O'REILLY®

Compliments of Qubole

Financial Governance for Data Processing in the Cloud

Managing Costs While Democratizing Data at Scale

Amit Duvedi, Balaji Mohanam, Andy Still & Andrew Ash

9 781492 049739

Financial Governance for Data Processing in the Cloud

Managing Costs While Democratizing Data at Scale

Amit Duvedi, Balaji Mohanam, Andy Still, and Andrew Ash

Beijing • Boston • Farnham • Sebastopol • Tokyo O'REILLY®

Financial Governance for Data Processing in the Cloud
by Amit Duvedi, Balaji Mohanam, Andy Still, and Andrew Ash

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

O'Reilly books may be purchased for educational, business, or sales promotional use. Online editions are also available for most titles (*http://oreilly.com/safari*). For more information, contact our corporate/institutional sales department: 800-998-9938 or *corporate@oreilly.com*.

Editor: Nicole Tache
Production Editor: Nan Barber
Copyeditor: Octal Publishing, LLC
Proofreader: Nan Barber

Interior Designer: David Futato
Cover Designer: Karen Montgomery
Illustrator: Rebecca Demarest

April 2019: First Edition

Revision History for the First Edition
2019-03-15: First Release

This work is part of a collaboration between O'Reilly and Qubole. See our statement of editorial independence.

978-1-492-04973-9

[LSI]

Table of Contents

Executive Summary

Companies of all sizes are riding a wave of democratization of data, fueled in part by the need for data-driven decision making and access to cutting edge cloud-native data-processing platforms with no need for upfront investment or physical infrastructure. Think targeted promotions, optimized marketing spend, A/B testing for R&D, and decision support for executives. Cloud platforms are increasingly used to facilitate data democratization, not only to minimize costs, but also to provide traceability and predictability.

However, the flexible and dynamic nature of the cloud combined with its usage-based billing policies means that it can lead to unexpected and unpredictable bills, if left unchecked. Effective financial governance, therefore, is essential.

A good financial governance plan does three things:

Understands
: It tracks usage over time, both short and long term, as well as forecasts future usage. In addition, it quantifies the types of use cases enabled by the platform and their corresponding business impact.

Controls
: It puts in place access controls to restrict the usage of the platform. Such controls are either proactive to prevent an action from occurring or reactive to alert when thresholds are reached.

Optimizes
: It uses the power of cloud-native platforms to automate activities that optimize the costs of your platform.

There are three ways to achieve financial governance for data processing:

Use tools provided by cloud service providers
: Examples are Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).

Use specialized cloud management platforms
: Examples are Flexera (Rightscale), CloudBold, and Scalr.

Use purpose-built financial governance capabilities available in cloud-native data platforms
: Examples are Qubole and Snowflake.

The tools offered by cloud service providers are generic, limited in scope, and will require creation of additional tooling to reliably interpret the data they provide into a financial governance policy.

Cloud management platforms offer an effective set of tools for building a generic cloud financial governance plan, but they offer limited tooling for the specific financial governance needs for data processing.

Cloud-native data platforms are purpose built for data processing needs. These platforms act as a wrapper around the cloud platform that understands what actions are being undertaken and why. They provide understanding, control, and optimization of the cloud system at a higher level, in light of the data processing workload that is being undertaken, the role that is undertaking it, therefore allowing for a much higher level of financial governance. Data platforms are therefore ideal to gain control over cloud usage and build an effective financial governance approach.

CHAPTER 1
Introduction

The advent of cloud service providers such as AWS, Azure, and GCP has changed the world of big data for the better, opening up the ability to build and run a best-of-breed big data–processing system to virtually every company, regardless of size.

The need for large upfront investment in hardware and specialist system administration staff to build and configure on-premises platforms for data processing has been removed as cloud platforms have evolved. The services offered are all on demand, pay-as-you-go services, meaning that investigations and proofs of concept (PoCs) will cost very little or even nothing.

The development of open source data processing engines such as Hadoop, Spark, Kafka, TensorFlow, Presto, and others as industry leading platforms has led to the widespread adoption of data processing and the development of a vibrant technology community.

However, with every revolution comes new challenges. With cloud platforms, although the initial investment is low, it is very easy for costs to get out of control without careful management.

This report provides guidance to effectively govern the costs associated with data processing in the cloud. Looking at the three areas of any successful financial governance plan—cost control, traceability and predictability—the following chapters provide some pointers to the tools, systems, and processes that you can employ to deliver a successful cloud-based data-processing platform with effective financial governance.

We focus on the financial challenges specific to running a data-processing platform, though many elements will be valid to running any kind of platform within the cloud.

This text is for those who are responsible for the operation of data-processing platforms in the cloud. It does not offer any guidance on how to set up these platforms or recommendation on preferred platforms or toolsets.

Tools and Platforms Discussed Herein

Many of the examples that follow discuss AWS offerings, Cloudability, and the Apache suite of products. You should not construe this as an endorsement for these products; it simply reflects that they are some of the most widely used platforms and tools and the ones we are most familiar with.

Notes on Terminology

Saying *cloud platform* can mean many different things to different people, from pure Software as a Service (SaaS) offerings such as Google Docs to virtual server platforms.

For the purpose of this report, we use *cloud service provider* to refer to the large service providers (AWS, GCP, Azure) that offer a range of Infrastructure as a Service (IaaS) and SaaS offerings aimed at being the building blocks for complex infrastructure and systems.

Some common features of cloud platforms include the following:

Instant elasticity
: Resources can be created and destroyed instantly.

Pay by usage
: No long-term commitments.

Programmatic control
: All elements of the system can be fully controlled via API.

Infrastructure and services
: All offer a combination of being able to create infrastructure and consume services.

Data Processing

The term *data processing* is a widely used term that can mean different things depending on background and context. For the purpose of this report, we use the terms *data processing* and *big data* in a very loose sense to mean any large-scale, complex data operation. This could range from a large Extract, Transform, and Load (ETL) process to machine learning processes. Typically, these tasks will use specialized software such as Hadoop, Spark, or Presto.

CHAPTER 2

Big Data and the Cloud

Even though the cloud can enable companies to compete with their bigger competition in the big data space, it is not a journey without risk or pain.

Most businesses taking their journey into the cloud will follow a pattern similar to that outlined in the following sections.

Stage 1: Adoption

A typical cloud journey begins with some experimentation and small-scale tests of functionality.

