Overview

- SAMGrid & DØ
- SAMGrid in a Nutshell
- More About “The Grid”
- Forwarding “The Grand Scheme”
- The Job
- Data Handling with SAM
- SAMGrid - Application Centric
SAMGrid & DØ

• SAMGrid (JIM + SAM)
  • DØ’s way of using computing resources on the grid
    • Monte Carlo
    • Reprocessing
    • Refixing
    • Skimming (Beta Testing)
SAMGrid in a Nutshell

Operational Tools (automc/d0repro_tools)

SAMGrid Client

SAMGrid Broker

SAMGrid Submission Site (Grid Job Queue)

The Grid
SAMGrid/OSG/LCG

SAM
The Grid
SAMGrid/OSG/LCG

Native Samgrid
Head Node

Globus Gatekeeper

Jobmanager Samgrid

Samgrid Batch Adapters

Batch System
(Local job queue)

Batch System

LOCAL JOB

LOCAL JOB

LOCAL JOB

CondorG (OSG)/LCG UI
(Local job queue)

OSG/LCG
Site #1

OSG/LCG
Site #2

OSG/LCG
Site #n

Samgrid Batch Adapters

FORWARDING NODE

Parag Mhashilkar
SAMGrid Overview
Forwarding “The Grand Scheme”

Network Boundaries
- FW: Forwarding Node
- C: OSG/LCG Cluster
- S: VO-Service (SAM)

Job Flow
- Offers Service

SAM-Grid Overview
The Job

- **Grid Job**
  - Single grid job splits into several local jobs based on the job type.
  - Start the project for data distribution using SAM based on the job type.
- **Local Job (Batch/OSG/LCG Job)**

<table>
<thead>
<tr>
<th>Tasks common to all job types</th>
<th>Tasks specific to all job types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bootstrapping script initiates the process and fetches the input sandbox from the SAMGrid head/forwarding node</td>
<td>Fetch and deploy the binaries (RTE + D0Runjob) from SAM based on the job type</td>
</tr>
<tr>
<td></td>
<td>Fetch the input data from SAM using job type specific data queues</td>
</tr>
<tr>
<td></td>
<td>Setup the environment and generate D0runjob macro. Process the input data.</td>
</tr>
<tr>
<td>Cleanup and exit.</td>
<td>D0runjob or SAMGrid stores the output back to SAM based on the job type (Tape v/s SAM)</td>
</tr>
</tbody>
</table>
Data Handling with SAM

- SAM - Data management in SAMGrid
- Jobs get input data from tapes or SAM disks
- Jobs store the output data to tapes or SAM disks
- Data Handling optimized using data queues.
  - SAM can be configured to use sam_fcp for data transfers
  - sam_fcp queues based on job types
  - sam_fcp queues based on the data transfer rate.
SAMGrid - Application Centric

- Application specific knowledge embedded in SAMGrid
  - Number of local jobs to split into
  - Fetching/Setting up of binary and input data
  - Generation of Macro and Metadata
  - Merging
- Above steps done in SAMGrid as an optimization step.
- Prevents us from doing business in a general manner.
- Supporting new job types requires changes to SAMGrid packages.
- Increases the development/integration/testing cycle.
- We could do better in splitting the application specific functionalities between SAMGrid and D0Runjob
  - 100% splitting may not be possible.
Wrapping Up

- DØ’s uses computing resources on the grid using SAMGrid (JIM + SAM)
- JIM does Job management while SAM does the Data management
- SAMGrid has been optimized at various levels to serve the experiments needs.
- Optimization, at the cost of losing generalization making SAMGrid more and more application aware.
- Time to split the application specific knowledge from SAMGrid into D0Runjob for ease of support and maintenance in future.
- 100% splitting of functionality not possible, but we can definitely do better.