



Financial Study for Sustainable Computing e-Infrastructures

Deliverable Title: D. 2.3 Computing e-Infrastructure cost estimation and analysis – Pricing and Business models

Partner Responsible: AUEB-RC

Work Package 2: Computing e-Infrastructure cost calculation

Submission Due Date: 31.1.2013

Actual Submission Date: 1.3.2013

Amendment Date: 3.7.2013

Distribution: Public

Nature: Report



Abstract: This document provides an overview of the European HTC and HPC costs calculated with the e-FISCAL hybrid methodology, while comparing these costs with commercial cloud offerings. It also presents the developments of business and pricing models for the use of commercial resources. Finally it provides an evaluation of the related estimations and concludes with recommendations for various stakeholders.

Document Information Summary

Deliverable number:	D2.3
Deliverable title:	<i>Computing e-Infrastructure cost estimation and analysis – Pricing and Business models</i>
Editor	<i>Sandra Cohen</i>
Contributing Authors:	<i>Sandra Cohen, Fotis Karagiannis, Costas Courcoubetis AUEB-RC, Kashif Iqbal ICHEC, Sergio Andreozzi EGI.eu, Matti Heikkurinen, ETL</i>
Reviewer(s):	<i>Sy Holsinger EGI.eu, Matti Heikkurinen, ETL</i>
Work package no.:	WP2
Work package title:	Computing e-Infrastructure cost calculation
Work package leader:	<i>Sandra Cohen</i>
Work package participants:	<i>Sergio Andreozzi, Panagiotis Antonopoulos, Owen Appleton, Costas Courcoubetis, Matti Heikkurinen, Sy Holsinger, Kashif Iqbal, Fotis Karagiannis, Steven Newhouse, George Thanos,</i>
Distribution:	<i>Public</i>
Nature:	<i>Report</i>
Version/Revision:	<i>v.am1</i>
Draft/Final	<i>Final</i>
Keywords:	<i>Cloud offering, pricing models, business models,</i>

DISCLAIMER

e-FISCAL is a Research Infrastructure Project co-funded by the GÉANT & e-Infrastructure Unit of the Information Society & Media Directorate General of the European Commission. e-FISCAL targets computing e-Infrastructure providers, national funding agencies, scientific communities, as well as European Union policy makers. e-FISCAL is supported by an external Advisory Board (AB).

This document has been produced with the co-funding of the European Commission. The content of this publication is the sole responsibility of the e-FISCAL consortium and cannot be considered to reflect the views of the European Commission.

Grant Agreement Number: 283449

Start Date: 1 August 2011

Duration: 18Months

Document Approval

Approval by	Date
Project Management Board	3/7/2013

EXECUTIVE SUMMARY

Over the last ten years the European countries and the EC have made significant investments in e-Infrastructures for scientific computing, notably High Throughput Computing (HTC) and High Performance Computing (HPC) services. Their contributions have made it possible for the European research to maintain most prominent role in solving key global challenges. The approach has been validated by a growing number of research initiatives and success stories such as the recent discovery of the Higgs boson. Sustainability of such services is essential, as the research supported by them is also more and more crucial for European competitiveness – however, sustainability can only be planned if the costs are known. The emerging commercial offerings (Cloud-based HTC and HPC solutions) pose additional opportunities and challenges for sustainability. It is thus important to understand the cost of the dedicated computing related e-Infrastructures, as primarily expressed by EGI and PRACE initiatives, both in their current state and in their evolution towards the Horizon 2020 framework. The calculation of such costs is not trivial; keeping detailed accounting data for the evolving mix of capital (hardware) and operational (personnel, energy) expenditures is only one of the challenges. This is because, each of the compared infrastructures among the HPC, HTC and Cloud offerings have their distinct characteristics and use-case applications. For example, high-end HPC systems tailored for capability computing deploy state of the art hardware and network as compared to HTC infrastructure.

The goal of the e-FISCAL project has therefore been to analyse the costs of the current European dedicated HTC and HPC computing e-Infrastructures for research and compare the service components provided with equivalent commercial leased or on-demand offerings. With regard to the comparison, it is crucial to keep in mind that the service palette provided by the computing e-Infrastructures – which includes advanced user support, engagement with user groups, cross-organisational authentication and authorisation mechanisms – extends beyond the basic Cloud computing service used in the comparison. Furthermore, the results should not be used to evaluate moving from in-house to commercial public clouds; this is a different exercise and is a next step requiring a different approach based on avoidable costs, while this is also related to the qualitative value of e-Infrastructure that cannot be easily measured. Still, this cost comparison allows a baseline quantitative analysis of the cost ratios that is more easily repeatable than comparison addressing broader set of services and their qualitative aspects.

