

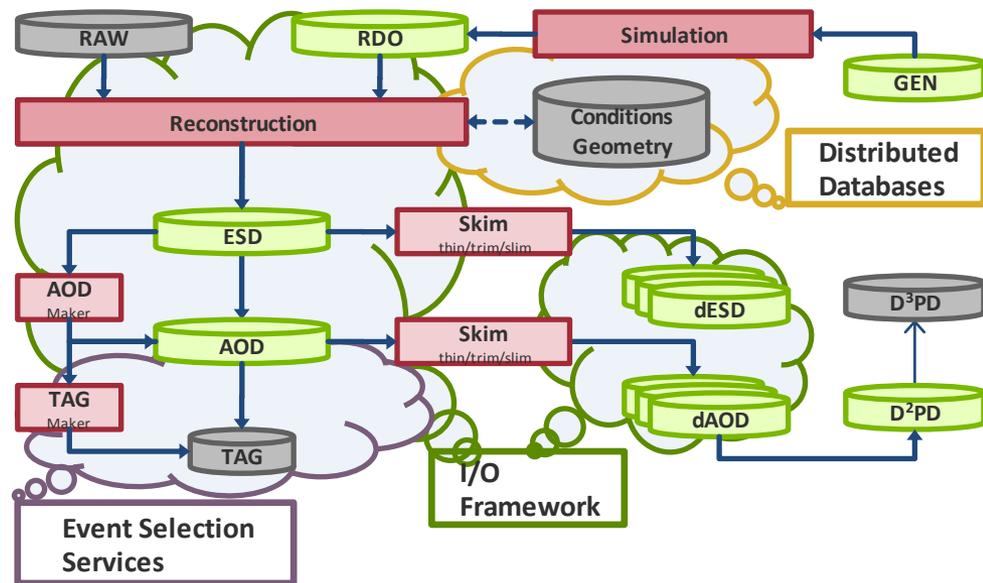
Software and Computing

D.O.E. Energy Frontier Review

Peter van Gemmeren

Outline

- The group and what we do
- Selected accomplishments
- Ongoing roles and activities
- Future
- Summary



The ATLAS Software Group

- Argonne has been an integral part of international and U.S. ATLAS computing programs since their respective inceptions, since 2005 is at the level of approximately 5 FTE (Malon, Cranshaw, van Gemmeren, Vaniachine, Zhang).
- Holds lead responsibility in the international ATLAS collaboration for: Design, Implementation and Operation of a **distributed data store and supporting infrastructure** for the ATLAS experiment with more than 100 PB of data, including:
 - Metadata,
 - Analysis I/O performance,
 - I/O Framework and persistence,
 - Event-level selection and navigation infrastructure
 - Persistence evolution and support of multi/many-core I/O for emerging architectures
 - Event Reprocessing,
 - Responsible for the job definition code in the ATLAS production
 - Web services and event-level metadata infrastructure,
 - Integration of metadata from a variety of sources
 - and integration of many of the above services with analysis tools.

Accomplishment: In-File Metadata

- Argonne proposed, designed and implemented an infrastructure to write metadata directly into event data files.
 - Went from 'would be nice' feature to mandatory even for files without events.
 - Metadata includes, sample summary, luminosity, detector conditions and more
- Used for:
 - optimization (reduce database lookups), **User-friendly**
 - automatic job configuration (to avoid user errors) and **Correctness**
 - E.g.: Job is configured for the correct center-of-mass energy automatically.
 - bookkeeping.
 - E.g.: so physicist know what luminosity their filtered event sample corresponds to.
- Metadata (an ANL responsibility) is becoming more important, as the size and complexity of the ATLAS event store increase.
- At the same time, metadata creates challenges for exploiting parallelism:
 - Metadata violates the 'embarrassingly parallel' structure of event processing.
 - During file merging, metadata needs to be summarized, not just appended (like event data).

Accomplishment: Transient / Persistent Separation of Data Model

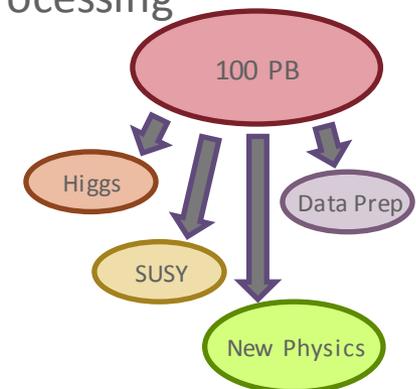
- Argonne advocated and ATLAS has adopted the principle that when one is building a scientific data store at scales of hundreds of petabytes, one should have a persistent data model.
 - ANL group developed a transient/persistent separation layer.
- ATLAS has seen the benefits of transient/persistent separation in several areas:
 - Support for a **transient model too complex** for our underlying persistence technology to handle directly **Robustness**
 - i.e. transient object could not be persisted by ROOT, because it relied on very advanced C++ features, but corresponding persistent representation could.
 - Schema **evolution** beyond the capabilities of the underlying technology **Longevity**
 - transient/persistent remapping layer allowed backward compatibility.
 - Today, ATLAS relies on that capability when reading old data with new software.
 - Improved I/O **performance**: **Performance**
 - storage footprint and read/write speed for every data type in every data product (RDO, ESD, AOD...).
 - Currently, many changes to persistent data model are driven by performance, not transient type. With 100 PB of total data, these improvements save important resources.

