What type of data processing organization are you?

Firat Tekiner and Susan Pierce

The data value chain
Page 3

Data processing: A decision based on your organizational landscape
Page 6

Building your data-driven organization with Google Cloud
Page 12

Introduction

We are generating more data by the day fueled by the increasing digitization of business and society. Further, the complexity and speed of data arrival is changing the ecosystem. This data, and its potential to drive better decision making, promises to deliver higher revenues, lower costs, and reduced risk to all types of businesses. However, the most valuable insights no longer come from straightforward sales, inventory, and personnel data. The next generation of insights is hidden across countless, often unstructured data points from myriad sources and systems. Extracting the insights requires the right blend of tools, skills, and strategy.

Fortunately, there are plenty of modern data processing tools to help you rise to the challenge. What your organization chooses depends on its culture. This article helps you succeed by navigating the options that enable a wide range of skill sets and personas to derive insights from data. Connect your culture to the right tools to ensure you succeed in delivering better value from data and hence gain competitive advantage. Success is now predicated on choosing the right strategy for your business based on your personnel and objectives.

This paper outlines the framework a technology leader can use to select the right approach for their organization. We discuss strategies for approaching data processing tasks within an organization based on its culture. Every organization is unique and requires a tailored approach to data processing. It would be naïve, for example, to compare a digital native which built all of its processes in the big data era with a brick and mortar store which has been running legacy systems for the past 50 years. The expected outcomes and success metrics for the organizations would be different. Recognizing this, we will define the types of organizations and map out realistic expectations for each.

Furthermore, we look at the accompanying data processing technologies that can be used to build a foundation for success in each type of organization. Once we develop a cultural framework that CIOs and CTOs can use to make decisions about their data processing stack, we illustrate how it can be applied using the range of Google Cloud data processing options. We review decision points in detail while offering a broad framing of issues based on high-level concepts like typical user roles and institutional tolerance for directly managing technical infrastructures.

We hope you will find this useful and that, in applying the framework, you will learn to think about your business and its technology requirements in new ways — less focus on bits and bytes, more focus on objectives and people skills.

The data value chain

When we think about data processing in particular, it's important to place it within the broader context of the data value chain. We can imagine data traveling along an assembly line, as a car might in a factory. This assembly line progressively adds parts and value to an object moving along it. Raw data at the beginning of the line is eventually transformed into actions taken by humans or machines. Let’s examine the steps in this data value chain:

1

Data genesis is the initial creation of a unit of data; this could be a click a website, the swipe of a card, a sensor recording from an IoT device, or countless other examples. It’s the raw material that will eventually be milled into an insight ready for action.

2

Data collection brings that initial unit of data to the insights assembly line. The data collection stage can be extended to include data extraction from various sources and ingestion to the target environment. This step tackles the process of gathering data relevant to the analysis being performed. In addition, data collection involves data pushed from certain systems, such as a sensor that could push events into a message bus and then process them. This collection step can have dramatically different requirements based on the volume, velocity, and variety of the raw data that’s required for a given analysis, as well as how fast the data needs to be analyzed.

3

Data processing is where the raw data we’ve collected is transformed into a form that is ready-made for deriving insights. This might involve any number of adjustments to the data, from merging datasets to advanced statistical methods and everything in between. It can be a single-stage operation, or it can be a complex tree of cascading procedures, more widely used as directed acyclic graph (DAG) in the traditional Extract Transform Load (ETL) world. Data processing might be in service of a long-term strategic analysis where speed isn’t a priority, or it may need real-time capabilities to capture time-bound opportunities like IoT maintenance or shopping cart cross-sell. In our manufacturing process analogy, this is the phase where raw materials take the shape of the pre-assembly parts of a manufactured product.

4

Data storage is where our data lands and lives, ready for analysis and action. Traditionally, in a model known as ETL, data moves into storage after data processing occurs. However, cost innovations in object storage-powered data lakes, along with more recent developments in enterprise data warehouses, make Extract Load Transform (ELT) viable. As with real-world manufacturing, where storage options vary depending on the type of product being processed, different types of data can be stored in different ways. For example, NoSQL is available for fast reads and writes, data warehousing for fast access to analysis, and object storage for unstructured data. There are also specialty blends of these standard stores.