The benefits of a cloud platform are usually immediately obvious, whether it be ease of use, scalability, flexibility, or simplicity.

There is usually a sense of excitement and a desire to escalate usage.

Stage 2: Expansion

Adoption is usually followed by expansion of services, often in a relatively unstructured way as different teams see different benefits that they can realize from the new technology.

The ease of creation and use of services also removes the need for structured change requests, meaning that the gatekeeper role offered by the IT department is often bypassed or loosened and old system controls are no longer applicable.

An inevitable outcome of this is that it is seen as a way of quickly getting things done without the constraints that used to be put in place by the IT department. Usually these constraints are seen as being unnecessary.

The expansion phase of many companies' journey to the cloud often leads to chaotic, uncontrolled, siloed implementations operating in isolation from one another, albeit as implementations that might be delivering value to the business.

Stage 3: Control

Expansion is usually stopped by a panic as the realization of the extent of expansion becomes obvious and an urgent demand for change is made.

This usually comes from one of two sources: cost or risk. With either, a chief financial officer, who sees the monthly bill and demands accountability or a chief information security officer who realizes that things are being used that do not meet security standards and demands that it is all turned off.

The cloud gives you great power and, as the saying goes, with great power comes great responsibility. Cloud data platform usage within companies needs appropriate control or the costs can easily get out of hand. Control and governance over cost and risk are both essential for any cloud platform.

It is very easy to run up large bills on cloud platforms, and cloud providers will happily let you do it.

The final stage in the journey to a mature cloud system is bringing that system under control so that costs and risk can be appropriately understood, managed, and predicted.

The rest of this report focuses on how your organization can complete this journey to stage three in terms of cost management by putting in place a robust financial governance process for your cloud data-processing platform. In an ideal world, this will be understood and put in place soon enough to control the expansion stage and avoid the chaos and overspending often seen at that stage.

CHAPTER 3
Financial Governance for Data Processing

Financial governance generally refers to the ability to collect, monitor, track, and control financial information.

When you're reading the following chapters, it is important for you to distinguish between simple *cost control* and *financial governance*. Cost control refers to the practice of taking action to minimize costs, whereas financial governance is a much wider topic that not only focuses on minimizing costs (though obviously that is a core objective), but also on providing fully traceable and predictable costs.

Any good financial governance process therefore includes three distinct elements:

Cost control
: Ensuring that no costs are being accrued that are not needed, and those that must be are minimized.

Traceability
: The ability to know who is spending what, when, and why and to be able to relate that back to the business value being delivered.

Predictability
: To understand and be able to predict what future costs will be.

Good financial governance will likely lead to discovery of potential cost control, but in some cases, a decision might need to be made to

increase cost in favor of traceability and predictability. As a simple example, a decision might be made to use multiple distinct pieces of infrastructure traceable back to specific departments or use cases. A shared platform might be cheaper to run, but individual usage can't be tracked, so costs could not be fully traceable.

Financial Governance in the Cloud

Financial governance is a fundamentally different challenge in the cloud versus on-premises solutions, which involve agreeing to costs up front for long-term commits. There might be small variable costs, but these are usually easily traced. This is a traditional model and very easy around which to build a governance process.

The cloud throws these methods out the window. Virtually all cloud services are on-demand, usage-based systems, meaning that at the end of the month you will receive an invoice for exactly the services that you have used, usually billed down to a very small unit (e.g., cost per second of infrastructure use or per request made to a service).

Rising Costs Are Not All Bad

In the cloud, rising costs are not necessarily bad; it means that you are using more services, which theoretically means you are doing more "good stuff" and hopefully delivering business value. Financial governance makes sure that wasteful spending is identified and eventually eliminated.

This billing model inevitably makes predictability very challenging, unless you happen to carry out exactly the same tasks every month (unlikely!).

Similarly, traceability can be challenging. As we discuss in the sections that follow, the cloud providers by default give limited breakdown of billing activity.

The following sections outline the methods that you can use to deliver financial governance on a cloud platform. Many of the challenges of reliable financial governance for data processing platforms in the cloud are the same as those for delivering any cloud-based

system. However, there are specific challenges related to data platforms, which we highlight in the sections that follow.

The Financial Governance for Data Processing Life Cycle

This report outlines a three-stage approach to building effective financial governance in your cloud platform. These three stages are not strictly sequential, but following this general order is a common-sense approach to take. In some cases, it might be necessary to introduce some basic *Control* elements ahead of the *Understand* elements in order to solve a pressing cost problem.

Here are the three stages:

Understand
: Know what is currently happening and build a financial profile of your cloud spending.

Control
: Put measures in place to control spending.

Optimize
: Take advantage of cloud data platform facilities to reduce costs and improve overall financial governance.

These three stages are explored in more detail in the following chapters.

Options for Delivering Financial Governance in the Cloud

Your chosen approach to enacting financial governance will naturally be driven by the requirements of your particular business and the preferences and requirements of your finance department. There is a wide range of tools available to help you get to the solution that meets your requirements.

Broadly, we can divide these tools into three categories: those provided by cloud service providers; those provided by cloud management platforms; and those provided by cloud-native data platforms. The sections that follow discuss the facilities that these toolsets offer

to aid financial governance at each stage of the financial governance life cycle.

These categories are not mutually exclusive, and your final approach will likely involve elements of all three.

Financial Governance Tools Provided by Cloud Service Providers

All cloud service providers offer a selection of tools that you can use to provide financial governance, usually in the form of a web interface or in raw data that you can download (this data is complex, low-level data that requires a degree of knowhow to interpret). These systems are constantly evolving, driven by customer demand, but they are still often limited and tend to be driven by a technical view of the systems that are accruing cost rather than a business-focused view.

Many companies have built their own tool sets on top of the datasets provided, for the purpose of improving the levels of information they can extract. These can be as simple as Excel sheets, or much more complex custom applications.

Financial Governance Tools Provided by Cloud Management Platforms

There is also a growing number of third-party cloud management platforms that will automatically import and process billing and usage data from the cloud service providers, presenting it back in a range of reports and prediction tools that aim to provide the traceability and predictability customers require. These services will also provide security validation of your cloud platform.

A leading example of these types of systems is Cloudability.

This category of tools tends to focus on building a business-centric view of the data, taking cloud management from the hands of technology departments, and exposing it to the wider business.

