Data and Analytics Services Overview

Prabhat
DAS Group Lead

New User Training, 2/24/2017
DAS Team

Debbie Bard
Cosmology, Particle Physics
Workflows, Analytics

Wahid Bhimji
Particle Physics
Management, Transfer

Shane Canon
Systems Biology, Genomics
Workflows

Shreyas Cholia
Access, Management

Lisa Gerhardt
Particle Physics
Management, Workflows

Annette Greiner
BioMedical Sciences
Access, Visualization

Burlen Loring
Visualization

Jialin Liu
Management

Jeff Porter
Nuclear Physics
Transfer

Evan Racah
Analytics

Rollin Thomas
Astrophysics, Cosmology
Analytics, Access

Quincey Koziol
Management

Prabhat
Climate, Neuroscience
Analytics, Management
DAS Goal: “Enable Data-Intensive Science at Scale”

Internal Goals

• Provide world-class, production quality software services for all major Data capabilities:
  – Analytics, Management, Workflows, Transfer, Access, Visualization

• Pioneer evaluation, research and deployment of Big Data technologies
  – Focusing on productivity and performance

• Engage with stakeholders to enable scientific discovery in a data-driven world
  – Users, Computing Sciences Staff, Vendors, Researchers
Cori Data Features

• NVRAM Flash Burst Buffer as I/O accelerator
• High bandwidth external connectivity to experimental facilities from compute nodes (Software Defined Networking)
• More login nodes for managing advanced workflows
• Data-friendly queues
  – Shared, Interactive, Real-time, High-mem, Transfer
• Virtualization capabilities with Shifter
# Data Software and Services

<table>
<thead>
<tr>
<th>Capabilities</th>
<th>Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Transfer + Access</td>
<td><img src="image" alt="globus online" />, GridFTP, <img src="image" alt="python" />, <img src="image" alt="django" />, <img src="image" alt="newt" /></td>
</tr>
<tr>
<td>Workflows</td>
<td><img src="image" alt="FireWorks" />, Swift</td>
</tr>
<tr>
<td>Data Management</td>
<td><img src="image" alt="HDF" />, <img src="image" alt="netCDF" />, <img src="image" alt="ROOT" />, <img src="image" alt="PostgreSQL" />, <img src="image" alt="SciDB" />, <img src="image" alt="mongoDB" />, <img src="image" alt="MySQL" /></td>
</tr>
<tr>
<td>Data Analytics</td>
<td><img src="image" alt="python" />, jupyter, <img src="image" alt="IPython" />, Tensorflow, Theano, Caffe, <img src="image" alt="Wolfram Mathematica" /></td>
</tr>
<tr>
<td>Data Visualization</td>
<td><img src="image" alt="VISIT" />, ParaView</td>
</tr>
</tbody>
</table>
How to get help

• Documentation:
  – www.nersc.gov
  – http://www.nersc.gov/users/data-analytics/

• File a ticket
  – send e-mail to consult@nersc.gov

• NUG monthly telecons/webinars, Data Day

• In-person engagement
# Training Schedule

<table>
<thead>
<tr>
<th>Time (PDT)</th>
<th>Topic</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>9:00 am</td>
<td>Data and Analytic Services Overview</td>
<td>Prabhat</td>
</tr>
<tr>
<td>9:20 am</td>
<td>Python and ipython</td>
<td>Rollin Thomas</td>
</tr>
<tr>
<td>9:40 am</td>
<td>Machine Learning</td>
<td>Evan Racah</td>
</tr>
<tr>
<td>10:00 am</td>
<td>Spark</td>
<td>Lisa Gerhardt</td>
</tr>
<tr>
<td>10:20 am</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>10:40 noon</td>
<td>Databases and Data Formats</td>
<td>Wahid Bhimji, Quincey Kozoil, Jialin Liu</td>
</tr>
<tr>
<td>11:10 pm</td>
<td>Burst Buffer</td>
<td>Wahid Bhimji</td>
</tr>
<tr>
<td>11:30 pm</td>
<td>Lunch</td>
<td></td>
</tr>
<tr>
<td>12:50 pm</td>
<td>Intro to ESNet</td>
<td>Eli Dart</td>
</tr>
<tr>
<td>1:20 pm</td>
<td>Moving and sharing data</td>
<td>Shreyas Cholia</td>
</tr>
<tr>
<td>1:40 pm</td>
<td>Shifter</td>
<td>Shane Canon</td>
</tr>
<tr>
<td>2:00 pm</td>
<td>Workflow Tools</td>
<td>Shane Canon</td>
</tr>
<tr>
<td>2:20 pm</td>
<td>Break</td>
<td></td>
</tr>
<tr>
<td>2:40 pm</td>
<td>Science Gateways</td>
<td>Annette Greiner</td>
</tr>
<tr>
<td>3:00 pm</td>
<td>Visualization</td>
<td>Annette Greiner</td>
</tr>
<tr>
<td>3:20 pm</td>
<td>Q &amp; A Session</td>
<td></td>
</tr>
<tr>
<td>3:30 pm</td>
<td>End</td>
<td></td>
</tr>
</tbody>
</table>
Asks...

• Please engage with the DAS team
  – Ask questions
  – Provide critical feedback on choice of tools
  – Let us know if we are missing capabilities

• Please let us know of your use cases
  – Productivity
  – Performance
    • Portfolio should work at 1TB / 10,000 core level
    • Scale to 100TB / 100,000 cores

• Have Fun!