fieldfisher

Cloud Financial Management: The Core Principles of FinOps

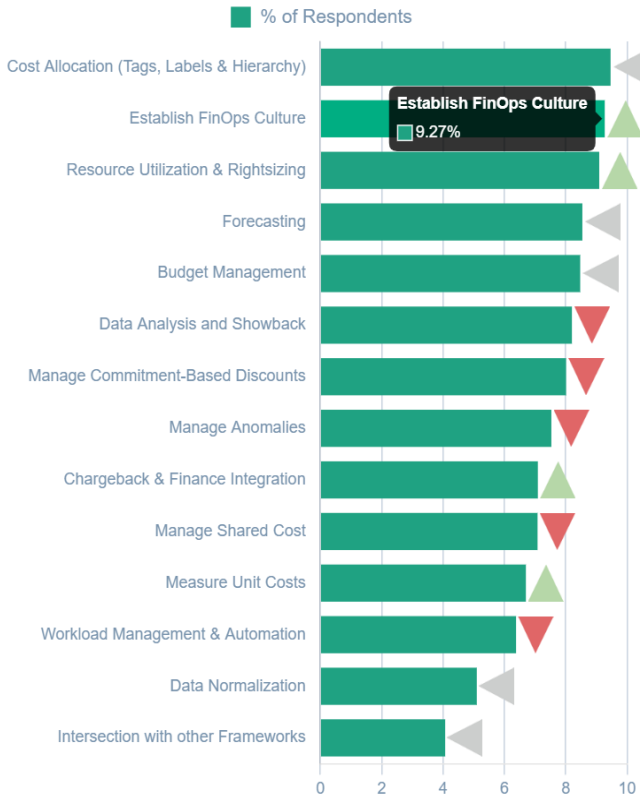
May 2023: Digital Transformation Group

Cloud Financial Management: The Core Principles of FinOps

There is little doubt that cloud services has been one of the most significant developments in the history of computing.

It is driven by an extremely powerful value proposition: IT infrastructure that is available immediately at the time and scale required by businesses thereby driving efficiencies in both IT operations and economics.

However, as the mainstream use of cloud services enters its third decade a more complex picture of the economic impact has started to emerge.[1] Although cloud services delivers on its promise early on in a business' life, use of cloud services can put pressure on margins as a company reaches significant scale and the pace of growth slows. In this article, we look at the emerging issues in FinOps (finance and devops) and discuss the key emerging priorities and practices for managing corporate spend on cloud services and the importance of establishing a collaborative culture between finance, procurement, IT and software engineering teams in order to optimise levels of spend.



Attribution: The State of FinOps by FinOps Foundation

Industry Forecasts and changes to finance and procurement processes

According to the latest Gartner study, enterprise IT spending on public cloud computing within the application software, infrastructure software, business process services and system infrastructure markets will overtake spending on traditional IT in 2025. By 2025:

- 51% of IT spending in these four categories will have shifted from traditional solutions to the public cloud, compared to 41% in 2022; and
- Almost two-thirds (65.9%) of spending on application software will be directed toward cloud technologies, up from 57.7% in 2022.

The emerging dominance of cloud services has been well publicised. One of the issues that has received less coverage is the shift in purchasing responsibility from finance teams to software engineering teams. Being self-service, scalable and on-demand allows an engineer to spend money with the click of a button or a line of code, without going through traditional finance or procurement approval processes. As a result, the roles of finance and procurement personnel have altered drastically.

Whereas many organisations used to pre-purchase large amounts of equipment in three-to-five year cycles, the emergence of cloud has meant that the new normal is buying very small units of resources for tiny amounts per hour, making fundamental changes to traditional procurement and finance processes. The graphic below from the 2023 FinOps Foundation illustrates that, whilst cost allocation remains the top priority, establishing a FinOps culture (i.e. a collaborative approach between finance and software engineering teams) is the second top priority (which we discuss in more detail below).

[1] *Andreesen Horowitz (the initial investors in Meta) highlight these issues in their paper "The Cost of Cloud, a Trillion Dollar Paradox".*

Cloud Financial Management: The Core Principles of FinOps

Cost Allocation/Resource "Rightsizing"

At its most basic level, there are two main ways to reduce/optimise cloud costs: using less and paying less for what you use:

(a) **reduce cost "run rate"** e.g. by a rate reduction by taking advantage of reserved instances (AWS and Azure) or Committed Use Discounts (Google Cloud Platform) or using spot instances; and

(b) **cost avoidance**: e.g. by terminating idle resources, "right-sizing" over-provisioned resources/ workloads, or reducing the number of resources/ workloads during off-peak times.

In the following sections, we have set out some of the possibilities (and the challenges!) for organisations to consider:

- Maximising use of the cost management and commercial offers of the various cloud platforms
- How to increase visibility of cloud services costs (with the objective being that improved real-time visibility of costs will drive better decision-making)?
- How can policies, governance and automation assist in managing cloud services costs?
- What additional challenges apply where software container technology is used?

