rescale

EBOOK

# HPC Cost Modeling

## Defining the Modern HPC Experience

# Table of Contents



For every $1 spent on HPC, businesses see $463 in incremental revenues and $44 in incremental profit (Hyperion Research)

If IT leaders are to realise the revenue and profit opportunities afforded by HPC, it is essential to view cost and ROI from a workload perspective. Using reductive costs structures such as price per core hour means you can actually increase overall cost of ownership by up to 30%

# Introduction

CIO's are under increasing pressure to create a disciplined, business oriented structure of accountability for the services they provide to their companies. This pressure has lead to a business like management philosophy across most aspects of the IT landscape. However HPC remains one area which has lagged entering the modern era of IT business management.

Assessing the cost of your HPC practice is a necessary and important part of providing support to HPC stakeholders without sacrificing a disciplined approach to business services. However, assessing the cost of HPC presents several difficulties in measurement.

This ebook will focus on why the historical methodology for determining cost, price per core hour is at best, incapable of providing effective cost optimization for HPC workloads. At worst, could increase your overall cost of ownership and slow innovation and productivity.

In conclusion this ebook will present a new model for assessing the cost of an HPC practice that helps align business level objectives to HPC investments. This model replaces cost per core hour with cost per workload to take advantage of an approach we call Full Stack Economics.


**KEY POINTS**

- HPC practices don't live up to modern IT standards
- HPC cost models are often misleading
- HPC resourcing should be aligned to business goals

## Why don't businesses optimise HPC costs to their most valuable assets?


In traditional operating models, IT sits isolated and separate to the lines of business, limiting their understanding of the workflows that will run on the infrastructure.

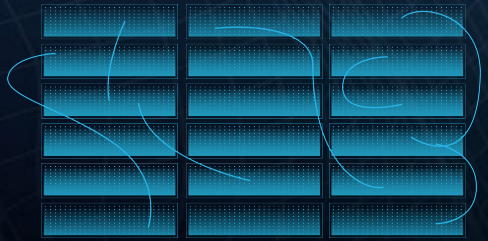
For HPC, the only information IT had to make procurement decisions was the number of cores and core hours. This led to an IT optimising to hardware specifications, in this case price per core hours.

To maximize the return from HPC investments, IT must begin to look at the total simulation costs and not price per core hour.

## **Traditional HPC cost models:** Focus on utilization and hardware

HPC cost models have traditionally focused on a pretty simple premise: **The core hour.**

Most cost models take the number of cores purchased, and divide by their runtime. This gives a unit cost for a single core hour. It represents how a job submitted might work at a given scale. For example a job run on 100 cores for one hour would be the same as a job run on one core for 100 hours.



While this model may accurately describe the relative cost of two jobs run on the same cluster, it is fundamentally flawed because it does not address key areas of HPC cost allocation, and therefore represents an inaccurate model for business level decision making.

# The problem with basing HPC decisions on core hour pricing

Core hour pricing represents a very limited view of why companies do HPC in the first place. HPC as a practice exists to serve business objectives, and therefore seeks to gain insight, improve designs, and make discoveries. This model of costing is flawed in 3 specific ways:

**1**

## Core Hours are not a Standard Unit

**2**

## Software Utilization is not Considered

**3**

## R&D Productivity is Ignored

# 1

# Core Hours are not a Standard Unit

Modern HPC resources are highly optimized, and their performance is sensitive to many factors. In some cases, changing chipset family can reduce the compute time required for a given simulation by more than 50%. Therefore an hour on one coretype might be double the performance of another. This means that a core hour does not represent the same thing when comparing diverse architecture.

**Example: How cost per core hour can lead to incorrect business decisions**
A company has standardized on coretype Alpha based on its low cost per core hour. The company now has a new project that require running FEA and CFD simulations. Because coretypes Gamma and Delta are better suited for those workloads, they cut the run time in half, and have lower overall costs.

