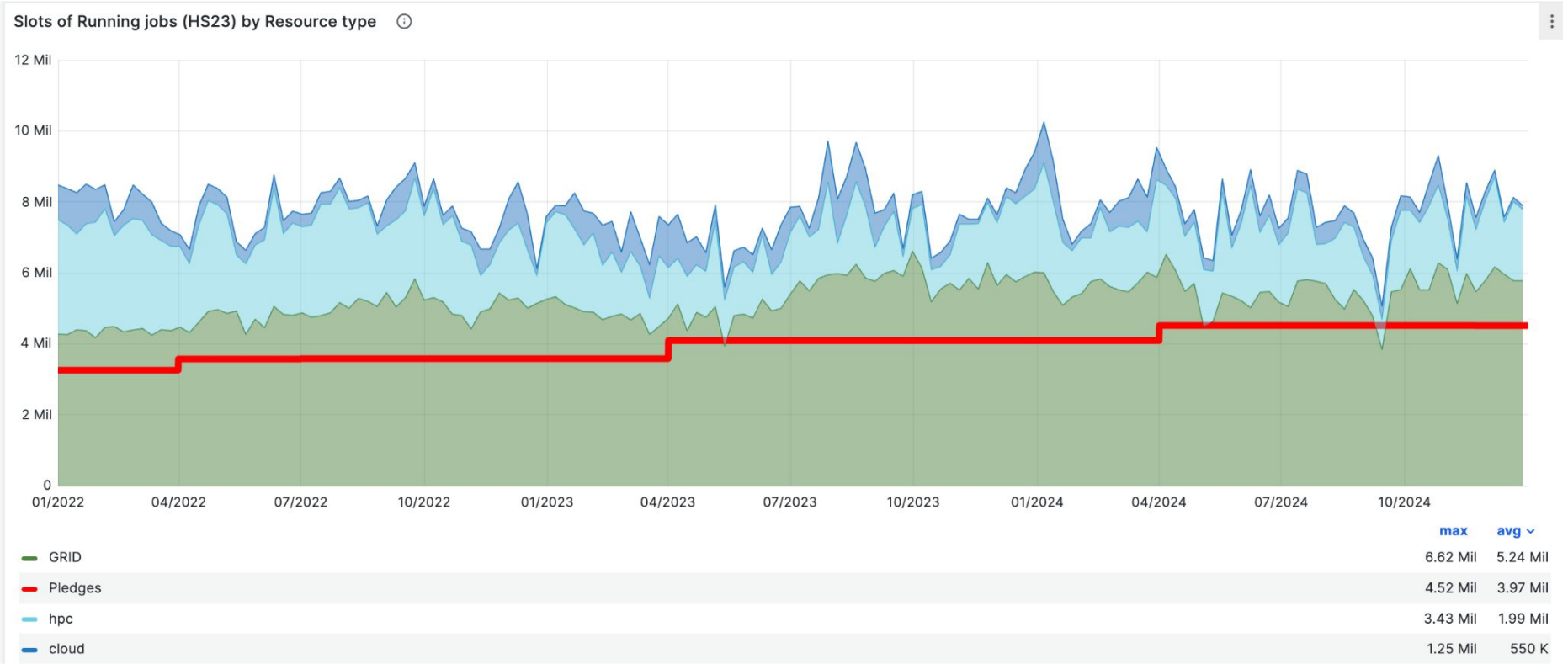
Where we are: 3-year pledged data volume on tape (source)



Where we are: 3-year pledged data volume on disk (source)



Where we are: 3-year normalized slots of running jobs (source)



- Centralised production and user analysis have been running at full steam
  - Multiple concurrent campaigns of various intensity and duration
  - Physics Validation, Production, Reprocessing, Derivation

Some ADC Topics in 2024

- ADC Coordination Board, ADC Weekly and ADC Operations daily meeting incl. an additional meeting in US time zone
- Communication with ATLAS clouds and sites
- WFMS modernisation triggered after the WFM review in 2023
- Alma 9 migration
- Transition to tokens
- DC24 Data Challenge (Feb 2024) and Mini-Data Challenges (Dec 2024-)
- Under-pledge grid usage
- Site evolution
- Unavailable data trapped at sites with long downtimes
- Memory optimization
- ARM and GPU architectures
- Opportunistic resources (Clouds & HPCs)
- Tape resources
- Effort for central operations

