LTIMindtree

POINT OF VIEW

Leveraging the Power of BigQuery Omni for Advanced Multi-Cloud Analytics

LTIMindtree

Table of Contents

01 Introduction

Multi-cloud and hybrid clouds are an accepted reality of 2023! As per a Statista report, over 90% of large enterprises use multi-cloud. The percentage is set to grow further to 94% by 2023. Forrester reported similar findings. In a survey of over 616 global IT decision-makers, 79% of respondents reported using multi-clouds. The primary reasons for investing in multiple public clouds include increased agility (48%), avoiding vendor lock-in (40%), needing specialized capabilities (42%), improved resilience (56%), and compliance/regulations (47%).

It's easy to see why: enterprises store information in multiple clouds to leverage the strong capabilities of individual service providers. However, one of the key challenges faced by enterprises today is seamlessly integrating the data residing in different cloud systems and making them work together. In a report titled Digital Insights Are The New Currency Of Business, Ted Schadler and Brian Hopkins talked of how "businesses are drowning in data but starving for insights." They are spot on. Global data is expected to rise exponentially to 180 zettabytes by 2025. Yet, despite the availability of so much data, only 29% of firms are able to derive measurable insights from their data.

When done correctly, the ROI of a multi-cloud strategy can be immensely beneficial. But the truth is getting these different multi-cloud environments to communicate with each other is a challenge in itself. What you have is a situation where it becomes difficult to operationalize data leading to missed business opportunities.

02 Data Analysis Challenges in Multi-cloud

Nevertheless, before we delve into the challenges of data analysis in multi-cloud, let us understand why companies are adopting multi-cloud in the first place. The prime reasons for choosing multi-cloud are:

Prevents Vendor Lock-in

Vendor lock-in is a key factor in influencing enterprises to opt for multi-cloud. Cloud customers don't want to be tied down into long-term software and hardware agreements with a single cloud provider. They want more options. Savvy customers are shifting workloads across vendors, who can match their workload requirements to a CSP that best meets their needs.

Protects Against Outages in a Particular Region

Cloud outages are not something to be taken lightly. According to the Uptime Institute's 2022 Outage Analysis Report, over 60% of outages result in at least $100,000 in total losses. That's a huge number! Multi and hybrid clouds offer high levels of redundancy. So, even if one of the cloud providers hosting your workloads experiences network downtime, your business can still continue functioning thanks to your workloads being hosted on other clouds.

Lower Costs

A multi-cloud strategy enables you to pick and choose between the different pricing models offered by cloud service providers (CSPs). You can choose the most cost-effective one for your business. Some providers may charge more for a particular service compared to others. Accordingly, you can select the most economical one amongst them. For example, you can use Google Cloud BigQuery to run your ML applications while choosing AWS Lambda to run event-driven code.

Although there are plenty of reasons to go multi-cloud, it cannot be denied that multi-cloud deployments lead to more siloed and complex environments. The inherent challenges are:

Data Sprawl

As per a survey by Osterman, large organizations have an average of 3,750 data stores and 7,750 identities in the public cloud. In an effort to avoid vendor lock-in, companies are going all out and embracing multi-cloud. Apart from companies choosing multiple vendors, an average company uses over 50 different cloud-native products, making it a challenge to 'talk' between these different systems. Data is stored across clouds, and there is no common platform to manage this vast data sprawl. Left unchecked, data sprawl can lead to costly security incidents.

Complex Environments

Multi-cloud complexity is real. In the pursuit of greater agility, enterprises are spreading workloads over multiple clouds. But what they don't account for is the increased complexity that comes with it. Managing a single cloud with multiple moving parts is hard enough. Adding multiple cloud environments brings in additional complexity, as every cloud vendor has its own unique way of working.

Data Analytics can be Expensive

In a multi-cloud environment, data is stored in tons of different platforms. If you need data stored with a different cloud provider for analytics purposes, you'll first have to bring back that data for computing. This can incur huge costs. Further, if you don't pay proper attention, your clouds can end up hosting duplicate data that not only take up storage space but also inflates your cloud bills.

03 What is BigQuery Omni?

Gartner estimates that by the end of 2024, 75% of organizations will shift from piloting to operationalizing AI. But the biggest impediment to achieving that is segmented data. Companies are moving to multi-cloud for greater agility and flexibility. But complex architectures and siloed data ownership prevent them from deriving valuable insights from their data that can help them to stay ahead of the curve.

