ATLAS Distributed Computing Technical Interchange Meeting, January 21 - 23, 2025 at Stony Brook University, NY

TIM live notes

Agenda page: https://indico.cern.ch/e/TIM2025
Zoom link : https://cern.zoom.us/j/64426397126?pwd=VrlBC0ZI4UPfnivtd34cadhivCaunq.1

Tue Jan 21

Present : Andreu, Rafael, Raees, Chris, Alexei, JohnDS, Hiro, Timo, Fred, Lincoln, Torre, Ofer, Mario, Rod, Riccardo, Ilija, Shawn, DavidS, Kaushik, Ivan, Judith, Rui, Carlos, Doug, Shigeki, Paul, Tasnuva, Xin, Horst, DavidR

Remote : AlexanderA, Verena, Petr, Tony, Srini, Tanya, Aidan, Fa-Hui, Will, Jammel, Aresh

14:00 coffee and group photo

14:20 Introduction and TIM goals

Alexei : welcome from BNL, SBU and US ATLAS
CDS Directorate at BNL and Projects
Chris : welcome and logistics

Andreu : TIM goals

- Rucio managed data volume 1.04 EB
- Selected 2024 ADC topics
- ATLAS HL-LHC TDR
- Questions/Comments
    - Tape transparent usage (Data Carousel) vs tape only as archive?
    - PIC experiences with tape difficulties or more?
        - overpledging, data ingress volume
    - Central operations lack of support
        - Is it too complex? Are we stuck in our old ways?
        - We simply don't have enough people
        - DDM Ops support in 2024
        - Mario : slide "Continuous loss of FTE in ADC”

                - The majority of 77 persons in ADC have contributions < 0.2 FTE
                - Monitoring is one of main worries
            - Automation. Ivan : Automation will need additional effort (which we don't have today)
            - Ed : how we can encourage new people to come in ?
                - Mario : we need to define R&D (also as ATLAS qualification projects)
    - DS: On production 2025/26
        - We have to be careful filing our resources next year. There will be no large scale "standard" simulation for 2025 (and 2026), mc23g, will only come at the start of 2026. What will or can come in 2025 is
            - Large (many billions) scale GEANT11 simulation samples
            - Extensions of existing Run 3 samples
            - More AF3 (is held up by delayed CP recommendations)
            - We *can* start the EVNT for mc23g in 2025, and some of it is slow
        - Then we expect in 2026 that **everything** comes at once
            - Large scale mc23g, and maybe mc23i (for 2026), depending on machine differences / bugs found/fixed for 2025 vs 2026
            - Potential new MC for all of Run 3
            - Several Run 3 reprocessings
            - HI processings
            - Stuff we haven't thought of yet

15:00 Distributed Computing Operations: Issues and Challenges

Mario : this is session - "looking back" (tomorrow "looking forward")

Fred : US ATLAS WBS 2.3 topics

- US ATLAS daily meetings (thanks to Ivan) help a lot to spot problems
- WORN / unused vs useful data
- constant failure rate ~11%
    - breakdown exists (but not shown)
- high number of repeats problem
- out of memory example that could be found by junior shifter
- storage overload by users payload (manual action to inform/block user)
    - potential way to automate this
- scout jobs for analysis users has the potential to turn people away from the grid
    - It will be discussed at automation session (Wed Jan 22 morning)
- some problems are so difficult that they require expert/developer intervention, impossible to fix/mitigate by shifters
- responsibility/chain-of-command unclear for some external projects/systems (CVMFS, HTCondor)
- Andreu : US ATLAS cloud support vs other clouds (there is no full Ops support in most of the clouds)

- Take away:
    - Automation may solve ~90% of our problems
    - If Monitoring will be restructured it will help sites

X.Zhao : Data Carousel and Archive Metadata
- tape collocation with "virtual containers" solely for writing purposes
    - RAW data is without containers
    - should be possible to estimate dataset/container size for production
- proposed project "analysis of atlas tape access patterns"
- Rod: can the tape system feed back (co)location of data to Data Carousel / Rucio?
- Repack can be tricky for sites, investing in repacking needs conscious effort

