Google Cloud

The data journey is a flexible path

So far, you've learned that the data journey is a structured process that includes locating data and presenting data insights to stakeholders. The data journey enables organizations to harness the power of data, and convert raw information into actionable knowledge. In this reading, you'll learn how the steps in the data journey optimize data's value. You'll also review a business example of how data travels from collection to activation.

Data journey steps

Cloud analysts transform data into actionable knowledge through five important stages:

1. Collecting
2. Processing
3. Storing
4. Analyzing
5. Activating

Collecting

The purpose of the collection stage is to identify and gather relevant data based on business needs. Stakeholders generally provide data in multiple formats, like file extracts (text files, CSV files, etc.) from their own databases, spreadsheets, or survey results. As part of the collection process, you must ensure that teams source data from reliable sources. It's also important to understand how often those data sources are refreshed, or updated, to ensure you're working with the most up-to-date information. The quality of data collected determines the work required in later stages, and the final output in the activation stage.

Processing

The processing stage, also known as the transformation stage, involves cleaning, organizing, and standardizing the data. As an analyst, it's your job to ensure data is in a format that's usable and ready for the analysis stage. In this stage of the data journey, you'll handle data errors and inconsistencies like gaps, null values, and formatting issues. Depending on the type of data, processing might include data enrichment and anonymization when handling sensitive information, like personally identifiable information (PII) or protected health information (PHI).

Storing

After processing, you'll need to determine where and how the data should be stored. Data may be stored locally or in the cloud, so that it's ready to be used for analysis. As an analyst, you won't store all data in the same way, because different data types and use cases have different storage needs. When choosing an option, make sure to consider the security requirements, accessibility needs, and cost for storage.

Some organizations will choose to implement a multi-tier storage strategy, meaning their data will be stored based on how often it needs to be accessed. Frequently-used data can be stored in faster, costly storage, and legacy, or archival data can be stored in cheaper, slower options.

Analyzing

After processing, you'll uncover valuable insights that will drive decision-making in the analyzing stage. The insights gathered in the analyzing stage will later be visualized and presented in the next stage of the data journey. As an analyst, you'll use a variety of tools to apply different techniques, like predictive or text analysis to detect patterns and trends in the data. The analyzing stage is where you'll get to know your data, and determine what it's trying to tell you.

You may need to make sense of aggregations, plot data on a graph to understand how it's changed over time, or review text data to gain insight on consumer sentiment. For example, you might find an increase of sales year-over-year, or a decrease in user engagement on a company website.

Activating

In the activating stage, your goal is to equip stakeholders with the relevant information they need for informed decision-making. Insights from the analysis stage are presented visually to stakeholders in the form of reports, dashboards, and presentations. As an analyst, you'll illustrate your findings through storytelling, along with graphs, charts, and tables to set the scene for stakeholders. Stakeholders make decisions based on both the findings and their needs established during the collection stage.

It's important to note that activation isn't just about presenting; it's about enabling real-time action. Creating real-time dashboards and visualizations with up-to-date data from the cloud ensures stakeholders can make immediate decisions based on current data insights. The activation stage is where all your work is realized and applied. It's the insights presented in this stage that businesses need to make tangible decisions.

Data's non-linear path

Although experts lay out the data journey in a logical order, it doesn't always follow a straight line. As a reminder, the nature and source of the data can impact the data journey, meaning some stages are skipped or repeated. For example, some data might be pre-processed, skipping the collection phase. Or, analyzed data might return to the processing stage for refinement. And, the storage phase may be revisited when moving data from local storage or the cloud.

The data journey is flexible, so be prepared to be adaptable depending on the type of data you're working with, and stakeholder requests. Understanding the data journey's flexibility will help you reduce the unrealistic expectation that every stage will run smoothly every time.

Key takeaways

The data journey is a structured process that transforms raw data into actionable insights for stakeholders. Despite the logical sequence of the data journey, your progression as an analyst will not always be linear. Some stages may be repeated or skipped based on the nature of the data and the context of your work. Remember, the data journey is flexible, so you should be too.

Resources for more information

Use the following resources to explore data collection methods for different data types:

- 7 Data Collection Methods in Business Analytics provides a breakdown of data types, why data collection is important, what to know before you begin collection, and seven data collection methods to leverage from Harvard Business School