ADC Risk register for R2R4

Category	Risk	Likelihood	Impact	Severity	Owner	Effect
Consensus	Stagnant interactions between key community projects	3	4	12	WLCG, CERN-IT, ADC	Long-winded discussions with no clear decisions and outcomes; Non-working APIs; Incompatible implementations
Catastrophes	Significant downtimes related to environmental disasters	2	5	10	WLCG	Unavailability of large volumes of data for extended periods; Cannot run as many jobs as we need
Capacity	Not enough available storage, CPU, network and/or dedicated hardware (GPUs)	2	5	10	WLCG, ATLAS	Cannot execute jobs; Cannot distribute data; Cannot store data
Dependencies	Loss and/or frequent changes in software dependencies	3	4	12	CERN-IT, WLCG, ADC	Failures in software execution; Additional development effort; Being stuck with old unsupported software
Procurement	Mismatch of ATLAS requested resources vs. site-provided resources	2	4	8	WLCG, ATLAS	Starvation of tasks due to limited resources; Additional costs for sites due to unclear schedules; Loss of allocations
Central support	Lack of persons to support analysis users, production managers, and to follow-up daily operational tasks	4	4	16	ADC	Prolonged task execution; Task submission delays; Storage resource overloads; Angry users
Personpower	Difficulty in retaining key personnel, impact on progress of ongoing projects	5	5	25	ADC	Impact on progress of important projects; Long delays; Potential cancellation of projects
Organisation	Insufficient number of suitable persons in leading roles	4	3	12	ADC	Destabilise ADC organisation, effectiveness, and follow-up of tasks

R2R4

The ATLAS Computing road to HL-LHC

- ATLAS Software and Computing HL-LHC Roadmap was published in March 2022 with concrete milestones (40 pgs; milestones and deliverables by area)(ADC)
- 2024 Software and Computing for ATLAS paper (157 pgs) was published in April 2024
- The next step in HL-LHC Planning: to write a **Phase-II Software and Computing TDR**
  - Decisions must be made.
- This Technical Interchange Meeting is opportunity to discuss
  - Focus on where we want to be for HL-LHC.
  - How to bring our community together and jointly work on activities.
  - Automating our day-to-day work to enable effort to enable future-looking activities.
- …and define a list of action items for 2025 and beyond

TIM 2025 Topics (Indico)

- Distributed Computing Operations:
  - Issues and Challenges (Tuesday afternoon)
  - Automations and Future Plan (Wednesday morning)
  - Improving our support tasks (Wednesday afternoon)
    - AI and GenML
    - Documentation
    - Communication
- Mini data challenges and system modelling (Thursday morning)
- ADC-REDWOOD mini-workshop on WMS and Distributed Computing Modeling in parallel with ATLAS SW&C and ADC coordinators meeting at BNL with SDCC and NPPS staff (Thursday afternoon at BNL)

We aim in this TIM

- to have **consensus** among experts in key areas for the **ATLAS Computing road to HL-LHC** (R2R4)
- to identify **ideas** for increasing **automation** and **introducing ML/AI** in **operational** areas such as job brokering, data distribution, and data recovery
- to develop **strategies** to **mitigate** potential **mismatches** between **requested and available resources**
- to find **ideas** for **retaining personnel** and **improving of support**
- to discuss **expanding data challenges** to **infrastructure stress-testing** in all ADC components for R2R4
- to give a valuable **input** to the **Phase-II Software and Computing TDR**

Support slides

3:00 PM → 6:00 PM **Distributed Computing Operations: Issues and Challenges** Room P118 (SBU Physics Buil...

Convener: Mario Lassnig (CERN)

3:00 PM	**USATLAS WBS 2.3 Topics** **Speaker**: Fred Luehring (Indiana University (US))	30m
3:30 PM	**Data Carousel & Archive Metadata** **Speaker**: Xin Zhao (Brookhaven National Laboratory (US))	30m
4:00 PM	**Tape and network R&D at NET2 [CANCELLED]** **Speakers**: Eduardo Bach (University of Massachusetts (US)), Rafael Coelho Lopes De Sa (University of Massachusetts (US))	30m
4:30 PM	**A Sysadmin's Point of View on Operations** **Speaker**: Judith Lorraine Stephen (University of Chicago (US))	30m
5:00 PM	**Discussion on site accounting** **Speaker**: David Rebatto (Università degli Studi e INFN Milano (IT))	30m

Wed, January 22

9:00 AM → 1:35 PM **Distributed Computing Operations: Automations and Future Plan** — Room P118 (SBU Physics Buil...