Judith Stephen: Site administrators view
- Memory management, communication, monitoring, partial downtimes
- Judith will summarize topics for S&C week

David Rebatto: Accounting
- Sudden drop of HS23 at INFN T1 (Site's understanding : possible inconsistency in CRIC or sudden change average HS23 value in CRIC for site)
- Natalia Szczepanek's studies
- We have actual HS values per cpu/model
- HS23 per node is important for scouts (ticket for workernode map exists)

Wed Jan 22

09:30 Distributed Computing Operations: Automations and Future Plan

Present : Fred, Rafael, Raees, Chris, Kaushik, Doug, Alexei, Shigeki, Shawn, JohnDS, Tasnuva, Andreu, Lincoln, Judith, Carlos, Ivan, DavidS, Ed, Xin, Ilija, Rui, Rod, Horst, Paul, Timo, Mario, Ofer, Riccardo, Hiro, DavidR

Remote : Fabio, Fa-Hui, Verena, Fernando, Eduardo, Eddie, Will, Tanya, Tadashi, Jammel

09:30 Ivan - Automation : ADC
Person power available @ADC ("Mario plot")
Automation is already available in several places f.e. Internal Rucio tools

Metrics, priorities, effort vs benefits
User scores
QT and SW grants (some topics are perfect for SW grants)

Sites involvement
CPU time wasting
Firefighting

10:20 Timo&Rod - DPA view

Task tails
GUI for users
Lxplus & NAF interactivity vs PanDA submission

10:40 DDM - Riccardo & Mario
ATLAS DDM reports automation
(lack of person power since T.Bermann has left)

11:15 Fernando (presented by Kaushik) WFMS view
11:25 Doug - site view
- AF portal
- “Ivan’s menu”
- Different monitoring portals with links for sites, vs ADC central

14:30 Improving our support tasks: AI and GenML, documentation, communication - Andreu -> Started at 12am

12:00 Mario - ADC Communications and Documentation¶

In 2024, Atlas Management, through the OAB, initiated Atlas Documentation Weeks to focus on documentation efforts. ADC created a documentation report, collecting information on existing ADC documentation. CERN IT now has a documentation service using mkdocs. new system is called atlas-computing-docs.ch, and its code is in GitLab. new system uses markdown, has version control, a functional search, and allows plugins (also to LLMs). There is a push to migrate all semi-static information from Twiki to the new system (Timo?). A web IDE is integrated for easier editing. marker for migrated Twiki pages should be agreed

The central communication point is the morning meeting at 9am CERN time. pre-filled minutes, focusing on ongoing issues. Content provided by CRC and DPA but everyone is welcome to contribute. "News of the Day" is sent after the morning meeting.

The ADC Weekly meeting is chaired by ADC Coordination and provides reports from various areas. place to learn about upcoming campaigns and large workloads. ADC Coordination tries to attend other meetings to stay informed, but they are not experts in every area. virtual control room is a public chat for best effort support. issues requiring follow-ups or with expected delays, use the mailing lists such as the user support list. DDM support is for data management issues, and DPA is for workflow management. If a ticket is filed, it is recommended to also send a

message to the mailing list. strategic communication or R&D projects, contact the relevant area coordinators directly. Security-related communication is handled at the experiment level, not by ADC. Security incidents should be reported to the CSIRT. Problems with non-ADC systems should be reported to the Fabrics Coordinator.

requested a single page showing all the contact information

concern about how action items from the morning meeting propagate to sites or individuals; the main channel is GGUS, or the WLCG helpdesk

request to have a dashboard where one can see ongoing campaigns or data rebalancing activities

need for a single place to send for help and report problems, as well as a single place for real-time updates

Mattermost for announcements and activity summaries? suggestion to encourage the use of threads in Mattermost

information in Mattermost can be difficult to find if you are not in the conversation. sometimes the ticketing system gets used for conversation rather than filing tickets.

moving documentation to mkdocs will help new sites and admins, and that a high level view of the distributed system would be useful

It was agreed that a page with a higher-level view of the distributed system with the relevant services, such as Panda, Rucio, and Frontier, would be helpful.

