Converging High-Throughput and High-Performance Computing: A Case Study

ABSTRACT

The computing systems used by LHC experiments has historically consisted of the federation of hundreds to thousands of distributed resources, ranging from small to mid-size resource. In spite of the impressive scale of the existing distributed computing solutions, the federation of small to mid-size resources will be insufficient to meet projected future demands. This paper is a case study of how the ATLAS experiment has embraced Titan – a DOE leadership facility in conjunction with traditional distributed high- throughput computing to reach sustained production scales of approximately 51M core-hours a years. The three main contributions of this paper are: (i) a critical evaluation of design and operational considerations to support the sustained, scalable and production usage of Titan; (ii) a preliminary characterization of a next generation executor for PanDA to support new workloads and advanced execution modes; and (iii) early lessons for how current and future experimental and observational systems can be integrated with production supercomputers and other platforms in a general and extensible manner.

KEYWORDS

Distributed computing, Workload management system

ACM Reference format:
. 2017. Converging High-Throughput and High-Performance Computing: A Case Study. In *Proceedings of The International Conference for High Performance Computing, Networking, Storage and Analysis, Denver, Colorado, USA, Nov 2017 (SC2017),* 13 pages.
DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The Large Hadron Collider (LHC) was created to explore the fundamental properties of matter for the next decades. Multiple experiments at LHC have collected and distributed hundreds of petabytes of data worldwide to hundreds of computer centers. Thousands of physicists analyze petascale data volumes daily. The detection of the Higgs Boson in 2013 speaks to the success of the detector and experiment design, as well as the sophistication of computing systems devised to analyze the data.

Historically, the computing systems used by LHC experiments consisted of the federation of hundreds to thousands of distributed resources, ranging in scale from small to mid-size resource [19]. Although the workloads to be executed are comprised of tasks that are independent of each other, the management of the distribution of workloads across many heterogeneous resources to ensure the effective utilization of resources and efficient execution of workloads presents non-trivial challenges.

Many software solutions have been developed in response to these challenges. CMS, one of the LHC experiments, devised a solution based around the HTCondor [40] software ecosystem. The ATLAS [1] experiment, utilizes the Production and Distributed Analysis (PanDA) workload management system [25] (WMS) for distributed data processing and analysis. The CMS and ATLAS experiments represent arguably the largest production grade distributed computing solutions and have symbolized the paradigm of *high-throughput computing*, i.e., the effective execution of many independent tasks.

As the LHC prepares for Run 3 in $\approx$ 2022 and the high-luminosity era (Run 4), it is anticipated that the data volumes that will need analyzing will increase by factors of 10-100 compared to the current phase (Run 2). Data will be larger in volume but will also require more sophisticated computational processing. In spite of the impressive scale of the ATLAS distributed computing system, demand for computing systems will significantly outstrip current and projected supply.

There are multiple levels at which this problem needs to be addressed: the utilization of emerging parallel architectures (e.g., platforms); algorithmic and advances in analytical methods (e.g., use of Machine Learning); and the ability to exploit different platforms (e.g., clouds and supercomputers).

This paper is a case study of how the ATLAS experiment has "broken free" of the traditional computational approach of high-throughput computing on distributed resources to embrace new platforms, in particular high-performance computers. Specifically, we discuss the experience of integrating PanDA WMS with a DOE leadership computing facility called Titan to reach sustained production scales of approximately 51M core-hours a year. We also discusses the investigation of a task execution runtime system on Titan, that is based on the pilot-abstractions and which allows advanced execution modes of the ATLAS workload as well as the enhanced support for heterogeneous workloads such as molecular dynamics.

This state-of-practice paper provides three main contributions: (i) a critical evaluation of the many design and operational considerations that have been taken to support

SC2017, Denver, Colorado, USA

DOI: 10.1145/nnnnnnn.nnnnnnn

the sustained, scalable and production usage of Titan for historically high-throughput workloads; (ii) a preliminary characterization of a next generation executor (NGE) for PanDA to support non-traditional heterogeneous workloads and execution modes; and (iii) early lessons and guidance as the community looks forward to designing the next generation of online analytical platforms [5], for how current and future experimental and observational systems can be integrated with production supercomputers in a general and extensible manner.

2 PANDA OVERVIEW

PanDA is a Workload Management System (WMS) [28] designed to support the execution of distributed workloads and workflows via pilots [41]. Pilot-capable WMS enable high throughput of tasks execution via multi-level scheduling while supporting interoperability across multiple sites. This is particularly relevant for LHC experiments, where millions of tasks are executed across multiple sites every month, analyzing and producing petabytes of data. The design of PanDA WMS started in 2005 to support ATLAS. PanDA went into production for the LHC Run 1 on 2009, and was then extended with new subsystems for Run 2 in 2015.

2.1 Design

PanDA's application model assumes tasks grouped into workloads and workflows. Tasks represent a set of operations performed on data sets stored in one or more input files. Tasks are decomposed into jobs, where each job consists of the task's operations performed on a partition of the task's data set. Since 2005, a certain amount of parallelism has been progressively introduced for job execution [16] but, so far, HEP applications have not required a message passing interface (MPI).

PanDA's usage model assumes multitenancy of resources and at least two types of HEP users: individual researchers and groups executing so called 'production' workflows. PanDA's security model is based on separation between authentication, authorization and accounting for both single users and group of users. Both authentication and authorization are based on digital certificates and on the virtual organization abstraction [20].

Currently, PanDA's execution model is based on four main abstractions: task, job, queue, and pilot. Both tasks and jobs are assumed to have attributes and states and to be queued into a global queue for execution. Prioritization and binding of jobs are assumed to depend on the attributes of each job. Pilot is used to indicate the abstraction of resource capabilities. Each job is thought to be bound to one pilot and executed on the site where the pilot has been instantiated.

In PanDA's data model, each datum refers to the recorded or simulated measurement of a physical process. Data can be packaged into files or other containers. As with jobs, data have both attributes and states, and some of the attributes are shared between events and jobs. Raw, reconstruction, and simulation data are assumed to be distributed across multiple storage facilities and managed by the ATLAS Distributed Data Management (DDM) [21]. When necessary, datasets required by each job are assumed to be replicated over the network, both for input and output data.

PanDA's design supports provenance and traceability for both jobs and data. Attributes enable provenance by linking jobs and data items, providing information like ownership or project affiliation. States enable traceability by providing information about the stage of the execution in which each job or data item is or has been.

2.2 Implementation and Execution

The implementation of PanDA WMS consists of several interconnected subsystems, most of them built from off-the-shelf and Open Source components. Subsystems communicate via dedicated API or HTTP messaging, and each subsystem is implemented by one or more modules. Databases are used to store eventful entities like tasks, jobs and input/output data, and to store information about sites, resources, logs, and accounting.

Currently, PanDA's architecture has five main subsystems: PanDA Server [27], AutoPyFactory [12], PanDA Pilot [29], JEDI [10], and PanDA Monitoring [24]. Other subsystems are used by some of ATLAS workflows (e.g., PanDA Event Service [13]), but we do not discuss them as they are not relevant to an understanding of how PanDA has been ported to supercomputers. For a full list of subsystems see Ref. [38]. Fig. 1 shows a diagrammatic representation of PanDA main subsystems, highlighting the execution process of tasks while omitting monitoring details to improve readability.

During LHC Run 1, PanDA required users to perform a static conversion between tasks and jobs: tasks were described as a set of jobs and then submitted to the PanDA Server. This introduced inefficiency both with usability and resource utilization [11]. Ideally, users should conceive analyses in terms of one or more, possibly related tasks, while the execution manager (i.e., PanDA) should partition tasks into jobs, depending on execution constraints. Further, the static partitioning of tasks into jobs does not take into account the heterogeneity and dynamicity of the resources on which each job will be executed, introducing inefficiencies.

Another problem of static job sizing is that PanDA instantiates pilots on sites with different type of resources and different models of availability of those resources. An optimal sizing of each job should take into account these properties. For example, sites may offer cores with different speed, networking with different amount of bandwidth, and resources could be guaranteed to be available for a defined amount of time or could disappear at any point in time as it happens with opportunistic models of resource provision.

JEDI was deployed for the LHC Run 2 to address these inefficiencies. Users submit task descriptions to JEDI (Fig. 1:1) that stores them into a queue implemented by a database (Fig. 1:2). Tasks are partitioned into jobs of different size, depending on both static and dynamic information about available resources (Fig. 1:3). Jobs are bound to sites with

resources that best match jobs' requirements, and submitted to the PanDA Server for execution (Fig. 1:4).

Once submitted to the PanDA Server, jobs are stored by the Task Buffer component into a global queue implemented as a database (Fig. 1:5). When jobs are submitted directly to the PanDA Server, the Brokerage component is used to bind jobs to available sites, depending on static information about the resources available for each site. Jobs submitted by JEDI are already bound to sites so no further brokerage is needed.