5

Data analysis is similar to data processing and features the same usage of compute resources to better understand and interpret data in general terms. In every application, the amount of computational resources can be reduced by applying advanced optimizations, but it does not always result in a higher return for the time invested. The big difference is the intention; data processing aims to set the stage for data analysis, whereas the analysis itself provides direction for business-oriented action. This analysis is frequently accomplished by data analysts using SQL in the data warehouse, but they can also use open source frameworks such as Spark and Flink or sophisticated statistical analysis through machine learning. To continue with our manufacturing line analogy, this is the stage where inputs from the data processing stage are assembled into a final product.

6

Data activation is the final step in our data value chain. Once an analysis is produced, it needs to be pushed to the relevant business procedures and decision makers so action can be taken and the value chain completed. The most common points of activation are applications making automated decisions and business intelligence dashboards guiding humans towards better, more informed decisions. In our manufacturing line example, this is the step where a fully produced widget is put to its intended use.

There is no one way to assemble a data value chain, as there's no one way to create a real-world manufacturing line. Similarly, as technologies progress, new inputs become available, your workforce evolves, or the desired output changes, the optimal value chain will also change. However, at its core, the value chain principles hold. We want to take raw data, however varied, and complex it may be, and change it into action that benefits the business.

The value chain is arguably evolving most rapidly with data processing. As a result, the key is to have an ecosystem whereby decisions are made based on an organization's capabilities rather than trending technologies.

These capabilities support your users — from data scientists to engineers, developers, analysts, and business users — with an open, intelligent, and trusted data platform. The data platform has been designed to solve for your use cases to collect, process, store, analyze, and act to turn your data into value. **The output enables you to bring rich data experiences into every process within the organization, bring data assets together with an intelligent data fabric, and ensure trust and reliability with a secure data foundation.**

Ingestion:

Setting the stage for data processing

It is difficult to have an in-depth conversation about data processing without addressing data ingestion in some detail. After all, we cannot process or analyze data that doesn't enter our environment. For this reason, we often see ingestion and processing coupled together in solutions (such as with Pub/Sub and Dataflow). Ingestion systems need to match the speed and scale of processing systems which in turn need to match the speed and scale of analysis driving better business decisions.

The basic function of ingestion is to extract data from the system in which it is hosted in order to bring it to a new system. The two goals for this step in the data value chain are to reduce complexity by making data accessible as soon as it is available and to deliver data with timeliness that meets the needs of the use case at the end of the value chain. Traditionally, similar systems relied on matching the schema of the target, which required complex processes to transform the data in the format expected for loading.

In choosing the right approach, users must examine the number of data sources, the required speed of ingestion, and the sheer volume of data required for the needed analysis. Therefore, solutions should be able to store any type of data, regardless of its structure and speed, and then run with any type of volume. Decisions around real time versus batch ingestion are typically a top priority, but the degree of scalability and automation are equally important considerations because use cases become more complex as the variety and volume of the data increases.

The process of tuning, configuring, monitoring, and provisioning ingestion resources can put a significant strain on data engineers and increase the total cost of ownership (TCO) of a solution as requirements shift and grow. For common pipelines, ETL tools (like Data Fusion) can be ideal. They cover multiple phases of the data value chain and often feature pre-built transformations that significantly simplify the process and remove the coding requirements typically involved with building data pipelines. These tools reduce the strain on a business' data engineers, can scale to thousands of pipelines, and lower the barrier to taking on ETL work.

While these preceding topics are important considerations for ingesting and collecting data, another ingestion consideration for our people-centric approach to data processing is the data's destination. Structured data has the potential to be ingested directly into the Enterprise Data Warehouse (EDW) in an ELT-focused approach powered by SQL. For unstructured data (such as XML, EBCDIC, or others), data need to be processed in an ETL pipeline before hitting a structured datastore. Still, further options exist to keep full-fidelity unstructured data in object storage within a data lake for an ELT process pioneered by the Hadoop and Spark ecosystem.