To perform the cost analysis, the project developed a hybrid costing methodology that builds on the two main methodologies used for cost assessment in e-Infrastructures: Total Cost of Ownership (TCO) and Full Cost Accounting (FCA) that have been used in most of the case studies identified in the state of the art analysis. However, neither of the methodologies was sufficient for reaching the e-FISCAL goals. While TCO is a useful concept in assessing the cost of a specific project over its useful life, it sets high demands for the analytical model and data used to make predictions. In a heterogeneous distributed e-Infrastructure it is very hard to come up with acceptably precise results by using a methodology that is almost completely forward-looking. On the other hand, FCA methodology relies on actual cost accounting data information that is dependent on the level of cost accounting systems sophistication by additionally inducing a backward looking stance, and gathering this type of information in Europe-wide infrastructure with sufficient accuracy and coverage is impossible.

The e-FISCAL model is a hybrid model that approximates the costs of maintaining services at their current level in the short to medium term. It does that without the need to identify funding sources or the exact time when the

actual infrastructure investments have been made. Therefore, rather than relying on the detailed financial data or projections to the future, it uses high level information about computing and storage hardware costs (including interconnection costs), auxiliary equipment costs (i.e. cooling, uninterruptable power supply devices - UPSs, power generators), software costs, personnel costs, and site operating costs. The e-FISCAL methodology is completed in two phases; firstly a simulation of the physical infrastructure is sketched and secondly the annualised cost of the simulated physical infrastructure and the operating cost of the physical infrastructure are added together. All the data necessary to feed the model (e.g. e-Infrastructure acquisition costs, personnel costs, electricity cost, depreciation rates) is retrieved from a properly developed questionnaire (the e-FISCAL survey instrument).

In the study, the e-FISCAL methodology has been applied to a sample of HTC and HPC centres in Europe making this the first initiative to gather and analyse the costs from a representative number of countries in a comprehensive and systematic way. As is evident by the state of the art review, the majority of costing studies either concentrate on a site or make a multiple site albeit in the same country comparison. The analysis revealed that the cost per core hour for 2011 lies in the range of €0.03 core hour to €0.07 core hour¹.

HOW TO READ THE RESULTS

The cost analysis in this project has been undertaken under three alternative approaches – basic case (costs for whole e-Infrastructure), basic case split (cost calculated for HTC and HPC services separate) and case by case analysis (per computing centre approach) – with results calculated for both yearly cost per core and cost per core hour. To preserve anonymity of the answers, results are discussed consistently using average and median values. As there is a significant difference between median and average values, it is useful to discuss this difference and help managers to interpret the results. Average is the mean value taking all sites equally into account regardless of how far from the “norm” they fall. In contrast, the median value has the inherent property of excluding the “extreme” values (high or low), and thus represents the “most likely value for a typical case” that is more appropriate for guidance and decision making. Median is thus a more relevant reference point for both computing centre managers and policy makers.

Average values are always higher than their median counterparts. This is an indication that there are computing centres that are less cost- efficient when measured purely in terms of cost per core hour, and that there is space for improvement in order a site to move towards the cost that corresponds to the median value. Sites that are close to (or below) the median value should be considered as cost efficient compared to the group of academic installations in our sample. Sites above the average value should try to analyse the issues that affect their cost efficiency, and try to separate activities providing genuine added-value (that may not be captured by the cost per core hour metric) from the areas where efficiency improvements would be possible without adverse consequences.

¹ Note that the above values takes into account all the personnel working directly or indirectly in the centre, namely for administration, operation, middleware and application development, user support, training, dissemination, as well as managers and policy makers

Some other underlying issues revealed by the analysis were:

- Hardware depreciation rates are in several cases well above the typical well-cited three-year period. The average depreciation period corresponds to 5 years,
- The Power Usage Effectiveness (PUE) rates are around 1.5 (median value) indicating rather efficient energy use, and
- Approximately 2 FTEs are required for each 1,000 cores (median value).

The analysis also provided evidence of a decreasing trend in costs from year 2010 to 2011.

After assessing the per core hour costs, they are compared with the prices charged by commercial cloud providers (such as Amazon EC2). The comparison is done in two variants: first without taking into consideration performance differences between in-house service provision and commercial cloud services, and second by adjusting for the differences in performance based on the benchmarking results (performance-adjusted cost comparisons). The small-scale benchmarking exercise included in the project was used to establish range of performance degradation percentages that correspond to the most relevant in-house infrastructure installations and their commercial service counterparts.