Financial Governance Tools Provided by Cloud-Native Data Platforms

Lastly, third-party vendors who offer cloud-native data platforms provide purpose-built capabilities to address the financial gover-

nance challenges related to managing data processing in the cloud. Although these data platforms are not exclusively designed for financial governance, they provide facilities for managing elastic data platforms, ensuring that platforms are provisioned in as timely, performant, cost-effective, and efficient manner as possible while also providing the financial tracking information necessary for understanding platform usage.

The on-premises method of running a data platform involved having a finite, fixed amount of resource to be shared by all users with no additional cost for when the system was inactive. Moving to the cloud involves a change of mindset, but many people will transfer the same mindset into the cloud, resulting in inefficiencies. The result is often either of the following:

- A similar number of platforms that sit there to be shared by all users resulting in a system that is shared for efficiency but sits idle for some of the time and isn't easily traceable back to users
- Distinct clusters created for each use case which are under capacity and have an overhead cost of the process that creates and destroys them

Cloud-native data platforms are working to change this mindset. They remove the need to understand the underlying infrastructure by providing an interface that is focused on understanding the data needs and then efficiently creating and destroying the amount of infrastructure needed to provide a shared environment that can deliver the required functionality.

The differentiator that this category of toolsets offers is that although the preceding category is focused on providing a business-accessible view of the spend data, this category is offering a view of financial data based on a much more nuanced understanding of the underlying usage of a data platform. Therefore, rather than just knowing that infrastructure is being used to run a big data platform, these tools will understand who is running queries on that platform and what those queries are. An example of this type of company is Qubole.

CHAPTER 4
Stage 1: Understand

The first stage in achieving effective financial governance is to *understand* your current cloud usage and therefore identify the gaps between your current position and effective financial governance.

To take this step, it is necessary to gather data to answer the following questions:

- What is being spent and how is that split across different services?
- Who is responsible for that spending?
- How does that spending relate to business objectives or value creation?

In a traditional on-premises infrastructure, data processing costs cannot be easily categorized into infrastructure and data processing buckets. However, in the cloud, billing is a mixture of on-demand infrastructure creation and usage-based service charges—and although cloud providers typically present a single bill for all usage at month end, a cloud platform can (if the right systems are used) break these down at a much more granular level to give very specific levels of traceability.

Financial Governance Tools Provided by Cloud Service Providers

All cloud service providers have billing dashboards that you can use to understand the costs across the cloud estate. Typically, these tools will allow infrastructure managers to report on costs associated with running the compute, network, storage, and services that make up their cloud environments.

The default dashboard views provided give you a useful "at a glance" understanding of where costs are being incurred. Usually, the out-of-the-box functionality allows for the following data to be revealed:

- The historic trend in monthly costs. This allows for broad-brush visibility of increases or decreases in costs and can be used to map changes in the environments to changing costs.
- Component costs as a share of total cost (compute, database, storage, etc.). This allows for visibility of high-spend areas or changes to specific infrastructure costs over time.
- A forecast for the current monthly bill.

Cloud service providers also maintain cost calculators that allow infrastructure managers to estimate the cost of ownership of services prior to deployment. Anticipated usage of cloud infrastructure for new or growing requirements can be mapped into the calculator to give an estimated service cost.

Managed service providers (MSP) or service providers who utilize multiple cloud accounts can combine billing data into a single dataset.

These tools are an invaluable aid to infrastructure managers looking to understand the detailed costs of their cloud environments at a technical level. This understanding, however, is initially limited to oversight of infrastructure. Out-of-the-box billing tools are focused on the usage of cloud components and provide a picture *of what* is being spent without the context of data specific to the services running on the infrastructure.

The Importance of Tagging

To understand, control, and optimize cloud infrastructure, standard tagging that persists across the cloud estate is required as the means

of defining business and technical usage of components. Furthermore, tagging enables the use of automated tools to improve financial, operational, and security governance.

NOTE

Tagging in Cloud Platforms

Cloud platforms allow you to create infrastructure and use services in an ad hoc, on-demand fashion, in many situations in an entirely automated manner. This means that traditional approaches of asset tracking in a manual register are no longer viable. Cloud platforms mitigate this gap by allowing metadata known as "tags" to be associated with elements as (or after) they are created. Multiple tags can be used to understand the purpose of the item.

Cloud service provider, third-party, and custom reporting all require the user to manage tags in a standard way to allow for clarity on usage and cost. This is the cornerstone of good practice. Regardless of the size of the estate, tagging should be standardized to give the appropriate visibility to all stakeholders. A standard set of cloud tags might contain the following:

Environment
: Identify production versus UAT/Dev environments

Service
: Identify which service this component is part of (should be multitiered in complex applications)

Function
: Identify what this component does

Technical/service/business owner
: Identifies the person or department that manages each aspect of the component or service

Operational tags
: Used to automate shutdown or other desirable technical functions relating to the automation of the service

Financial Governance Tools Provided by Cloud Management Platforms

The limitations in the reporting offered by cloud service providers has led to many companies creating their own systems to interpret and display the data in a more business-accessible manner. As we discussed earlier, many third-party cloud management platforms have been created to fill this gap and to provide management and configurable financial reports based on cloud service provider data.

Use of these services is generally regarded as essential for any company looking to effectively manage billing for anything beyond the most basic cloud platform.

Using a standard tagging structure (as just described) embedded in the provisioning process allows third-party reporting services such as Cloudability to reveal technical and business insights into spending across the estate.

Having a business-centric view of billing, clustered by applications, customers, or lines of service, will feed into strategic thinking and decision making. Reporting targeted at the business or service owners as well as the technical management moves costs out of the technical sphere of influence and into the business lines that manage the services. Being able to show the value, or lack thereof, in any hosted application or service allows for accurate decision making in strategic planning for success.

Previously complex multiaccount setups or multicloud platforms in which services or customers are spread across multiple domains or service providers can be consolidated by using standard tagging and third-party tooling, again allowing for a view of costs by application, customer or service in a single-pane dashboard, such as that shown in Figure 4-1.