These destination-oriented decisions map to enabling specific personas, such as data analysts, developers, data engineers, and data scientists within an organization to drive data analysis. Understanding the type of data organization you have will help you make the right decisions for your data architecture which, in turn, will help your data users to make the right decisions for your business.

Data processing:
A decision based on your organizational landscape

The idea that there is a "one size fits all" or "best" approach to data analytics and data processing is a legacy from a time before today's data-centric, cloud-first world. In today's data landscape, just as in professional sports, successful organizations adopt a strategy that makes sense for the people and skill sets they have. If a sports team has a great defense, they should try to win through defense, not by copying the offensive strategy of a different team with different players. Similarly, if an organization has a strong bench of data analysts, they should lean into their people instead of trying to transform into an organization full of data engineers.

What has changed? New tools, like Google's BigQuery, have opened new options for how data processing and data analysis are done. They reduce the data life cycle considerably, allowing analysts to carry out data engineering either in batches or in real-time while simultaneously experimenting with data science solutions. Furthermore, the potential surface area of any given persona — their capabilities and responsibilities — has expanded. Existing data workers can take on new tasks and address the data value chain without the bottlenecks associated with the traditional ***persona value chain***.

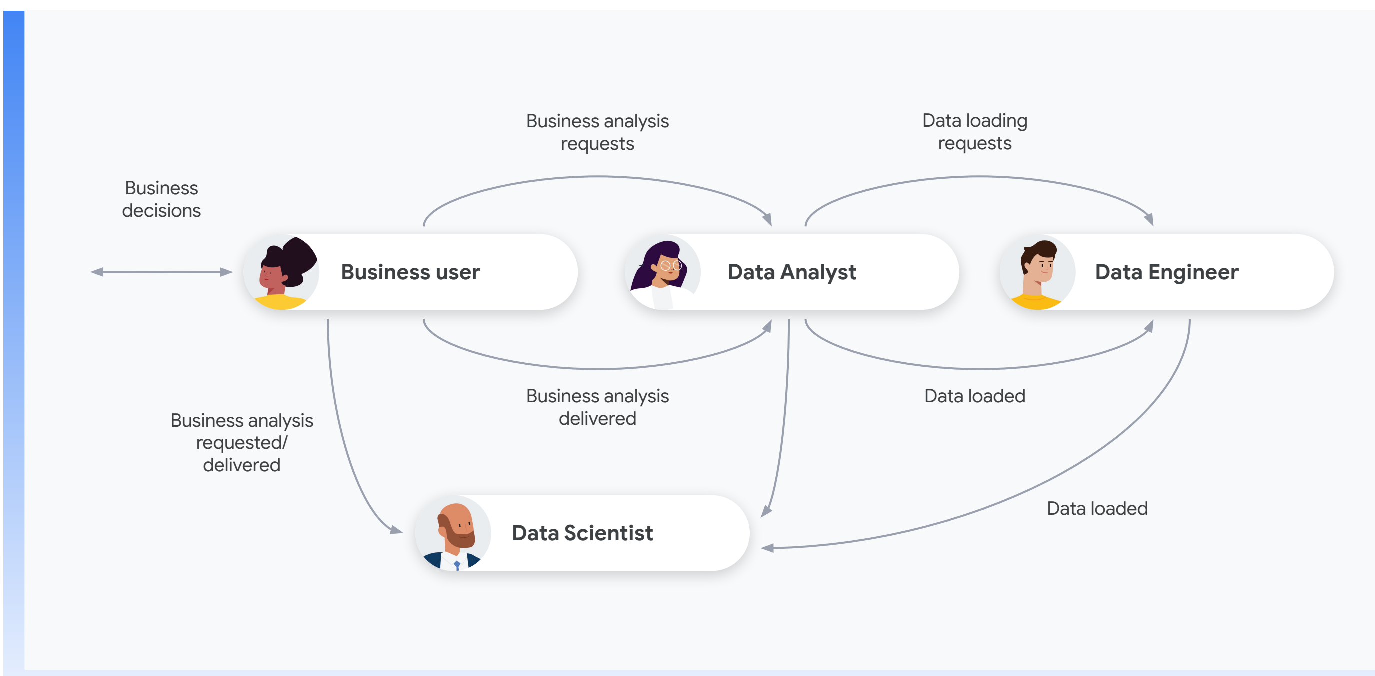


Figure 1: Data processing: Traditional persona value chain

In the traditional persona value chain (Figure 1), every data user in an organization has a small and specialized set of technical skills. If an organization wants to increase the scope of their data analysis team, they also have to scale the size of their data engineering and data science teams. This will make sure that enough people have the right technical skills to support the data analysts. Now that the technology is more accessible, we see a blending of skills across roles, allowing existing teams to more easily scale to additional scope. Today, the technology can match the staff, which changes the landscape of data processing.

As described in Figure 2, traditional data processing is handled either by data analysts or data engineers (or other personas such as data scientists acting as data engineers). The distinction between them is how the ELT and ETL patterns are handled, as described above. However, a blended organization can take advantage of the best of both ETL and ELT patterns. Most of the organizations we see today fall into this blended model, though the balance of roles and how much of the data processing is to be done either through ETL or ELT varies depending on the type of the organization.

Figure 2: Data processing: Persona framework

Every organization is unique, and very few teams are defined by a homogenous group of data workers. Most organizations have a mix of roles and skill sets. But to determine an overarching data processing strategy, leaders should lean into where the bulk of their data workers' skills are strongest. This leads to data processing strategies that are broadly data analyst driven, data engineer driven, or truly blended.

