



<u></u>

# DATA AS A SERVICE (DAAS)

### CONTENT

DaaS: Foundational aspects Key architecture aspects of DaaS

Benefits and Advantages

### INTRODUCTION



Data as a Service (DaaS) is the latest kind of services being explored in various Cloud communities. The main purpose of DaaS is to overcome limitations of approaches in data technologies, according to which data is stored and accessed from source repositories whose location is known and its relevance for sharing and processing the data.

Besides, DaaS approach is used for intelligent sharing and processing of large data collections with the purpose of abstracting the data location and to fully decoupled the data and its processing. In various cloud computing platforms, DaaS supports large communities of users that need to share, access, and process the data for collectively building knowledge from data.

In today's world, DaaS comes with two trends as Big Data and Cloud Computing defining the emerging Enterprise Computing Platform, and has a lot of future for a new era of combined applications. The facilities of Data-as-a-Service capabilities using multiple cloud delivery techniques could ease adoption for many organizations & companies, and it's important for cost savings, which could help to get useful insights that could provide them with various ways of competitive advantage.

The evolution of ML/AI, IoT and Sensors further drives the need for DaaS as it makes it operationally efficiently manage and scale. Data as a Service (DaaS) provides an approach to an extensible platform that can provide cloud-based capabilities such storing, accessing and processing of the data over a variety of industries and use cases.

From a solutions perspective, DaaS covers the end-to-end capabilities of an analytical solution, from data acquisition, data storing, data processing and to the end-user visualization, reporting and interaction. In addition, the DaaS supports the needs of the different users who interact with it, including those of the emerging 'Data Engineer' and 'Data Analytics' role. Importantly, due to the inherent problems of analytical processes, the implementation of DaaS represents an important set of challenges, as it is mostly similar to a flexible Platform as a Service (PaaS) solution than a more Software as a Service (SaaS) application.



Aspects like the PaaS internal architecture, the distinction between real-time vs. non realtime data processing, the specific characteristics of the Data Analytic Services, the needs for data storage and data modeling on the cloud platform, the delivery over hybrid cloud services and other models make this a complex design challenge. This scenario explores some of the most important points that need to be solved in a DaaS.



## DEFINING DATA AS A SERVICE

Data as a Service (DaaS), is enabled by software as a service[1]. DaaS builds on the concept of its data product which can be provided to the end user on demand, regardless of location or organizational separation between provider and consumer. Software as a Service (SaaS) is used in every organization and business. Companies like SalesForce or an email marketing service provide customer relationship management. Software as a Service products were launched globally in the 90s, where most of the organizations and businesses have taken root both in SaaS and in the "as a service" family in general.

Recently, "Data as a service" was added in a cloud-based strategy "as a service" family. In today's world, Data as a Service wrangles an exponentially growing large amount of information and makes it available across the globe anywhere and anytime. Data as a Service providers (GCP, AWS, Azure, IBM) served as centralized data warehouses to gain data-centric insights through the cloud securely and affordably. DaaS service that provides you an option to log into whenever you have spare time and immediately get access to a variety of data-driven insights. Lots of planning, engineering and structuring that goes into democratizing datasets so that it is understandable and actionable. Visual shown below.



### KEY ELEMENTS IN DAAS



Data as a Service (DaaS) provides a data analytical platform using a cloud-based solution model. There are multiple key elements involved in DaaS to draw out information from actionable insights. Following are the key elements of Data as a Service.

#### DATA COLLECTION

Data collection is the process of gathering and measuring structured and unstructured data, information from various data sources by identifying best and optimizing methodology. The main purpose of all data collection is to identify data quality and patterns that can easily transform the data analysis, potentially answering the questions that have been posed. Once the data is collected, it is stored and loaded into data warehouses on the cloud. This first step is important as it ensures the integrity of the data collected. Additionally, the selection of the best data collection methodology help clarify if the data collected is correct while reducing the probability of errors.

### 2

#### DATA AGGREGATION

Data aggregation is the process of assembling and presenting the data in summarized manner. The structured and unstructured data collected from various data sources is combined and merged into a summarized metadata for rich analysis. The most important process is to identify the accuracy of hidden patterns from the data analysis which in turn depends



upon the volume, integrity and quality of the data collected. It crucial to collect high volume and high quality accurate data to get the expected results. Finally, the data aggregation process is a helpful use case for finance, business decisions, market planning, operations, and pricing.

# 3

#### DATA CORRELATION

Data correlation is the process of using statistical analysis to find a strong relationship between two or more data points. Its purpose is to determine how the data is associated and mapped. A strong correlation implies a strong relationship between two data points, improving the data analysis and decision making processes.

# 4

### DATA VISUALIZATION

Data Visualization is the process of identifying patterns and hidden insights by crunching big data and creating a visual representation or business intelligence dashboard for any organization, stakeholders, or end user. The data is analyzed and represented in interactive and dynamic dashboards, charts, graphs, and so on. These visualizations give business teams the ability to formulate marketing, sales, finance and management strategies.

### 5

#### DATA ANALYTICS

Data analytics is the process of building complex models around the data to find hidden patterns and gain insights from the large and complex data. This analysis simplifies big data problems, unlocking new insights and avoiding analysis paralysis.