Once jobs are bound to sites, the Brokerage module communicates to the Data Service module what data sets need to be made available on what site (Fig. 1:6). The Data Service communicates these requirements to the ATLAS DDM (Fig. 1:7) that, when needed, replicates data sets on the required sites (Fig. 1:8).

Meanwhile, AutoPyFactory defines PanDA Pilots, submitting them to a Condor-G agent (Fig. 1:9). Condor-G schedules these pilots wrapped as jobs or VMs to the required sites (Fig. 1:10).

When a PanDA Pilot becomes available, it requests the Job Dispatcher module of the PanDA Server for a job to execute (Fig. 1:11). The Job Dispatcher interrogates the Task Buffer module for a job that is bound to the site of that pilot and ready to be executed. Task Buffer checks the global queue (i.e., the PanDA database) and, upon availability, returns a job to the Job Dispatcher. The Job Dispatcher dispatches that job to the PanDA Pilot (Fig. 1:12).

Upon receiving a job, a PanDA Pilot starts a monitoring process and forks a subprocess for the execution of the job's payload. Input data are transferred from the stage-in location (Fig. 1:13) and the job's payload is executed (Fig. 1:14). Once completed, output is transferred to the staging-out location (Fig. 1:15).

The Data Service module of the PanDA Server tracks and collects the output generated by each job (Fig. 1:16), updating jobs' attributes via the Task Buffer module (Fig. 1:17). When the output of all the jobs of a task are retrieved, it is made available to the user via PanDA Server. When a task is submitted to JEDI, task is instead marked as done (Fig. 1:18) and the result of its execution is made available to the user by JEDI (Fig. 1:19).

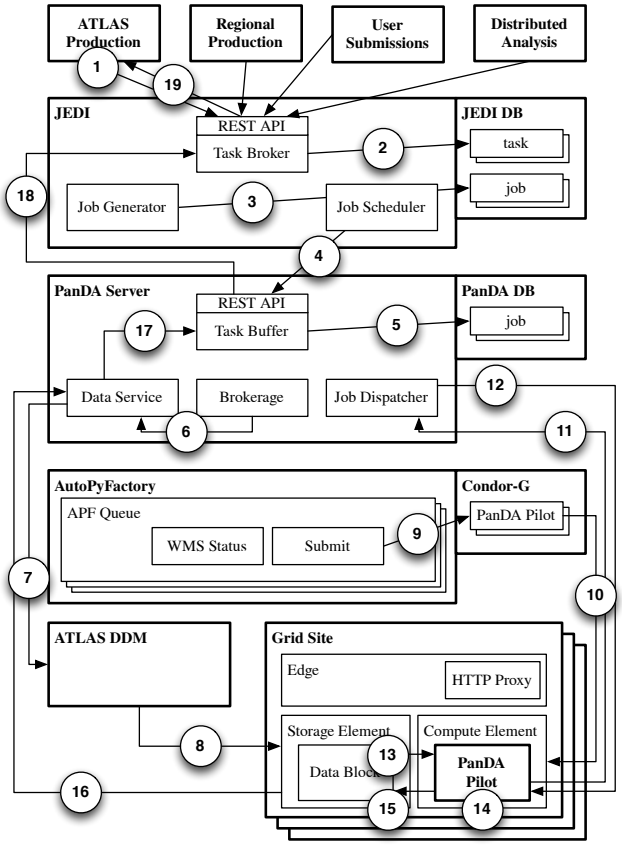

Figure 1: PanDA WMS architecture. Numbers indicates the JEDI-based execution process. Several subsystems, components, and architectural and communication details are abstracted to improve clarity.

3 DEPLOYING PANDA ON TITAN

The upcoming LHC Run 3 will require more resources than the Worldwide LHC Computing Grid (WLCG) can provide. Currently, PanDA WMS uses more than 300,000 cores at over 100 Grid sites, with a peak performance of 0.3 petaFLOPS. This capacity will be sufficient for the planned analysis and data processing, but it will be insufficient for the Monte Carlo production workflow and any extra activity. To alleviate these challenges, ATLAS is expanding the current computing model to include additional resources such as the opportunistic use of supercomputers.

Generally, supercomputers are designed to support parallel computation that requires runtime communication. Jobs are executed across multiple cores, each core calculating a small part of the problem and communicating with other cores via MPI. Accordingly, supercomputers have large number of worker nodes, connected through a high-speed, low-latency dedicated network. Each worker node has multicore CPUs, usually augmented with Graphics Processing Units (GPUs) or other specialized coprocessors.

PanDA WMS has been designed to support distributed Grid computing. Executing ATLAS workloads or workflows involves concurrent and/or sequential runs of possibly large amount of jobs, each requiring no or minimal parallelization and no runtime communication. Thus, computing infrastructure like WLCG have been designed to aggregate large amount of computing resources across multiple sites. While each site may deploy MPI capabilities, usually these are not used to perform distributed computations.

We developed and deployed a single-point solution to better understand the problem space of enabling a WMS designed for HTC to execute production workflows on resources designed to support HPC. The PanDA team developed a job broker to support the execution of part of the ATLAS

production Monte Carlo workflow on Titan, a leadership-class supercomputer managed by the Oak Ridge Leadership Computing Facility (OLCF) at the Oak Ridge National Laboratory (ORNL).

3.1 Architectures and Interfaces

The Titan supercomputer, current number three on the Top 500 list [18], is a Cray XK7 system with 18,688 worker nodes and a total of 299,008 CPU cores. Each worker node has an AMD Opteron 6274 16-core CPU, a Nvidia Tesla K20X GPU, 32 GB of RAM and no local storage, though a 16 GB RAM disk can be set up. Work nodes use Cray's Gemini interconnect for inter-node MPI messaging. Titan is served by the Spider II [30], a Lustre filesystem with 32 PB of disk storage, and by a 29 PB HPSS tape storage system. Titan's worker nodes run Compute Node Linux, a run time environment based on SUSE Linux Enterprise Server.

Titan's users submit jobs to Titan's PBS scheduler by logging into login or data transfer nodes (DTNs). Titan's authentication and authorization model is based on two-factor authentication with a RSA SecurID key. Login nodes and DTNs have out/inbound wide area network connectivity while worker nodes have only local network access. Fair-share and allocation policies are in place both for the PBS batch system and shared file systems.

Titan's architecture, configuration and policies poses several challenges to the integration with PanDA. The default deployment model of PanDA Pilot is unfeasible on Titan: PanDA Pilot is required to contact the Job Dispatcher of the PanDA Server to pull jobs to execute, but this is not possible on Titan because worker nodes do not offer outbound network connectivity. Further, Titan does not support PanDA's security model based on certificates and virtual organizations, making PanDA's approach to identity management also unfeasible. While DTNs offer wide area network data transfer, an integration with ATLAS DDM is beyond the functional and administrative scope of the current prototyping phase. Finally, the specific characteristics of the execution environment, especially the absence of local storage on the worker nodes and modules tailored to Compute Node Linux, require re-engineering of ATLAS application frameworks.

Currently, very few HEP applications can benefit from Titan's GPUs but some computationally-intensive and non memory-intensive tasks of ATLAS workflows can be offloaded from the Grid to Titan's. Further, when HEP tasks can be partitioned into independent jobs, Titan worker nodes can be used to execute up to 16 concurrent payloads, one per each available core. Given these constraints and challenges, the type of task most suitable for execution at the moment on Titan is Monte Carlo detector simulation. This type of task is mostly computational-intensive, requiring less than 2GB of RAM at runtime and with small input data requirements. Detector simulation tasks in ATLAS are performed via AthenaMP [2], the ATLAS software framework integrating the GEANT4 simulation toolkit [4]. These tasks account for $\approx 60\%$ of all the jobs on WLCG, making them a primary candidate for offloading.

Detector simulation is part of the ATLAS production Monte Carlo (MC) workflow [17, 32, 33]. The MC workflow consists of four main stages: event generation, detector simulation, digitization, and reconstruction. Event generation creates sets of particle four-momenta via different generators, e.g., PYTHIA [37], HERWIG [15] and many others. The detector simulator is called Geant4 [4] and simulates the ATLAS detector and the interaction between that detector and particles. Each interaction creates a so-called hit and all hits are collected and passed on for digitalization, where hits are further processed to mimic the readout of the detector. Finally, reconstruction operates local pattern recognition, creating high-level objects like particles and jets.

3.2 PanDA Broker

The lack of wide area network connectivity on Titan's worker nodes is the most relevant challenge for integrating PanDA WMS and Titan. Without connectivity, Panda Pilots cannot be scheduled on worker nodes because they would not be able to communicate with PanDA Server and therefore pull and execute jobs. This makes impossible to port PanDA Pilot to Titan while maintaining the defining feature of the pilot abstraction: decoupling resource acquisition from workload execution via multi-stage scheduling.