By concentrating only on financial considerations, our findings provide strong evidence that **cloud prices are not necessarily** and unanimously **lower than** the costs calculated by e-FISCAL for **in-house service** provision. This holds true even without taking into account the differences in the actual service portfolios discussed above. Parameters like utilization rates, infrastructure size, length of time committed to use cloud services (for example by using “reserved instances” offered by some of the Cloud vendors), efficiency in manpower utilization and service performance are important factors affecting the basic cost relation between cloud and e-Infrastructures. Thus cost is not synonymous to value; on the contrary **cost is only one of the parameters that should be considered in the value creation process.**

Main achievements and findings

During the final part of the e-FISCAL project, the consortium consolidated its findings focusing on the cost estimations of the European computing e-Infrastructures based on survey answers and other relevant data (market and literature). Estimates of basic metrics, such as the cost per core hour, have been updated based on a slightly extended sample (basic case²) as well as based on a new methodology (the case-by-case analysis³), which narrowed down the result ranges. Furthermore, the project calculated separate values for HPC and HTC (basic-case split⁴) based on the corresponding sub-samples (using the respondents own feedback regarding their categorisation of HPC or HTC (or both), we call this analysis basic-case split). We used the number of EGI cores in order to approximate the total yearly cost of the dedicated European HTC e-Infrastructure and the number of PRACE Tier-0 centres' cores together with other national Tier-1 centres' cores in order to approximate costs of the dedicated European HPC e-Infrastructure. The number of cores is based on figures reported in public deliverables or project websites, and analysis is based on the values 2011 (the latest year in the survey scope)⁵. The benchmarking results with the NPB and HEPSEPC06 benchmarks have been integrated in the costing exercise by introducing the performance degradations through weight factors, and updated (and fairer) comparisons have been made with commercial cloud providers, adding OVH.com besides Amazon. Sensitivity analyses have been performed showing interesting results, in particular illustrating the crucial role of the utilisation rate of the in-house infrastructure and comparing the results with the different instances from Amazon. Moreover, the project extended to other areas such as business models for the use of commercial resources serving the research community and possible pricing schemes for leasing resources. The main results achieved during the second reporting period can be summarised as follows:

- **Update of the State of the art:** The project identified and *reviewed* a total of *50 related publications and articles (another 11 since the first period)* with emphasis on business and pricing models. The state of the art review is published on the project website and has proven to be an important tool for engaging with the researchers and broader audience. The link to the e-FISCAL state of the art repository from Wikipedia page on Cloud Computing⁶ remains an important source of visitors to the project website.
- **Updated cost estimations for European HTC/HPC e-Infrastructure:**

The **sample** for the e-FISCAL financial study has been *extended with an additional 4 responses including 2 new countries (Italy and the UK)* bringing the total sample up to *28 responses from 16 countries*. However, some of the largest HPC centres (such as PRACE Tier 0s) are not included in the survey sample due to confidentiality reasons (such as non-disclosure agreements between the vendors and these centres). The *e-FISCAL values have been calculated as follows:*

 - o **Basic case:** The updated median values (which reduce the impact of the so-called “outliers” that are either very high or low compared to typical values) are around €0.03 per core hour in 2011, while averages are around €0.07 per core hour in 2011. The detailed values are presented in the section 4 of this document.

² Results presented on **Error! Reference source not found.** on page 43

³ Results on **Error! Reference source not found.** on page 45

⁴ Results on **Error! Reference source not found.** and **Error! Reference source not found.** (page 45)

⁵ However the performance of PRACE Tier-0 systems has not been taken into account in the study

⁶ http://en.wikipedia.org/wiki/Cloud_computing

- **Basic case split (to HPC and HTC):** The corresponding values for the HPC centres (self-identified as providing HPC services) are slightly lower than €0.03 per core hour for the median and €0.05 per core hour for the average. For the HTC centres (also self-identified) the median is between €0.03 per core hour and €0.04 per core hour, while the average is close to €0.08 per core hour.
- **Case-by-case analysis:** In this case, the e-FISCAL costing model is run on each of the centres based on their reported input values. In case of missing values the average values from the whole sample were used. Based on this approach, the median cost per core hour is roughly € 0.06, while the median cost for HPC core hour is slightly below €0.04 per core hour and for HTC between €0.06 and €0.07.