Organizations driven by **data analysts**

Data analysts are the most common data persona at many large Fortune 500 companies. They are experts in data warehouses and use SQL to analyze data. These organizations have built teams structured around the traditional persona value chain, where data engineers clean and prepare data for ingestion into the enterprise data warehouse (EDW), at which point data analysts take over.

However, a few shifts in the broader landscape of data have both stressed that common model and presented new opportunities.

The first shift occurred on the data front. There is more data than ever, originating from more sources than ever, and data workers want access to it all. Business users want analysis derived from these varied sources, and individual departmental leads want different insights based on the same data. That means data workers need access not only to processed data, but specifically to raw data that allows them to build bespoke analysis from the ground up. In a world where data analysts rely on data engineers to preprocess data for them, bottlenecks slow down innovation and rapid iteration.

The second shift occurred on the technology front. Traditional EDWs intrinsically coupled compute and storage, resulting in what has historically been an expensive appliance. Even if businesses wanted to store more data, cost constraints meant they couldn’t. ETL was a necessary pattern because it helped reduce the volume of data needed for analysts to drive to the same results.

In a world where data analysts rely on data engineers to preprocess data for them, bottlenecks slow down innovation and rapid iteration.

The legacy systems that have worked well for the past 40 years have proven to be expensive and pose significant challenges around data freshness, scaling, and high costs. Furthermore, they cannot easily provide artificial intelligence (AI) or real-time capabilities without bolting that functionality on after the fact. These issues are not only present in on-premise legacy data warehouses; we even see this with the newly created cloud data warehouses. Modern cloud data warehouses have changed these traditional economics by decoupling storage and compute. However, many do not offer integrated AI capabilities, despite their claims. These new data warehouses are essentially the same legacy environments but ported over to the cloud.

On the other hand, BigQuery is truly serverless, so compute and storage are separate. All of the above applications, like AI capabilities, can run without impacting the performance of any other jobs accessing BigQuery at the same time. In addition, BigQuery introduces:

- **Real-time analysis over streaming data**
- **Centralized storage to feed advanced analytics systems like Apache Spark or Python**
- **Comprehensive security and trust through a defense-in-depth approach that secures hardware from the lowest level of the stack through all the software layers**
- **Data accessibility through flexible and secure data sharing**
- **Predictive analytics by using ML & AI accessed through built-in SQL extensions**

As a result, users can affordably store as much data as they like within their data warehouse while consuming only the amount of computing resources necessary to carry out their analyses. This has given rise to the concept of a "structured data lake" that has made the ELT pattern of traditional data lakes, previously only used by data scientists and data engineers, possible for data analysts.