The unavailability of pilots is a potential drawback when executing distributed workloads like MC detector simulation. Pilots are used to increase the throughput of distributed workloads: while pilots have to wait in the supercomputer's queue, once scheduled, they can pull and execute jobs independent from the system's queue. Jobs can be concurrently executed on every core available to the pilot, and multiple generations of concurrent executions can be performed until the pilot's walltime is exhausted. This is particularly relevant for machines like Titan where queue policies privilege parallel jobs on the base of the number of worker nodes they request: the higher the number of nodes, the shorter the amount of queue time (modulo fair-share and allocation policies).

The backfill optimization of Titan's Moab scheduler allows to avoid the overhead of queue wait times without using pilot abstraction [3]. With this optimization, Moab starts low-priority jobs when they do not delay higher priority jobs, independent of whether the low-priority jobs were queued after the high-priority jobs.

When the backfill optimization is enabled, users can interrogate Moab about the number of worker nodes and walltime that would be available to a low-priority job at that moment in time. If a job is immediately submitted to Titan with that number of worker nodes and walltime, chances are that Moab will immediately schedule it, reducing its queue time to a minimum. In this paper, we call this number of worker nodes and walltime an available 'backfill slot'.

Compared to pilots, backfill has the disadvantage of limiting the amount of worker nodes that can be requested. Pilots are normal jobs: they can request as many worker nodes and

walltime as a queue can offer. On the contrary, jobs sized according to an available backfill slot depend on the number of worker nodes and walltime that cannot be given to any other job at that moment in time.

At any point in time, the size of an available backfill slot is typically a small fraction of the total capacity of a resource. Notwithstanding, given the size of Titan this translates into a substantial capacity. Every year, about 10% of Titan's capacity remains unused [7], corresponding to an average of 30,000 unused cores (excluding GPU cores). This equals to roughly 10% of the overall capacity of WLCG.

Given the communication requirements of PanDA Pilots and the unused capacity of Titan, PanDA pilot was repurposed to serve as a job broker on the DTN nodes of Titan (Fig. 2). Maintaining the core modules of PanDA Pilot and its stand-alone architecture, this prototype called 'PanDA Broker' implements functionalities to: (i) interrogate Titan about backfill availability; (ii) pull ATLAS jobs and events from PanDA Server; (iii) wrap the payload of ATLAS jobs into MPI scripts; (iv) submitting MPI scripts to Titan's PBS batch system and monitor their execution; and (v) staging and preparing input/output files. Backfill querying, payload wrapping, and scripts submission required a new implementation while pulling ATLAS job and events, and file staging were inherited from PanDA Pilot.

Backfill querying is performed via a dedicated Moab scheduler command while a tailored Python MPI script is used to execute the payload of ATLAS jobs. This MPI script enables the execution of unmodified Grid-centric, ATLAS jobs on Titan. Typically, a MPI script is workload-specific as it sets up the execution environment for a specific payload. This involves organization of worker directories, data management, optional input parameters modification, and cleanup on exit. Upon submission, a copy of the MPI script runs on every available worker node, starting the execution of the ATLAS job's payload in a subprocess and waits until its completion.

MPI scripts are submitted to Titan's PBS batch system via RADICAL-SAGA [39], a Python module, compliant with the OGF GFD.90 SAGA specification [23]. The Simple API for Grid Applications (SAGA) offers a unified interface to diverse job schedulers and file transferring services. In this way, SAGA provides an interoperability layer that lowers the complexity of using distributed infrastructures. Behind the API façade, RADICAL-SAGA implements a adaptor architecture: each adaptor interface the SAGA API with different middleware systems and services, including the PBS batch scheduler of Titan.

The data staging capabilities of the PanDA Broker are implemented via a file system that is shared among DTNs and worker nodes. The input files with the events of the ATLAS jobs are downloaded on the shared filesystem from the data center of Brookhaven National Laboratory (BNL). The MPI script setup process includes making the location of these files available to the payload of the ATLAS's jobs. The PanDA Broker can locate the payload's output files on the shared filesystem and transfer them from Titan BNL.

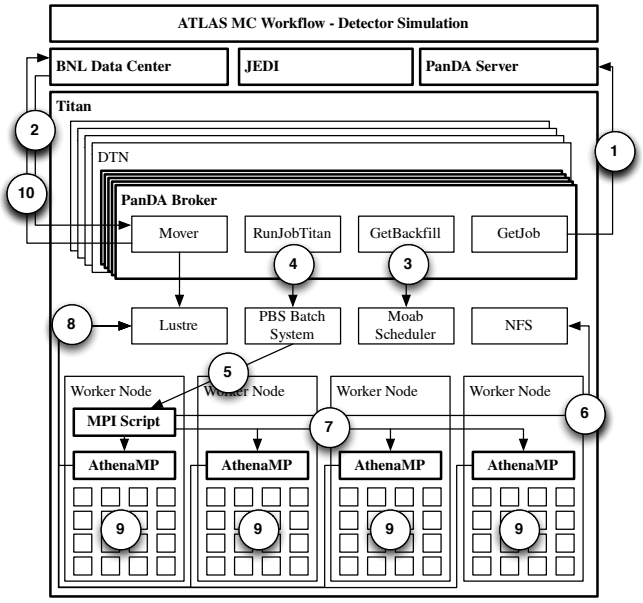

Figure 2: PanDA Broker architecture as deployed on Titan. Numbers indicates the execution process of a detector simulation job, part of the production ATLAS MC workflow.

Once deployed on Titan, every PanDA Broker supports the execution of MC detector simulations in 9 steps. PanDA Broker queries the Job Dispatcher module of the PanDA server for ATLAS jobs that have been bound to Titan by JEDI (Fig. 2:1). Upon receiving jobs descriptions, PanDA Broker pulls jobs' input files from BNL to the OLCF Lustre file system (Fig. 2:2). PanDA Broker queries Titan's Moab scheduler about the current available backfill slot (Fig. 2:3) and creates an MPI script, wrapping enough ATLAS jobs' payload to fit the backfill slot. PanDA Broker submits the MPI script to the Titan's Torque batch system via RADICAL-SAGA (Fig. 2:4).

Upon execution on the worker node(s) (Fig. 2:5), the MPI script initializes and configures the execution environment (Fig. 2:6), and executes one AthenaMP for each available work node (Fig. 2:7). AthenaMP retrieves events from Lustre (Fig. 2:8) and spawns 1 Geant4 event simulation process on each of the 16 available cores (Fig. 2:9). Upon completion of each MPI script, PanDA Broker transfer the jobs' output to BNL (Fig. 2:10), and performs cleanup.

While PanDA Broker implementation is resource specific, it was successfully ported to other supercomputers, including the HPC2 at the National Research Center "Kurchatov Institute" (NRC-KI) [9], Edison/Cori at the National Energy Research Scientific Computing Center (NERSC) [8], and SuperMUC at the Leibniz Supercomputing Centre (LRZ) [8].

4 ANALYSIS AND DISCUSSION

Currently, ATLAS deploys 20 instances of the PanDA Broker on 4 Titan's DTNs, 5 instances per DTN. Each broker submits and manages the execution of 15 to 300 ATLAS jobs,

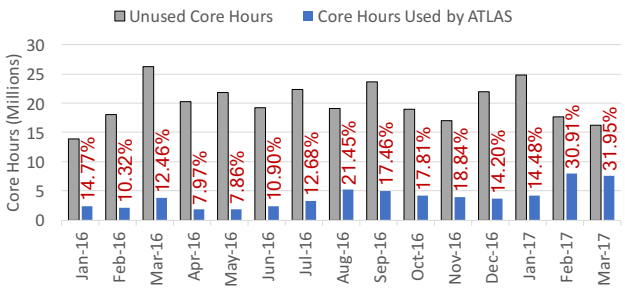

Figure 3: Titan's total backfill availability: CPU core-hours (gray) and CPU core-hours used by ATLAS (blue). GPU core-hours unaccounted for as they cannot be used by ATLAS. Efficiency of PanDA Brokers defined as percentage of total Titan's backfill availability used by ATLAS (Red labels).

one job for each Titan's worker node, and a theoretical maximum concurrent use of 96,000 cores. Since November 2015, PanDA Brokers have operated only in backfill mode, without a defined time allocation, and running at the lowest priority on Titan.