12:30 Kaushik De - AI and GenML and Distributed Computing Operations and SW stack

Challenges of managing large-scale operations with numerous collaborations, people, and data, and suggests leveraging Gen-AI to address these challenges.

primary focus is on using Gen-AI to improve information access and integration across various systems

Kaushik proposes several short-term projects, roughly six months in duration

Current monitoring systems are fragmented, making it difficult for non-experts to find relevant information. Instead of building new monitoring systems, Kaushik suggests training a Gen-AI model to find the necessary information to debug problems. The system would allow users to ask questions in natural language, potentially in any language, rather than using GUIs or forms.

The system could retrieve data from Grafana, for instance, and present it in a user-friendly format. answer questions like "Show me the error rate for the last hour compared with 24 hours ago". "What does error XXX mean?". or "Show me the network usage for the last hour and 24 hours ago for these two sites"

Documentation Integration: Similar to monitoring, documentation is spread across many locations. integrate existing documentation, rather than rewriting it or moving it to a new platform. weights to different documentation and answers to prioritize the most relevant results. Concerns about outdated informatio. Mario. Plug-ins that automatically update as documentation changes already exist.

Knowledge Base Unification: massive knowledge base across various systems, including user support, site information, and ticketing systems, but this information is difficult to find outside of specific communities. integrate this knowledge base into a single system.

Database Query Interface: natural language query interface to databases. most challenging. allow users to query databases using spoken language, eliminating the need to learn query languages. LLMs good at creating complex SQL or Elasticsearch queries.

These projects involve fine-tuning existing LLMs rather than building them from scratch. better results than searching through the systems manually. need for people to work on these projects,. Ed proposed to email individuals to join the efforts. Gabriel Faccini possibly.

Kaushik proposed to test Google fine-tuning.

13:00 AskPanDA: Paul Nilsson

Trained ChatGPT to answer questions from Panda documentation and instructed to answer only questions on Panda except BNL and CERN. The problem with OpenAI is that you can only send text chunks of 15 kilobytes, which is nothing. The Panda documentation is 600 kilobytes. So, the script divides the text into smaller chunks. Paul tried two different models. GPT 3.5 och GPT4. The system can generate the proper URL for users. He imagines an ecosystem of LLMs training for different things (Panda, Rucio, etc.) and talking about each other. Paul would like to also look into other technologies like this fine-tuning of Gemini. The current implementation only uses documentation and does not query Panda.

We cannot use the regular LLMs because the problem is that regular LLMs have a lot of information that's randomly picked up from random places. Because of that, it constructs answers by juxtapositioning answers from different sources. So the general LLM models are helpful to start a search, but they're not very good at giving precise and correct answers.

The current ChatGPT implementation cannot be learned from the interaction with the user. Gemini can do it.

14:30 Ilya: Experience with UChicago AF support using ChatGPT

Ilya discussed using a chatbot to answer user questions about analysis facilities.

Existing AI chatbot for SLAC, BNL, and UC:

Users often have questions about submitting jobs, logging in, and storage space.

The chatbot uses documentation to answer common questions and can also answer runtime questions using Elasticsearch

Documentation cleaning and organization was a significant effort. Documentation was in multiple formats and locations, and some of it was clashing.

Real-time questions are answered using retrieval augmented generation, with data remaining in Elasticsearch.
Functions are provided for the AI to call to access Elasticsearch data. The system uses examples of Elasticsearch queries and data to help the AI understand how to create queries.

Debugging the system can be difficult, as the LLM might try to work around bugs

Users may not know what kind of questions to ask the system.

The system needs to be able to handle non-trivial questions. For example, if a user asks who the biggest user is, it's not straightforward to define "biggest."

Plans include creating a user specifically for the assistant so it can run code and behave like a regular user and test for system issues
.
Questions on a future development of a chatbot

The presentation also discussed the questions they would like AI to answer: Is my site running smoothly? Are there any problems with my site? Are my jobs running correctly? What's the status of this task? What problems are associated with this task? How many slots with 32 GB per core are available?