Accomplishment: Storage Layout Optimization

- By optimizing the physical data layout in root to better match the transient retrieval granularity, Argonne achieved improvements in I/O performance:
- Read **speed** up by 20 - 30% for sequenced reading
- Write **speed** up by almost 50%
- Saves more than 100 MB of **memory (RAM)** per core.
 - Memory is a critical resource limitation especially for multicore processing.
- Individual event retrieval (e.g. via direct navigation but also in multi-core event processing) 4 - 5 times **faster**.

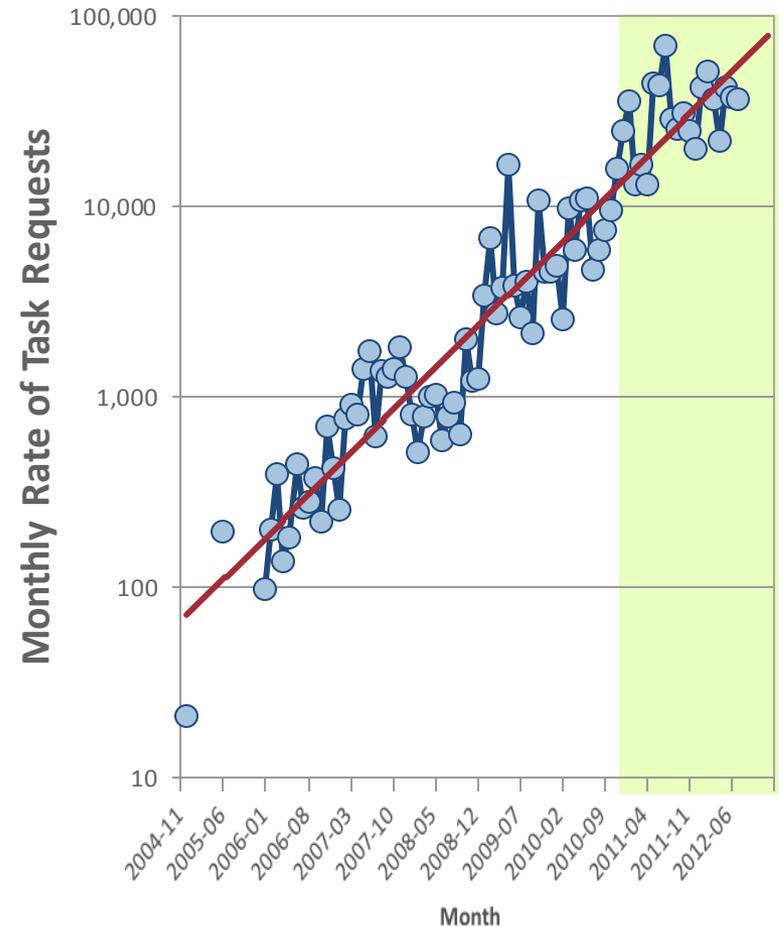
Event-level Metadata and Selection Services

- Argonne has led the development of ATLAS event-level metadata infrastructure and associated services
- Supports identification and selection of events of interest based upon key quantities, along with sufficient navigation information to locate and retrieve those events at any production processing stage
- Underlies event picking undertaken routinely in ATLAS for a wide range of purposes
 - Interesting or anomalous or problematic events, selections for event displays, ...
- Routinely used in monitoring by Data Preparation, and for fast physics monitoring
- Used in production for heavy ion skimming of raw data for reprocessing
 - To select which 5% of events to reprocess, without even decompressing raw data
- Can demonstrably support physics skims (Higgs, SUSY, ...)
 - Increasingly viable in practice as quantities upon which selections are based stabilize



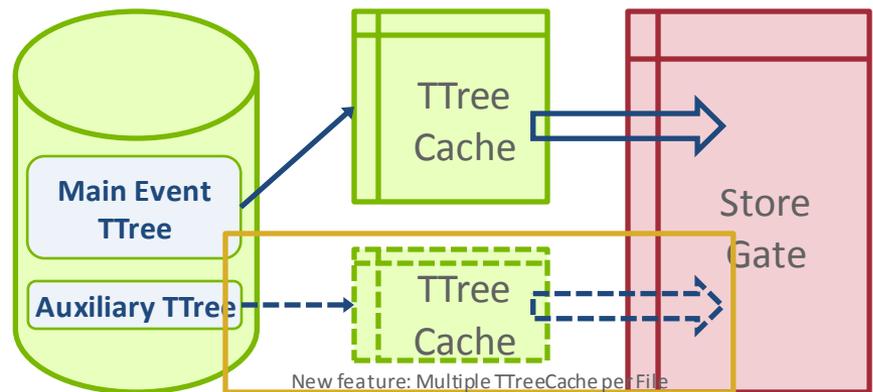
Ongoing Activity: Event Store Processing