We evaluate the efficiency, scalability and reliability of the deployment of PanDA WMS on Titan by characterizing the behavior of both PanDA Broker and AthenaMP. We discuss challenges and limitations of our approach at multiple levels arising from the specifics of workload, middleware and methods. All the measurements were performed between January 2016 and February 2017, hereafter called 'experiment time window

4.1 Characterizing the PanDA Broker on Titan

We calculate the total amount of backfill availability over a period of time by: (i) polling the available backfill slots at regular intervals during that time window; (ii) converting the number of worker nodes available and their walltime into core-hours; (iii) summing the number of core-hours. We call this number of core-hours 'total backfill availability'.

Fig. 3 shows the total backfill availability on Titan (gray bars) and the number of core-hours of that availability used by ATLAS (blue bars) during the experiment time window. ATLAS consumed a total of 51.4M core-hours, for an average of ≈3.7M core-hours a month, with a minimum of 1.7M core-hours in April 2016 and a maximum 7.9M core-hours in February 2017.

PanDA Brokers' efficiency (Fig. 3, red labels) is defined as the fraction of core-hours utilized by the PanDA Brokers to Titan's total backfill availability during the experiment time window. The brokers reached 18% average efficiency, with a minimum 7.8% efficiency on May 2016 and a maximum 30.9% efficiency on February 2017 (excluding the preliminary results of March). The total backfill availability was 20.3M in April 2016, and 17.6M in February 2017. This shows that the measured difference in efficiency did not depend on a comparable difference in the total backfill availability.

During the experiment time window, about 2.25M detector simulation jobs were completed on Titan, for a total of 225M events processed. This is equivalent to 0.9% of all the 250M detector simulations performed by ATLAS in the same period of time, and 3.5% of the 6.6B events processed by those jobs. These figures confirms the relevance of supercomputers' resource contribution to the LHC Run 2, especially when accounting for the amount of unused total backfill availability and the improvement of PanDA efficiency across the experiment time window.

On February 2017, PanDA Brokers used almost twice as much total backfill availability than in any other month (preliminary results for March 2017 displayed in Fig. 3 confirm this trend). No relevant code update was made during that period and logs indicated that the brokers were able to perform faster. This is likely due to hardware upgrades on the DTNs. The absence of continuous monitoring of those nodes does not allow to quantify bottlenecks but spot measurements of their load indicate that a faster CPU and better networking were likely responsible for the improved performance. Investigations showed an average CPU load of 3.6% on the upgraded DTNs. As such, further hardware upgrades seem unlikely to improve significantly the performance of PanDA Brokers.

Every detector simulation executed on Titan process 100 events. This number of events is consistent with the physics of the use case and with the average duration of backfill availability. The duration of a detector simulation is a function of the number of events simulated but not all events take the same time to be simulated. 1 event simulation takes from ≈2 to ≈40 minutes, with a mean of ≈14 minutes. Considering that each worker node process up to 16 events concurrently, 100 events takes an average of 105 minutes to process. As such, PanDA brokers do not use backfill availability with less than 105 minutes walltime.

Fig. 4 shows the distribution of backfill availability on Titan as a function of number of nodes and the time of their availability (i.e., walltime). We recorded these data by polling Titan's Moab scheduler at regular intervals during the experiment window time. The mean number of nodes was 691, and their mean walltime was 126 minutes. Detector simulations of 100 events, enable to use down to 5/6 of the mean walltime of backfill availability. As such, it offers a good compromise for PanDA Broker efficiency.

PanDA Broker could fit the number of events to the walltime of each available backfill slot on the base of the distributions of the time taken by one event to be simulated. That specific number of event could then be pulled from the PanDA Event service [13] and given as input to one or more simulations. Once packaged into the MPI script submitted to titan's PBS batch system, these simulations would better fit their available backfill slot, contributing to increase the efficiency of PanDA Brokers.

The transition from a homogeneous to a heterogeneous number of events per detector simulation has implications for the application layer. An even number of events across simulations makes it easier to partition, track and package

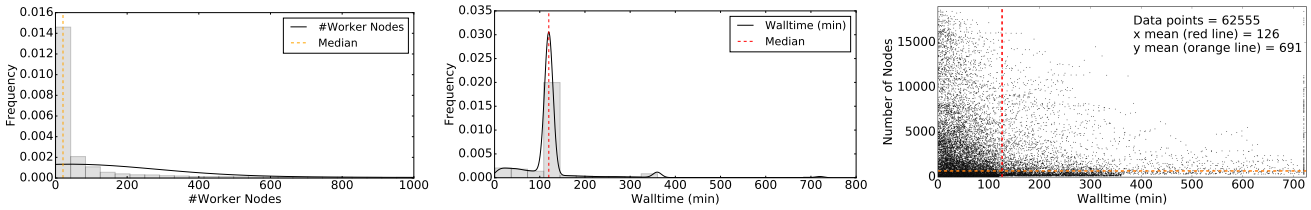

Figure 4: Backfill availability: distribution of number of work nodes (left) and of walltime in minutes (center). Scatter plot of the two variables (right). 62175 measures: mean number of work nodes available 691; mean walltime available 126 minutes.

events across simulations, especially when they are performed on both the Grid and Titan. A homogeneous number of events also helps to keep the size and duration of other stages of the MC workflow (§3.1) more uniform. Further analysis is needed to evaluate the trade offs between increased efficiency of resource utilization and the complexity that would be introduced at the application layer.

Currently, each PanDA Broker creates, submits, and monitors a single MPI PBS script at a time. This design is inherited from PanDA Pilot where a single process is spawn at a time to execute the payload. As a consequence, the utilization of a larger portion of Titan's total backfill availability depends on the the number of concurrent PanDA Brokers instantiated on the DTNs: When all the 20 PanDA Brokers have submitted a MPI PBS script, further backfill availability cannot be used.

In September 2016, increasing the number of concurrent PanDA brokers from 4 to 20 markedly improved efficiency (see Fig. 3) but further research is undergoing to understand whether an even larger number of brokers would yield similar results. This research focuses on evaluating the overheads of input/output files staging, including its impact on DTNs, and on an alternative design of PanDA Broker that enables the concurrent submission of multiple MPI scripts [8].

The current design and architecture of the PanDA Broker is proving to be as reliable as PanDA Pilot when used on the WLCG. Between Jan 2016 and Feb 2017, the overall failure rate of all the ATLAS detector simulation jobs was 14%, while the failure rate of jobs submitted to Titan was a comparable 13.6%. PanDA Brokers were responsible for around the 19% of the failures, compared to the 29% of failures produced by the JobDispatcher module of the PanDA Server, and the 13% failures produced by the Geant4 toolkit. The current failure rate of the PanDA Brokers confirms the benefits of reusing most of the code base of the PanDA Pilot for implementing the PanDA Broker. It also shows that adopting third-party libraries like RADICAL-SAGA did not have a measurable adverse effect on reliability.

4.2 Characterizing the Detector Simulation on Titan

We use two main parameters to measure the performance of the detector simulation jobs submitted to Titan: (i) the time taken to setup AthenaMP; and (ii) the distribution of the time taken by the Geant4 toolkit to simulate a certain number of events.

AthenaMP has an initialization and configuration stage. At initialization time, AthenaMP is assembled from a large number of shared libraries, depending on the type of payload that will have to be computed. Once initialized, every algorithm and service of AthenaMP is configured by a set of Python scripts. Both these operations result in a large number of read operations on the filesystem shared between the worker nodes and the DTNs, including the operations required to access small python scripts.

Initially, all the shared libraries of AthenaMP and the python scripts for its configuration were stored on the Spider 2 Lustre file system. However, the I/O patterns of the initialization and configuration stages degraded the performance of the filesystem. This was addressed by moving the AthenaMP distribution to a read-only NFS directory, shared among DTNs and worker nodes. NFS eliminated the problem of metadata contention, improving metadata read performance from ≈6,300 seconds on Lustre to ≈1,500 seconds on NFS.

Once initialized and configured, AthenaMP is used to execute 16 concurrent Geant4 simulators on a Titan's worker node. Geant4 requires to read events descriptions from a filesystem and simulate them as they would happen within the ATLAS detector. We characterized both the compute performance of the simulation and the impact of acquiring event descriptions on the filesystem.

The AMD Opteron 6274 CPU used on Titan has 16 cores, divided into 8 compute units. Each compute units has 1 floating point (FP) scheduler shared between 2 cores. When using 16 cores for FP-intensive calculations, each pair of cores competes for a single FP scheduler. This creates the overhead shown in Fig. 5: the mean runtime per event for 8 concurrent simulations computing 50 events is 10.8 minutes, while for 16 simulations is 14.25 minutes (consistent with the measured distribution of the duration of event simulation). Despite an inefficiency of almost 30%, Titan's allocation policy based on number of worker nodes used instead of number of cores does not justify the use of 1/2 of the cores available.