Three levels of AI:

Using a "band-aid" approach, detecting and understanding issues but not solving them. The "band-aid" approach for monitoring questions is easy to implement. It involves creating a web front end and functions for the AI assistant. This approach would not require new infrastructure and is cheap and easy. The main drawback is that users might not know what questions to ask.

The "band-aid" approach for answering questions about problems with tasks is much more difficult. It would require ingesting and processing logs from multiple systems. It would require time synchronization, noise reduction, error code mapping, and anomaly detection. This would require a custom-tuned model, significant effort, and hardware.

High-level AI would interpret what physicists want and deliver only the needed data. This approach would require a filtering system and a formal query language. It may be hard to get users to adapt.

Ilya estimates that the first version of the monitoring system can be done in two weeks, the system for understanding tasks and why they fail would take three years, and agents that help systems would take two years. The high-level AI can be done in several months.

Discussion about whether users should be informed about the storage of their queries.

Human sign-off is needed for big destructive operations

Torre's Presentation on EIC Software and Computing:

Torre discussed the EIC (Electron-Ion Collider) and the EPIC (Experiment at the EIC) experiment. The EIC will have two host labs: Brookhaven and Jefferson Lab. All collision event data will be streamed to computing facilities at JLab and BNL. The computing model is called the "butterfly model," with symmetry in how computing is handled at the two labs. Echelon 0 is the detector and DAC (data acquisition system). Echelon 1 is the pair of facilities at the host labs that perform prompt processing and archiving. Echelon 2 and 3 are like Tier 2 and Tier 3 in the LHC. The computing scale is comparable to ATLAS, with a million core years to process a year of data. Storage scale is expected to be 350-400 petabytes per year.

Both Echelon 1 facilities will receive 100% of the data. Data will be sent in time frames (0.6 milliseconds) aggregated into super time frames. Data will be stored in object stores. Rucio will be used for prompt processing. Panda will be used for workload management. The super time frame concept was discussed with Cedric and Dan Vanderster, and it works well with Rucio and object stores. Metadata for the objects will come from Rucio.

AI/ML is integrated into EPIC from the start. AI/ML is used to understand the difference between real data and simulation. AI is being used for detector design optimization. AI/ML is also being used to maximize proton polarization. Redwood will deliver the first AI/ML applications in Panda.

Rucio is now operating at JLab, and BNL is close to completing an instance for EIC use. A Panda instance is now at BNL.

Thu Jan 23

09:00 Mini-data challenges and system modelling

Present : FredS, Doug, Shawn, Chris, Alexei, Shigeki, Raees, Xin, FredL, Ofer, Rafael, Tasnuva, Torre, Paul, Timo, Rui, DavidS, Mario, JohnDS, Ed, Ivan, Andreu, Rod, DavidR, Riccardo, Horst, Lincoln, Judith, Ilija, Hiro

Remote : Eddie, AlexanderA, Tanya, Fa-Hui, Petr, Aidan, Johannes, Tony

Simulation HEP workloads with SimGrid - Fred Suter (ORNL)

SimGrid project in a nutshell
SimGrid and HEP workloads
Mario : today Rucio managers 15M file transfers/day (MartinWillSImGrid - 2009 - modelling for 1.5M files transfer)
DCSim - attempt to simulate CMS workflows (for KIT)

- Doug : is it extendable for cache simulation and network simulation ?
    - Yes, it is possible
- Did you try this tool for data placement simulation ?
    - Yes, if you can define your data placement strategy
        - Mario : this was MartinB PhD thesis

SimGrid and REDWOOD - jobs brokering strategy
WRENCH - toolkit to write simulators (for instance, to simulate HTCondor or XROOTD)
DCSim - planning a computing and storage infrastructure for HEP (KIT contribution)

All components are available and can be used

Xin : How SimGrid in comparison with other simulation products ?
SimGrid stopped to do it, story about SimGrid and GridSim

System Modelling - Paul

Latest development in REDWOOD
Simulator for ATLAS using SimGrid is in progress (WRENCH is not the option)
Simulation of errors : using real PanDA data
Status and plans : move from simple to realistic brokering, errors injection

DavidR : how sites are simulated ? on the level of WN; Shawn : how LAN topology is described (and simulated) Fred : it is possible to do it, if topology is described + logs

DavidS : how can this work impact TDR ?
Discussion about what level of simulation details should be done for sites
Rod : can simulation be done for 2 sites ?