Maximising use of cost management and commercial offers tools - AWS

There are various cost management tools within AWS' pricing model that can assist in controlling cloud services costs, the main ones being:

(a) reserved instances; and

(b) AWS Savings Plans.

For the purposes of this article, we have ignored EC2 spot instances as AWS can reclaim the capacity at short notice (but the discounts are higher, so for non-critical workloads it may be worth exploring).

Reserved Instances

In respect of EC2, since 2009 AWS has offered a pricing model called "reserved instances" whereby, in exchange for an upfront commitment of between one to three years, customers can obtain savings of between 30% to 70% on cloud services bills (when compared to the existing on-demand rates). This has obvious commercial advantages and means that EC2 can extend beyond short -term projects. Reserved instances can be paid upfront fully, partially or in arrears with the level of savings increasing with the amount of up-front commitment.

Categories of reserved instances

There are two main classes of reserved instances categories:

(a) **standard class** (which can be resold on the AWS reserved instance marketplace, but cannot be applied to a different AWS instance "family type" than the one originally specified); and

(b) **convertible reserved instances** which cannot be resold, but can be modified and applied to different family types.

AWS reserved instances can be applied beyond EC2 to additional services such as AWS DynamoDB, RDS (Relational Database Service), Redshift and Elasticache (whereas Savings Plans typically can only be applied to EC2 and Fargate). Reserved instances are typically associated with Linux and Windows servers.

Reserved instances can either be regional or zonal in nature (zonal reservations allow for guaranteed capacity, but no flexibility in terms of varying the reserved instances zone, family and size). Regional reservations are "size flexible" – i.e. they can be applied to different instance sizes within the same instance "family" (e.g. if a customer purchased one c5.x-large reserved instance but decided to run two c.5 large instances, the c5.x-large reserved instance could be applied to the 2 smaller instances).

Cloud Financial Management: The Core Principles of FinOps

AWS Savings Plans

AWS Savings Plans (e.g. EC2 Savings Plans and Compute Savings Plans) require a commitment in terms of minimum spend per hour over a number of years. EC2 Savings Plans provide slightly higher discounts, but Compute Savings Plans can be used across EC2, Fargate and AWS Lamda (i.e. Amazon's serverless offerings).

Any spend that is within the minimum committed amount will be charged at a significant discount (typically between 65-70%) whereas any spend in excess of the minimum committed amount is charged at on-demand rates.

Convertible reserved instances (see above) can accommodate increased commitment (eg. additional reserved instances to cover more EC2 instances) without needing to increase the term of the commitment, but Savings Plans are not as flexible. Our understanding of the Savings Plans regime is that any addition to the original contract is implemented with a new contract that "re-starts the clock" in terms of the minimum period. Standard reserved instances can be bought and sold on the AWS Marketplace but Savings Plans cannot – the customer is required to keep its committed spend at the initially defined level.

EC2 Instance Savings Plans will apply usage across any given instance family regardless of the operating system used. The EC2 Instance Savings Plans offer up to 72% savings compared to On-Demand pricing on a customer's Amazon EC2 Instances usage.

Microsoft Azure has similar offerings in respect of reserved instances (aside from Azure VMs, these can also be used for storage capacity, MySQL, compute component of Azure Database for MariaDB, PostgreSQL and other Azure services).

Increasing visibility of Cloud Costs/Cost Avoidance: Tagging and Account hierarchies

As an initial step, companies need to review hourly usage data and resource-level reporting in order to ensure that usage of reserved instances is going to exceed the break-even point for not reserving the resources. From a financial reporting perspective, the key objectives for an organisation are likely to include:

- Improving the organisation's ability to provide real-time financial forecasting and planning and process cost data as soon as it becomes available
- Analysing cost, usage and performance history by cost centre, project, department etc
- Making continuous small adjustments in cloud usage/optimisation

- Ensuring faster feed-back loops which ought to result in more efficient infrastructure provisioning choices and behaviour
- Consistent visibility into cloud spend is provided to relevant departments
- Each department receives trending and variance analysis information to help explain why costs have increased

At a basic level, cloud costs need to be assigned to the correct business units within an organisation in order to understand where anomalies are occurring. The major cloud providers have a range of mechanisms to assist with cost allocation strategies:

- **Tagging conventions**: tags are text-based keys/ values that track the department or cost centre for the cloud resources that have been used;
- **Account hierarchies**: individual accounts can be used to isolate costs from one environment or application from another;
- **Account naming conventions**: organisations can consider whether to have a centralised function to open accounts and to ensure that naming conventions are utilised in order to include details of the specific workloads, products or environments that will be using the account.

