REDTOP Computing Model

Pascal Paschos

April 6, 2020

I. Introduction

The proposed REDTOP experiment[1] at Fermilab will study rare η decays and η' meson decays in search of discrepancies in the Standard Model at energies below 1 GeV. The physics goal of REDTOP is to collect statistics at a factor ~10^4 compared to the world η sample. This will allow exploring Beyond Standard Model (BSM) with couplings at the 10^{-8} level or lower.

In this document, we summarize a model to accommodate the computational and data management workflow required by the experiment. The proposed schema reflects the current computing activity of REDTOP on the Open Science Grid (OSG). The collaboration consumed nearly 1.9 million core-hours and generated approximately 100 TB of Monte Carlo data alone during the past year as part of a campaign to validate the detector design.

II. REDTOP computing requirements

The REDTOP experiment operates in a typical beam environment where the inelastic interaction rate of the proton beam with the target is ~1GHz. Level-0 and Level-1, implemented in hardware, will reduce such a large rate by a factor of ~10^4 before sending the data to a compute-farm for the Level-2 trigger and preliminary reconstruction.

In terms of storage requirements, the experiment is expected to generate approximately 2.5 PB of production data and approximately 2 PB of processed data each year. It is proposed that Fermilab houses the production data on tape storage each year of operation and provides an allocation on dCache for staging data. In addition, the collaboration can leverage OSG Connect which provides ephemeral storage to stage data for grid jobs.

In terms of compute requirements, REDTOP's single-core computational workflow has proven to be well suited for the distributed High Throughput Computing (DHTC) environment of the OSG. The proposed computing scheme here aims to accommodate the dataflow from the full experimental apparatus. Extrapolating from available data, from REDTOP jobs currently running on OSG, it is estimated that the collaboration would need approximately 90 million core hours annually; 55 million core-hours for Monte Carlo jobs and 35 million core-hours for data reconstruction jobs.

III. **The computing model**

With the above considerations in mind, we assume that the output DataStream from the Level-2 farm will be staged at Fermilab's (FNAL) dCache storage and, eventually, preprocessed on site. The process will require an allocation on the General Purpose Grid (GPGrid). Local access to the files in dCache is enabled via a POSIX-like interface over an NFS mount. In the present contest, dCache will serve as a high speed front-end ephemeral storage to provide access to FNAL's tape system. Since direct access to tape is limited to on-site machines, staging to dCache first is required for off-site access. In order to increase the flexibility of the model and to offer more optimal implementations to the participating institutions. several options are discussed

The collaboration plans to process the bulk of the experimental results with jobs submitted from FermiGrid or OSG Connect submission nodes. An example of a job submission for REDTOP from FNAL is shown Figure 1. A submission node sends jobs to the grid while data are delivered to grid jobs by first staging them on dCache and then transporting them to the remote sites via a common protocol such as GridFTP. It is proposed for Fermilab maintain an origin (stratum-0) repository for REDTOP data in the distributed CernVM File System (CVMFS) which will then deliver data and software to remote compute sites over the HTTP protocol. Reconstruction or analysis data designated for long term storage can then be archived back to tape at FNAL by first transferring them back onto the dCache storage.

The collaboration will continue to use the OSG Connect submission nodes to launch jobs to the OSG. OSG Connect provides a collaborative environment for a multi-institutional collaboration like REDTOP in order to have access to a single point of submission to the OSG grid. Researchers from the individual institutions, not necessarily affiliated with the laboratory hosting the experiment, can get an account on OSG Connect for job submission and access to transient storage space. A dedicated submission node, independent from the OSG Connect servers, can also be provisioned for the exclusive use of the collaboration if required and funds are available.