Deploying an ELT pattern within a data warehouse puts data analysts in control of data processing and allows analyst-driven organizations to drive data into every corner of their business. This starts by directly loading (and even streaming) data into the EDW as it gives full fidelity access to data analysts. Rather than transforming the data outside of the EDW, analysts no longer need to wait for data engineers to prepare the data — they can exploit their SQL skills to transform and analyze it. This helps them generate the bespoke analysis that business and departmental leaders need to better inform their decisions. The traditional bottlenecks involved with the persona value chain are therefore removed, because analysts can now perform tasks that would have required data engineers or ETL developers in the past. **This reduces operational costs and increases the timeliness of the data.**

In short, the new paradigm combines the familiarity of the enterprise data warehouse (SQL, stored procedures, scripting, etc.) with the ability to effectively and efficiently process data, all in the same place. For businesses with large data analyst populations, this is a nimble and powerful way to process data.

Organizations driven by **data engineers**

Data engineers can be found in every data-driven organization, but this persona is the dominant presence at many digital native businesses in particular. Built upon data and the ease with which online platforms can be instrumented, successful digital natives have been at the forefront of using data to deliver a better experience to customers.

Building complex data engineering pipelines allows for the creation of repeatable processes and the rapid scaling of inputs to data analysis. Data never lives in isolation and it needs to be analyzed alongside other datasets to deliver the type of analysis that can lead to better decision making. Data engineers are skilled at bringing datasets together and then automating both the analysis of data and the subsequent actions that are taken. This is done either directly by the platform (such as shopping cart recommendations) or by using a visualization that prompts a person to take action (e.g. ride hailing services).

The past decade has seen the capabilities of data engineers grow by leaps and bounds. The big data revolution brought about by new technologies and open source projects like Apache Hadoop meant it was possible to store and analyze more data than ever. Raw, unstructured data could be refined and put to action, either directly or by teeing it up for data analysts. Automated, on-the-fly decisions for businesses could be powered by real-time analytics, made possible by increased digital signals. But managing these environments, handling their growing scale, and finding the people capable of handling these tasks has been difficult in an on-premises world.

Today, data engineers are increasingly turning to public cloud providers so they can focus on building the data processing pipelines their businesses thrive on instead of managing infrastructure. Modern cloud solutions provide autoscaling resources that adjust the data environment to exactly what is needed for processing jobs. This makes it fast and easy to iterate new pipelines, new analyses, and new capabilities. Data engineers can easily move back and forth between batch and stream analysis. They can even add machine learning capabilities through data processing extensions or easy to use APIs. Furthermore, leading cloud providers provide deep integration between ingestion, processing, and other analytic tools giving even more speed to the development of data processing pipelines and analysis.

Blended organizations

and addressing data scientists

The third type of organization is a blended organization, featuring a balanced mix of data engineers and data analysts. We should be careful to note that "balanced" does not necessarily infer a 1:1 ratio. Instead a ratio that has been built to support the traditional model of ETL where the data processing is handled by data engineers (or ETL developers) and then landed in a zone where data analysts can access this cleaned data. The ratio is not set but is specific to a given company. These blended organizations typically started as data analyst-driven companies that began to hire more data engineers as big data and digitization increased the number of signals a business could collect and analyze.

The varied types of personas that need data access in a blended organization form the concept of an interoperable, connected, and complete platform critical to data-driven decision making. Without a modern data platform, differences in skills within the organization will lead to natural data silos, likely resulting in a data lake for engineers and a data warehouse for analysts.

