On optimising cost and value in eScience

Chris Broekema, P.; Allan, Verity L.; Bal, Henri E.

published in
2018 IEEE 14th International Conference on e-Science (e-Science)
2018

DOI (link to publisher)
10.1109/eScience.2018.00118

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Chris Broekema, P., Allan, V. L., & Bal, H. E. (2018). On optimising cost and value in eScience: Case studies in radio astronom. In *2018 IEEE 14th International Conference on e-Science (e-Science): Proceedings* (pp. 396-396). Article 8588737 Institute of Electrical and Electronics Engineers Inc.. https://doi.org/10.1109/eScience.2018.00118

On optimising cost and value in eScience: case studies in radio astronomy

P. Chris Broekema†
ASTRON

Verity L. Allan
University of Cambridge

Henri E. Bal
Vrije Universiteit Amsterdam

***Abstract*—Large-scale science instruments, such as the LHC and recent distributed radio telescopes such as LOFAR, show that we are in an era of data-intensive scientific discovery. All of these instruments rely critically on significant eScience resources, both hardware and software, to do science. Considering limited science budgets, and the small fraction of these that can be dedicated to compute hardware and software, there is a strong and obvious desire for low-cost computing. However, optimizing for cost is only half of the equation, the value potential over the lifetime of the instrument should also be taken into account. Using a tangible example, compute hardware, we introduce a conceptual model to approximate the lifetime relative science merit of such a system. With a number of case studies, focused on eScience applications in radio astronomy past, present and future, we show that the hardware-based analysis can be applied more broadly. While the introduced model is not intended to result in a numeric value for merit, it does enumerate some components that define this metric.**

Modern large-scale science instruments rely heavily on eScience technologies to turn instrument data into useful science results. Considering limited science budgets, of which only a small fraction can be dedicated to computing, there is a strong desire to use these expensive systems in an optimal way.

In this work we discuss the cost and value of eScience technologies, and how we can optimise the combination of these two for maximum science impact. Since these are difficult to measure for the complex combination of hardware, middleware and software that are generally required, we focus our detailed analysis on a relatively tangible component, compute hardware. We enumerate some of the factors that impact the total cost of a system. However, we propose that total cost over the lifetime of a system is only part of the equation: different systems may have radically differing values for the applications in question. A more accurate metric would look at the useful output of a system per invested Euro. For example, the Distributed ASCI Supercomputer (DAS) consortium tracks the effectiveness of its distributed cluster infrastructure via the number of awarded PhDs per cluster generation, as shown in Table I[1]. Considering the nearly constant budget for these systems, between 1.2 and 1.5 M€, discounting inflation, the cost per supported PhD has dropped considerably over the lifetime of the DAS consortium. Alternatively, we can argue that the relative science value per Euro invested in the DAS clusters has dramatically increased over time.

[1]source: https://www.cs.vu.nl/das4/phd.shtm, https://www.cs.vu.nl/das5/phd.shtml and historical data

	Year	PhDs	€/PhD	Research agenda
DAS-1	1997	7	€ 214.285	Wide-area computing
DAS-2	2002	22	€ 68.181	Grid computing
DAS-3	2006	36	€ 41.666	Optical grids
DAS-4	2010	33	€ 45.454	Clouds, diversity, green IT
DAS-5	2015	40	€ 37.500	Harnessing diversity & complexity

Table I
AWARDED PHDS PER DISTRIBUTED ASCI SUPERCOMPUTER GENERATION

More generally we assert that the usefulness of modern data-intensive science instruments, and the eScience resources that are part of these instuments, are a function of their cost and value. In other words, the relative usefulness of a system, its relative science value (M_S), depends on its total aggregate value over the lifetime of the system (Total Value of Ownership, TVO) and aggregate cost over the lifetime of the system (Total Cost of Ownership, TCO):

$$M_S = \frac{TVO}{TCO} \tag{1}$$

Whereas Total Cost of Ownership is a well known and studied concept, the same can not be said of the Total Value of Ownership of an instrument or compute system. Considering the often subjective nature of value, this is perhaps not surprising. While we do not attempt to define value as such in this work, we do explore some of the components that define the total value of an instrument or system. Using compute hardware as an example, we enumerate both cost and value components and we propose two possible measures to define the *relative science value* of a system: *total lifetime computational value* (V_c) and *total lifetime scientific value* (V_s).

We study a number examples in radio astronomy, past present and future, and show that the proposed methodology has been applied in various ways to most effectively utilise the limited available resources.

978-1-5386-9156-4/18/$31.00 ©2018 IEEE
DOI 10.1109/eScience.2018.00118

IEEE computer society