ON THE ECONOMICS OF HUGE REQUIREMENTS OF THE MASS STORAGE
A Case Study of the AGATA Project

Víctor Méndez Muñoz*, Mohammed Kaci, Andrés Gadea and José Salt
IFIC - a mixted intitute CSIC and Universitat de València, Apt. Correus 22085, E-46071, València, Spain

Keywords: Mass storage systems, Grid, Cloud, Costs analysis.

Abstract: The AGATA is a shell detector for gamma-ray spectroscopy. At the present stage of the project the AGATA collaboration is running an AGATA-demonstrator, which is a small part (only 12 Germanium crystals) of the future full AGATA spectrometer (180 crystals). The AGATA-demonstrator is producing a huge amount of raw-data with a high throughput. This paper focuses on the economics study regarding various options of data storage for the AGATA spectrometer. We discuss the raw-data storage requirements on the demonstrator and the forecasted storage size requirements for the full AGATA spectrometer. We also analyze the data communication requirements. The case study focuses in costs analysis of three options: a dedicated storage, a Grid storage and a Cloud storage service. In this manner, we explain how a huge size mass storage can be affordable, and why the costs savings depends more in the particularity of the problems than in general estimations. The results show a lower total costs for the Grid option.

1 INTRODUCTION

The AGATA collaboration (http://www-win.gsi.de/agata/) is using a high purity Germanium crystal based multi-detector array as a γ-ray spectrometer in multiple experimental configurations. The AGATA spectrometer array produces a huge amount of raw-data that would be reprocessed later in order to reduce them by a factor of about 20. The obtained reduced data are used by the physicists for their analysis. The raw-data reprocessing is based on the novel concepts of Pulse Shape Analysis (PSA) and γ-ray tracking (Kaci et al., 2010). The raw-data analysis have to be reprocessed off-line, since the quality of the response of the AGATA detector array depends on the performance of the PSA and data tracking.

Nowadays, the AGATA collaboration is running a demonstrator detector-array which is composed of 4 triple highly-segmented Ge semiconductor detectors (12 Ge crystals). The full AGATA γ-ray spectrometer, with 180 Ge crystals, will run in the next few years. With the demonstrator configuration, the raw-data production is in average of 10TB by experiment. Thus, each triple detector throughput is 5 TB in average. Depending on the experimental configuration, with more detectors the filters of the data taking are less effective, a complete ball can increase to 5 TB in average for a triple detector. More than 30 experiments a year are forecasted in the AGATA collaboration. With this premises we are talking about a mass storage requirement range from 5PB/year up to 10PB/year for the AGATA raw-data production. To particularize the case study, the off-site storage for backup and recovery system is not contemplated.

When the mass storage scales up to a huge size, the storage system becomes critical regarding the adoption of an overall computing solution. As the storage system can be separated from the rest of the computing solution, in this paper we focus on the mass storage requirements, regardless of the other computing requirements of the problem. Another important factor of this case study is, the collaborative environment between the distributed research groups of the AGATA experiments. This feature introduces some constrains to the problem, related with the raw-data accessibility while permissions are granted.

We make this case study in order to know how a huge mass storage can be affordable, and why the costs savings depends more in the particularity of the problem than in the general estimations. For this reasons, in this paper we focus in a detailed

*Currently at PIC: *vmendez@pic.es*

Méndez Muñoz V., Kaci M., Gadea A. and Salt J..
ON THE ECONOMICS OF HUGE REQUIREMENTS OF THE MASS STORAGE - A Case Study of the AGATA Project.
DOI: 10.5220/0003380505070511
In *Proceedings of the 1st International Conference on Cloud Computing and Services Science* (CLOSER-2011), pages 507-511
ISBN: 978-989-8425-52-2

analysis of the costs. In Section 2 we analyse the different cost factors. We identify the important cost factors for a mass storage provider. Furthermore, we qualify the costs analysis within the constrains of the AGATA project. In Section 3, we focus in the AGATA costs estimations of the storage service. Particularly, we analyse a dedicated storage system, a Data Grid solution within the European Grid Initiative (http://www.egi.eu/projects/egi-inspire) infrastructure, and a solution with the Amazon Simple Storage (S3) (http://www.amazon.com/s3) service. Section 4 discusses the results in the AGATA context. The case study is summarized in Section 5.