What enterprises need is a cloud analytics solution that provides a seamless experience and doesn't involve a lot of specialist requirements. To address these challenges, Google Cloud has launched BigQuery Omni. It is a multi-cloud analytics product that enables customers to access and query data stored in Amazon Web Services (AWS) and Azure without copying or moving datasets.

Data analytics is a complex process. It involves collecting, storing, categorizing, performing ETL (extract, transfer, and load), and finally analyzing data to extract valuable insights for business users. Apart from the time taken, moving data across clouds is cumbersome and prohibitively expensive.

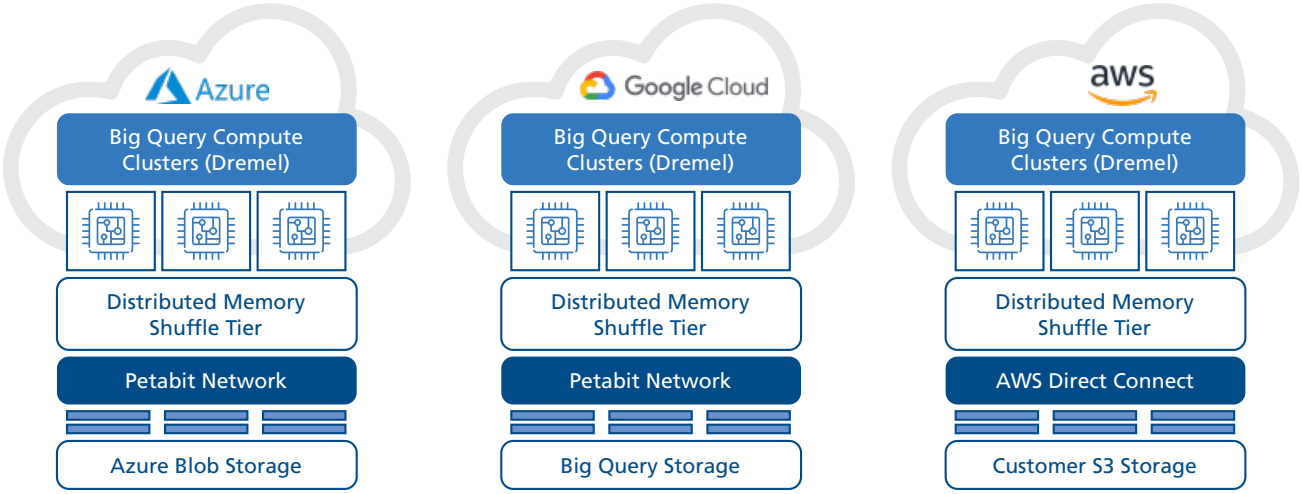
BigQuery is Google's fully managed enterprise data warehouse that allows organizations to query petabytes of data in real time. While this serverless service can be scaled up and down to meet an enterprise's needs and allows you to create machine learning models in minutes, it can only work with data stored in Google Cloud. So, if your customer wanted to query data stored in another cloud, they would first have to bring it to GCP. The high egress fees incurred in such cases can derail your data analytics goal.

This is where Omni gets to exercise its superiority over other solutions. It allows businesses to perform cross-cloud analytics from a single pane of glass without having to move/copy datasets to Google Cloud BigQuery. This ensures democratizing access to data in the true sense of the term. With BigQuery Omni, you can simply use standard SQL and BigQuery APIs to analyze strategic information stored in different clouds like AWS or Azure without having to move them to GCP.

The clear advantage here is that your query gets executed where the data is stored (other clouds), and only the results of that query are moved back to GCP. Now that you don't have to move all the data back to Google, you don't need to pay exorbitant egress charges. As a result, it turns out to be more cost-efficient as you pay egress fees only for the result set, i.e., if at all you decide to move query results to BigQuery.

Under the hood, Omni is powered by Anthos technology, a hybrid and multi-cloud application platform based on GKE (Google Kubernetes Engine). While Anthos handles orchestration and securely connects GCP with other clouds, the compute clusters (Dremel) run directly in Azure or AWS. In addition, you can use the Cross-cloud Transfer feature and transfer data from other clouds to BigQuery for more advanced analytics.

04 BigQuery Omni Architecture



05 How Does BigQuery Omni Work?

In the above image, the data flows between Google Cloud and AWS or Azure in the following sequence:

- Customers fire query jobs via the BigQuery CLI/API or the Cloud console. These are received in the BigQuery control plane.
- Next, these query jobs are sent to the BigQuery data plane (on Amazon AWS/Microsoft Azure) for processing via BigQuery routers.
- BigQuery data plane queries data from the Amazon S3 /Azure Blob storage tables.
- The table data is processed in the select AWS or Azure region.
- The BigQuery data plane then sends the query results to the BigQuery control plane through the BigQuery routers.
- The query results are then displayed in the BigQuery control plane and made available to the user.
- A similar process is followed for exporting the data with the exception that the data is written to a specified customer-defined destination instead of the data being sent back to the BigQuery control panel for display.