All of these findings are in-line with the published results (that have been included in state of the art repository). Some key observations are listed below:

- The breakdown between CAPEX and OPEX in 2011 in our calculations is around 30%-70% respectively (median) to 26%-74% average. 51% total costs (median) is dedicated to personnel costs.
 - The utilisation rate used to calculate the average and median cost per core hour for the above results for 2011 is 65% and 75% respectively.
 - Other findings include the slow depreciation rates typically used for the hardware (average 5 years), the PUE values of around 1.5 median value (lower than typical rates quoted for the industry average) and the percentage of electricity cost (around 16-17% median value of all costs).
- **Approximated annual total cost of the in-house HPC and HTC e-Infrastructure in Europe lies between 175 and 295 M€.** In order to approximate the cost of HPC and HTC centres we use the number of cores available in 2011 and we multiply with the average and median cost per core per year under option 2 (basic case split) and option 3 (case by case analysis) summarised above. The numbers of cores in the dedicated HTC e-Infrastructure (EGI) is roughly 300.000, while the dedicated HPC e-Infrastructure (PRACE Tier 0's and other national Tier 1's) consists of approximately 550.000 cores (both in 2011). Hence the yearly cost of the dedicated HTC e-Infrastructure lies between 70-126 Million Euros, and between 103-170M€ for the dedicated HPC e-Infrastructure that includes Tier 0 and Tier 1 HPC systems⁷. The latter estimate is based on the e-FISCAL sample which is populated by Tier 1's and not Tier 0's.
- **Comparison with commercial on-demand prices** was informed taking into account the results of the benchmarking efforts of the project on the latest Amazon EC2 pricing information⁸. The comparison focuses on pure computing costs and omits some of the other costs categories: network and storage costs are not included (they depend on usage patterns of the application), personnel (such as application developers and administrators) for configuring and operating EC2 instances and for the adaptation of the application code

⁷ The difference in the performance between HPC systems has not been taken into account and findings are again based on e-FISCAL models and data

⁸ Amazon dropped EC2 prices an additional 2 times, with latest change on 1st February 2013. However, the pricing on the 31st January was retained as the basis for comparison, as this was the last day of the project and already gave considerable advantage since the prices were compared with costs based on 2011 information.

that would be needed if researchers migrate to Amazon. Comparing in-house costs with EC2 prices⁹ brings in an inherent uncertainty (not to mention costs from 2011 with prices in 2012/2013).

The comparison is not straightforward as there are several types of instances such as reserved (for 1 and 3 years) and on-demand, or low-end and high-end, which have significant differences among them. Furthermore, the utilisation of the reserved instances needs to be taken into account, since it might not be possible to absorb all of the Amazon reserved capacity in all occasions. A sensitivity analysis has also been performed having as a variable the utilisation rate of the e-FISCAL in-house infrastructure (supposing all the rest remain constant) and comparing them with the Amazon ones. Some basic conclusions (once again taking into account the benchmarking results) are:

- EC2 “on-demand” instances are always more expensive than e-FISCAL in-house findings (even “standard” Amazon instances like L/XL (Large/Extra Large).
- EC2 “Reserved” instances are much more competitive than “on-demand” ones and are comparable to e-FISCAL (and in fact towards the e-FISCAL low end, i.e. the € 0.03 per core hour).
- However EC2 “cluster compute” (HPC in the cloud instances) are 1.5 to 2 times more expensive than the e-FISCAL costs per core hour

Regarding the sensitivity analysis on the utilisation the main conclusions are (again there are variation among the different EC2 instances) summarised as follows. The in-house utilisation rates above which it is more cost-efficient to keep the infrastructure in-house are:

- 30-60% (median-average) compared to EC2 “standard on-demand” instances.
- 15-25% (median-average) compared to “on-demand cluster compute”.
- 40-95% compared to “standard reserved ones” and in one case (the 3 year standard reserved ones compared to e-FISCAL average) Amazon is always more cost-efficient from in-house even with a 100% utilization for the in-house.
- 20-40% for 1 year reserved “cluster compute” ones.

This can be further summarized that the cluster compute instances are not yet comparable and it is more cost-efficient to keep them in-house (even with a low utilisation 15-40%). While for the standard EC2 ones, on-demand are again less cost-efficient (but not at the level of cluster compute ones), while 1-3 year reserved ones are comparable to e-FISCAL and sometimes can be always cheaper (even compared to 100% of in-house utilisation).