Although coretypes Gamma and Delta have higher cost per core hour than Alpha, they provide greater overall savings on the new project and deliver faster time to answer.

| Coretype | Simulation type | Number of cores | Hours to complete | Total core hours | Price per core hour | Total cost per job |
|---|---|---|---|---|---|---|
| Coretype Alpha | FEA | 64 | 100 | 6,400 | $0.03 | $ 192 |
| Coretype Gamma | FEA | 64 | 50 | 3200 | $0.05 | $ 160 |
| Coretype Alpha | CFD | 1024 | 200 | 204,000 | $0.03 | $ 6,144 |
| Coretype Delta | CFD | 1024 | 100 | 102,400 | $0.05 | $ 5,120 |

# 2

# Software Utilization is not Considered

In many HPC practices simulation software makes up a large line item. Typically this cost is hidden from IT and held within the line of business. In many industries such as EDA, Automotive and Aerospace, software costs can be as much as 5x that of the cluster on which it runs. In this case, faster operating hardware is easily justified if it provides optimization on software costs.

In an environment of heterogeneous hardware and varying demand, teams are left with too many licences, or unpredictable needs. Many companies waste as much as 30% of their software spend due to suboptimal deployment across hardware.

The gaps in unified reporting across both software and hardware produce cost models that are not representative of actual cost of the real world workloads the company is running.

For many customers of Rescale, optimization of software licenses is the biggest opportunity for cost reduction but will never be realised if cost per core hour is the main metric considered.

## Impact of faster hardware on software utilization

Acme Company needs to run 25 simulations per month to support their team of engineers. These simulations take an average of 48 hours each to complete on their cluster. With a maximum of 744 hours in a month, Acme needs two software licenses to run their jobs on coretype Alpha.

| Core Type | Simulation Runtime | Sims per month | Total Sim time per month | Licenses Needed |
|---|---|---|---|---|
| Alpha | 48 hrs | 25 | 1200 hrs | 2 |
| Delta | 24 hrs | 25 | 600 hrs | 1 |

**By moving to a more performant coretype, Acme is able to half their simulation runtime bringing the total monthly simulation time down to 600 hours a month. This allows Acme to reduce the number of software licenses they need in order to run their standard jobs.**

25%
Savings

On average, Rescale customers are able to optimize their software spend by 25%

# 3

## R&D productivity is ignored

No matter the cost of the hardware or software, science & engineering talent will always be the most valuable asset to any company. As covered so far, companies are not considering the opportunity cost associated with engineering projects being stuck in queues or reduced productivity due to slower simulation runtimes.

By optimizing their hardware and software resources to match the simulation workloads, and not the limitations enforced by cost per core hour, Rescale customers have found productivity improvements of up to 20%.

This allows engineering teams to run more scenarios and higher fidelity models leading to better products and lower overall product lifecycle costs.

**72.8%**

Engineers report delayed or cancelled HPC jobs *(Hyperion research 2018)*

**20%**

Productivity improvements

"While 20% of the project cost has been accrued, 80% of the total lifecycle cost is determined by the test phase"

*data reported by the U.S. Department of Defense*

Due to frontloading, companies can improve overall product ROI by running more scenarios and higher fidelity simulations earlier on in the development stage of a project.

# Conclusion

Companies that adopt a **full stack economics** view of their HPC practice will achieve savings in hard expenses, improvements in the utilization of software, and increase R&D productivity. Although price per core hour is a useful data point, in order to maximize the return on HPC investments it is essential to instead shift investment decision criteria to cost per workload.

This represents a new way to consider the cost of HPC and how it is tied to business outcomes. Companies that succeed in this can shift the primary efforts of HPC teams from tactical maximization of fixed assets to the strategic alignment of business goals.

### Moving to cost per workload

Moving to a cost per workload model requires companies to benchmark their HPC systems with multiple workloads on different software applications and multiple hardware coretypes. This can be done manually, but with millions of workloads and benchmarks having been run on Rescale, our platform can intelligently determine which coretypes are most suitable to your specific requirements. Optimizing by either scalability, value or speed.

Click the button below to learn more about optimizing your HPC workloads in the cloud with Rescale. https://info.rescale.com/contact_sales

**Learn more about Rescale**