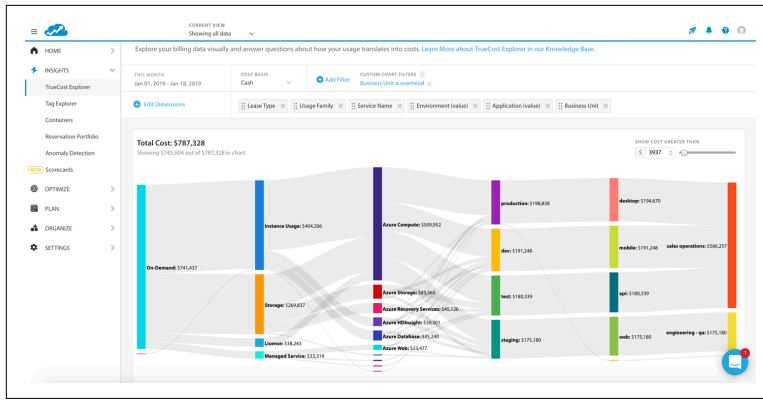

Figure 4-1. Cost breakdown of a typical application stack

These tools drive a business-centric view of the traceability of costs, as illustrated in Figure 4-2.

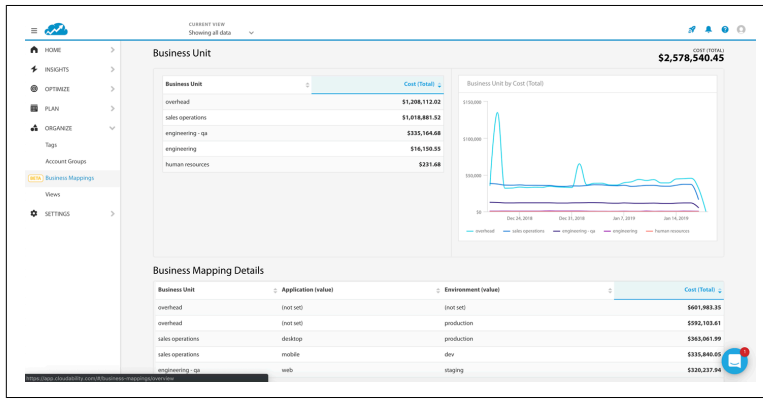

Figure 4-2. Cost data shown against business function

They also encourage users to begin considering predictability of costs, both in the short and long term and these tools provide insight and reporting at both levels:

Short term
: What do we expect this month's bill to be, allowing cash flow to be planned and action to be taken if higher than anticipated?

Long term
: How have costs varied over time and can those trends be used to predict future costs?

Tools in this category will form the basis of a good financial governance process for a general cloud platform; however, this does not come out of the box. There will be work to do to ensure that your cloud systems are effectively tagged and the tools configured to understand that tagging strategy.

Financial Governance Tools Provided by Cloud-Native Data Platforms

Generic cloud management tools will do a good job of extracting business intelligence from the raw data supplied by the cloud platforms. However, these are aimed at general business use cases, rather than those specifically related to a big data business.

Cloud-native data platforms will sit on top of your cloud platform to provide big data-specific intelligence and management. In real terms, this means that a layer of control above that provided by the cloud platform is added. As a simple example, rather than having to create two separate clusters to run two queries so that usage could be easily traceable, a cloud-native data platform will create a single shared cluster and track the partial usage by each query.

These data platforms will also provide an enhanced level of reporting, allowing you to understand the costs at a much more detailed level. Typically, they move the traceability to thinking in terms of *people* (who is executing the query) and *workload* (what that query is being executed for), as demonstrated in Figure 4-3, rather than infrastructure or services.

Results Logs

Showing partial result set (100 rows). Download to see full result.

account_id	user_id	ec2_cost	qcuh	percentage_consumption	total_ec2_cost	total_QCUH
689	8728	$0.0	0	0.0%	6946	9495
689	4540	$16.29	22	0.2345%	6946	9495
689	5196	$30.13	41	0.4337%	6946	9495
689	3840	$3.19	4	0.0459%	6946	9495
689	5216	$0.7	1	0.0101%	6946	9495
689	6094	$8.58	12	0.1236%	6946	9495
689	2392	$1.45	2	0.0209%	6946	9495
689	2502	$0.58	1	0.0084%	6946	9495
689	4567	$0.72	1	0.0104%	6946	9495
689	10995	$54.39	74	0.783%	6946	9495
689	6691	$0.96	1	0.0138%	6946	9495
689	1511	$6.52	9	0.0938%	6946	9495
689	3625	$11.84	16	0.1705%	6946	9495
689	5776	$22.78	31	0.3279%	6946	9495
689	10888	$23.34	32	0.3359%	6946	9495

Figure 4-3. Examples of reporting of costs per user/business activity on Qubole platform

This means that you can achieve a very accurate view of the cost of all pieces of data analysis and execution. This offers some real advantages when looking at financial governance:

- Costs can be related directly to the business value being, or expected to be, achieved.
- Costs can be tracked and related to budgets or recharged to departments.
- Future costs can be predicted at a much more granular level, allowing for up front assessment on the cost/value decision on activities.

CHAPTER 5
Stage 2: Control

The ethos of cloud platforms is to make everything frictionless, available, and easy to use, particularly programmatically.

The entire system is designed to free developers and Ops people from the constraints of old-world datacenter and on-premises systems: the world of purchase orders, change-control forms, work schedules, and everything else that restricts the efficient and reactive creation of new infrastructure and functionality. The cloud is designed to be an on-demand system with zero resistance to creation and usage of services.

This is one of the reasons why the cloud services are game changers in the industry. They have freed companies from the constraints of traditional infrastructure management and allowed them to be much more dynamic and reactive. Any CFO in control of finances for a cloud-centric company will tell you that this freedom is nice in theory but in practice some control is needed.

Although this is a brave new world for many reasons, it creates a nightmare for financial governance. New infrastructure can be created at will with no human intervention; on-demand, usage-based systems can be integrated into applications to be called on an ad hoc basis. In short, costs cannot be easily controlled, traced, or predicted.

This second stage in the financial governance life cycle—control—at a high level looks to control or put limits on who (whether human, virtual service, or other system) can do what within the platform.

There are two types of control that can be introduced: *proactive* and *reactive*.

Proactive controls look to restrict actions that can be undertaken before they are undertaken. These can range from the very draconian ("no one can create any new infrastructure without requesting it from the Ops team, which will carry out the work, and telling it the details") to much more fluid (users are given limited permissions to be able to create specific elements).

Reactive controls have much looser restrictions on what can be done but effective monitoring and alerting systems are put in place to catch where controls need to be applied.

NOTE

Infrastructure versus Services

As just mentioned, cloud services are broadly divided into infrastructure and services. Infrastructure is much more naturally suited for controls at this level. Access can be allowed or denied to services, but any more nuanced controls will usually require a specific control software to be developed or a third-party service to be employed. The ability to get specific control of usage of data-specific services is one of the facilities offered by analytic data platforms.