- **Review of business models** highlighted the trend to combine the usage of public-funded federated infrastructures with commercial cloud providers to expand the offered capacity, to enable access to different resources or service levels or to enable efficient access to research data for commercial exploitation. Furthermore, pricing models characteristics from cloud services show the way to define pricing schemes for leasing in-house e-Infrastructure services (if legally permitted). The identified pricing models and classification have been derived by analysing the real-world offering from the major commercial cloud providers. Usually, all providers offer at least usage-based and subscription-based models, while only one provider digs into the market-based model. Strategy-based models may appear for certain customer domains (e.g. education, not for profit). Some key conclusions are the following:

⁹ Pricing is influenced by the profit margins, which in some cases can also be negative.

- High utilisation is key to maintain economic efficiency and thus strongly related to the business model that will be selected in a hybrid in-house and leased infrastructure.
- A broker role is essential to facilitate demand meeting the right suppliers and to increase utilisation
- Need to evolve funding streams and new pricing models.
- **Green IT:** Green IT is a very wide topic, with multitude of approaches. However, the PUE ratio is at the moment the most universally known single metric, and in addition to indicating technical efficiency of operations it may also indicate the degree of attention given to the broader Green IT issues.
- **e-FISCAL workshops:** The above results have been presented and discussed initially at the e-FISCAL workshop in **Prague** and then the final e-FISCAL workshop in **Amsterdam**¹⁰, where very positive feedback has been received from financial experts on the results.
- **e-FISCAL tools and sustainability:** e-FISCAL prepared not only the promised downloadable spreadsheet tool, but also an on-line version of the tool (only marginally funded by e-FISCAL) that will help any computer centre manager to self-assess its centres' costs, **new or existing**, and compare them with the e-FISCAL average and median values. Both are available under www.efiscal.eu/tools

Regarding the e-FISCAL outputs sustainability:

- The tools will remain available on the website under free type licenses.
- State of the art repository will also remain available and updated on a best effort basis.
- LinkedIn group on ICT cost assessment¹¹ continues discussions.
- Cost collection/estimation to continue in part through the EGI compendium (go.egi.eu/EGI-Compendium-2011).
- Consortium members can provide consultancy upon request.
- **Regarding the e-FISCAL potential impact and use,** e-FISCAL will continue to contribute to the continuity and sustainability discussions for e-Infrastructures as was done with EGI and its “Evolving EGI” workshop, with which the final e-FISCAL workshop was combined. The identified costs calculations and business models contribute in this direction. Better understanding of the financial realities related to computational e-Infrastructure services allow more efficient strategic shaping of incentives on national and European scales to drive innovation related to e-Infrastructures and the research activities supported by it. e-FISCAL claims that it laid the ground for better resource utilisation, provided the related financial backing for growing new markets and expanding investments.
- **The sustainability of the e-FISCAL effort** depends to a high degree on the number of individuals and organisations that adopt it in the future activities. The success in this area is based on the continued use of the approach in the EGI Compendium effort as well as choosing a licensing scheme that maximises the reuse and visibility of the results.
- **Satisfaction surveys** were also sent to the teams of the computing centres that answered and up to the time of the writing all answers received (19) indicate the e-FISCAL effort as useful (11) or very useful (8). The scale was from (1) major waste of effort to (5) very useful. Some of the comments received are highlighted below:

¹⁰ Reports from the e-FISCAL workshops are available: <http://www.efiscal.eu/events>

¹¹ LinkedIn group on ICT cost assessment: linkd.in/VqEth0

- The e-FISCAL approach to analyse trends and compare different market segments makes it a useful tool in understanding strengths and weaknesses of our own organisation and its role in the marketplace.
- We used the e-FISCAL work to compare with our local systems (and they were very comparable)
- I was really glad to see that a couple of beta computation tools have been included in project's website <http://www.efiscal.eu/tools>.
- The online tool is great.
- Carrying out personal interviews would probably make information gathering more efficient and complete.
- It would be useful to have structured method for HPC centres to follow so they can calculate their operational costs. This may be a set of guidelines which HPC centres have to go through (check)
- The outputs were useful. However it is very difficult to make a step forward towards sustainability for the distributed grid infrastructures due to various reasons, e.g. no clear funding path at national level, no clear sustainability path at European level (for both EGI and the "evolution" of EMI), no concrete actions to make the current distributed infrastructures available (and appealing) to new user communities (in particular the ESFRI projects).
- I would like to highlight the importance of maintaining this information updated in the coming years, in order to follow the evolution of "this market". The methodology developed under e-FISCAL, if continued in time, would provide relevant information to guide investments in e-infrastructures. Supercomputing centres need clear, reliable, and updated information to optimise investments in infrastructures like the one that e-FISCAL attempted to provide.