



3 Steps to Justify & Reduce The Cost of Your Data Lake

How to Position Your Data Lake Expenditure to Finance

Uncertain Times Bring New Cost Challenges

Data processing costs are under the microscope more than ever before with Finance leaders asking tough questions of data teams to reign in cloud costs that cannot be qualified or quantified. Data Lakes and data warehouses are quickly becoming prime targets as they are a source of driving data decisions but often lack financial governance and have unnecessary cost leakage. This is where Qubole can help you to confidently position your data team, processes and technology in a way that talks directly to your CFO.

Build Or Sharpen Your Plan

CFOs are driven by data and processes, just like your data team. The challenge is to frame your approach and to justify your expenditure in a way that removes the technology and purely focuses on commercial outcomes and strategy.

It all starts with a good financial governance plan that 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 open data lake 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 open data lake platforms to automate activities in an order that not only optimize the costs of your platform but also keeps delivering on the promise of Data lake or data warehouse.

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. By segmenting your financial governance plan into these three distinct buckets, you are able to display a prudent set of measures to prioritize spend against the most optimal outcomes of the data team. You are also providing safeguards and predetermined thresholds where you can 'negotiate' with your finance team.

Understand The Past, Present & Future

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?

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.

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

Having a business-centric view of billing, clustered by applications, customers, or lines of business, 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.

Proactive & Preventative Controls

Any CFO in control of finances for a cloud-centric company will tell you that the freedom of cloud is nice in theory but in practice more control is needed.

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”) 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.

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

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.

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.

Optimizing For The Long Term

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.

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.

Optimizations also means a heavy focus on reducing the amount of waste within the system. This can include the following:

1. Removing orphaned or unused infrastructure
2. Resizing underutilized infrastructure
3. Starting/stopping infrastructure based on pre-determined SLAs which includes schedules and priorities

Conclusion

Effective financial governance is essential, and this must not only minimize costs, but also provide a good degree of traceability and predictability. As a summary, a good financial governance plan includes three core 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.

For more information:

Contact:
sales@qubole.com

Try Qubole Open Data Lake Platform For Free.
Visit www.qubole.com/testdrive

469 El Camino Real, Suite 205
Santa Clara, CA 95050
(855) 423-6674 | info@qubole.com

WWW.QUBOLE.COM