Overview of Scientific Computing at BNL

Eric Lançon
DUNE visit - 03/01/19
The Scientific Data and Computing Center

- Scientific Data and Computing Center (SDCC) a component of CSI
- It includes RACF (RHIC & ATLAS Computing Facility) from Physics Department
- Operation oriented with limited resources for development
The Scientific Data and Computing Center

• **Service Operation For:**
  RHIC, LHC ATLAS, Belle II, BER ARM, LQCD, RIKEN, BES Center for Functional Nano Materials, National Synchrotron Light Source II, National Nuclear Data Center, Simons Foundation,…

• **Analysis Facility:** ~1500 users from 20 projects (<10 to 100+ users/project)

• **R&D lead:** Containerization, Active tape storage, Jupyter analysis platform, Object storage technology
SDCC and large experiments

• The RHIC Tier 0
  • Store and process data from RHIC experiments
  • Provide analysis means for 1’200 users
  • Long term data preservation
  • Simulation resources for future programs (sPHENIX & EIC)

• The US ATLAS Tier 1
  • ~25% of ATLAS Tier 1 computing capacity worldwide
  • Store RAW data from LHC and from simulation
  • Distribute data to the 4 US Tier 2 + analysis site (SLAC)
  • Analysis center for US physicists, from 41 institutes
    (incl. 4 Nat. Labs)

• The Belle II data center outside Japan
  • Store and process Raw data
  • Operate key services: databases and data distribution

24/7 availability
US ATLAS computing @ BNL

- Largest and most reliable of ATLAS Tier-1s
  - Used by ATLAS for time constrained and critical workloads
- Coordination of the US ATLAS Facility Complex: 4 Tier-2 sites
- Analysis facility for 200+ US physicists
- Responsibilities
  - ATLAS computing facilities co-convener
  - WLCG archival group co-convener
- Member of WLCG MB & OB

CPU delivered by ATLAS T1s in 2018

- USA: 22%
- Germany: 13%
- France: 12%
- Canada: 11%
- Netherlands: 8%
- UK: 7%
- Italy: 6%
- Nordic: 4%
- Russia: 4%
- Taiwan: 4%
- Spain: 4%
- USA: 22%
- Germany: 13%
- France: 12%
- Canada: 11%
- Netherlands: 8%
- UK: 7%
- Italy: 6%
- Nordic: 4%
- Russia: 4%
- Taiwan: 4%
- Spain: 4%
US ATLAS T1 within WLCG

CPU delivered in 2018 by WLCG Tier-1s

- DE-FZK: 15.3%
- FR-IN2P3-CC: 7.4%
- IT-INFN-T1: 7.4%
- US-BNL-ATLAS: 10.2%
- UK-RAL: 10.9%
- US-USCMS-FNAL: 13.0%
- RU-T1: 10.1%
- NL-NIKHEF-SARA: 9.7%
- CA-TRIUMF: 5.2%
- ES-pic: 3.9%
- NDGF-T1: 3.8%
- TW-Taiwan-LCG2: 1.7%
- KR-KISTI: 1.5%

LHC experiments only
ATLAS grid jobs at BNL: 2018

20.5M ATLAS jobs in 2018

- 61% Analysis
- 13% MC Simulation
- 13% Group Production
- 13% MC Reconstruction
- 7% Data Processing
- 4% Others

218 PB data processed by ATLAS jobs in 2018

- 39% Analysis
- 43% MC Simulation
- 2% Group Production
- 2% MC Reconstruction
- 15% Data Processing
Belle II data center

- Capitalize on experience demonstrated for RHIC and ATLAS Experiments
- 15% of Belle II computing capacity
- Hosting Belle II Raw Data:
  - 100% until 2020,
  - 30% afterwards
- BNL responsible for critical services
  - Conditions Database
  - Data Management

Second Copy of Raw Data

- First copy at KEK
- Second copy shared among Raw Data Centers
- Criteria:
  - First copy in Asia, better to have second copy elsewhere
  - Use sites that are already WLCG Tier1

<table>
<thead>
<tr>
<th>Year</th>
<th>2018</th>
<th>2019</th>
<th>2020</th>
<th>2021</th>
<th>2022</th>
<th>2023</th>
<th>2024</th>
</tr>
</thead>
<tbody>
<tr>
<td>BNL</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Canada</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>France</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Germany</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>Italy</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

Proposed sharing of second copy of raw data (%)
SDCC in numbers

• 90+k CPU cores — 4 PFlops
  • 3 HPC Institutional Clusters (GPU, KNL, Skylake)