Financial Governance Tools Provided by Cloud Service Providers

Cloud service providers offer a range of facilities to put controls in place to implement the cost optimization and traceability necessary for improved financial governance.

Core to all cloud service providers is a granular user structure that can be used to manage access appropriately to enforce "least privileged" permissions and help reduce infrastructure sprawl. This permission system is controlled by system administrators and can be managed manually or programmatically. Although granular, the permissions are very black and white. They only allow or disallow an action, there is no concept of reasonable use or time-based usage.

Utilization of out-of-the-box tooling can also help predict and save costs across the estate, taking advantage of reserved instances and savings around underutilized resources. For example, the AWS cost

explorer allows users to analyze and purchase reserved instances for compute, data, and caching services over a one- or three-year period with all or partial upfront costs. For situations in which these components are likely to be in service for the duration of the reservation, this is a sound strategy for cost reduction.

Combining these controls is critical to managing cloud estates effectively and typically falls within the governance of the Ops teams and managers. With a greater understanding of billing and asset management, infrastructure managers can bring control to an estate with budgeting, alerting, and *rightsizing* (more on this shortly) of components.

Reporting and controls typically exist at the infrastructure level, although the use of standard tags across the estate (as discussed earlier) will give stakeholders greater visibility at the application or service layer.

Budgeting and Alerting

Cloud service providers generally offer a range of tools to enable reactive control, typically based around budgets that can be set and alerts fired upon hitting those budgets.

Based on previous spend and projected growth in the estate, infrastructure managers can use built-in tooling to set budgets that allows reporting and alerting on either direct infrastructure costs or costs associated to grouped service or application tags.

These alerts can be triggered based on hard values, or they can be used to alert when unexpected patterns of growth are seen. Forecasting tooling allows for future costs, typically for the calendar month, to be compared with previous months' spend patterns. Thresholds can be set for alerting to identify where costs have increased based on a forecast generated by the cloud provider.

Understanding existing spending across the services and applications and setting budgets that generate alerts using forecasting are core to the control of the estate.

Forecasting alerts enable stakeholders at all levels to intervene and make corrective changes quickly when costs change or, more often, to compare the costs being generated with the projected costs of changes. Checks against projected growth of costs can be relayed to the business in a timely manner, and decisions around the value of

changes can be made ahead of the final bill hitting the finance team. The granularity of these forecasts and alerts depends on the granularity of tagging across the estate.

Tools of this type are an essential element for ensuring predictability of cloud costs.

Rightsizing

One of the major selling points of cloud platforms is the ability to scale elastically on demand rather than having to size platforms up front to meet requirements. However, making this a reality involves careful management and oversight of system usage, known as *rightsizing*.

Rightsizing is the practice of applying appropriate resources to current or projected workloads. In many cases, predicted usage of an infrastructure stack can be wrong. This is true for both *under provisioning* and *over provisioning*.

Under provisioning

In many ways, and despite the disruption that poor performance brings, under provisioning is the easier proposition to manage. As long as budgets allows for extra resources, cloud services are designed specifically to allow for quick remediation. Compute, I/O, or memory can be easily provisioned and deployed to rightsize the estate.

Over provisioning

Over provisioning is the less understood of the two states. Cloud service providers and third-party vendors offer a range of tools to counteract the over provision of services. Here are some of these tools:

- Identification of idle resources such as server instances, database, or storage volumes
- Autoscaling for changeable or burstable workloads
- Recommendations for reserved instances
- Recommendations for spot instance suitability for server instances

These controls typically exist at the infrastructure layer. Looking at historic usage of components of the estate, recommendations are delivered to the user through the administrator console or as part of the alerting configuration. Infrastructure managers can act on these recommendations to manage cost across the estate.

NOTE

Scaling in the Cloud

Over and under provisioning are important considerations in the cloud (as in any other platform). Cloud IaaS systems will equally allow you to over or under provision systems but will provide you with facilities to very quickly rectify the problem. You should give consideration at the outset to system sizing, but this is not as important as in on-premises systems, much more important is using the techniques described in this section to monitor and adjust platforms to optimize sizing.

SaaS within the cloud are shared systems; therefore, they are driven entirely by usage, rightsizing is not an issue for these services.

Financial management of cloud estates is as much about understanding and controlling over provision as it is making sure that there are enough resources to manage the workload. In the majority of cases, under provisioning is one of the key drivers for change in organizations, with poor performance or availability issues driving change across the business.

Most organizations, therefore, closely monitor and alert on performance metrics of services and applications; many employ teams of people or third parties to understand and control the performance of their value creating assets. Services that are *not fast enough* or, worse, *not always there* are the two conditions that all infrastructure managers seek to avoid.

Over provisioning is a relative newcomer to the dynamic. Traditional colocated or managed estates were deliberately designed with $n + x$ architecture to avoid a lengthy provisioning process of additional resource if demand for the service suddenly increased.

Cloud computing allows for a much quicker infrastructure life cycle, and there is no need to run at a significant excess. The downside to this approach is the ease in which cloud services can be created.

Combined with the understandable obsession with performance it is not always a priority for organizations to monitor and report on unnecessary costs and therefore require a new way of thinking to maintain control.

Financial Governance Tools Provided by Cloud Management Platforms

Cloud platforms offer a host of solutions to impose control and policy that provide good financial governance. However, as with reporting, they tend to be aimed more at technical than business users. Third-party cloud management platforms exist to fill this gap.

Recommendations

Cloud service providers will offer recommendations on how costs can be optimized, but cloud management platforms will translate and add to these recommendations, putting them into business context making them actionable and, if possible, create policy around them.

Recommendations will typically focus on making sure that underutilized infrastructure is removed and that optimal decisions are being made to get the best value infrastructure (for example, when instances can be purchased as reserved or spot instances).

Additional Controls

You can define and enforce additional types of control using these third-party platforms. You can define policies around usage patterns, creating or destroying infrastructure when it is not needed, (e.g., test platforms outside office hours). This helps to ensure not only cost optimization, but cost predictability.

You can also use cloud management platforms to help enforce controls around traceability, adding alerting or even forced termination rules where tagging policy is not met.

Financial Governance Tools Provided by Cloud-Native Data Platforms

As discussed, when considering reporting, data platforms operate at a deeper, data-specific level of understanding, having a much better appreciation for not only what infrastructure and services are being used, but also how and why those systems are being used.

