ATLAS Metadata Handling and AMI Wokshop Highlights

Alexei Klimentov

BNL

Introduction

- The workshop was organized by ATLAS Distributed Computing in consultation with Data preparation and Physics Coordination.
- External members were invited, Jean-Jacques Blaising (LAPP), Predrag Buncic (CERN), Jérome Lauret (BNL);
- The workshop was chaired by J-J Blaising.
- The agenda and presentations are available at:
  - http://indico.cern.ch/conferenceOtherViews.py?view=cdsagenda&confId=99048

- Excellent organization and friendly atmosphere
  - Many thanks to Solveig

Workshop agenda. Day 1.

Meta-data: point of view of physics (20')

Borut Paul Kersevan (JSI) , Junji Tojo (KEK)

Meta-data experience in other collaborations (40')

Jerome LAURET (BNL) , Predrag Buncic (CERN)

Meta-data and data preparation (20')

Andreas Hoecker (CERN)

Meta-data: technology overview (20')

David Malon (ANL)

Meta-data and distributed analysis (20')

Johannes Elmsheuser (Ludwig-Maximilians-Universität München)

ATLAS computing overview (10')

Alexei Klimentov (BNL)

AMI and its place in the ATLAS metadata infrastructure. (20')

Solveig Albrand (LPSC)

Data and metadata in distributed computing (20')

Simone Campana (CERN/IT/GS) , Vincent Garonne (CERN)

Workshop Agenda. Day 2.

- After the presentations, a summary session involving the external members, Solveig Albrand, Dario Barberis, Alexei Klimentov and Massimo Lamanna established the item list for the day-2 discussions.
- The second day was dedicated to list findings, requirements, technical requests and wishes.

Workshop Findings, Observations, Recommendations.

- Three groups
  - Metadata TaskForce Document (MDTF)
  - System architecture
  - Coordination issues

Observations. MDTF

- The last collection of Meta-Data requirements was done in 2006; it was documented in the MetaData Task Force Document (MDTF).
- The list of requirements was not updated; there is no recent requirements list document which gives a coherent picture of ATLAS Meta-Data sources and destination.
  - The MTDF document is not up-to-date:
    - For instance, end user data analysis use case isn't addressed in it
    - There are many new requirements from Distributed Analysis, Data preparation, DDM and they need to be addressed with the high priority

Recommendations. MDTF.

- MDTF document should be updated and it should be a life document
- Collecting requirements should be separated from tools development
- Define procedure for
  - Meta-data requests submission and the way they are documented
  - Meta-data requests implementation
    - Update tools and preserve overall coherence

Observation. Architecture

- Requirements collection should be separated from tools development and implementation
- There are many tools, projects, repeated attempts to handle meta-data
- It is not clear if the meta-data system architecture have been thought enough; there is no architecture description document.
  - There is no clear separation between data collector and data presentation and API. As a result there is no a coherent view
- There is no a single tool (gateway, interface) allowing a user to query and navigate between meta-data
  - Experienced user is getting information from many sources using various tools.
  - Sometimes information is incoherent
  - Often users and groups use Twiki pages as source of meta-data
  - Meta-data interface looks a bit technical for the end user

Recommendations. Architecture

- System architecture
  - Architecture should be user centred
    - One should be able to navigate to all **essential** meta-data related to a data sample (dataset).
  - Separation of interface and presentation layer is recommended
    - A single point of entry for data access and navigation
- Data aggregation in one database versus integration and data caching need to be addressed. It is a technical task to find a balance between aggregation and integration
- Meta-data system architect is needed (within ATLAS Distributed Computing)

Observations and Recommendations: Metadata Coordination

- Meta-data coordination issues
  - It looks communication channel between Physics community and developers is missing
  - Coordination is considered as a part-time job
  - Co-coordination is tricky

Recommendation :

Meta-data coordinator mandate and role should be defined in more details; the coordinator should not be responsible of implementation.

Post-workshop Steps.

- Conveners report was sent to the TOB and discussed on TOB session 15 November
- Data preparation and Computing should write a draft proposal of Metadata Coordinator and Metadata Architect mandate.
  - Drafts were written by Beate and Alexei
    - Comments from SW&C , ADC coordination are collected
    - Both drafts will be sent to the TOB chair