• ~90 PB of disk storage
  • of various technologies

• 150+ PB of tape storage
  • Largest/second HPSS tape library in the US, 3rd/4th worldwide [1]

• 2x100 Gbps connection to ESnet
  • Onsite ESnet support

Institutional Clusters

• Interconnected nodes (HPC)
  • CPU-GPU cluster with 200+ compute nodes (2x36 CPU cores Xeon Broadwell and 2 GPUs each: K80 & P100)
  • KNL cluster: 144 nodes (64 cores@1.3GHz)
  • Skylake cluster: 64 nodes (72 cores each)

• In production since Jan. 2017
  • >180 registered users
  • Over 95% utilisation

MoUs (describing level of resources and services) organized with each user community
Mass Storage on Tape

- 9 x Oracle SL8500 (most of them are 10,088 slots)
- Currently deployed: LTO-7 (6TB, 300 MB/s)
- Older drive & media: LTO-4,5,6 and T10K-D
- Currently 130+ PB on ~60k cartridges
High Throughput Parallel Archiving

RHIC: Ingest data of year N, recall data of year N-x

Data Injection: Average 1.6 GB/s (Green)

Data Restore: Average 1.0 GB/s (Blue)

2 days
Active tape storage: 2017 statistics

- Archived: 19M files, 21 PB
- Restored: 12M files, 25 PB
Data transfer expertise

- Last 12 months
  - Data import: 57 PB
  - Data export: 53 PB
- File Transfer Service (FTS) service for US ATLAS and Belle II operated at BNL
- ESnet support on site
BNL and HPC Centers

- Increasing use of HPC centers
  - Allocations on DOE Leadership Class Facilities
  - Opportunistic on NSF (Bluewaters, Comet, Stampede,...) via OSG CONNECT
- High performant network backbone provided by ASCR through ESnet
- Data produced at DOE HPCs are transferred and stored at BNL
- Increasingly decoupled storage and CPU

For ATLAS (2018)
>1B simulated events
>500 M-hour
Up to 600k simultaneous CPU cores
Support for collaborations

- Collaborative tools: JIRA/Confluence, BNLBox (Cloud storage), Indico, Git, Invenio, …
- CVMFS
- Documentation, Web sites
- User support (incl. VO mgt)
- Ticketing system
- Interactive nodes & batch farm
- Local storage & archive
- Single Sign On & Federated Identity in deployment
- Rucio (being tested for non HEP projects)
Addressing future challenges

- CPU needs:
  - HPC facilities
  - Convergence between US-ATLAS and CSI expertise on software for heterogenous platforms

- Storage needs:
  - Increased usage of tapes
  - Reduced duplication of data
Support for new use cases

- Increasing demand of Jupyter based analysis platforms
  - With needs of GPUs for ML
  - Belle II analysis framework
- Development of a JupyterHub platform with GPU institutional cluster as backend

Example: ML Applications for ATLAS
- Tensorflow/Keras

Jupyterhub Training

Teraflops of Jupyter: A Notebook Based Analysis Portal at BNL
Ofer Rind
Spring HEPiX, Madison, WI
May 17, 2018

In collaboration with: Doug Benjamin, Costin Caramarcu, Zhihua Dong, Will Strecker-Kellogg, Thomas Throwe

Example: ML Applications for ATLAS
- Tensorflow/Keras

Jupyterhub Training

Teraflops of Jupyter: A Notebook Based Analysis Portal at BNL
Ofer Rind
Spring HEPiX, Madison, WI
May 17, 2018

In collaboration with: Doug Benjamin, Costin Caramarcu, Zhihua Dong, Will Strecker-Kellogg, Thomas Throwe
Leading expertise in tape system

- BNL: 3rd/4th HPSS site worldwide
- Efficient tape management acquired through RHIC program applied for leading the R&D on ‘tape carousel’ concept
- Co-chair of WLCG tape archival group
- RFI: Future DOE/SC archival system
New Data Center

NSLS-I Building

CD-2/3A August 2018

Available for ATLAS early CY21 (after LHC Long Shutdown 2)
Summary

• Leading expertise in supporting computing for data driven projects
  • From data transfer and processing
  • To data analysis
• Actively engaged in R&Ds to address future challenges
• Expertise in medium-scale HPC computing
DUNE: What SDCC can bring

- Expertise in
  - Distributing computing development and operation (including data transfers and data management)
  - Platform for software development and management
  - Databases
  - Analysis center
- Operating a DUNE computing facility at low cost, leveraging from BNL’s existing infrastructure
- Synergy with computing and software efforts and R&Ds for LHC and other programs through CSI