US Mini Data Challenge - Shawn, Hiro
Shawn :

Sites reports and status; OU participation in the next test and RSE

Analysis of files in BNL Tape system & NANO Data Challenges - Hiro
Xin – one question asked was “how much optimization should we do?” I would say this is eventually up to site to decide. Our ultimate goal is to meet the target *delivered* tape bandwidth (BW) ATLAS wants from sites. The goal of smart writing is to improve BW on a per tape basis, (while keeping balance with parallelism), so that sites can reach that goal with fewer drives. But, if a site can afford over-subscription of drives, then maybe they can meet the goal without high BW efficiency on a per tape mount, if that’s the site decision, that is fine. In the latter scenario, instead of higher per tape mount efficiency, site may want to try to keep all available drives engaged all the time, even though some tape mounts will just read a small amount of data. Again, if that’s how a site chooses, fine with ATLAS as long as the target BW is met.

TIM Preparation (notes, thoughts, participants)

We will keep this as a reference for future TIM organizations.

Focus topics and invited talks (preliminary)

Mario: Focus on where we want to be for HL-LHC. How to bring our community together and jointly work on activities (rather than many separate silos)?

Mario: Automating our day-to-day work to enable effort to enable future-looking activities

Agenda

======
1st day (½ + ½) (WG: Rod, Fred, Ivan, Judith, Paul,..)

Distributed Computing issues and challenges
    Errors studies
Operations and Automation
Outcome: list of qualification projects (with timeline and needed effort)

2nd day (afternoon) (WG: Andreu, Ilija, Timo,...)
    Documentation: Role of ML and GenAI to detect ATLAS sites issues (GPT4ADC)

3rd day (morning)
    WG (Shawn, Hiro) : DC(26) mini-challenges / ramp-up challenges
        - US and UK mini-challenges
        - NANO tests available to other sites and clouds on demand.

    (FredS, Mario, Alexei) : Modeling and HT workflows tracks (splinter meeting)

- Distributed computing and site operations
    - Monitoring TF first report [?]
    - Site evolution
        - ARM, Cloud, …
        - Memory optimization (RSS experiences)
    - Role of ML and GenAI
- <selected> HL-LHC demonstrators and R&Ds
    - Tokens, future HPC integration, smart writing, and data grouping
- DC(26) mini-challenges / ramp-up challenges
    - US and UK mini-challenges
    - NANO tests are available to other sites and clouds on demand.
- Monitoring/Analytics
- REDWOOD: Modeling and HT workflows tracks (maybe for plenary session talk or/and splinter meeting ?)
- Analysis Facilities
- Storage to tape
    - Metadata to facilitate tape grouping
    - Setting limits to tape storage to avoid writing over the pledge (PIC, TRIUMF,...)
    - MoU weights to data stored on tape in addition to T0 export.
- ARC-CE/aCT token support and the future of aCT
- Invited talks
    - EIC and ePIC SW&C - Torre
    - AI/ML @BNL and SUNY

Timeline

Day 1: Tue Jan 21

- 14:00 - 17:30 Plenary session (with Zoom connection)

Day 2: Wed Jan 22

- 09:00 - 13:00 Plenary session (with Zoom connection)
- 13:00 - 14:00 lunch
- 14:00 - 17:30
    - Splinter meeting 1 : Distributed Computing and Sites Operations
    - Splinter meeting 2
- 19:00 TIM dinner