Unlike cloud management platforms or cloud service platforms, a cloud-native data platform provides a much more granular level of can be applied over activity on the platform. For example, you are able to create policies that define permitted levels of usage at a workload or user level, as shown in Figure 5-1, therefore putting a restraint on the queries that can be executed.

```
<allocations>
  <queueMaxResourcesDefault>12000 mb, 2 vcores</queueMaxResourcesDefault>
     <clusterMaxAMShare>0.67</clusterMaxAMShare>
          <!-- Set root queue maxResource definition to a large value if jobs of different users are going to have
          their own queue, otherwise queueMaxResourcesDefault would be considered as the root queue's maxResources.-->
      <queue name="root">
        <maxResources>1000000 mb, 100000 vcores</maxResources>
      </queue>
      <queue name="etl">
      <queue name="prod">
        <maxResources>45000 mb, 5 vcores</maxResources>
      </queue>
     <queue name="dev">
       <maxResources>16000 mb, 3 vcores</maxResources>
      </queue>
  </queue>
</allocations>
```

Figure 5-1. Examples of creating usage limitation policies on Qubole platform

Generally, you can manage these policies in two ways:

- By setting hard limits and requiring anyone who wants to operate above those limits to justify the additional spending before adjusting the policy.
- By tracking additional costs for later, internal track back.

The benefit of the cloud-native data platforms is that they allow this decision making and control to be implemented at a data level, and that level to be related back to the business level, rather than at an infrastructure level.

Some specialized tools will also allow you to set more specific limitations, such as a budget for a specific query, to ensure that there are no unexpected high-cost queries executed.

There is also a movement toward self-learning systems that use machine learning to predict what your financial governance policy setup should be and provide recommendations for change.

Controlling the Underlying Cluster

Cloud-native data platforms also provide control over how the underlying data cluster should be optimized, again looking to not only minimize costs but also to ensure that costs are as predictable as possible.

Cluster sizing is a juggling act between cost optimization and speed of query execution (or amount of time to wait for a query to begin executing), which is influenced both by how busy an existing cluster is and whether a cluster already exists or must be created or resized.

Data platforms provide control over how both these factors can be managed, setting minimum and maximum cluster sizes and the amount of time a cluster can be idle before being terminated, as illustrated in Figure 5-2.

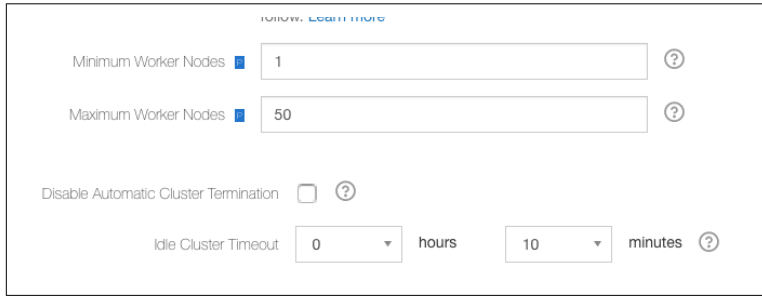

Figure 5-2. Example of cluster management; for example, setting cluster scaling limits and idle time on the Qubole platform

Additional controls are delivered by providing reactive alerting when cluster limits are reached.

Serverless Data Platforms

The next evolution of data platforms, which is currently at the very cutting edge of the industry, is the move away from infrastructure-based data platforms and toward a more service-based or serverless approach.

Serverless systems mean that the user has no awareness of the underlying infrastructure and simply pays by usage; in this case, per query, putting all responsibility for managing the underlying infrastructure in the hands of the platform providers.

The charging model would therefore provide full traceability as each specific query would have a single cost, directly triggered by a specific user or workload.

In this model, financial governance would be enforced by setting a limit per user on the level of spending they could make, and then each query they executed would be tracked against that limit.

These systems are very new and the charging model is still evolving, so it is too early yet to understand the relative overall cost of serverless versus managed clusters as a means of running a data analytics platform.

CHAPTER 6

Stage 3: Optimize

Although traceability and predictability are important elements in financial governance policies, cost control and cost reduction are typically the focus of any financial governance exercise.

Having fully understood the nature of your platform and implemented sufficient controls, the next step is to see how you can take advantage of the cloud platform in order to optimize your usage and therefore minimize cost without affecting the quality of service or the traceability and predictability put in place.

As discussed earlier, the fluidity and ease of control of cloud platforms can cause real difficulties for maintaining financial governance. However, when used correctly, you can use these same elements as a tool of cost reduction.

Cloud platforms are designed for automation, to be dynamically created and destroyed on demand, and careful use of these facilities can result in a highly optimized system that is extremely cost effective while always meeting the requirements of the business.

Optimizing for Performance

Part of the optimization process is to try to ensure optimal performance. However, when we optimize for performance, it is important to remember that we are optimizing not only the speed of query execution, but also the timeliness of the execution. One of the costliest resources to the business is data scientist time; this can often be a

hidden cost of running a data-processing platform. So, the least amount of waiting these people must do, the better.

However, timeliness does not always mean "as quickly as possible." It is more a matter of understanding when the results are needed and ensuring that they are available by that time and optimizing the cost of delivery to have them ready by then.

Reinventing Capacity Management

Moving to a cloud platform requires you to fundamentally reinterpret what is meant by capacity management. As discussed, capacity management was traditionally a matter of planning what capacity was going to be needed during the lifetime of the infrastructure being purchased, allowing some extra capacity for unexpected growth, and then building your system to meet that capacity level. In other words, the objective was always to have spare capacity.

In the cloud world, the opposite view should be taken, because you can create and destroy infrastructure on demand and pay only for what you use. The objective should be *never* to have any spare capacity.

Your goal during the optimization stage should be building systems that are constantly providing sufficient capacity to be slightly above that needed (cloud capacity versus real capacity, as demonstrated in Figure 6-1), while maintaining the traceability and predictability put in place in earlier stages.