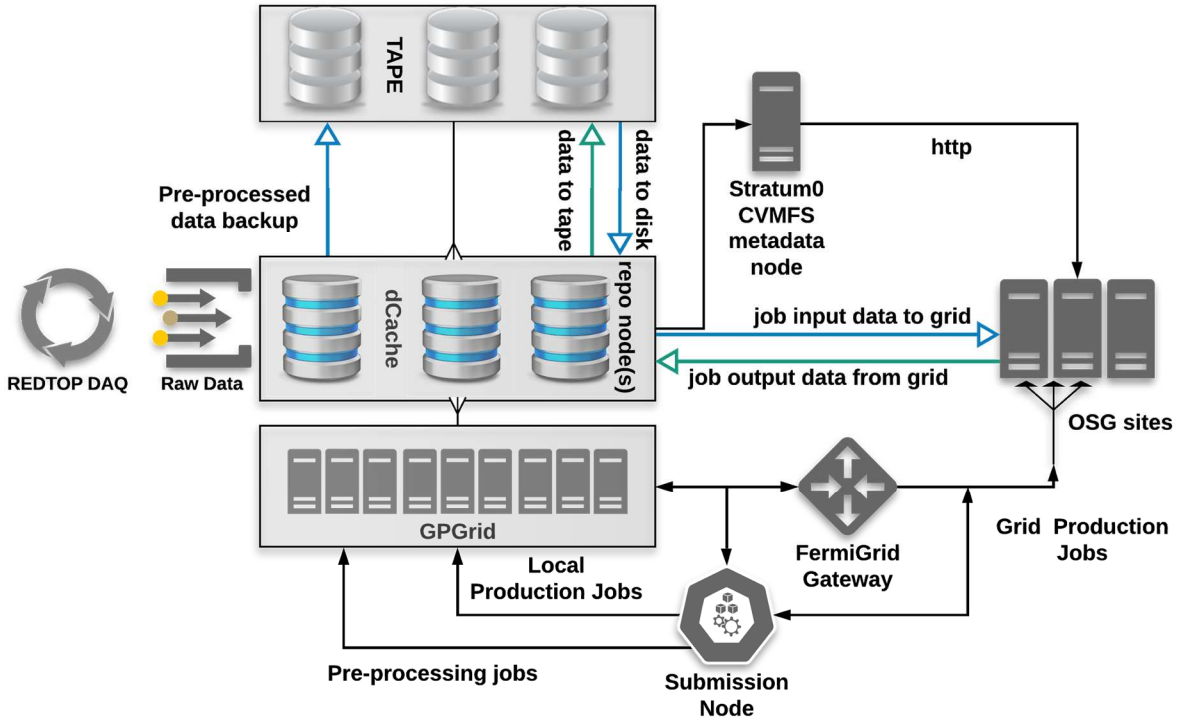

Figure 1: REDTOP workflow at Fermilab.

Collaborative shared storage for OSG Connect is provided by Stash, a Ceph storage system. Stash provides capacity for temporary storage for OSG projects. During production, REDTOP will make use of Stash to store the processed output from the jobs before distributing them to various endpoints of the collaboration. Processed data can also be stored back to tape at FNAL. Figure [2] shows such an example of a data processing workflow using OSG Connect. Users will be able to submit jobs to the grid from the login server. A Stratum-0 server hosts a CVMFS repository of the REDTOP software stack, based on the GenieHad[2] and Geant4[3] Monte Carlo software as well as the reconstruction framework. The server's role is to distribute the collaboration's software stack and tools to the remote sites where the job is running. The CVMFS software repo can alternatively be hosted and managed by the OSG Application Installation Service (OASIS).

Production quality raw data will need to be distributed to the grid site running the job via GridFTP from FNAL's dCache as in Figure 1 or directly from OSG Connect Stash if a copy of the input data is staged there. Due to the expected decline in the use of GridFTP - Globus support has ended - alternative methods to transfer data include the XrootD[4] and WebDAV/HTTPS protocols. Alternatively, raw data could also be served to the compute site from the data origin at FNAL using OSG's StashCache infrastructure. StashCache[5] is the OSG data caching service that stores data needed by a project in

proximity to the compute sites. Requesting data from the nearest StashCache instead of the Origin minimizes the flight path to the worker node at the remote site.