The best thing about BigQuery Omni being a single pane of glass is that the manner of querying remains the same irrespective of the cloud your data is stored in. So the way you would run a query in AWS will be the same as when you run a query on BigQuery or Azure data. Essentially this means that your analysts do not need to learn any additional skills like cloud-specific tools or data warehouse architectures for managing data in disparate platforms. This frees up their time and allows them to focus on their core work, which is deriving insights from data.

In short, BigQuery Omni:

- **Analyzes all data no matter where it is stored**
- **Securely share data across clouds**
- **Is incredibly fast**
- **Saves on egress fees**
- **Doesn't involve infrastructure management**

06 BigQuery Omni Features

Cross-cloud Analytics

Cross-cloud Transfer is one of the most important features of BigQuery Omni. Analysts don't always need to move/copy data. Sometimes, they need to copy small subsets of data for their analytic needs, and sometimes not at all. A true cross-cloud solution does just that: access to data and the flexibility to move it whenever required. With **GCP BigQuery Omni**, you can securely analyze data stored in other public clouds through a fully managed infrastructure. Not just that, you can also load data from AWS buckets or Azure Blobs without involving any ETL pipeline.

Single Pane of Glass View

Google Cloud BigQuery Omni breaks down the data silos and lets you view all your data from a single pane of glass. The unified management interface imports data from multiple clouds and presents it to the analyst as a single source of truth. Getting all the metrics in one place provides them with a complete picture of the key KPIs and helps them make better decisions.

A Fully Managed Serverless Infrastructure

BigQuery Omni is a fully managed serverless infrastructure, meaning Google is in charge of provisioning, scaling, and managing your servers. By using a serverless architecture, your data analysts can focus on data analysis instead of worrying about operating the servers. Additionally, BigQuery is highly scalable, meaning it can easily be scaled up from a handful of files to hundreds of petabytes in minutes.

Perform ML for Advanced Analytics

With BigQuery Omni, you can get query results in real-time and perform more advanced analytics. This is made possible because you can build and operationalize ML models on structured/ unstructured data directly inside BigQuery with zero latency and high concurrency.

07 The Benefits of Using BigQuery Omni

Quick Insights

GCP Bigquery Omni delivers faster results as data does not move across clouds. All computation takes place in the local cloud, and you get the results from a single interface. In BigQuery, data is stored in columnar format, and the tree architecture format helps aggregate results from thousands of queries in seconds. Additionally, your analysts don't need any advanced skills to analyze the data, so that also saves time.

Cost Effective

By ESG estimates, BigQuery provides a **three-year TCO that is 26-34% lower** than other cloud solutions. Add to it the fact that in GCP Bigquery Omni, you don't need to move data from its base location, and you can well imagine the savings on network costs. When you use Google Cloud Bigquery Omni, you're only billed for the queries that run on the BigQuery UI. Consistent security controls and lower operational costs encourage analysts to perform more permutations and combinations of data to uncover more unique insights.

Better Security

Using BigQuery Omni is more secure as you don't have to move or copy the raw data out of the local cloud. Further, GCP provides fine-grained access control at the column or row level to ensure that data always remains secure regardless of where and how you access it. With fine-grained role-based access control (RBAC), developers grant authenticated users access to only those tables they require access to.

Ease of Management

BigQuery Omni provides a single unified management interface to deliver query results, thereby significantly simplifying the entire process for analysts. All you have to do is write a standard SQL query in the BigQuery console to query data in local clouds and see the results displayed in real time in the BigQuery console. There is no need for custom integrations, data replication, or provisioning of infrastructure.

08 Use Cases

Example 1

Let's say you are a Project Manager and are tasked with optimizing project planning and hiring in your company. For that, you would need details of your HR, Sales, Finance, and Operations teams. But you have access to only the Operations teams' data stored in GCP. The rest of the data is stored in other hyperscalers like AWS and Azure. So how do you go about it?

Of course, you can combine the different tables residing in two different clouds by doing periodic data synchronization, but doing so can be expensive. You can also limit yourself to only the data on GCP, but then you won't get the complete picture.