Organisations also need to be provided with the right reporting data to enable them make decisions on:

- Terminating unused compute resources (e.g. any unused virtual machines, unattached storage volumes, aged snapshots or disassociated IP addresses) or those compute resources that fall outside any organisation-set "guardrails";
- Right-sizing underutilised compute resources in order to improve efficiency;
- Exchanging existing convertible reservations to eliminate wasted costs; and
- Purchasing AWS Savings Plans and Reserved Instances in order to optimise future cloud costs.

Organisations are likely to need to invest in third-party tools (e.g. tools such as Apptio Cloudability or Cloud Health) in order to assist in optimising cloud services costs. This will help to report on (and potentially charge back) cloud spend by cost centre or other business groupings.

Reserved Instances arguably need more active monitoring to ensure that the savings are automatically being applied, whereas AWS Compute Savings Plans will automatically apply to any EC2 instance regardless of region, instance family, operating system, or tenancy, including EC2 instances that are part of ECS, or EKS clusters, or launched by AWS Fargate. However, there are certain features of reserved instances that have not historically been made available in Savings Plans (e.g. the ability to trade on AWS Marketplace).

Using policies/governance and automation to manage cloud services costs

Once an organisation has clearer visibility of, and has taken steps to optimise, its cloud services costs, it can start to consider establishing policies/guardrails to govern its IT environment and trigger automated alerts when these parameters or policies are violated, and automate actions to prevent and/or correct policy deviations.

By way of example, alerts can be set up when:

- The monthly cost for any department is projected to exceed the budget by say 5%;
- The monthly cost for any IT environment is projected to increase by more than a certain percentage (say 10 to 15%) compared to the previous month; or
- Reserved instances or Savings Plans are expiring in the next 90 days.

Beyond alerting, organisations can also explore whether to establish policies that automatically terminate a particular cloud resource when certain thresholds have been met.

Cloud Financial Management: The Core Principles of FinOps

Does containerisation and Kubernetes lead to different challenges in respect of visibility of cloud costs?

Container technology ought to have some significant cost advantages in that it can be possible to run multiple applications on a host machine each within its own container (so that packing multiple applications on a single host machine reduces the overall number of compute instances and reduces overall IT infrastructure costs).

In terms of cost allocation, containerisation may need different tool-sets and methods in order to accurately forecast how much a particular workload costs. When you run a single application on a single virtual machine an organisation can associate a single item on its monthly cloud bill with a single application (with the potential for one line item per instance on a cloud bill with each of the instances contributing to the total cost of the application).

However, with container orchestration tools (such as Kubernetes) a single workload can be split across multiple VMs, so you have to look inside the cluster to calculate how much of each VM's cost should be apportioned to a particular application. Kubernetes does not contain cost constraints, but you can set guardrails in respect of CPUs and GB of memory (one problem to be overcome is the tendency of developers to set limits arbitrarily high in order to get an application up and running).

Conclusion

Organisations that are under cost pressure are likely to need to adapt their procurement and finance processes in order to establish an effective FinOps function. One potential solution is to establish a centralised "cloud finance"/operations management function which is responsible for setting policy on cost optimisation and which will centrally govern and control committed use discounts, reserved instances with cloud providers (and will help to influence commercial assessments in respect of issues such as the cost of capital deployed to purchase reserved instances).

The reporting processes also need to be considered in order to monitor cost and usage reports which will give details of whether a reserved instance has been applied and which reserved instance has been applied in a specific hour of the month (which should give a clearer understanding of the utilisation of reserved instances and how they affect additional purchasing and resource modification decisions).

Detailed billing reports with account resource details and tags can give details of every compute resource used in every time period with a column for each and every tag key which should enable an assessment of which specific resource ID in which specific time period was driving spending increases. In time, this can help to lead to real-time data-driven decision making which should lead to more effective use of "cloud spend".

Fieldfisher LLP
17 May 2023

Paul Graham
Partner
+44 330 460 7094
paul.graham@fieldfisher.com

Contacts

fieldfisher

Paul Graham
Partner
+44 330 460 7094
paul.graham@fieldfisher.com

Olivia Woolston Morgan
Partner
+44 330 460 7092
olivia.woolstonmorgan@fieldfisher.com

John Brunning
Partner
+44 330 460 7085
john.brunning@fieldfisher.com

Nikhil Shah
Director
+44 (330 460 6346
nikhil.shah@fieldfisher.com

Laura Witherspoon
Director
+44 330 460 7090
laura.witherspoon@fieldfisher.com

Michael Butterworth
Director
+44 330 460 6122
michael.butterworth@fieldfisher.com

Lynton Brooks
Senior Associate
+44 330 460 6603
lynton.brooks@fieldfisher.com

Eleanor MacIntyre Clampitt
Associate
+44 330 460 7079
eleanor.macintyreclampitt@fieldfisher.com

Kristina Holm
Solicitor
+44 330 460 6538
kristina.holm@fieldfisher.com