The idea of a connected and interoperable platform is important for another persona, the data scientist. Often armed with data science notebooks, like open source Jupyter Notebooks, these users drive machine learning and advanced statistical analysis to bring forward new insights across datasets. The nature of data science means these users need access to the domains familiar to analysts (structured data) and engineers (unstructured data) to perform their analyses. When these environments are separate and disconnected, data scientists have to burn cycles moving data, merging separate analyses, and gaining access to datasets. In a unified environment, data scientists can use the same tools to analyze various sets of data seamlessly.

Clearly, a unified and interoperable environment is important for data processing and analysis. This comes together mostly clearly in today's landscape through the blending of data warehouses and data lakes.

Building your data-driven organization

with Google Cloud

In theory, optimizing data processing systems based on organizational skill sets sounds reasonable. But what does it look like in reality? In the earlier sections, we took a mostly platform-independent approach to describe data processing styles. In the following sections, we're going to provide examples of what these architectures would look like using Google Cloud products and services.

One key consideration is how many different data sources you need to manage. Are you looking to scale thousands of sources using generic pipelines? Or do you want to create one generic pipeline and apply data quality rules and governance as the data is being ingested into the data warehouse? ETL tools are ideal for these use cases as generic data pipelines can be written and then parameterized, one of the capabilities of Data Fusion. It is also important to understand the speed and time of arrival of the data. What SLAs and time durations/windows are relevant for your data ingestion plans? This not only drives the ingestion profiles but also dictates which framework to use. As discussed, velocity requirements drive the decision-making process.

Figure 3 demonstrates an integrated platform from a Google Cloud point of view:

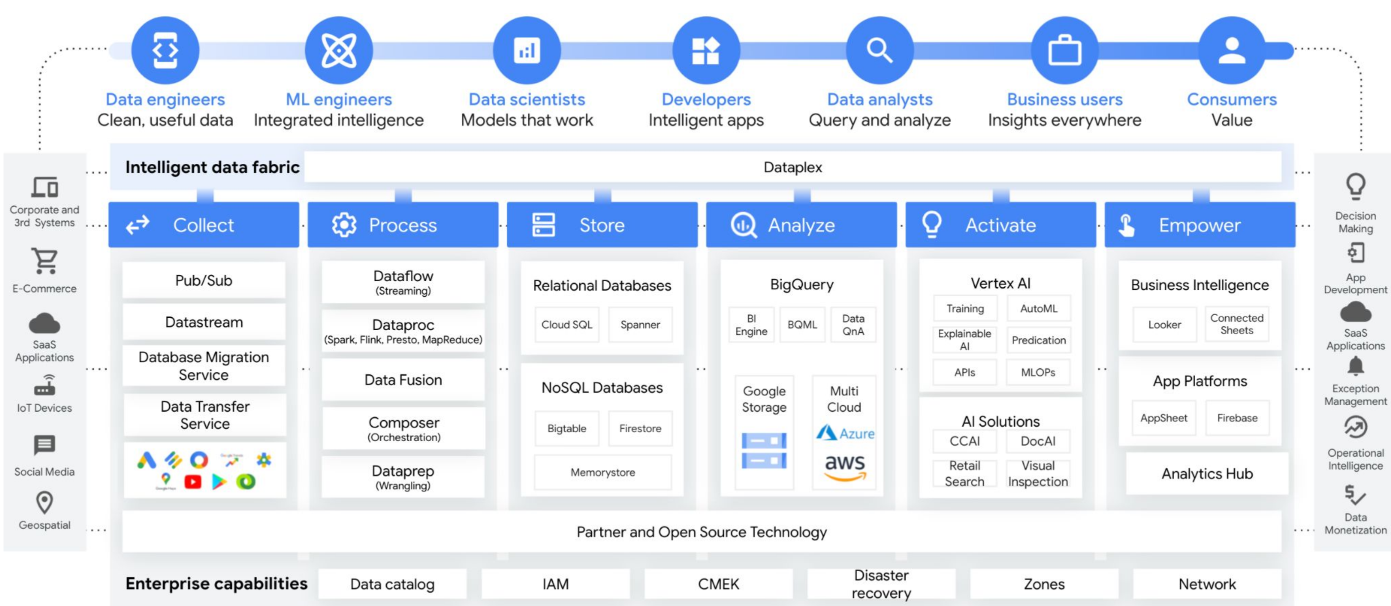


Figure 3: Data platform

Painting the picture:
Data analyst organization

We've talked about the processing layer, but where is the data coming from? Can the data be directly ingested without transforming and formatting the data? If the data does not need to be transformed and can be ingested directly into BigQuery, a managed solution such as Data Transfer Service (DTS) would be ideal. This not only reduces the operational costs but also allows for timely delivery of data to the data warehouse.

Why is ELT important? In order to enrich, aggregate, and cleanse data for analysis, one needs to have access to all applicable data sources. With ever increasing requirements, one needs to do this quickly and with as few steps as possible in a data warehouse. Modern data warehouses, such as BigQuery, offer unparalleled ability to store any volume of data economically and process complex queries in little time.

This significantly reduces the time and effort spent managing applications. In this way, data is extracted from the external system into a temporary staging area. Whether it is a real-time data stream, like from Pub/Sub or a managed Kafka cluster, or batch process, it can be loaded into a Cloud Storage bucket within the data lake. Data is then transformed before it is loaded into the data warehouse.

Although this model describes a more traditional ETL pipeline, it is only used when required due to the business and technical processes involved. On the other hand, this is not a necessity with BigQuery as it reduces operational overhead and increases efficiency. It is possible to build complex ELT pipelines with the use of tools such as DataForm and leverage the SQL capabilities of the data analysts. In doing so, the need for managing and learning a separate ETL tool is eliminated.

Painting the picture: Data engineering/ science organization

If data is coming in an unstructured format such as XML or EBCDIC and needs to be transformed, then a tool with ETL capabilities such as Data Fusion or Dataflow is useful. In a data engineering-centric organization, the choice between Data Fusion and Dataflow can be based on the speed of the data arrival, with Dataflow as a better option to handle real-time workloads. On the other hand, if the application is being migrated from an existing Hadoop-based pipeline, like Spark code, then a tool like Dataproc would be the best approach.

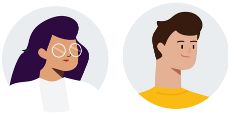
Furthermore, some applications require data in flight to be managed by some type of buffer and act with low latency as the data arrives. Data may be unstructured and may require calling external APIs in a programmatic way. For example, data in flight needs processing immediately, such as in an application that monitors and alerts on data usage for mobile networks. The data is streamed in real time, then an application built in Dataflow uses window aggregation functions and accumulates the amount of data used by the user while roaming. This is then actioned once the user reaches certain thresholds. The action in this example can be a simple text message alert to the user or an automated throttling of their data speeds.

Traditionally, data scientists either needed to choose between Spark and their favorite ML libraries or spend time setting up multiple environments. This has proven cumbersome and often repetitive. Anytime spent on environment configuration is better spent exploring interesting data instead. Administrators historically were able to provide users with ready to use environments but had little means to customize the managed environments based on specific users or groups of users. This led to unwanted costs and security management overhead. By combining those technologies in Dataproc Hub, we now offer a way for data scientists to swiftly select the Spark-based predefined environment that they need without having to understand all the possible configurations and required operations.

Painting the picture: Blended organization

In an ideal scenario, data engineers bring data in and focus on resolving data quality and data governance issues. For an evenly balanced organization (for instance, strong data engineering and data analyst capabilities), data engineers can create reusable data pipelines using either Data Fusion or Dataflow templates. The choice between tools depends on the use case, whether it is real-time or batch, and where the organization resides on the spectrum. The ideal solution is to have data engineers develop generic data pipelines that data analysts can “configure” by using parameters. This effectively leads to a repeatable process and more streamlined data operations. Depending on the use case, a rule engine on top of Data Fusion can be used by analysts. Consider a scenario where, for a marketing application, rules need to be entered into the system daily for thousands of campaigns. Data analysts use a rule engine interface to enter parameters and values to configure the application like they do in spreadsheets. They can also write transformations for individual parts of their work.