Historically, you had to move the data from AWS/Azure to GCP before starting data analysis. That would have incurred a lot of egress costs. Now, since you have Omni computing locally on Azure or AWS, you can query the data through the user interface without moving it and thus avoid paying costly 'egress' fees.

The result: The Omni dashboard gathers and collects data from all the clouds and provides a clear picture of all the metrics, allowing you to make better business decisions.

Example 2

Let's take another example of the retail sector. Suppose you want to find out how advertising and/or audience response to your product impact your sales. How do you achieve that? The problem is that your retail data, which includes a mix from both your online and offline stores, is stored on different cloud systems. Now, if a brand manager wants to understand how advertising affects your sales, they would need to get a complete overview of all the retail data.

But doing so is not easy, especially when you need specialized skills to connect to individual cloud environments. What if you had to build an ML model based on that data?

Thankfully, with Omni, you can quickly set up a connection between the BigQuery interface and the data residing on other cloud platforms, query it, and receive results in minutes. Once you have the query results, you can build a dashboard with Looker, visualize the data, and present audience behavior along with purchases in an easy-to-understand format.

Now once you have your commerce data, you can tie that data back to your advertising data and do better targeted advertising to generate more sales.

09 Opinion

Popular Integrations with Other Google Services

GCP provides a portfolio of solutions to help users solve their business problems.

BigQuery ML

With BigQuery ML, analysts can create and execute ML models in BigQuery using standard SQL queries. By allowing developers and business users to build models using existing SQL skills, Google has democratized machine learning. The development speed also increases as analysts don't waste time exporting data from other applications. Instead, they can quickly deploy ML models on both structured and unstructured data and perform predictive analysis.

.

Looker

Looker is a business intelligence (BI) tool that business users can effectively harness for data preparation, training, running predictions, and model management. Looker's agnostic database architecture lets you connect to hybrid clouds for cross-database queries, allowing developers to access apps through Apigee or the BigQuery API, irrespective of where the data is stored. With Looker, you can aggregate data from multiple clouds and explore and share visualizations with other users.

10 Conclusion

Studies show that multi-cloud is for the long haul. According to RightScale's "State of the Cloud" survey, organizations use almost an average of 5 different clouds to store their data. It's not hard to understand why. Google Omni is ideally suited to meet the challenges of multi-cloud head-on, as it makes true workload interoperability a reality.

At LTIMindree, our mission is to help companies embark on their digital transformation journey across multiple areas, such as Infrastructure Modernization, SAP on GCP, Data Warehouse Migration, Advanced Analytics, and Big Data. As a Strategic Google Cloud Partner, we can help you execute end-to-end migration of your data warehouse system to the Google Cloud Platform. Our Google-certified experts can guide you through the entire journey and design a customized migration plan that exclusively meets your business requirements.

Interested to learn more about LTIMindtree and Google solutions? Reach out to our **experts** here.

References

https://cloud.google.com/bigquery/docs/omni-introduction

https://cloud.google.com/blog/products/data-analytics/introducing-bigquery-omni

11 Author Profile

Chaitanya Pande

Principal – Data Engineering Google Data CoE
@ LTIMindtree

Chaitanya Pande is the Principal – Data Engineering associated with Google Data CoE @ LTIMindtree. He is seasoned professional in the Data & Analytics leadership space with more than two decades of experience.

Data technologist at heart, he played key leadership roles delivering strategic and tactical high-end solutions, consulting assignments, solutioning RFI / RFP with reputed IT firms across global regions. He helped strategizing GTM data strategies for building unified data foundations and Advanced Analytics use cases on Google Data Cloud.

He is deeply involved into building Cloud Native Data specific competencies, accelerators and providing technical depth to various initiatives on Google Data Cloud. Prior to LTIMindtree, Chaitanya was Data Head @ CloudCover. Previously he served Schlumberger, SAS, Cognizant in various Data-Leadership roles out of USA and India.

LTIMindtree is a global technology consulting and digital solutions company that enables enterprises across industries to reimagine business models, accelerate innovation, and maximize growth by harnessing digital technologies. As a digital transformation partner to more than 700+ clients, LTIMindtree brings extensive domain and technology expertise to help drive superior competitive differentiation, customer experiences, and business outcomes in a converging world. Powered by nearly 90,000 talented and entrepreneurial professionals across more than 30 countries, LTIMindtree — a Larsen & Toubro Group company — combines the industry-acclaimed strengths of erstwhile Larsen and Toubro Infotech and Mindtree in solving the most complex business challenges and delivering transformation at scale. For more information, please visit www.ltimindtree.com.