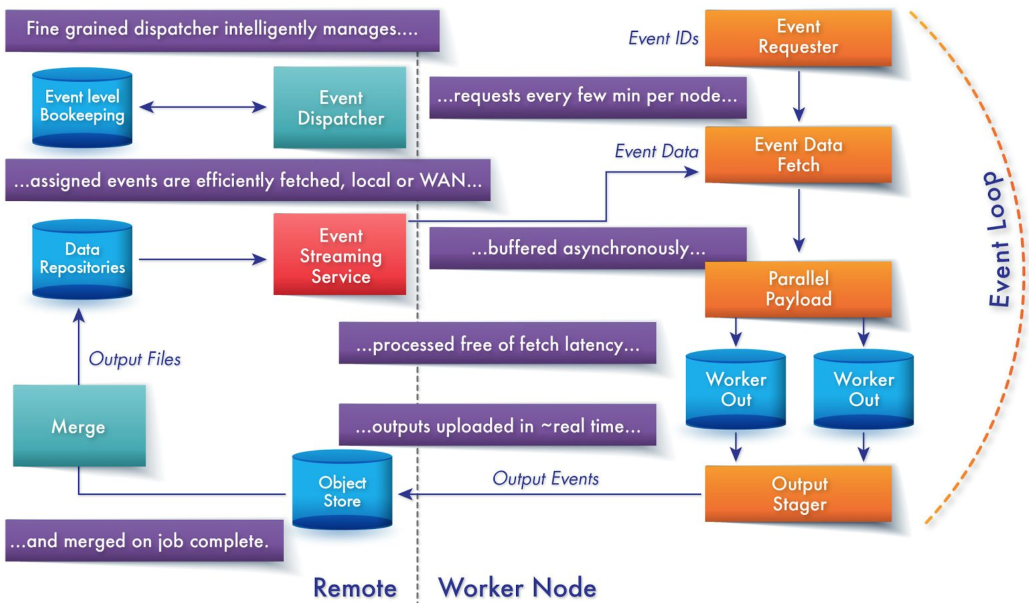
Science-Aware Dynamic Data Delivery at the Exascale

M Ernst, A Klimentov, N Malitsky, T Wenaus (BNL), P Calafiura (LBNL), D Malon (ANL), R Mount (SLAC), S Jha (Rutgers), K De (UTA)

The ATLAS Experiment at the LHC has represented the extreme limit of data intensive scientific computing since its physics program began in 2010. Its growing data set is approaching 200 PB, with data processing consuming about 4M CPU hours/day, over an exabyte of data processed per year at about 140 computing centers globally. ATLAS' compute limited science is driving it to extend its computing to new resources including LCFs and HPCs, cost-effective commercial clouds, and volunteer computing. Using such resources fully and efficiently demands rethinking the processing workflow, which led ATLAS to develop the ATLAS Event Service (ES) supporting agile, fine grained workflows that can flexibly shape themselves to available resources, depositing their outputs in fine grained object stores in near real time.



While the fine grained Event Service enables large gains in the efficiency and flexibility with which processing resources are utilized, it also opens new potential for the efficient use of storage and networking. Fine grained, quasi continuous workflows offer the possibility to move beyond conventional file based data handling to a similarly fine grained approach that better relates the data to the science, streaming data from distributed stores to scientific computation in a 'science aware' manner. Data delivery can be informed by a deep knowledge of data attributes imparted by the scientists, and by dynamic resource knowledge to optimize efficient access and storage use, lowering costs in what can be the most expensive component of data intensive computing. Such an 'Event Streaming Service' (the red component in the figure), or more generally a Data Streaming Service (DSS), can provide an intelligent, dynamic bridge between distributed data and the scientific computing workloads consuming them. Its intelligence arises through its pairing with a 'Data Knowledge Base' (DKB) capturing the scientists' knowledge of data types, associations, contexts, life cycles, and analysis roles.

We are beginning to develop such a system for ATLAS to provide science aware dynamic data delivery at the exascale, comprising an Event Streaming Service coupled with a Data Knowledge Base. Should we receive support to do so, we will pursue the development in a more general fashion as a new science data streaming infrastructure engaging science communities beyond ATLAS, such as the new generation of light source facilities including BNL's NSLS II.

This distributed streaming infrastructure will be novel in its ability to deliver fine grained and optimally selected data efficiently and transparently to scientific applications and clients, applying knowledge of the nature, structure, usage characteristics and availability of exabyte scale scientific data samples. Data requests will be expressed in the abstract terms of scientific data objects, independent of physical storage particulars. Requests will be resolved through the DKB together with external data management and cataloging services into file and object requests. The requested data can then be optimally delivered by using the DSS' dynamic awareness of available resources to determine the most appropriate resource and mechanism. Data streaming will be the key mechanism, mixing local and remote streaming based on sensor driven estimation of computing speed vs network bandwidth, mixing in also dynamic caching and selective use of replication. The approach can also support 'virtual data' computed on demand, streaming data that is recreated on the fly.

Designing a DSS that achieves near optimal resource usage and requires little supervision, while providing each scientist fast access to his/her data, is a challenging optimization problem in a high dimensional space that includes user queries, data and metadata patterns, resource classification, and resource status. By applying machine learning to collected information, we will identify usage patterns that will inform how we optimize data clustering and replication, anticipate and prepare for requests, and suggest data sets relevant for the scientist's research.

The work leverages our team's leadership and experience at the largest scales of data intensive distributed scientific computing at the LHC, together with systems of proven scalability found in open source and commercial web based services, to combine leading edge innovation with proven approaches. Our experience confers knowledge of the requirements and problem space such that we can evaluate and select technologies and approaches in an informed manner. Our extreme-scale science enables at-scale testing and deployment throughout our tightly iterated development cycle.

Such a system can be the Big Data scientist's agent and agency by which heterogeneous, distributed, dynamic federated resources become readily usable at a massive data scale, with a low barrier of entry, low operational complexity as seen by the user, and lowered costs through the system's favoring network-heavy streaming over storage-heavy replication. A first prototype is foreseen in 2016, with a scope that will depend on the supportable effort level.