If your organization is coming from a more traditional data warehouse that relies on stored procedures and scripting, you need to think about how you can leverage your existing skill sets with a different data warehouse. The question that one may ask is, “Do I continue leveraging these skills and expertise and use these capabilities that are also provided in BigQuery?” Cloud data warehouses such as BigQuery expand on 40 years of development in traditional on-premise data warehouses and take it further by providing an architecture where storage and computation are separated, streaming of data is allowed, and machine learning is done within the data warehouse itself. As a result, ELT with BigQuery is more natural, continuing the trend with similar architectures to what is already in Teradata BTEQ, Oracle Exadata. This reduces the time for organizations to adapt to a new, modern architecture and enables them to exploit advanced capabilities through familiar languages such as SQL. Therefore, within a blended organization, data analysts rely heavily on ELT and exploit the capabilities of managed and capable services like BigQuery.



Machine learning and breaking silos

As companies become more data driven, they are organizationally more ready to adapt a data first solution to business problems such as improving customer experiences, preventing fraud, and increasing manufacturing efficiency. These can all be carried out in ways that make them act like a technology vendor and extend beyond their core business. This can be done with the help of Google's leading AI solutions portfolio at scale — scale of impact and scale of access across different functions within an organization, not just the number of AI experts and data scientists. Democratizing the business value of AI across an organization is a foundational principle of the Google Cloud.

For example, data scientists become another user of BigQuery by using notebooks and access data directly from the data warehouse. Data often does not live in isolation — it needs to be analyzed alongside other data with some context and then joined appropriately. For example, federating data access simply by running SQL across analytical store (BigQuery), relational store (Cloud SQL or Spanner), and key store (NoSQL DB BigTable) allows data scientists to focus most of their time on ML workloads. Hence, by bringing data in location to query and processing it with other data, you can extract more value from it. As a result, unified and integrated ML capabilities increase productivity. Powerful ML solutions such as BigQuery ML ensure that raw data is kept in an easily accessible location. Then, using ML routines within SQL, the ML framework is carried out by accessing enriched and transformed datasets, allowing data scientists to get to the raw data and automate processes. In other words, traditional enterprises can use existing frameworks and scales. On the other hand, by using the BigQuery Storage API, machine learning engineers can access the same data directly from their favorite ML frameworks such as PyTorch or TensorFlow.

Real-time and semi-structured data

Bringing semi-structured data, such as XML/JSON, has its issues. The data requires cleansing and enrichment, as it usually does not follow as strict a schema as a traditional data warehouse would use. Data can be brought with external APIs or the enrichment process may need to happen using data from external APIs before loading into the data warehouse.

Real-time analytics enables immediate responses, as there are specific use cases where low latency applications need to run. For example, anomaly detection applications that detect credit card fraud require identifying unusual usage patterns and responding with an action immediately upon detection (for instance, in this example, blocking the transactions). In other words, business requirements demand that data has to be acted upon as it arrives on the fly. Processing this type of data or application usually requires transformation done outside of the warehouse as the data coming from the system would be in its raw format to make sure that no further latency is introduced by the processing stages.

Conclusion

When undertaking a large digital transformation project, organizations frequently look at technical requirements that inform which architecture to implement. But a key and frequently overlooked component needed to truly become a data-driven organization is the impact of the architecture on your data users. When you take into account the responsibilities, skill sets, and trust of your data users, you can create the right data platform to meet the needs of your IT department and your business.

Google Cloud has the tools, services, and expertise in data analytics to help you decide which analytics data platform is right for your organization. The time for digital transformation is now, and Google Cloud is the data analytics partner to guide you through your journey. Contact us to get started.

When you take into account the responsibilities, skill sets, and trust of your data users, you can create the right data platform to meet the needs of your IT department as well as your business.

What type of data processing organization are you?

September 2021

Interested in getting started? Contact us to learn more.

Google Cloud