Day 3: Thu Jan 23

- 09:00-11:00 Plenary session (with Zoom connection)
    - Splinter meetings reports
- 11:00-12:30 Wrap up and action items
- 14:00-17:00:
    - SW&C / ADC / US ATLAS SW&C coordinators visit to BNL
        - Visit BNL and meet with the SCDF division and NPPS staff
            - R. Di Maio (CERN) - guest number: 1031C (temporary number: GR234327)
            - M.Lassnig (CERN) - valid BNL access card
            - A. Pacheco Pages (IFAE & PIC) - Guest registration (**5863B**) is valid until March 2026.
            - D.South (DESY) - *Guest Number **5913B***
            - Rod Walker (LMU) - Guest Number **GR234487**
    - List of discussion topics and presentations
        - M.Lassnig / A.Pacheco: ADC plans, milestones and R&D projects
        - SDCC topics
            - Fabrics (Tom Smith)
                - Communication with HT Condor team
            - Storage (Carlos Gamboa)
                - Analysis Workflows and Tier 1 Storage Resources
                - Dynamic Deletion Rate for Saturated Usage of Storage Space Tokens
                - ADC Experience with EOS
                    - SDCC is looking into EOS as an R&D Activity
                - ADC/ Tier 1 envisioned activities for 2025
                    - HPSS
                    - dCache
            - Analysis Facilities (Ofer Rind)
            - Operations (Ivan Glushkov, Doug Benjamin)
        - NPPS topics
        - Discussion
            - ADC R&D projects and sites participation

- Visit to BNL to meet with REDWOOD Track 4 team
  - P.Nilsson (BNL)
  - F.Suter (ORNL) - ORNL badge
  - Raees Ahmad Khan (U Pittsburgh)

Info

- No registration fees
- BNL will pay for coffee breaks
- From Chris :
  - Lunch - I suggest that attendees avail themselves of the campus dining facilities next door to Wang - alternatively, someone goes out with a sandwich order. Ordering lunch through conference services is painful.
  - Dinner: The Taj - Crown of India is opposite campus & certainly big enough. I think they will give me a deal. Other possibilities exist in Port Jeff
  - After meeting, beers - Several bars around, including the Bench next to Stony Brook station, and the Taj also has a lovely bar.

Participants (preliminary, final list will be limited to 35 (?) persons max, registered)

1. Chris Bee
2. Doug Benjamin (yes)
3. Lincoln Bryant (yes)
4. Rafael Coelho Lopes De Sa
5. Kaushik De
6. John DeStefano (JD)
7. Carlos Gamboa
8. Ivan Glushkov (yes)
9. Hironori Ito
10. Alexei Klimentov
11. Mario Lassnig (yes)
12. Fred Luehring (yes)
13. Verena Martinez Outschoorn
14. Shawn McKee (yes)
15. Shigeki Misawa
16. Paul Nilsson
17. Andres Pacheco Pages (yes)
18. Ofer Rind (yes)
19. Horst Severini
20. Oxana Smirnova
21. David South
22. Judith Stephen (yes)
23. Ilija Vukotic

24. Rod Walker(yes)
25. Torre Wenaus (yes)
26. Xin Zhao (yes)
27. Fred Suter (ORNL) - if we want to have distributed computing modeling as a focus topic or splinter meeting
28. Riccardo Di Maio
29. Timo Wilken

Please confirm participation:

1. Chris Lee (need to ask the powers that be, and check my visa)
2. Tom Smith

Remote participants

1. Sasha Alekseev (via Zoom only - no US visa)
2. Paolo Calafiura
3. Alessandra Forti
4. Tania Korchuganova (via Zoom only - no US visa and no hope to get it)
5. Fernando Barreiro
6. Philippe Laurens (no, but can connect remotely)
7. Eddie Karavakis (not 100% sure due to pregnancy due date but I might be able to make it to the meeting if needed)
8. Fa-Hui Lin
9. Fabio Luchetti
10. Tadashi Maeno
11. Rob Gardner (cannot attend)

Not coming

1. Misha Borodin
2. Martin Barisits
3. Dimitrios Christidis