The SLAC Scientific Computing Recharge Model

Introduction

As it makes the transition to a multi-program laboratory, SLAC addressed how scientific computing would be funded. The decision was made that most of the cost would be provided out of lab-wide funds accounting for shared services. Customers would be asked to pay for support of their specific hardware and special services at a discounted rate. This model makes the customer cost attractive, while applying a little impedance to customer appetites requiring some payment.

In broad brush, the common funding handles shared knowledge base, routine networking, installations, shared services (like interactive login servers provided by 4 FTEs). Customer costs are based on the amount of hardware they have supported and any unique services they require. Additionally, the common funding covers the first 5.5 FTEs that support clusters and storage systems. The recharge covers the remaining FTEs needed to support customer systems. At the time of the FY2011 model, that number was also 5.5 FTEs.

It was decided that support effort scales with the number of hardware nodes: for example there is not significantly different effort involved in supporting a 4-core vs 8-core computer. Experts were interviewed to estimate hours of effort needed to support each type of system: then time is money. For systems with the concept of a cluster, the charges are separated into base and per unit costs.

At present, the hardware counts will be done every 6 months and costs applied to recover the required number of FTE (less the base 5.5). Costs will be normalized each time to balance the books. At the time of writing, an average FTE cost $260k, with the customer discount being the fraction of direct charged FTEs to total customer system support FTEs. This is 50% at this time.

Types of Systems and Costs

Computing has identified several types of supported systems: Lustre, xrootd and nfs for storage; batch and standalone compute nodes; customer-specific database servers; and heavy use of the tape silos.

Service		Unit		Cost/unit (K$)
Storage	Lustre	1	Cluster	51.45
		10	Server	5.57
	xrootd	1	Cluster	17.15
		10	Server	9.95
	NFS	10	Server	7.03
	HPSS Storage Tier	10	Tape Drive	34.30
Scheduler	LSF	100	User	0.86
Database	Oracle, MySQL	1	Landscape	12.86
Standalone Server	Unix, MAC, Windows	100	Server	51.45
Batch Cluster		1	Cluster	17.15
		100	Server	21.44

Description of Fields

Storage:
Various types of storage providing room for large amount of scientific data to be stored and retrieved at various speeds and costs

Scheduler:
Software that distributes pieces of a large scientific computation to many servers to speed up the computation

Database:
Data stored in relational structures allowing complex and fast queries. A landscape is an interrelated set of database instances often including a development, quality assurance and production instance

A single instance is considered a landscape even if it has no supporting instances. Instances with multiple schemas are considered a single instance. Instances spread across several physical servers (RAC etc.) are considered a single instance.

Standalone Server:
Miscellaneous standalone servers with various operating systems

Batch Cluster:
Servers organized in clusters running scientific computations

Data Collection Methodology:

HPSS

The HPSS Data Flow information for the Recharge Report is obtained by processing the PFTP log files in HPSS (PFTP stands for Parallel File Transfer Protocol). PFTP is used to transfer data into or out of the HPSS disk cache. For each project using HPSS, the byte counts for every transfer during the sample period are extracted and summed. This is currently a manual process but the OCIO Storage team intends to automate it and publish a periodic report.

LSF

A table of LSF user groups and their corresponding fair shares is obtained by querying any one of the general queues. The list of hosts running general queue jobs is obtained by recursively expanding the "genfarm" LSF machine group. This group of servers is then divided into several "virtual compute clusters" proportional to their groups' fair shares. To obtain the number of users for each of virtual cluster, the corresponding user groups are recursively expanded and sorted into a list of unique usernames.

Private Queues

The complete list of LSF queues is obtained and the general queues are removed. The remaining private queues are then correlated manually with SLAC departments. The number of servers associated with each department is obtained by recursively expanding each machine group associated with a queue belonging to that department, merging the resulting host lists and sorting them into a list of unique hostnames. Similarly, the number of users associated with a department is obtained by recursively expanding the user groups associated with each of the department's queues, merging the resulting user lists and then sorting them into a list of unique usernames.