Convener: Alexei Klimentov (Brookhaven National Laboratory (US))

9:05 AM	**Central ADC proposals**	30m
	Speaker: Ivan Glushkov (Brookhaven National Laboratory (US))	
9:35 AM	**Proposals from DPA**	30m
	Speakers: Rodney Walker (Ludwig Maximilians Universitat (DE)), Timo Wilken (University of Texas at Arlington (US))	
10:05 AM	**Proposals from DDM**	20m
	Speaker: Riccardo Di Maio (CERN)	
10:25 AM	**Proposals from WFMS**	15m
	Speakers: Fernando Harald Barreiro Megino (University of Texas at Arlington), Tadashi Maeno (Brookhaven National Laboratory (US))	
10:40 AM	**A site view on potential ops automation**	30m
	Speaker: Doug Benjamin (Brookhaven National Laboratory (US))	

2:25 PM → 6:00 PM **Improving our support tasks: AI and GenML, documentation, communication**

Room P118 (SBU Physics Buil...

Convener: Dr Andreu Pacheco Pages (Institut de Física d'Altes Energies - Barcelona (ES))

2:30 PM	**AI and GenML and Distributed Computing Operations and SW stack** **Speaker**: Kaushik De (University of Texas at Arlington (US))	30m
3:00 PM	**AskPanDA** **Speaker**: Paul Nilsson (Brookhaven National Laboratory (US))	30m
3:30 PM	**ADC Communications and Documentation** **Speaker**: Mario Lassnig (CERN)	30m
4:00 PM	**ePIC/EIC Software & Computing, and AI/ML at BNL** **Speaker**: Torre Wenaus (Brookhaven National Laboratory (US))	30m
4:30 PM	**Experience with UChicago AF support using ChatGPT** **Speaker**: Ilija Vukotic (University of Chicago (US))	30m

Thu, January 23

9:00 AM → 12:00 PM **Mini data challenges and system modelling** — Room P118 (SBU Physics Buil...

Convener: Shawn Mc Kee (University of Michigan (US))

- 9:00 AM — **ARC/ACT/ND Insights [REMOTE, TBC]** — 30m
  Speaker: Oxana Smirnova (Lund University (SE))
- 9:30 AM — **US Mini Data Challenges** — 30m
  Speaker: Shawn Mc Kee (University of Michigan (US))
- 10:00 AM — **Analysis of files in BNL Tape system & NANO Data Challenges** — 30m
  Speaker: Hironori Ito (Brookhaven National Laboratory (US))
  BNL HTPSS Tape Te...
- 10:30 AM — **System modelling** — 30m
  Speaker: Paul Nilsson (Brookhaven National Laboratory (US))
- 11:00 AM — **Simulating HEP workloads with SimGrid** — 30m
  Speaker: Frederic Suter
- 11:30 AM — **Distributed analysis with GPU** — 30m
  Speakers: Eduardo Bach (University of Massachusetts (US)), Rafael Coelho Lopes De Sa (University of Massachusetts (US))

11:30 AM → 12:30 PM **Wrap up and action items** — Room P118 (SBU Physics Buil...

- 11:30 AM — **Project planning** — 1h
  Speaker: Mario Lassnig (CERN)

2:00 PM → 5:00 PM **Mini-workshop on WMS and Distributed Computing Modeling (BNL)** ¶ 3h Conference Room A (BNL 725)

Workshop to discuss REDWOOD topics.
BNL Computing and Science Directorate. Bldg.725 Conference room A

Speakers: Frederic Suter, Ozgur Ozan Kilic (Brookhaven National Laboratory), Paul Nilsson (Brookhaven National Laboratory (US)), Raees Ahmad Khan (University of Pittsburgh (US)), Sairam Sri Vatsavai (BNL), Sankha Dutta (BNL), Dr Tasnuva Chowdhury (Brookhaven National Laboratory (US)), Tatiana Korchuganova (University of Pittsburgh (US)), Yihui Ren (Brookhaven National Laboratory (US))

- SimGrid news — 20m
  Speaker: Frederic Suter
- Grid simulation status and plans — 20m
  Speakers: Paul Nilsson (Brookhaven National Laboratory (US)), Raees Ahmad Khan (University of Pittsburgh (US)), Sairam Sri Vatsavai (BNL)
- Error classification/analysis — 20m
  Speakers: Paul Nilsson (Brookhaven National Laboratory (US)), Sankha Dutta (BNL), Tatiana Korchuganova (University of Pittsburgh (US))
- ML-driven resource requirement prediction for WMS — 20m
  Speaker: Dr Tasnuva Chowdhury (Brookhaven National Laboratory (US))
- ModSim news and plans — 20m
  Speakers: Ozgur Ozan Kilic (Brookhaven National Laboratory), Yihui Ren (Brookhaven National Laboratory (US))
- Upcoming papers — 20m
- Discussion — 20m