- Since Argonne acquired major responsibilities in ATLAS Event Store processing, the demand for PanDA production system tasks has tripled.
- Argonne accomplishments in Grid Data Processing:
 - 2010: Scalable database access technologies for petascale data processing on the Grid
 - 2011: Leading to completion four ATLAS re-processing campaigns
 - 2012: Migration of ATLAS Trigger Reprocessing to PanDA



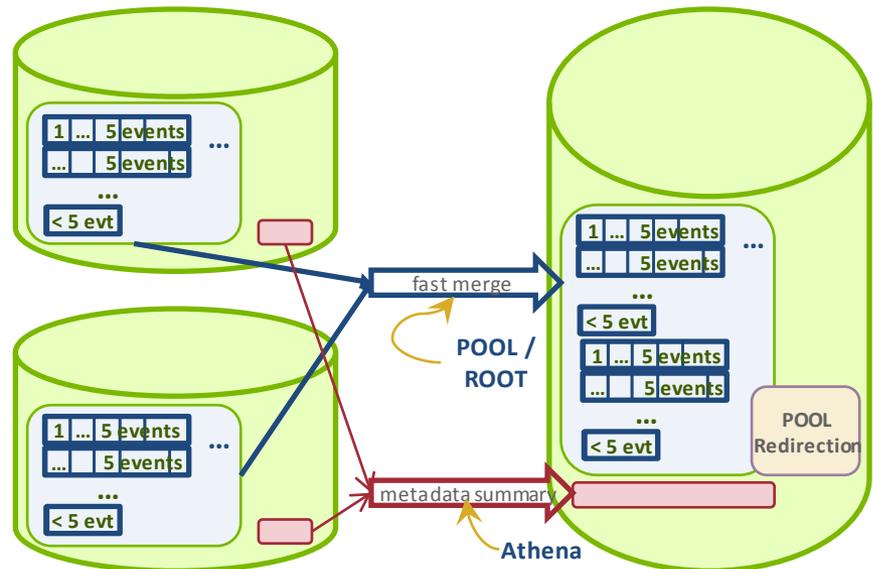
Ongoing Activity: Analysis and ROOT I/O

- Physics analysis often uses D³PD (flat ROOT ntuple) as input.
- In Fall 2011, a group led by Argonne was formed to measure and increase analysis I/O.
- Objectives:
 - Provide ‘benchmarking’ tools for I/O performance for analysis code.
 - Develop expertise in tuning I/O for analysis code.
 - Provide tools to facilities for analyzing option to improve analysis I/O.
- Argonne is working directly with the ROOT I/O experts, making code contributions for features which are of special interest of ATLAS.
 - E.g.: TTreeCache, read optimization
- Helps broaden ROOT expertise within the group.
 - ROOT is used by most physicists to do their analysis.



Ongoing Activity: I/O Components for Parallel Event Processing

- ANL implemented **extensions to the existing I/O framework** to support the first implementation of a multi-core event processing framework.
 - The initial deployment of relies on serial I/O components being run in parallel.
 - Each worker process produces its own output file, which need to be merged after all workers are done.
 - Huge bottleneck, reduced the event throughput by about 30% (vs. no merge).
- Just the **first step**, wide scale deployment will require dedicated parallel I/O components.



- New utility using:
 - “Fast Event Data Append” and reference redirection layer
 - Athena in-file metadata propagation
- This reduces the merge time by almost an **order of magnitude**.

Software evolution and upgrades

- Core counts are increasing in computing platforms, and I/O infrastructure must support these increasingly-many-core architectures
 - even as I/O bandwidth has not been scaling at comparable rates
- Storage technologies and data hosting models are evolving
 - And optimization of data organization and access depend upon such factors
- ATLAS computing must evolve both to take advantage of evolving computing technologies and to support the challenges of LHC and ATLAS upgrades
 - Trigger rates are increasing with each succeeding upgrade
- I/O, data handling and persistence, and the metadata to discover, identify, locate, navigate to, and access data of interest will play an ever more vital role
 - This is where the Argonne software and computing focus and recognized expertise lie, and where the group holds leadership responsibility within the collaboration
 - The group is actively engaged both technically and editorially in ATLAS computing upgrade planning, and in associated R&D

Summary: Contributions Past, Present, and Future

- The Argonne group has made **fundamental contributions to ATLAS core software**, well recognized within the collaboration, leading design, development, delivery, and support of a scalable, distributed, multi-petabyte ATLAS event store and the I/O, persistence, and metadata infrastructure that underlie it.
- Ongoing Argonne efforts and expertise are **vital to ATLAS physics**, supporting this infrastructure and its evolution, adapting it to evolving computing platforms, and continually delivering improvements both in functionality and in performance.
- The group is integrally engaged in **defining the future of ATLAS software** and computing
 - In which the data infrastructure will only grow in importance