The performance analysis of Titan's AMD CPUs for detector simulations helps also to compare Titan and Grid site performances. Usually, Grid sites exposes resources with heterogeneous CPU architectures and a maximum of 8 (virtual) cores per worker node, while Titan's offer an homogeneous 16 cores architecture. We used the rate of events processes per

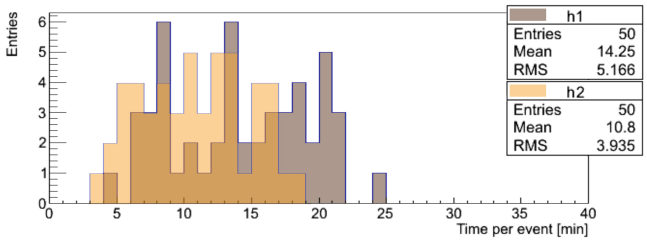

Figure 5: Distributions of the time taken to simulate one event when placing 2 simulations (h1) or 1 simulation (h2) per Titan's CPU. 2 simulation use 16 cores per node, 1 simulation 8. 50 Events; 1 Titan worker nodes; 16 work threads per node; 100 events per node.

minute as a measure of the efficiency of executing the same detector simulation on Titan or Grid sites. Comparisons of the efficiencies of Titan to the BNL and SIGNET Grid sites, normalized for 8 cores, show that the effective performance per-core at Titan is ≈0.57 event per minute, roughly 1/2 of BNL and 1/3 of SIGNET performances.

The differences in performance between Titan and the BNL and SIGNET Grid sites are due to the FP scheduler competition and the availability of newer processors. The CPUs at the Grid sites have one FP scheduler per core and are on average newer than the CPU of Titan. The heterogeneity of the Grid sites' CPUs explain the higher performance variance we observed compared to the performance consistency we measured on Titan.

We studied the impact of acquiring event descriptions on Lustre by analyzing 1,175 jobs ran on the week of 10/25/2016, for a total of 174 hours. Table 1 shows the overall statistical breakdown of the observed file I/O. ATLAS used between 1 and 300 worker nodes, and 35 on average. 75% of the jobs run by ATLAS consumed less than 25 nodes and 92% less than 100. During the 174 hours of data collection, 6.75 ATLAS jobs were executed on average per hour, each job running for an average of 1.74 hours. Every job read less than 250 GB and wrote less than 75 GB of data and, on average, each job read 20 GB and wrote 6 GB of data.

ATLAS jobs are read heavy: On average, the amount of data read per worker node is less than 400 MB, while the amount of data written is less than 170 MB. Distributions of read and written data are different: The read operation distribution per job shows a long tail, ranging from 12.5 GB to 250 GB, while the written amount of data has a very narrow distribution.

The metadata I/O breakdown shows that ATLAS jobs yield 23 file *open*() operations per second (not including file *stat*() operations) and 5 file *close*() operations per second, with similar distributions. On average, the maximum number of file *open*() operations per job is ≈170/s and the maximum number of file *close*() operations is ≈39/s. For the 1,175 ATLAS jobs observed, the total number of file *open*() operations is 172,089,760 and the total number of file *close*() operations is 40,132,992. The difference between these two values is still under investigation: One possible explanation is that ATLAS jobs don't call a file *close*() operation per every file *open*() issued.

Overall, the average time taken to read events from input files stored on Lustre is 1,320, comparable to the time taken to read the file required by assembling AthenaMP from NFS. Preliminary investigation shows that this time could be reduced to 40 seconds by loading the event descriptions into the RAM disk available on each worker node. Events descriptions could be transferred from Lustre to the RAM disk while configuring and initializing AthenaMP, almost halving the time currently required by initiating a Genat4 simulation.

5 PANDA: THE NEXT GENERATION EXECUTOR

As seen in §3.2, PanDA Broker was deployed on the DTNs as Titan's worker nodes lack connectivity to the wide area network The lack of pilot capabilities impacts both the efficiency and the flexibility of PanDA's execution process. Pilots could improve efficiency by increasing throughput and enabling greater backfill utilization. Further, pilots makes it easier to support heterogeneous workloads.

The absence of pilots imposes the static coupling between MPI scripts submitted to the PBS batch system and detector simulations (§4). This makes the scheduling of multiple generations of workload on the same PBS job impossible: once a statically defined number of detector simulations are packaged into a PBS job and this job is queued on Titan, no further simulations can be added to that job. New simulations have to be packaged into a new PBS job that needs to be submitted to Titan based upon backfill availability.

The support of multiple generations of workload would enable more efficient use of the backfill availability of walltime. Currently, when a set of simulations ends, the PBS job also ends, independent of whether more wall-time would still be available. With a pilot, additional simulations could be executed to utilize all the available wall-time, while avoiding further job packaging and submission overheads.

Multiple generations would also relax two assumptions of the current execution model: knowing the number of simulations before submitting the MPI script, and having a fixed number of events per simulation (100 at the moment). Pilots would enable the scheduling of simulations independently from whether they were available at the moment of submitting the pilot. Further, simulations with a varying number of events could be scheduled on a pilot, depending on the amount of remaining walltime and the distribution of execution time per event, as shown in §3.2, Fig. 5. These capabilities would increase the efficiency of the PanDA Broker when there is a large difference between the number of cores and walltime.

Pilots can offer a payload-independent scheduling interface while hiding the mechanics of coordination and communication among multiple worker nodes. This could eliminate the need for packaging payload into MPI scripts within the

	Num. Nodes	Duration (s)	Read (GB)	Written (GB)	GB Read/nodes	GB Written/nodes	*open*()	*close*()
Min	1	1,932	0.01	0.03	0.00037	0.02485	1,368	349
Max	300	7,452	241.06	71.71	0.81670	0.23903	1,260,185	294,908
Average	35.66	6,280.82	20.36	6.87	0.38354	0.16794	146,459.37	34,155.74
Std. Dev.	55.33	520.99	43.90	12.33	0.19379	0.03376	231,346.55	53,799.08

Table 1: The Statistical breakdown of the I/O impact of 1,175 jobs ATLAS executed at OLCF for the week of 10/25/16

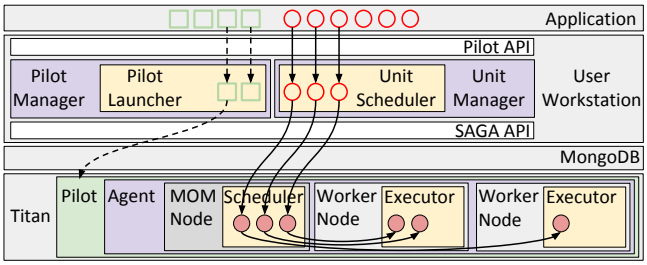

Figure 6: NGE Architecture: The PilotManager and UnitManager reside on a Titan DTN while the Agent is executed on its worker nodes. Color coding: gray for entities external to NGE, white for APIs, purple for NGE's modules, green for pilots, yellow for module's components.

broker, greatly simplifying the submission process. This simplification would also enable the submission of different types of payload, without having to develop a specific PBS script for each payload. The submission process would also be MPI-independent, as MPI is used for coordination among multiple worker nodes, not by the payload.

5.1 Implementation

The implementation of pilot capabilities within the current PanDA Broker require quantification of the effective benefits that it could yield and, on the base of this analysis, a dedicated engineering effort. We developed a prototype of a pilot system capable of executing on Titan to study experimentally the quantitative and qualitative benefits that it could bring to PanDA. We called this prototype Next Generation Executor (NGE).

NGE is a runtime system to execute heterogeneous and dynamically determined tasks that constitute workloads. Fig. 6 illustrates its current architecture as deployed on Titan: the two management modules (Pilot and Unit) represent a simplified version of the PanDA Broker while the agent module is the pilot submitted to Titan and executed on its worker nodes. The communication between PanDA Broker and Server is abstracted away as it is not immediately useful to evaluate the performance and capabilities of a pilot on Titan.

NGE exposes an API to describe workloads (Fig. 6, green squares) and pilots (Fig. 6, red circles), and to instantiate a PilotManager and a UnitManager. The PilotManager submits pilots to Titan's PBS batch system via SAGA API (Fig. 6, dash arrow), as done by PanDA Broker to submit MPI scripts. Once scheduled, the pilot's Agent is bootstrapped on Titan's MOM node and the Agent's Executors on the on the worker nodes, and the UnitManager schedules units to the Agent's Scheduler (Fig. 6, solid arrow). The Agent's Scheduler schedules the units on one or more Agent's Executor for execution. The Agent's executors can manage one or more worker nodes, depending on performance evaluations. The UnitManager and the Agent communicate via a database that is instantiated on one of Titan's DTN so as to be reachable by both modules.