### ARCHITECTURE ASPECTS OF DAAS



Sample Architectures of Enabling Talend on GCP

Use Case 2: A Medi	a Publisher Improving Online Advertising Revenue Streams by Building a Single Customer View in the Cloud			
Clickstreams Clickstreams JSON Files	Cloud Platform Cloud Pub/Sub → Cloud Storage → Cloud Storage → Cloud Storage → Cloud Big Data Platform → Cloud Storage → Cloud Big Data Platform → Cloud Big Data Platform → Cloud Big Data Platform → Cloud Big Data Big			
Figure 2: Real-Time Processing Data Workflow using Talend and Google Cloud Platform				

		Google Cloud Platform	
Clickstreams	Google Analytics	Talend Big Data Platform → BigQuery → BigQuery → BigQuery →	Analysts
JSON Files		Apache HAWQ HAWQ Greenplum	Analysts
		Figure 3: Cloud BI Lake using Talend and Google Cloud Platform	



There are various aspects of DaaS and options of components to enable the overall DaaS framework. DaaS Architecture can be built on various cloud computing platforms. The above figure showcases the Data as a Service with Talend on Google Cloud Platform, where the structured, unstructured, real-time and batch data is collected from various sources such as on-premise databases, IoT devices, APIs, social media platforms, and so on. Data collection plays a major role in DaaS architecture where the data is gathered in accurate ways in given formats such as Json, text, image, csv files. As we can see in the above figure, the data is collected from various sources. There are two approaches to process data on Google Cloud DaaS.

#### • Real-time Data Processing Workflow on GCP

Here, real time data is collected using the pub/sub mechanism provided by the google cloud. Pub/Sub service is used as messaging-oriented middleware, data ingestion and delivery for streaming data pipelines. The pub/sub pipeline collects the data and stores it into Cloud Storage bucket. Cloud storage is used for content delivery, backup and data warehouse where the raw data is stored. Further the data is transformed, refined and processed using Data flow and stored it into Bigquery. The Bigquery services on GCP are serverless, highly scalable and cost effective data warehouse designed for data analytics to gain business insights. In Bigquery, we do data aggregation, data correlation and data analysis to identify insights and hidden patterns inside the data. These insights and patterns are visualized in dashboards as graphs, charts which help organizations and companies to plan their business activities, marketing strategy and low risk in decision making. And can be exposed as web services and RestAPI to the end users.

#### • Batch Data Processing Workflow on GCP

In batch processing, the data collected from on-premise databases, social media, rest api, across different cloud platforms and so on. Here, the data pipeline is orchestrated based on the end user or has per the customer need. In Google Cloud Platform, we orchestrate data pipelines using cloud composer, or via cloud scheduler. In the data pipeline, data is collected and stored into Cloud Storage as backup or data warehouse. Based on the data structure and volume, the data is processed and transformed using Dataflow or Data Proc services. These GCP services build on top of Apache Beam and Spark, to process large volumes of structured and unstructured data. This data is further stored into Bigquery for data analysis. Where the data analytical team runs data aggregation and correlation processes to find out insights, information and hidden patterns from the data. This information is further visualized on a dashboard with help of graphs, charts, etc and exposed as web applications and rest api to the end user to build their business strategy and goals for the organizations.





### Benefits & challenges of DaaS

#### **Benefits of DaaS**

#### • Agility:

Customers can make quick and better decisions with the consolidation of access to data in a single place. The decisions could be made without the customer having the knowledge and details of the underlying data layer managed by DaaS. DaaS also helps in simplifying the access to data through a set of APIs.

#### • Utilization of data

A DaaS provider can help build an architecture to meaningfully derive data and utilize it. The organization can utilize the findings and insights from data to understand the customer better, find patterns in the behavior, and personalize customer experience.

#### • Data Quality and Management

The insights gained out of any data store is limited by the data quality. The quality of data improves as the updates and further maintenance of data goes through the strict adherence of rules as set by the organization and the updates happen via a single point. Also, DaaS providers can provide options to clean data as it is ingested and later merge with existing data.

#### • Cost effectiveness

The utilization of data to derive insights and refine customer experience adds value to the organization. Using DaaS, not only the data services layer is managed well, it also brings up value to the organization by serving customers better. The DaaS model will connect and give access to data across hybrid cloud and multi-cloud environments, bringing in insights and intelligence from data,effectively monetizing the data.





#### • Security and Data governance regulations

Data, at rest or in-transit, if exposed to any external source without authentication could result in a data security breach. A DaaS provider will provide the advanced and sophisticated data security. Along with data security, data governance regulations like GDPR compliance will be provided by the DaaS provider.

#### Challenges on DaaS

#### • Privacy

Privacy is a major concern for any organization. As the data belonging to an organization is handled by third party providers with a DaaS solution, the concerns are valid and significant.

#### • Security

Although the DaaS providers install the best-in-class security solutions, the DaaS solution brings in security concerns as it may leave mission-critical datasets to vulnerabilities. The concerns also arise with the capabilities of the service provider handling such a scenario.

#### **References:**

- <u>https://en.wikipedia.org/wiki/Data as a service</u>
- <u>https://cloud.google.com/blog/products/data-analytics</u>
- <u>https://cloud.google.com/blog/products/data-analytics/introducing-bigquery-omni</u>
- <u>https://cloud.google.com/storage</u>
- <u>https://cloud.google.com/blog/products/data-analytics/cloud-data-governance-collibra</u>
- <u>https://cloud.google.com/blog/products/data-</u> <u>analytics/data-lakes-in-cloud-with-kafka-and-confluent</u>