2 RELEVANT MASS STORAGE COSTS OF THE AGATA STORAGE REQUIREMENTS

In this section we follow the general guidance in (Alek Opitz and Szamlewska, 2008) to estimate the costs of a Grid resource provider. Several parameters have to be taken into account. Some of them are directly assumed by the AGATA project, other costs are assumed as a best effort by the different partners of the AGATA collaboration. Both of them have to be taken into account in the costs analysis. Therefore, we classify the typical parameters as follows

- Storage Total Costs of Ownership (TCO):
  - Hardware costs.
  - Business premises costs.
  - Software costs.
  - Personel costs.
- Costs of data communication with the computing center.

The following subsections analyse the costs breakdown differences between the three options: the dedicated storage, the Grid storage and the Cloud Infrastructure as a Service (IaaS).

2.1 Storage Total Costs of Ownership

The TCO is the aggregated costs of the hardware acquisition, the replacement of defective components, the business premises including electricity, the software and the personnel costs. The dedicated storage and the Grid storage require the acquisition of the hardware, software installation and the maintenance in the datacentre. The TCO in Cloud is included in the storage service price.

Some papers about storage TCO have already been published. Some of them are vendor reports focused in consulting optimization of the TCO (IBM, 2003; Marrill, 2008). Such models are based on general parameters for a wide range of storage systems. For our particular case we have found in (Moore et al., 2008), real costs analysis of the storage infrastructure at the San Diego Computing Center (SDSC). The SDSC storage features are valid requirements for the AGATA storage.

- The storage size of the SDSC costs analysis is 7PB, while in AGATA is about 10PB/year.
- The storage is a system of an archival in tapes and disk-caching.
- The storage is standard range since the data are ”write-once-read-rarely”.
- The service is 24x7.

The SDSC storage reports a TCO breakdown in the metric of $/TB/year. Table 1 shows these costs in €/TB/year. The upper part of the table are translations from the SDSC costs analysis. The Grid specific costs are explained below.

Table 1: Estimated normalized annual costs of storage.

TCO Breakdown Costs TB/year	SATA Disk 1 PB	Archival Tape 9 PB
Disk/tape media	396	74
Other capital	174	122
Maintenance	170	81
Facilities	118	18
Personel	251	74
Dedicated TOTAL	1110	370
Other capital	116	81
Maintenance	0	0
Personel	168	49
Grid TOTAL	798	228

Hardware costs are the disk/tape media costs, and also the other capital costs referring the replacement components, the file system servers and the storage area network. The maintenance costs are referred to the support and the software licensing. Facilities costs are the business premisses including electricity, estimated of the ”floor space” cost in the datacentre. The personnel costs are 3 Full-time Person Equivalent (FPE). The above are the translation in Euros from the SDSC datacentre, to apply in the dedicated storage option.

The Grid option has reductions in some of the costs. In general terms, all the fixed costs are reduced

because of the economy of scale, sharing resources with other Virtual Organizations (VOs). In our particular case study, for simplicity, minor fixed costs differences, like building amortization, are not considered. There are also vendor discounts for capital purchases and maintenance, which are difficult to estimate, because it is not unusual that the negotiated pricing is confidential. Such discounts can be important in the Grid option compared with the dedicated datacentre, but they are not taken in consideration for Grid costs estimation.