The NGE Agent uses the *Open Run-Time Environment (ORTE)* for communication and coordination of the execution of units. This environment is a critical component of the OpenMPI implementation, developed to support distributed high-performance computing applications operating in a heterogeneous environment [14, 35]. ORTE provides a mechanism to create a *"dynamic virtual machine"* (DVM) that spans multiple nodes. The DVM transparently provides support for interprocess communication, resource discovery and allocation, and process launching across a variety of platforms. Libraries enable the interaction between users and ORTE to enable the submission, monitoring and managing of tasks, avoiding filesystem bottlenecks and race conditions with network sockets. As a consequence, ORTE is able to minimize the system overhead while submitting tasks.

5.2 Experiments

We designed experiments to characterize the performance of the NGE on Titan, with an emphasis on understanding its overhead and thus the cost of introducing new functionalities. We perform three groups of experiments in which we investigate the weak scalability, weak scalability with multiple generation, and strong scalability of the NGE.

Each experiment entails executing multiple instances of AthenaMP to simulate a pre-determined number of events. All the experiments have been performed by configuring AthenaMP to use all the 16 cores of Titan's worker nodes.

We measured the execution time of the pilots and of AthenaMP within them, collecting timestamps at all stages of the execution. Experiments were performed by submitting NGE's pilots to Titan's batch queue. The turnaround time of an individual run is determined by queue waiting times. Since we are interested only in the performances of the NGE, we removed queue time from our statistics.

5.2.1 Weak scalability. In this experiment we run as many AthenaMP instances (hereafter referred to as tasks) as the

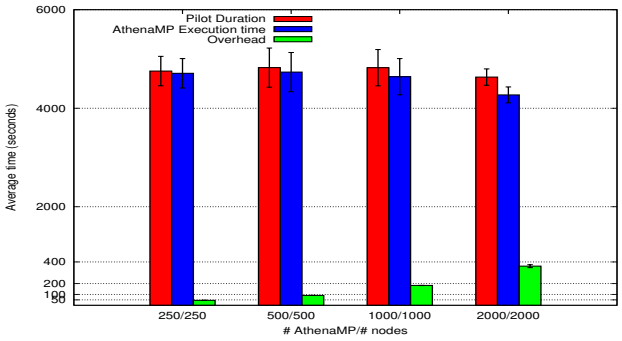

Figure 7: Weak scalability: average pilot duration, average duration of one AthenaMP execution, and pilot's overhead as a function of pilot sizes (200, 500, 1000 and 2000 nodes).

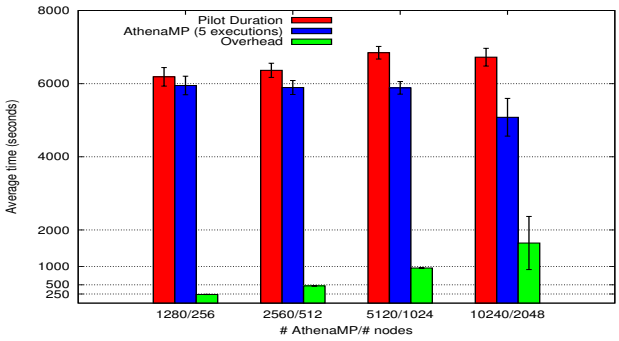

Figure 8: Weak scalability with multiple generations: average pilot duration, average duration of sequential AthenaMP executions, and pilot's overhead for pilot with 256, 512, 1024 and 2048 nodes.

number of nodes controlled by the pilot. Each AthenaMP simulates 100 events, requiring $\sim$ 4200 seconds on average.

Tasks do not wait within the NGE Agent's queue since one node is available to each AthenaMP instance. Overheads in task execution are consequence primarily of the three other factors: (i) the initial bootstrapping of the pilot on the nodes; (ii) the UnitManager's dispatching of units (tasks) to the agent; and (iii) time for the agent to bootstrap all the tasks on the nodes.

We tested pilots with 250, 500, 1000 and 2000 worker nodes and 2 hours walltime. The time duration is determined by the Titan's walltime policy. Fig. 7 depicts the average pilot duration, the average execution time of AthenaMP, and the NGE pilot overhead as function of the pilot size.

We observe that, despite some fluctuations due to external factors (e.g., Titan's shared filesystem and the shared database used by the NGE), the average execution time of AthenaMP ranges between 4200 and 4800 seconds. We also observe that in all the cases the gap between AthenaMP execution times and the pilot durations is minimal, although it slightly increases with the pilot size. We notice that NGE's overhead does grow linearly with the number of units.

5.2.2 Weak scalability with multiple generation . The NGE provides an important new capability of submitting multiple generations of AthenaMP to the same pilot. In order to investigate the cost of doing so, we performed a variant of the weak scalability experiments. This stresses the pilot's components, as new tasks are scheduled for execution on the Agent while other tasks are still running.

In these experiments, we run five AthenaMP instances per node. As these experiments are designed to investigate the overhead generated by the scheduling and bootstrap of AthenaMP instances, we reduced the number of events simulated by each AthenaMP task to sixteen in such a way that the running time of each AthenaMP is, on average, $\sim$ 1200 seconds. This experiment design choice does not affect the objectives or accuracy of the experiments, but allows us to scale experiments to large node counts by being conservative with allocation.

We ran pilots with 256, 512, 1024 and 2048 worker nodes and 3 hours walltime. Fig. 8 depicts the average pilot duration, the average execution time of five sequential generations of AthenaMP, and the corresponding overhead. We observe that the difference between the two durations is more marked than in the previous experiments. Despite this, we notice that the growth of the overhead is consistent with the increment of the number of tasks per node for pilots with 256, 512 and 1024 worker nodes, and less than linear for the pilot with 2048 worker nodes.

5.2.3 Strong scalability. The last experiments study strong scalability by running the same number of tasks for different pilot sizes. We used 2048 AthenaMP instances and pilots with 256, 512, 1024 and 2048 nodes. Thus, the number of AthenaMP generations is equal to eight times the size of the smallest pilot and corresponds to the size of the largest pilot. As a consequence, the number of consecutive generations of AthenaMP decreases with the pilot size by generating different dynamics within the pilots. These experiments are designed to investigate whether pilot overhead is affected by the degree of concurrency within the pilot and/or the number of tasks. Each AthenaMP instance simulates sixteen events as in the previous experiment.

Fig. 9 shows the average pilot duration and the average execution time of possibly sequential AthenaMP instances. We notice that the difference between the pilot duration and the AthenaMP execution times is almost constant for all the pilot sizes, although the overall duration of the pilot decreases linearly with the pilot size.

6 RELATED WORK

Several pilot-enabled WMS were developed for the LHC experiments: AliEn [6] for ALICE; DIRAC [31] for LHCb; GlideinWMS [36] for CMS; and PanDA [26] for ATLAS. These systems implement similar design and architectural principles: centralization of task and resource management, and of monitoring and accounting; distribution of task execution across multiple sites; unification of the application

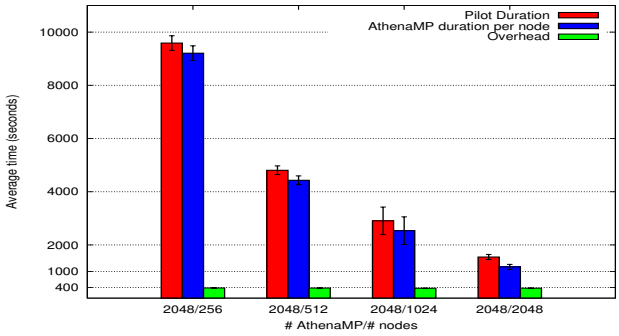

Figure 9: Strong scalability: average pilot duration, average duration of sequential AthenaMP executions, and pilot's overhead for pilots with 256, 512, 1024 and 2048 nodes.

interface; hiding of resource heterogeneity; and collection of static and sometimes dynamic information about resources.

AliEn, DIRAC, GlideinWMS and PanDA all share a similar design with two types of components: the management ones facing the application layer and centralizing the capabilities required to acquire tasks' descriptions and matching them to resource capabilities; and resource components used to acquire compute and data resources and information about their capabilities. Architecturally, the management components include one or more queue and a scheduler that coordinates with the resource modules via push/pull protocols. All resource components include middleware-specific APIs to request for resources, and a pilot capable of pulling tasks from the management modules and executing them on its resources.

AliEn, DIRAC, GlideinWMS and PanDA also have similar implementations. These WMS were initially implemented to use Grid resources, using one or more components to the Condor software ecosystem [40] and, as with GlideinWMS, contributing to its development. Accordingly, all LHC WMS implemented Grid-like authentication and authorization systems and adopted a computational model based on distributing a large amount of single/few-cores tasks across hundreds of sites.