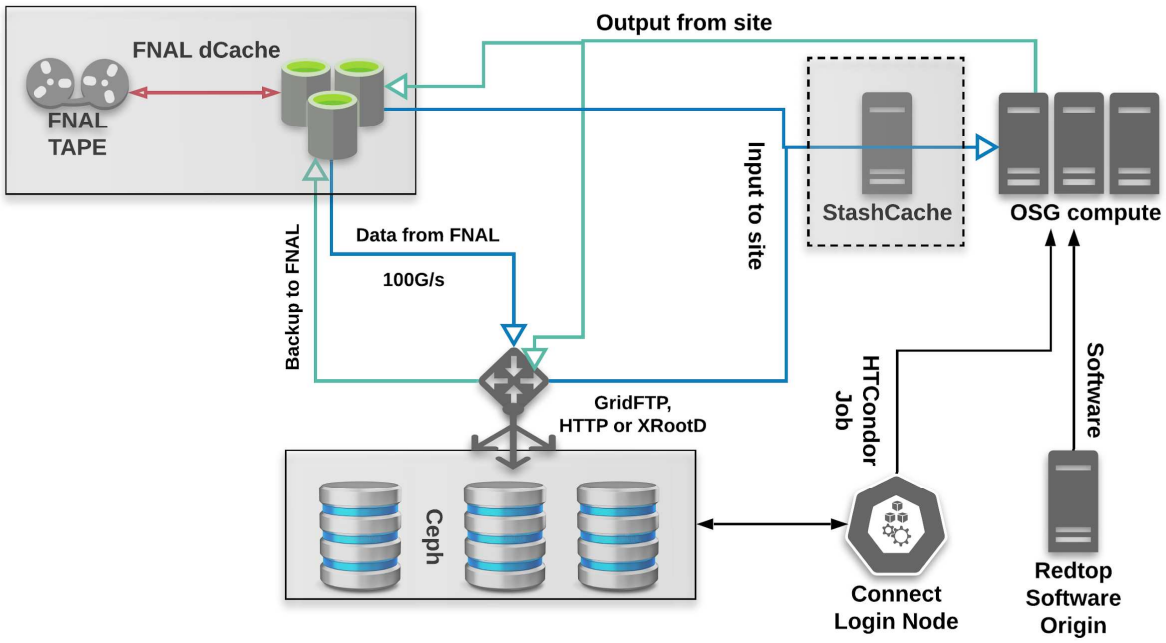

Figure 2: REDTOP workflow using OSG Connect.

While not in the critical path to the processing campaign of the experimental data, the REDTOP collaboration will pursue the decentralization of job submissions to the OSG by having member institutions provision and deploy their own submission hosts. Furthermore and contingent upon the availability of local computing resources, member institutions can join the OSG federation and accept jobs from OSG's GlideinWMS[6] job factory via a HostedCE[7,8] deployment.

IV. Optional Facilities

In addition to the baseline model discussed above, REDTOP is also investigating the deployment of a distributed storage scheme for the collaboration. In this case, member institutions will share the cost of provisioning storage systems to house portions of the REDTOP data, primarily processed datasets delivered by the jobs running on OSG. Rucio[9], a data management system already adopted by several High Energy Physics (HEP) experiments (ATLAS, CMS, IceCube, XENON and others), can then provide management across storage end-points. In the proposed plan, a Rucio Server and a File Transfer Service (FTS) will automate the distribution of data among end points using replication rules defined for each of the Rucio Storage Elements (RSEs). An RSE is a logical representation of a physical storage location. A proposed setup is illustrated in Figure 3. The Rucio server has an RSE defined for FNAL tape and a number of RSEs for

the storage end-points at the collaboration sites. The proposed implementation is described in In the top panel of Fig. 3, where data received from the experiment are ingested into Rucio and, then, replicated across the other storage end-points according to a set of replication rules defining the partitioning of the volume of data to each site.