In other capital costs of the Grid, the file system servers and the storage area network have costs reduction because they are shared with other VOs. We have estimate a reduction of $2/3$, which is easy to reach in the Grid context of EGI. Regarding the maintenance and licensing costs, we take the example of our institute complex Grid infrastructure. We have a site belonging to the federated Tier-2 for the ATLAS experiment (http://atlas.ch/), and a site belonging to the Grid-CSIC infrastructure (http://www.grid.csic.es), with a total storage near of 1 PB. These infrastructures are integrated in the National Grid Initiative NGI/EGI, and there is an operational collaboration with the rest of Grid site teams, for coordination and self-support. The operation of this complex Grid site includes not only the datacentre, but also computing resources administration. This complex infrastructure is supported by a team of 3 FTE persons working without third party maintenance support. Additional support is obtained from the collaboration with the NGI-EGI operation groups. This is possible in the collaborative environment of the Grid communities, with a know-how sharing context. About licensing, the gLite middleware is provided by the European Middleware Initiative (EMI) (http://www.eu-emi.eu/) with opensource license cut off cost. The mass storage systems supported by gLite middleware are dCache and CASTOR (Burke et al., 2009). In our Grid complex infrastructure we also use Lustre (SunMycrosytems, 2009), which neither have licensing costs. Other third party mass storage systems can have licensing costs. If we consider these premises, we can take a similar scheme for the AGATA storage Grid option and cut off the maintenance and licensing costs. Finally in the costs breakdown analysis, Table 1 assigns to the Grid storage of AGATA a 2 FPE, 1 FPE less than the dedicated datacentre, because the economy of scale in the operational tasks which can be done for many VOs.

Since most of the Cloud providers offer a wide range of storage service, the AGATA storage requirements are in the standard range of the Cloud storage.

2.2 Data Communication Costs

In what concerns AGATA, the data source is the Data Acquisition system (DAQ). The DAQ includes a premium range storage system, which is able to deal with the raw-data throughput of the AGATA detector. For cost reasons, the DAQ storage size is reduced to the space required for processing the data of the active experiment. Therefore, the DAQ needs to transfer the produced raw data in quasi-real time to the mass storage system. In the following we analyse the transfer requirements.

In the Introduction Section, we have shown that the experimental data produced by the AGATA demonstrator with 4 triple detectors is 10TB on average. For network requirements analysis we take into account not the average but the peak throughput of the experiments, which is about 20TB for the AGATA demonstrator. The peak experiments produce 5TB for a triple detector throughput. If we considerer the mentioned filtering factors, this throughput in the complete AGATA ball can reach 10TB, which gives a total of 600TB for the peak experiments.

In our transfer tests we get 170MB/s of effective transfer rate. Therefore, for a peak experiment 600TB of raw data is transferred in 42 days. This is clearly unscalable since there are about 30 experiments planned for each year, and the AGATA DAQ has storage space for only one experiment. For this peak size experiments the AGATA project can book some extra off-time before the next experiment. The transfers can start during the data taking, so the peak transfers can take two weeks. For our peak network requirement, 600TB in 14 days, it is necessary an effective transfer rate of 520MB/s, equivalent to 4,160 Mbps, which requires a dedicated network of 5Gbps rate.

This analysis illustrates a premium network requirements, not only at physical layer but also in the transfer software, to scale to the 60 triple detector data transfer. Private leased lines are dedicated circuits, with price depending basically on speed and distance. A good option for our purpose can be two circuits of OC-48 (Optical Circuit at 2,448 Mbps) to reach the required 5Gbps of full time dedicated connection. A estimation in (NortelNetworks, 2009) says that a 18 months leasing is 2,500 £per fibre mile, including installation and the rest of the costs. This is equivalent to 1,281 €/km for a year channel leasing. Vendors discount are usual for multiple channels after the first one, but it is difficult to estimate, for our purpose we take the costs of two complete channels. For the distance estimations, we take the distance between the AGATA demonstrator DAQ and the data storage, ab-

out 100 Km.