All the experiments at LHC produces and process large amount of data both from actual collisions in the accelerator and from their simulations. Dedicated, multi-tiered data systems have been built to store, replicate, and distributed these data. All LHC WMS interface with these systems to move data to the sites where related compute tasks are executed or to schedule compute tasks where (large amount of) data are already stored.

7 CONCLUSION

The deployment of PanDA Broker on Titan enabled distributed computing on a leadership-class HPC machine at unprecedented scale. In the past 13 months, PanDA WMS has consumed almost 52M core-hours on Titan, simulating 3.5% of the total number of detector events of the ATLAS production Monte Carlo workflow. We described the implementation and execution process of PanDA WMS (§2) and PanDA Broker (§4), showing how they support and enable distributed computing at this scale on Titan, a leadership-class HPC machine managed by OCLF at ORNL.

We characterized the state of practice by evaluating the efficiency, scalability and reliability of both PanDA Broker and AthenaMP as deployed on Titan (§4). Our characterization highlighted the strengths and limitations of the current design and implementation: PanDA Brokers enable the sustained execution of millions of simulations per week but further work is required to optimize its efficiency and reliability (§4.1). PanDA Brokers support the concurrent execution of multiple AthenaMP instances, enabling each AthenaMP to perform the concurrent execution of up to 16 Geant4 simulators. Nonetheless, our characterization showed how improving I/O performance could reduce overheads (§4.2), increasing the overall utilization of Titan's backfill availability.

More generally, this paper highlights the fundamental role of WMS for experimental science. HEP was amongst the first, if not the very first experimental community to realize the importance of using WMS to manage their computational campaign(s). As computing becomes increasingly critical for a range of experiments, the impact of WMS on HEP foreshadows the importance of WMS for other experiments (such as SKA, LSST etc.). campaigns. These experiments will have their own workload characteristics, resources types and federation constraints, as well metrics of performance. The design and experience captured in this paper will prove invaluable for designing WMS to support computational campaigns and will provide a baseline to evaluate the relative merits of different approaches.

The 52M core hours used by ATLAS, via PanDA, is over 2% of the total utilization on Titan over the same period, bringing the time-averaged utilization of Titan to be consistently upwards of 90%. Given that the average average utilization of most other leadership class machines is less (e.g., NSF's flagship Blue Waters the average utilization fluctuates between 60-80% (see XDMoD[42])) there is ample headroom for similar approaches elsewhere. These unprecedented efficiency gains aside, this work is just a starting point towards more effective operational models for future leadership and online analytical platforms [5]. These platforms will have to support ever increasing complex workloads with varying models for dynamic resource federation. Recent efforts with Google Cloud for the CMS experiment – HEPCloud [22, 34] represent similar yet different approaches to dynamic resource federation. As the type, heterogeneity and complexity of future platforms (e.g., Exascale, Commercial Clouds) for science increases, there will be a greater need and emphasis on abstractions to ensure WMS can both utilize as well as support evolving platforms. Our experience demonstrated the advantages arising from an abstractions based approach – in particular scalable implementations of pilot abstraction.

REFERENCES

[1] G. Aad and others. 2008. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST* 3 (2008), S08003. DOI:http://dx.doi.org/10.1088/1748-0221/3/08/S08003

[2] Georges Aad, B Abbott, J Abdallah, AA Abdelalim, A Abdesselam, O Abdinov, B Abi, M Abolins, H Abramowicz, H Abreu, and others. 2010. The ATLAS simulation infrastructure. *The European Physical Journal C* 70, 3 (2010), 823–874.

[3] Inc Adaptive Computing Enterprises. 2014. Maui Scheduler Administrator's Guide: Backfill. (2014). Retrieved Apr 2, 2017 from http://docs.adaptivecomputing.com/maui/8.2backfill.php

[4] Sea Agostinelli, John Allison, K al Amako, J Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G Barrand, and others. 2003. GEANT4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506, 3 (2003), 250–303.

[5] Argonne National Laboratory Argonne National Laboratory. 2017. Future Online Analysis Platform: BUILDING A RESEARCH ROADMAP FOR FUTURE SCIENCE PLATFORMS. (2017). Retrieved Apr 3, 2017 from https://press3.mcs.anl.gov/futureplatform/

[6] S Bagnasco, L Betev, P Buncic, F Carminati, F Furano, A Grigoras, C Grigoras, P Mendez-Lorenzo, A J Peters, and P Saiz. 2010. The ALICE workload management system: Status before the real data taking. *J. Phys.: Conf. Ser.* 219 (2010), 062004. 6 p. https://cds.cern.ch/record/1353182

[7] Ashley D. Barker, David E. Bernholdt, Arthur S. Bland, Jeff D. Gary, James J. Hack, Stephen T. McNally, James H. Rogers, Brian Smith, T.P. Straatsma, Sreenivas R. Sukumar, Kevin G. Thach, Suzy Tichenor, Sudharshan S. Vazhkudai, and Jack C. Wells. 2016. *High Performance Computing Facility Operational Assessment 2015: Oak Ridge Leadership Computing Facility*. Technical Report ORNL/SPR–2016/110. Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). Oak Ridge Leadership Computing Facility (OLCF). DOI:http://dx.doi.org/10.2172/1324094

[8] Fernando Harald Barreiro Megino, Siarhei Padolski, Danila Oleynik, Tadashi Maeno, Sergey Panitkin, Kaushik De, Torre Wenaus, Alexei Klimentov, and Paul Nilsson. 2016. *PanDA for ATLAS Distributed Computing in the Next Decade*. Technical Report ATL-COM-SOFT-2016-049. The ATLAS collaboration. http://cds.cern.ch/record/2218080

[9] A Belyaev, A Berezhnaya, L Betev, P Buncic, K De, D Drizhuk, A Klimentov, Y Lazin, I Lyalin, R Mashinistov, and others. 2015. Integration of Russian Tier-1 Grid Center with High Performance Computers at NRC-KI for LHC experiments and beyond HENP. In *Journal of Physics: Conference Series*, Vol. 664. IOP Publishing, Bristol, UK, 092018. Issue 9.

[10] Mikhail Borodin, K De, J Garcia, Dmitry Golubkov, A Klimentov, T Maeno, A Vaniachine, and others. 2015. Scaling up ATLAS production system for the LHC Run 2 and beyond: project ProdSys2. In *Journal of Physics: Conference Series*, Vol. 664. IOP Publishing, Bristol, UK, 062005. Issue 6.

[11] Mikhail Borodin, Kaushik De, Jose Garcia Navarro, Dmitry Golubkov, Alexei Klimentov, Tadashi Maeno, David South, and others. 2015. Unified System for Processing Real and Simulated Data in the ATLAS Experiment. (2015). arXiv:1508.07174

[12] J Caballero, J Hover, P Love, and GA Stewart. 2012. AutoPyFactory: a scalable flexible pilot factory implementation. In *Journal of Physics: Conference Series*, Vol. 396. IOP Publishing, Bristol, UK, 032016. Issue 3.

[13] Paolo Calafiura, K De, W Guan, T Maeno, P Nilsson, D Oleynik, S Panitkin, V Tsulaia, P Van Gemmeren, and T Wenaus. 2015. The ATLAS Event Service: A new approach to event processing. In *Journal of Physics: Conference Series*, Vol. 664. IOP Publishing, Bristol, UK, 062065. Issue 6.

[14] R. H. Castain, T. S. Woodall, D. J. Daniel, J. M. Squyres, B. Barrett, and G .E. Fagg. 2005. The Open Run-Time Environment (OpenRTE): A Transparent Multi-Cluster Environment for High-Performance Computing. In *Proceedings, 12th European PVM/MPI Users' Group Meeting*. Springer-Verlag, Berlin, Heidelberg, 225–232.

[15] Gennaro Corcella, Ian G Knowles, Giuseppe Marchesini, Stefano Moretti, Kosuke Odagiri, Peter Richardson, Michael H Seymour, and Bryan R Webber. 2001. HERWIG 6: an event generator for hadron emission reactions with interfering gluons (including supersymmetric processes). *Journal of High Energy Physics* 2001, 01 (2001), 010.

[16] D Crooks, P Calafiura, R Harrington, M Jha, T Maeno, S Purdie, H Severini, S Skipsey, V Tsulaia, R Walker, and others. 2012. Multi-core job submission and grid resource scheduling for ATLAS AthenaMP. In *Journal of Physics: Conference Series*, Vol. 396. IOP Publishing, Bristol, UK, 032115. Issue 3.

[17] J De Favereau, C Delaere, P Demin, A Giammanco, V Lemaître, A Mertens, and M Selvaggi. 2013. DELPHES 3, A modular framework for fast simulation of a generic collider experiment. (2013). arXiv:1307.6346

[18] Jack Dongarra, Hans Meuer, and Erich Strohmaier. 2016. Top500 Supercomputing Sites. http://www.top500.org. (2016).