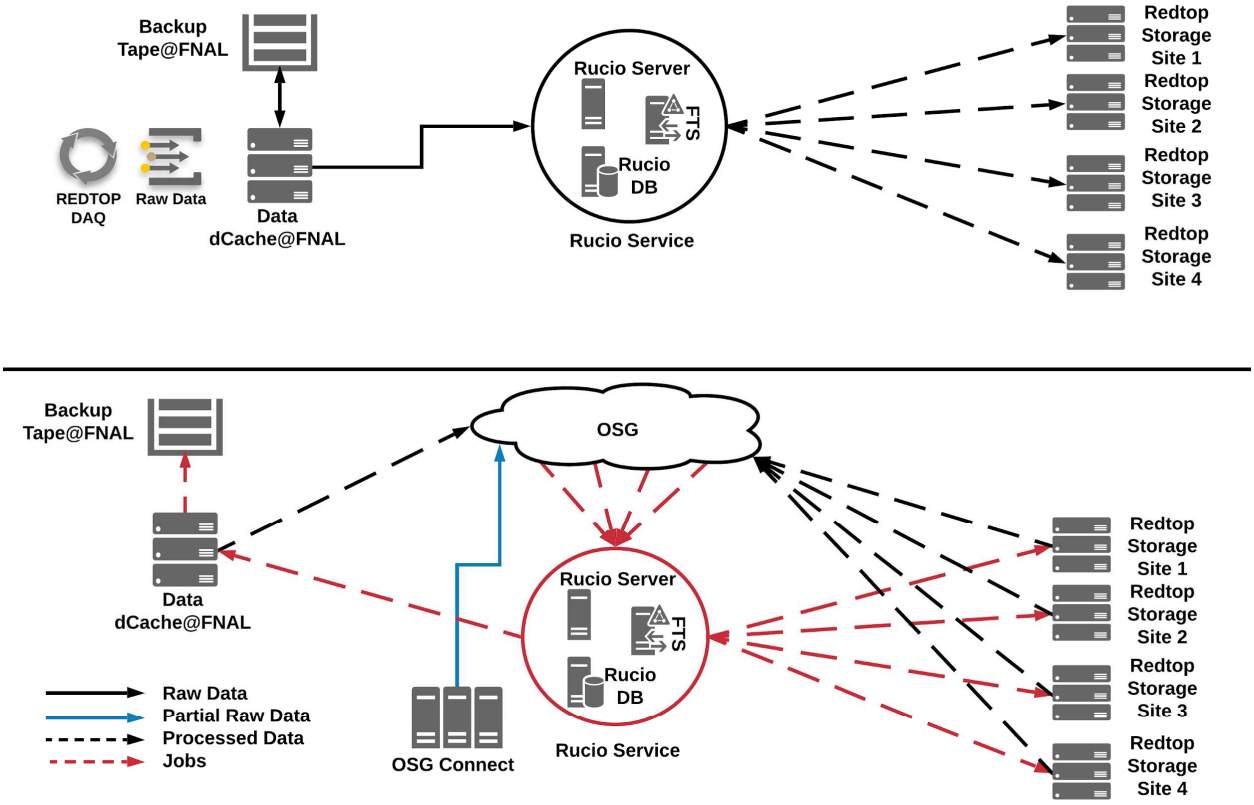

Figure 3: A Rucio data management scheme for REDTOP.

Jobs are submitted by users from OSG Connect submit-hosts to the OSG grid (bottom panel of Figure 3). Data can then be downloaded from the individual RSEs via Rucio at execution time. When the job completes, output data can be registered into Rucio and uploaded to the designated storage end-points. The FTS server - shown in Figure 3 - establishes the connections between the RSEs and ensures that files are correctly transferred.

V. **Conclusions**

We have outlined here a Computing Model for the REDTOP collaboration. We have indicated optimal storage solutions for the acquired data and briefly described the access patterns to the distributed computational resources required for the offline.

In conclusion:

- The REDTOP collaboration proposes the use of the Open Science Grid to access the working nodes running the simulation, reconstruction and data analysis.
- The collaboration recognizes the importance of leveraging FermiLab infrastructure for the Level-2 trigger farm and the primary repository for storage
- The collaboration also recognizes the added versatility of a distributed storage federation among the participating institutions and the opportunities afforded if they provision computational resources for the use of REDTOP workflow

References

[1] https://redtop.fnal.gov

[2] https://redtop.fnal.gov/geniehad-users-guide/

[3] https://geant4.web.cern.ch

[4] https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookXrootdService

[5] https://iopscience.iop.org/article/10.1088/1742-6596/898/6/062044/pdf

[6] https://glideinwms.fnal.gov/doc.prd/index.html

[7] https://opensciencegrid.org/docs/compute-element/hosted-ce/

[8] https://slateci.io/blog/2020-01-22-deploy-hosted-ce.html

[9] https://link.springer.com/article/10.1007/s41781-019-0026-3