In the dedicated storage option, the storage system can be directly endorsed to the DAQ. In this case, for the data transfer from the source to the storage system, the cost is included in the hardware costs of the datacentre local network. Traditionally, the physics experiments have used this schema. In these cases, the raw-data are stored in tapes at the DAQ, and each research group took the tapes and transport them to their institute, where they process the raw-data on their own. Such schema is not possible in the AGATA collaboration for two reasons. The first one is: the raw-data are owned by the entire collaboration for security and scientific reasons, with the appropriate access rights to the different experimental raw-data. The second reason is: the AGATA raw-data size of each experiment becomes the processing a not trivial matter, so research groups have to join forces for an affordable computing process. Thus, in AGATA we are talking of raw-data transfer and processing on appropriate computing performance. In this manner, we have to consider the transfer costs between the standalone storage system and the computing center of the raw-data processing.

The Grid option has to estimate the transfer costs from the DAQ to the Grid storage. Grid Computing technologies offers the possibility to send the processing jobs to run at the sites where the data are located, therefore, no additional remote transfers are needed.

Cloud companies incurs higher network charges from their service providers for storage of terabytes (Myerson, 2008). Since AGATA is a petabytes storage, a way to obtain better network costs is to negotiate with the cloud provider a third-party network leasing. In this case study, we consider a such third-party network leasing, without Cloud provider charges, for the costs estimations. The Cloud option requires, at least, 1 raw-data transfer from the DAQ. Additional transfer of the raw-data can be needed if the processing is performed out of the Cloud.

3 ESTIMATED COSTS OF THE MASS STORAGE OPTIONS

In this section we analyse the storage models of the different options, and we get the cost estimations in year basis following the premises of Section 2. The storage costs on delivery to the Computing Center is the addition of the storage TCO and the data communication costs.

The storage TCO/year for the dedicated storage and the Grid storage options is in the Equation 1, considering the costs of Table 1.

$$TCO = Disk\,€/TB * 1PB + Tape\,€/TB * 9PB \quad (1)$$

Where *Disk* and *Tape* have different values for the dedicated storage and the Grid storage.

For the Cloud storage option we have choosen a standard storage service, good enough for AGATA storage requirements shown in Subsection 2.1. A Cloud storage option which fulfil those requisites is the Amazon Simple Storage Service(S3). The price list depends on the location of the Amazon S3 bucket, (the European Union in our case study), and the storage used, obtaining discounts for bigger sizes of storage used. The S3 prices are in $/GB/month. The storage size requirements in a year is about 10PB for the full AGATA spectrometer, giving in average 833 TB per month. This affects to the price range, in the first year different prices have to be applied until reach the lower unit price for the bigger S3 storages over 5,000TB. Equation 2 shows the total year TCO in function of the month i charge.

$$TCO = 833 * t_1 + (\sum_{i=2}^{4}(833 * i)) * t_2 + (\sum_{i=5}^{12}(833 * i)) * t_3 \quad (2)$$

Where $t_1 > t_2 > t_3$ are the different S3 monthly price-list for the accumulated storage size. Nowadays, in November of 2010, these S3 price-list is for our case: $t_1 = 71.99, t_2 = 60,62, t_3 = 41,68$ in €/TB/month. The next years storage costs would be accumulated with the previous years costs. This is the same for the three evaluated options. Considering these premises the next years costs addition to the previous TCO is in Equation 3.

$$TCO = (\sum_{i=1}^{12}(833 * i)) * t_3 \quad (3)$$

Regarding the data communication explained in Subsection 2.2, the standard connection is a 2 channels OC-48 of a leased private line of 100km distance, with a year costs of 256,200 €/year. The dedicated storage option, endorsed to DAQ, requires a connection from dedicated storage to the Computing Center. The Grid option requires a connection from DAQ to the Grid storage. The Cloud option requires a connection from DAQ to Cloud and other from Cloud to the Computing Center.

Table 2 summarizes the year costs of the three options in €, to be accumulated to the previous years storage costs. Dedicated and Grid is from Equation 1, Cloud year 1 is from Equation 2 and Cloud of the next years (+1) is from Equation 3.

Table 2: Total acumulative for €/10PB/year on delivery.