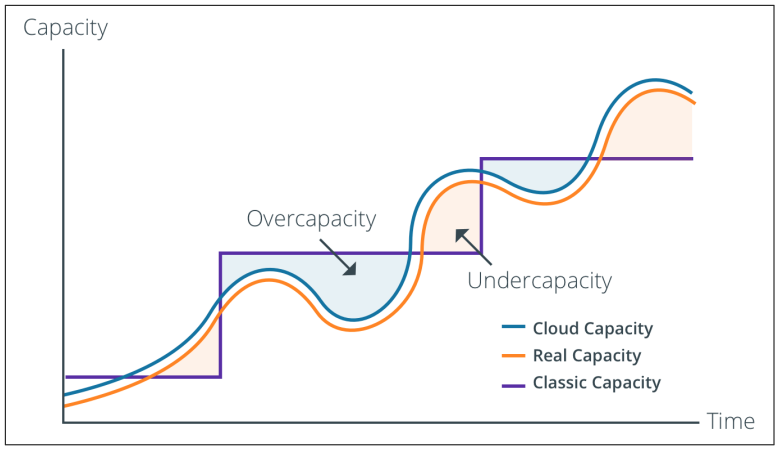

Figure 6-1. Traditional capacity management versus cloud capacity management

Financial Governance Tools Provided by Cloud Service Providers

Cloud service providers offer very limited functionality in the area of optimization. In general, their position is that they provide reporting information with full alerting and the ability to programmatically react to those alerts to manage infrastructure—any action that could be taken automatically to optimize cost becomes your responsibility.

Many companies will create their own custom scripts to carry out automation. For simple tasks, this can be a good solution; it is generally relatively quick to create and allows you to tailor very specific requirements. The downside is that the scripts then need managing and maintaining as cloud platforms evolve.

Another approach is to build optimization into the application that you have running on it, making it aware of its capacity or availability requirements and adjusting the platform in real time to meet those requirements. If you do this well, this can be a very sophisticated solution but it is a complex development task and carries higher risk and overhead than the aforementioned scripting approach.

Financial Governance Tools Provided by Cloud Management Platforms

In general, optimization is where people turn to third-party tools as a solution. Tools such as CloudCheckr or Cloudability offer a wide range of optimization tasks that you can easily configure and manage, which we cover momentarily.

Again, these tools look to move control of complex tasks from a technical to a management level, removing the risk of developing and managing automation scripts on an ongoing basis, bringing continuous improvements in the functionality they can offer, and integrating with the reporting and alerting solutions the platforms offer.

Waste Reduction

Optimizations usually focus on reducing the amount of waste within the system. This can include the following:

Removing orphaned or unused infrastructure
: Removing infrastructure that has been left behind when other infrastructure was terminated (e.g., disk volumes, ideally combined with auto-snapshot before deletion) or infrastructure that has sat idle for a specified amount of time.

Resizing underutilized infrastructure
: Adjusting the size of infrastructure that has had spare resource to an appropriate level. This requires careful policy creation because capacity must take into account expected spikes in usage.

Starting/stopping infrastructure based on schedules
: Automating the creation and destruction of systems to fit around usage patterns. For example, creating development environments for use during office hours or extending production platforms during peak trading hours.

Cost Optimization

You can undertake other automation tasks to minimize the cost of the infrastructure being used. The varying types of cloud charging models are discussed in more detail in a moment, but tooling can

apply automatic system management to ensure that the type of infrastructure being used is the best value while meeting the levels of resilience and availability required by the system.

Effective use of practices such as reserved instances or spot instances could reduce costs by up to 80%.

Traceability Management

You can also use automation to apply rules that will ensure that levels of traceability defined are being met. For example, you can configure policies to automatically destroy any elements that are created that do not meet the tagging policy in place.

Financial Governance Tools Provided by Cloud-Native Data Platforms

It is in the area of optimization that cloud-native data platforms come into their own. This was often the original objective in their creation, to remove the complexity and inefficiency of running a diverse set of big data activity on a cloud platform.

Cloud-native data platforms optimize your cloud usage by taking two approaches: ensuring the most efficient use of resources, and ensuring that resources are bought at optimal cost.

Resource Efficiency

Cloud-native data platforms ensure that the minimum amount of resources is being used by doing the following:

Ensuring efficient start up and shut down of infrastructure
: Many cloud providers charge by the second, so there are savings to be made by ensuring that infrastructure is destroyed as soon as it is not needed. Analytic data platforms manage this by ensuring that infrastructure is destroyed as soon as any workload is completed. An argument against destroying infrastructure immediately can be that there might be requirements to access that infrastructure later to retrieve additional information. Analytic data platforms reduce the risk of this problem by providing capture and backup of logs from any systems that are destroyed.

Ensuring efficient sharing of resources that might be underutilized
: Minimizing the need for creation, management, and termination of many platforms. Using an existing platform also speeds up processing time because there is no need to wait several minutes while the platform is created.

Appropriate sizing of infrastructure
: Ensuring rightsized infrastructure is used in order to meet performance needs at optimal cost. Analytic data platforms will include *workload-aware autoscaling*; that is, the dynamic scaling up of infrastructure specifically to meet the needs of the workload being carried out, scaling down the infrastructure as soon as it is completed. This is more efficient than standard cloud autoscaling, which is driven by physical metrics such as memory usage or CPU usage and therefore has no concept of the work being undertaken on the platform.

Resource Cost Optimizations

Cloud-native data platforms offer various costs depending on the level of commitment you want to make. There are three basic models; some platforms might use slightly different terminology, but the models are the same:

On demand
: Available immediately when you request it with no commitment; ability to destroy when no longer needed

Reserved
: Upfront payment is made in return for an agreed reduced on-demand cost

Spot
: Reduced cost given based on spare capacity being available within the cloud platform, typically done via an auction type system

Each of these cost models is best suited for a different use case. On demand suits immediacy without the need for commitment; reserved suits situations in which you know the infrastructure will be in use for the majority of the time; spot works for situations in which cost is a driver and the workload is not time sensitive. Spot instances (Figure 6-2) can be terminated at any point when your bid is below the current price so a level of resilience needs to be built in

to handle this. Cloud-native data platforms handle that resilience by building in systems to replace any spot instances that are terminated (in some cases that can include looking in alternative regions for right-priced spot instances) or by building a platform that is a mix of spot and on-demand instances to ensure that the core of the platform will never be terminated.