[19] Ian Foster and Carl Kesselman. 2003. *The Grid 2: Blueprint for a new computing infrastructure*. Elsevier, Amsterdam, Netherlands.

[20] Ian Foster, Carl Kesselman, and Steven Tuecke. 2001. The anatomy of the grid: Enabling scalable virtual organizations. *International journal of high performance computing applications* 15, 3 (2001), 200–222.

[21] Vincent Garonne, Graeme A Stewart, Mario Lassnig, Angelos Molfetas, Martin Barisits, Thomas Beermann, Armin Nairz, Luc Goossens, Fernando Barreiro Megino, Cedric Serfon, and others. 2012. The atlas distributed data management project: Past and future. In *Journal of Physics: Conference Series*, Vol. 396. IOP Publishing, Bristol, UK, 032045. Issue 3.

[22] Gabriele Garzoglio and Burt Holzman. 2016. HEP Cloud: How to add thousands of computers to your data center. (2016). Retrieved Apr 3, 2017 from https://goo.gl/IP8Jqi

[23] T Goodale, S Jha, H Kaiser, T Kielmann, P al Kleijer, A Merzky, J Shalf, and C Smith. 2008. A Simple API for Grid Applications (SAGA). OGF Document Series 90. (2008).

[24] A Klimentov, P Nevski, M Potekhin, and T Wenaus. 2011. The ATLAS PanDA Monitoring System and its Evolution. In *Journal of Physics: Conference Series*, Vol. 331. IOP Publishing, Bristol, UK, 072058. Issue 7.

[25] T Maeno. 2011. Overview of ATLAS PanDA workload management. *J. Phys.: Conf. Ser* 331, 7 (2011), 072024. http://stacks.iop.org/1742-6596/331/i=7/a=072024

[26] T Maeno, K De, A Klimentov, P Nilsson, D Oleynik, S Panitkin, A Petrosyan, J Schovancova, A Vaniachine, T Wenaus, and others. 2014. Evolution of the ATLAS PanDA workload management system for exascale computational science. In *Journal of Physics: Conference Series*, Vol. 513. IOP Publishing, Bristol, UK, 032062. Issue 3.

[27] Tadashi Maeno, K De, T Wenaus, P Nilsson, GA Stewart, R Walker, A Stradling, J Caballero, M Potekhin, D Smith, and others. 2011. Overview of atlas panda workload management. In *Journal of Physics: Conference Series*, Vol. 331. IOP Publishing, Bristol, UK, 072024. Issue 7.

[28] Cecchi Marco, Capannini Fabio, Dorigo Alvise, Ghiselli Antonia, Giacomini Francesco, Maraschini Alessandro, Marzolla Moreno, Monforte Salvatore, Petronzio Luca, and Prelz Francesco. 2009. The glite workload management system. In *International Conference on Grid and Pervasive Computing*. Springer Publishing, New York City, NY, USA, 256–268.

[29] P Nilsson, J Caballero, K De, T Maeno, A Stradling, T Wenaus, Atlas Collaboration, and others. 2011. The ATLAS PanDA pilot in operation. In *Journal of Physics: Conference Series*, Vol. 331. IOP Publishing, Bristol, UK, 062040. Issue 6.

[30] Sarp Oral, David A Dillow, Douglas Fuller, Jason Hill, Dustin Leverman, Sudharshan S Vazhkudai, Feiyi Wang, Youngjae Kim, James Rogers, James Simmons, and others. 2013. OLCF's 1 TB/s, next-generation lustre file system. (2013).

[31] Stuart Paterson, Joel Closier, and the Lhcb Dirac Team. 2010. Performance of combined production and analysis WMS in DIRAC. *Journal of Physics: Conference Series* 219, 7 (2010), 072015. http://stacks.iop.org/1742-6596/219/i=7/a=072015

[32] A Rimoldi, A Dell'Acqua, A di Simone, M Gallas, A Nairz, J Boudreau, V Tsulaia, and D Costanzo. 2006. Atlas Detector Simulation: Status and Outlook. In *Astroparticle, Particle and Space Physics, Detectors and Medical Physics Applications*, Vol. 1. World Scientific Publishing, Singapore, 551–555.

[33] Elmar Ritsch. 2014. *ATLAS Detector Simulation in the Integrated Simulation Framework applied to the W Boson Mass Measurement*. Ph.D. Dissertation. Innsbruck U.

[34] Paul Rossman. 2016. Google Cloud, HEP Cloud and probing the nature of Nature. (2016). Retrieved Apr 3, 2017 from https:

//goo.gl/MLKjsv
[35] Mark Santcroos, Ralph Castain, Andre Merzky, Iain Bethune, and Shantenu Jha. 2016. Executing dynamic heterogeneous workloads on Blue Waters with RADICAL-Pilot. In *Cray User Group 2016*. cug.org, online, 1–14. https://cug.org/proceedings/cug2016_proceedings/includes/files/pap130.pdf
[36] Igor Sfiligoi. 2008. glideinWMS—a generic pilot-based workload management system. In *Journal of Physics: Conference Series*, Vol. 119. IOP Publishing, Bristol, UK, 062044. Issue 6.
[37] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. 2006. PYTHIA 6.4 physics and manual. *Journal of High Energy Physics* 2006, 05 (2006), 026.
[38] ATLAS PanDA Team. 2017. The PanDA Production and Distributed Analysis System. (2017). Retrieved March 25, 2017 from https://twiki.cern.ch/twiki/bin/view/PanDA/PanDA
[39] RADICAL Team. 2017. RADICAL-SAGA Software Toolkit. (2017). Retrieved March 21, 2017 from https://github.com/radical-cybertools/radical-saga
[40] Douglas Thain, Todd Tannenbaum, and Miron Livny. 2005. Distributed computing in practice: the Condor experience. *Concurrency and computation: practice and experience* 17, 2-4 (2005), 323–356.
[41] Matteo Turilli, Mark Santcroos, and Shantenu Jha. 2015. A Comprehensive Perspective on Pilot-Job Systems. (2015). arXiv:1508.04180
[42] XDMoD Blue Waters. 2017. XDMoD - Blue Waters. (2017). Retrieved Apr 3, 2017 from http://xdmod.ncsa.illinois.edu/#tg_usage:statistic_Jobs_none_utilization

A ARTIFACT DESCRIPTION — CONVERGING HIGH-THROUGHPUT AND HIGH-PERFORMANCE COMPUTING: A CASE STUDY

A.1 Abstract

The computing systems used by LHC experiments has historically consisted of the federation of hundreds to thousands of distributed resources, ranging from small to mid-size resource. In spite of the impressive scale of the existing distributed computing solutions, the federation of small to mid-size resources will be insufficient to meet projected future demands. This paper is a case study of how the ATLAS experiment has embraced Titan – a DOE leadership facility in conjunction with traditional distributed high- throughput computing to reach sustained production scales of approximately 51M core-hours a years. The three main contributions of this paper are: (i) a critical evaluation of design and operational considerations to support the sustained, scalable and production usage of Titan; (ii) a preliminary characterization of a next generation executor for PanDA to support new workloads and advanced execution modes; and (iii) early lessons for how current and future experimental and observational systems can be integrated with production supercomputers and other platforms in a general and extensible manner.

A.2 Description

A.2.1 Check-list (artifact meta information).

- **Program: ATLAS Monte Carlo Workflow, Geant4, AthenaMP, GROMACS**
- **Data set: Available at URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW**
- **Run-time environment: PanDA Workload Management System, NGE**
- **Hardware: OLCF Titan Cray XK7**
- **Output: Available at URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW**
- **Experiment workflow: Raw data acquisition, data wrangling and filtering, plotting, analysis**
- **Publicly available?: Yes**

A.2.2 How software can be obtained (if available).

- PanDA Workload Management System: URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
- NGE: URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW

A.2.3 Hardware dependencies. Access and allocation on OLCF Titan Cray XK7, workstation with at least 8GB of RAM.

A.2.4 Software dependencies. Python, jupyter, pandas, matplotlib, gnuplot, excel.

A.2.5 Datasets.

Section 4.1 Figure 3 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 4.2 Figure 4 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 4.2 Figure 5 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 7 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 8 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 9 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW

A.3 Installation

Section 4.1 Figure 3 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 4.2 Figure 4 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 4.2 Figure 5 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 7 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 8 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 9 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW

A.4 Experiment workflow

Section 4.1 Figure 3 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 4.2 Figure 4 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 4.2 Figure 5 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 7 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 8 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 9 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW

A.5 Evaluation and expected result

Section 4.1 Figure 3 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 4.2 Figure 4 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 4.2 Figure 5 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 7 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 8 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW
Section 5.2 Figure 9 URL_REMOVED_FOR_DOUBLE-BLIND_REVIEW