Operating Dedicated Data Centers – Is It Cost-Effective?

M. Ernst, R. Hogue, C. Hollowell, W. Strecker-Kellog, A. Wong, A. Zaytsev
Brookhaven National Laboratory, Department of Physics, P.O. Box 5000, Upton, NY 11973, USA
E-mail: tony@bnl.gov

Abstract. The advent of cloud computing centres such as Amazon’s EC2 and Google’s Computing Engine has elicited comparisons with dedicated computing clusters. Discussions on appropriate usage of cloud resources (both academic and commercial) and costs have ensued. This presentation discusses a detailed analysis of the costs of operating and maintaining the RACF (RHIC and ATLAS Computing Facility) compute cluster at Brookhaven National Lab and compares them with the cost of cloud computing resources under various usage scenarios. An extrapolation of likely future cost effectiveness of dedicated computing resources is also presented.

1. Introduction
Traditional HENP computing models mostly relied on dedicated data centers until recently, but budgetary realities and the economies of scale compel the community to evaluate alternatives. Cloud computing services are offered by for-profit companies such as Amazon and Google, and academic virtual organizations such as OSG and EGI that seek to harness the power of non-dedicated resources. This paper describes the cost of computing of a dedicated facility like the RHIC and ATLAS Computing Facility (RACF [1]) at Brookhaven National Lab (BNL) and compares it to the cost of computing at for-profit cloud providers.

The RACF has operated a large-scale multi-purpose dedicated computing facility since the mid 1990’s, serving a geographically diverse, worldwide scientific community that participates in various projects in which BNL is involved. The major components of the RACF are the 23,000-computing cores Linux Farm (see Figure 1), a 15 PB disk storage system, a 35 PB robotic tape storage facility, and the high-availability general computing infrastructure, all connected together by a high-speed, 100-Gbs capable backbone with over 3500 active ports.

Figure 1: Total number of RACF computing cores
1. Cloud Computing

Because Monte Carlo simulation jobs are long-running, cpu-bound tasks with low I/O requirements and minimal local dependencies, the RACF chose to run Monte Carlo jobs on Amazon’s EC2 cloud system for an initial evaluation. EC2 virtualized cloud resources are available in two basic instances: a) spot, and b) on-demand. Table 1 summarizes the EC2 virtual instances relevant for this discussion.

| Type       | ECU   | RAM (GB) | Storage (GB) | Network I/O | Cost/hr (US$) |
|------------|-------|----------|--------------|-------------|--------------|
| spot       | m1.small | 1    | 1.7          | 160         | low          | 0.007        |
| spot       | m1.medium | 2          | 3.75         | 410         | moderate     | 0.013        |
| on-demand  | m1.medium | 2    | 3.75         | 410         | moderate     | 0.12         |

Table 1: Amazon EC2 instances (prices current as of August 23, 2013 for Eastern US)

According to Amazon EC2, 1 ECU = 1.2 GHz Xeon processor (HS06 is ~8/core) from 2007. For comparison, the RACF purchased the 2.2 GHz Xeon (Sandybridge) processor in 2013 with an HS06 rating of ~38/core.

Jobs run until the end and are never evicted with an on-demand instance. Prices for spot instances are dynamic and are dependent on supply and demand. Job submitter must choose maximum price one is willing to pay (target price), and jobs run if target price exceeds current price. Jobs are terminated if current (dynamic) price exceeds target price.

The RACF ran ~5000 Monte Carlo jobs in a 3-week period on EC2’s m1.small spot instances in early 2013. Based on EC2’s historical data on spot prices, the RACF set its target price at 3 times the current price ($0.021/hr) for this test. Jobs ran reliably, and unsurprisingly it took 50% longer to execute. Efficiency was low even at the chosen target price. Additional tests with EC2 are on-going.

An alternative to EC2 is Google Compute Engine (GCE). The price of $0.132 for standard instance (equivalent to EC2’s on-demand m1.medium instance) is similar to EC2. Unlike EC2, GCE does not allow jobs to upload a custom OS image. The application must run on Ubuntu or CentOS Linux image supported by Google. A description of the GCE tests and results can be found in [3].

2. The RACF Dedicated Computing Clusters

RACF costs can be broken down in two categories: direct and indirect. Direct costs cover hardware (servers, network, etc) and software (licenses, support, etc) components. Major indirect costs include staff and infrastructure (power and space). Costs include the institutional overhead, where applicable. Major contributors are discussed below individually.

3.1 Servers

Intel-based servers are the major contributing factor to the overall costs. Figure 2 shows the evolution of cost (in US$) per physical core over the past 5 years for the RHIC and USATLAS programs. Costs include servers, racks, pdu’s, switches, installation and warranty. It should be noted that server configuration changes (for example, 1-U vs. 2-U, additional memory or storage, hardware differences between RHIC and USATLAS, etc) have not been decoupled from the costs and are partly responsible for the fluctuations. Over the 2009 to 2013 period, the average price per core was US $474 (RHIC) and US $365 (USATLAS), respectively. The average core at the RACF acquired during the 2009 to 2013 period has 4.0 GB of RAM, 18 HS06, 700 GB of storage and a 1 Gb line-rate I/O capacity.
3.2 Network

The RACF network infrastructure is based on 1GbE connectivity, and it is expected to remain so for a few more years. To estimate the cost per physical core, we divided the cost of the operational network infrastructure (switch, line cards, cabling, warranty support, etc) by the maximum number of hosts per switch times the average number of cores per host (11.2 for RHIC and 10.6 for USATLAS):

\[
\text{Cost per core} = \frac{\text{network switch}}{(\text{number of hosts}) \times (\text{average number of cores per host})} \approx \frac{\$450,000}{900 \times 11} \approx \$45
\]

It is further assumed the network hardware remains operational throughout the lifetime of the hosts.