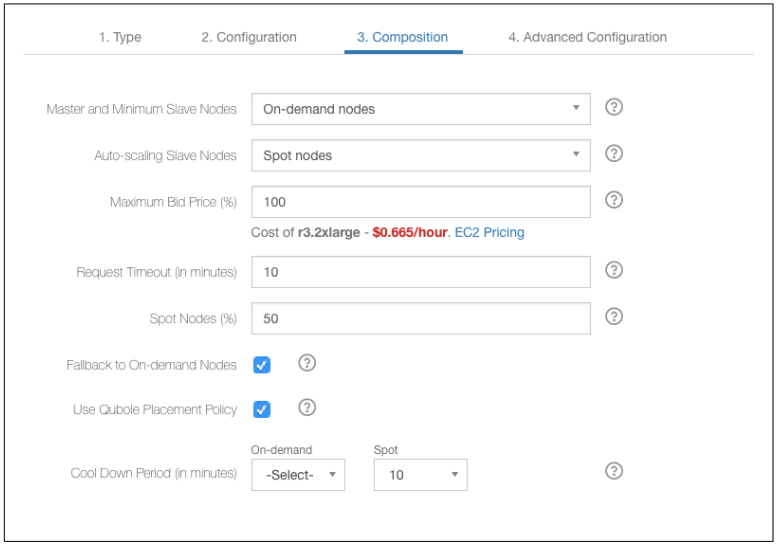

Figure 6-2. Example of spot instance management from Qubole platform

Cloud-native data platforms understand cloud platform cost models and, in line with the cost policies that you set in the analytic data platform, ensure that the infrastructure created is done so at as optimal a cost as possible while still achieving your performance needs, therefore allowing you to take advantage of the reduced cost models offered by the cloud platforms without having to understand the details or manage the process.

CHAPTER 7
Summary

There can be no doubt that the tools and technologies available today are leading to a true democratization of data that is allowing all companies to benefit from the competitive advantages that effective use of big data provides. Any company not taking advantage of these tools and technologies will soon be left behind by their more reactive and forward-thinking competitors.

The cloud is one of the fundamental pieces of this modern big data world. It gives any company, big or small, access to an industry-leading platform with no need for upfront investment, specialist skills, or physical infrastructure.

However, the power of cloud platforms—the ability to dynamically and easily create infrastructure, use services on demand, and be charged entirely by usage—can lead to chaotic and untracked systems and therefore unexpected and unpredictable bills. This can cause a backlash against the benefits of cloud platforms if you do not bring it under control.

Thus, effective financial governance is essential if cloud platforms are to be used, and this must not only minimize costs, but also provide a good degree of traceability and predictability.

A good financial governance plan includes three stages:

Understand
: Through detailed reporting, you should be able to have a full understanding of what is being undertaken on your platform, by whom, and how it relates to business objectives and value.

Your reporting should also be able to track usage over time, both short and long term, in order to be able to forecast future usage.

Control
: Put in place controls over who can do what and to what level within your platform. These can be proactive, stopping people before the action can be taken, or reactive, driven by alerting when thresholds are reached.

Optimize
: Use the power of cloud systems to begin automating activities that will optimize the costs of your platform, minimizing waste, and ensuring best value is being achieved from cloud charging patterns.

The cloud service providers offer some basic tools to implement your financial governance plan, but, in general, you will need to look at other specialized solutions in order to maximize the level of governance that you want to achieve.

There are cloud management platforms that will look to provide a less technical interface to cloud platforms, allowing management to start introducing levels of governance into the platform. These solutions operate across multiple cloud platforms but are aimed at general cloud users and so offer limited appreciation of how the cloud systems are being used for data processing.

However, cloud-native data platforms have been created that offer a big data–specific management platform. The key differentiator offered by these platforms is that they act as a wrapper around the cloud system that understands what actions are being taken and why. This means that they can provide understanding, control, and optimization of the cloud system at a higher level, appreciating the workload that is being undertaken and the person undertaking it, and therefore allowing for a much higher level of financial governance.

We summarize the various options for delivering financial governance in the cloud in Figure 7-1.

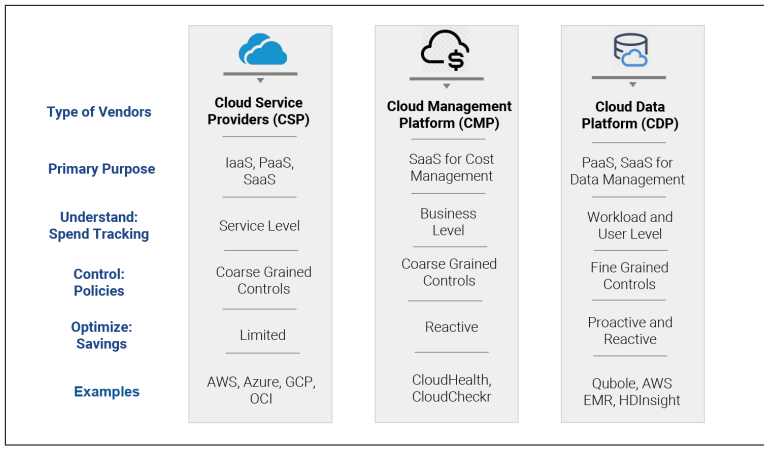

Figure 7-1. Summary of options for delivering financial governance in the cloud

If you're running a cloud-based big data platform, using a cloud-native data platform could be seen as an ideal approach to gaining control over cloud usage and building an effective financial governance approach.

About the Authors

Amit Duvedi is Vice President of Business Value Engineering at Qubole. He has 23+ years of experience in helping companies across industries discover and realize differentiated business value and competitive advantage by deploying technology. Amit's areas of focus include board ready business cases, business process performance benchmarking, business strategy and thought leadership. Prior to Qubole, Amit worked at Coupa, SAP and McKinsey. He has an MBA from University of Chicago and degrees in Engineering from University of Connecticut and the Indian Institute of Technology.

Balaji Mohanam works on Qubole's product management team and focuses on the cloud management platform. He has over 12 years of product development and management experience across big data, financial services, customer-relationship management (CRM), and ecommerce. Prior to Qubole, Balaji worked at Oracle, eBay/Paypal, Google, and MapR. He holds an MBA from Duke University.

Andy Still has worked in the web industry since 1998, leading development on some of the highest traffic sites in the UK. After 10 years in the development space, Andy cofounded Intechnica, a vendor independent IT performance consultancy to focus on helping companies improve performance on their IT systems, particularly websites. Andy focuses on improving the integration of performance into every stage of the development cycle with a particular interest in the integration of performance into the CI process.

Andrew Ash is Head of Operations at Netacea, where he helps customers detect and mitigate account takeover attacks and malicious bot traffic on their websites. Andy's main focus is to provide operational leadership for enterprise-scale deployments of the Netacea platform.