Storage Option	Storage TCO	Transfer Costs	TOTAL
Dedicated	4,546,560	256,200	4,802,760
Grid	2,918,400	256,200	3,174,600
Cloud year 1	2,875,357	512,400	3,387,757
Cloud year +1	2,708,116	512,400	3,220,516

4 DISCUSSION OF THE RESULTS

It is important to comment a particular issue of the Cloud solutions, which is related with the costs . With a traditional datacentre the costs are up-front, it is a Capital Expenditure (CapEx), while Cloud Computing is *pay-as-you-go*, it is an Operating Expenses (OpEx). CapEx refers to an investment for a long period of time. CapEx assets are depreciated in value over time on the accounting. OpEx refers to expenses incurred excluding cost of goods sold, taxes, depreciation and interest. Cloud Computing moves CapEx to OpEx, but there is no reason to think that there is a financial benefit from here. Anyway, they are hardware costs to be considerated, both in the storage acquisition and in the price of the IaaS.

It is also important, the difference between the storage acquisition options and the Cloud storage option. The Cloud costs are the Amazon S3 prices, therefore, it is true cost value, while in this sense, the storage acquisition options have estimated costs. This is an economical forecast advantage to the Cloud option.

In this context, Table 2 shows the Grid option with the lowest total costs of storage and data communications. The lowest TCO/year without transfer costs, is for the Cloud option.

Another consideration of Cloud storage is about the possible integration with Cloud Computing for raw-data processing. In this case, the storage upload is raw-data and the storage download is reduced data in a factor of 20. The communications requirements would be a connection of two channels OC-48. With this scheme, regardless of Cloud Computing costs, the Cloud storage on delivery is lower costs than the options in Table 2.

A final consideration in the Grid option is about the integration of a complete Grid solution for storage, data management and computing processing. In this case, communications from DAQ can be integrated in the Grid infrastructure to cut off the costs, because the economy of scale, sharing higher performance networks with others VO.

5 CONCLUSIONS

This case study analyses three options for a huge requirement mass storage system, in the context of the AGATA collaboration. We have shown how such a storage system can be affordable, focusing on the particularity of the problem. The estimated metrics have been taken from the AGATA demonstrator detector array (12 Ge crystals) throughput testing, and also from the data transfer tests. Considering the results, this case study supports the adoption of Grid mass storage system for the AGATA collaboration.

ACKNOWLEDGEMENTS

We are greatly indebted to the founding agency Spanish National Research Council-CSIC for their support from project Grid-CSIC.

REFERENCES

Alek Opitz, H. K. and Szamlewska, S. (2008). What does grid computing costs? *Journal of Grid Computing*, 6(4):385397.

Burke, S., Campana, S., Lanciotti, E., Lorenzo, P. M., Miccio, V., Nater, C., Santinelli, R., and Sciab'a, A. (2009). glite 3.1 user guide. http://glite.web.cern.ch/glite/documentation/.

IBM (2003). Ibm storage solutions: A total cost of ownership study. Technical report, IBM.

Kaci, M., Rechina, F., Mendez, V., Salt, J., and Gadea, A. (2010). Testing the agata pulse shape analysis and g-ray tracking on the grid. In *4th Iberian Grid Infrastructure Conference Proceedings*, page 482484. Universidad do Minho, Netbiblio.

Marrill, D. R. (2008). Storage economics: Four principles for reducing total cost of ownership. Technical report, Hitachi Data Systems.

Moore, R. L., DAoust, J., McDonald, R. H., and Minor, D. (2008). Disk and tape storage cost models. Technical report, San Diego Supercomputer Center, University of California.

Myerson, J. M. (2008). Cloud computing versus grid computing: Service types, similarities and differences, and things to consider. Technical report, IBM - developerWorks.

NortelNetworks (2009). Managed wavelength services: The new wave bandwith options, commercial applications and values. http://www.datasheetarchive.com/datasheet-pdf/010/DSA00164117.html.

SunMycrosytems (2009). Sun lustre storage system featuring the lustre file system. http://www.sun.com/servers/hpc/storagecluster/datasheet.pdf.