3.3 Software

Most of the software used for the computing cluster is open-source. The two exceptions are the annually-renewable contract for KSplice [4] and the one-time cost for Synapsense [5] monitoring infrastructure (software and hardware). The combined cost for both is negligibly small ($3/core each for RHIC and USATLAS).

3.4 Staff

The cost of staff includes salaries and employee benefits (health insurance, retirement program, defined vacation accrual, travel and training allowances, etc) and for the purposes of this study, we assume the staff is made up of highly-skilled individuals with at least 5 years of work experience. The average annual compensation package is US$200,000 per FTE (full-time equivalent). The RACF computing cluster is staffed by 4 FTE, which translates to $34/core each for RHIC and USATLAS in 2013.

3.5 Electrical power

The cost of electrical power to BNL has hovered around US$0.05 per kWh over the past few years. Figure 3 below shows the cost of electrical power per physical core for the RHIC and USATLAS computing clusters. RHIC costs are somewhat higher than USATLAS because of differences in hardware configuration and usage patterns. Although the number of physical cores per cpu has risen steadily, the actual power usage per server has remained almost flat. As a result, the electrical cost per core has dropped over the years. In 2012, the average instantaneous power consumption per core was \( \sim 25 \text{ W} \), and the computing clusters were responsible for \( \sim 60\% \) of all electrical power costs at the RACF.
3.6 Space Charges
In BNL lingo, space charges is an overhead cost charged to program funds to pay for data centre physical infrastructure (cooling, UPS, building lights, cleaning, security, repairs, etc) maintenance and is based on floor space usage (measured in $ft^2$) and other factors. Space charges also cover most of the labour costs associated with the maintenance of the physical infrastructure. The rate can be reset on a yearly basis and is subject to fluctuations due to external forces. Space charges are higher for USATLAS because it occupies a larger fraction (~60%) of the floor space. Figure 4 shows the space charges per core for the computing clusters, which occupy ~50% of the data centre floor space. Although space charges per core have declined, it has risen considerably in dollar value, since the number of computing cores has risen sharply over the past 5 years (see Figure 1).

3. Summary of average costs
To calculate the cost of the dedicated computing clusters at the RACF, we averaged the cost per core of the software, staff, electrical and space components over the 2009 to 2013 period. The costs for servers and network were first averaged over the 2009 to 2013 period and then amortized over the expected lifetime (4 years for USATLAS and 6 years for RHIC) of the hardware.
Table 2: Summary of major costs at the dedicated computing cluster at the RACF

|               | USATLAS ($/core/yr) | RHIC ($/core/yr) |
|---------------|---------------------|------------------|
| Servers       | 228                 | 277              |
| Network       | 28                  | 26               |
| Software      | 3                   | 3                |
| Staff         | 34                  | 34               |
| Electrical    | 12                  | 16               |
| Space         | 27                  | 13               |
| Aggregate total | 332 (0.038/hr)     | 369 (0.042/hr)   |

Note that these numbers assume physical cores only. In practice, the RACF uses hyper-threaded cores, which doubles the number of available computing cores and cuts the aggregate cost in half (~$0.02/hr).

4. Near-term trends and other considerations

This study has revealed several near-term trends regarding future costs. While the number of cores per server has gradually increased, the cost per core seems to be flattening out, and there doesn’t seem to be much more savings to be realized unless alternative technologies are considered, such as AMD-based processors. Since servers are the single largest contributor to computing costs, cost-control efforts will be prioritized in this area first. The outlook for staff cost is stable and predictable, but space charges have been trending higher due to internal organizational dynamics. Electrical rates have been stable, but can be volatile. The near-term outlook for network depends on the choice of technology as we approach the end of life for 1 GbE connectivity. The RACF is now investigating Infiniband (IB) and 10 GbE technologies as possible replacements for 1 GbE in the next few years, but if costs for IB or 10 GbE remain high, network costs may rise in the short-term owing to higher expected bandwidth requirements.

A recent development at the RACF is the issue of high-availability of data systems, which can be affected by hardware failure or loss of physical infrastructure (power and/or cooling) due to man-made or natural (for example, hurricanes) events. The RACF is investigating data duplication to address this issue. A second set of the data can be placed on tape, NAS, worker nodes, or in the Cloud, among other choices. A preliminary calculation shows that a 15 PB deployment (equivalent to our current storage capacity) on worker nodes would add ~$40/core/yr in hardware costs to servers. Additional software may be necessary but has not been factored in yet.

5. Summary

A dedicated computing facility such as the RACF is competitively priced. The RACF (~$0.04/hr) compares favourably with the similar EC2’s ($0.12/hr) on-demand instance. If we compare virtual cores (~$0.02/hr), we approach prices ($0.013/hr) for EC’s spot instance, even though jobs on RACF’s virtual cores are not subject to eviction as with spot instance.

Near-term trends indicate significant challenges to control costs at the RACF. While controlling hardware costs (the largest contributor to the overall costs) may benefit from technology advances, other fringe costs over which the RACF has no control over (such as space charges and electrical rates) are on the rise. Possible new requirements (such as data duplication) will raise costs too.

References
[1] RACF: www.racf.bnl.gov
[2] EC2 prices: www.amazon.com/ec2/pricing
[3] Sergey Panitkin et al., “ATLAS Cloud Computing R&D”, CHEP 2013
[4] Ksplice: www.ksplice.com
[5] Synapsense: www.synapsense.com