Building an Analytics and AI Architecture Using Google Cloud Platform

Published 10 January 2023 - ID G00765723 - 43 min read

By Analyst(s): Christopher Long, Prasad Pore

Initiatives: Analytics and Artificial Intelligence for Technical Professionals; Evolve Technology and Process Capabilities to Support D&A

Google Cloud Platform is emerging with innovative analytics and AI capabilities, but is it ready for your use case? Data and analytics technical professionals should evaluate the capabilities covered in this research to successfully build end-to-end analytics and AI architecture in Google.

Overview

Key Findings

- GCP simplifies the analytics architecture by providing distinct toolsets with minimal functional overlap. However, these services are still maturing, and clients are closely following Google's analytics and AI roadmap.
- Google is integrating distinct BI capabilities to enable enterprise-governed BI via Google Looker, self-service BI via Looker Studio, and spreadsheet experiences in Google Sheets based on the shared semantic models built in the Looker Modeling Service.
- Google Vertex AI provides a central platform to build, train and operationalize custom and AutoML models. However, some enterprise features, such as its integration with Google's Cloud Monitoring, are still in preview.
- The consolidation of Google Data Catalog into Dataplex delivers a single interface to organize and govern data. However, this is largely focused on data within Google Cloud Platform (GCP) and may have limited capabilities for incorporating information outside of GCP.

Recommendations

Data and analytics technical professionals who are considering using Google Cloud Platform-based solutions while modernizing their data analytics solutions should:

- Encourage analytics and BI developers to evaluate LookML and decide if the necessary upskilling provides business value and, for business users and developers, common, governed data through the Looker Modeling Service.
- Leverage Google's diverse machine learning options to augment structured (tabular) and unstructured data (text, video and image), and evaluate Vertex AI's capabilities to deliver value through low-code implementation.
- Use Google BigQuery's low-latency query performance, real-time streaming and in-database ML capabilities to deliver complex analytical workloads.

Analysis

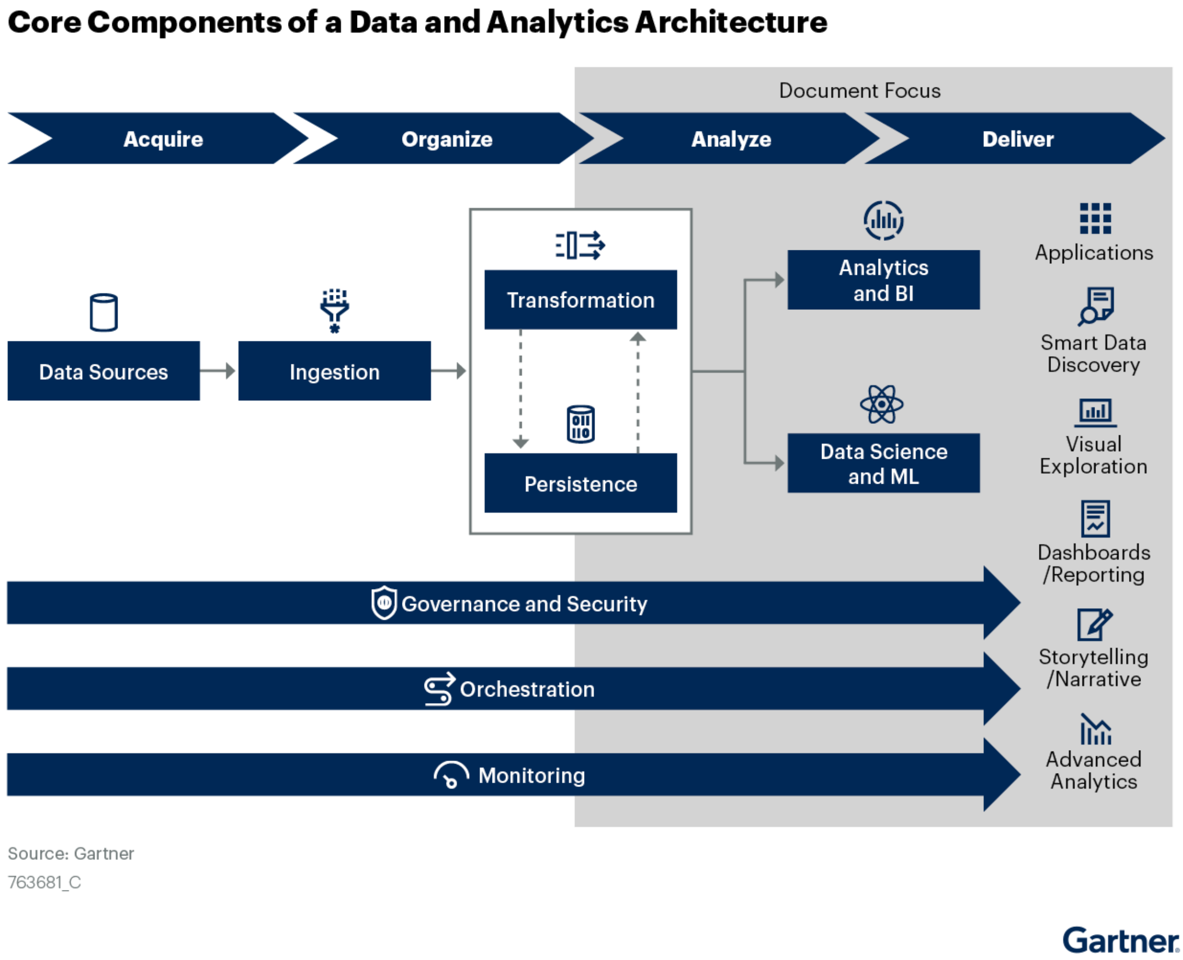
The cloud continues its forward momentum as the favored destination for deploying modern data and analytics (D&A) workloads, with many organizations already leveraging hybrid and multicloud architectures. Cloud providers offer continuous innovation at a pace impossible to match in traditional on-premises environments. At the same time, more analytics tool providers are arising as cloud-only solutions to compete with traditional reporting and business intelligence (BI) solutions.

> **74% of organizations use or plan to use cloud for analytics, BI and data science.**
>
> — *2020 Gartner Cloud Data and Analytics Survey*

This research provides an overview of the capabilities within the analytics and AI areas of the Google Cloud Platform (GCP). It is the second part of a combined research collection to outline GCP's data and analytics capabilities and how they can fulfill the needs of organizations today. The companion document, Building a Data Management Architecture in Google Cloud Platform, provides technical professionals with the overview of GCP's data management capabilities and their functionality — each category of the pipeline often has multiple offerings.

To simplify the complexity of data and analytics workloads, Gartner has disaggregated the overall D&A architecture into nine sections, as shown in Figure 1.

Figure 1. Core Components of a Data and Analytics Architecture



The building blocks of the modern D&A architecture include ingestion, data transformation, persistence and analytics layers, with governance and security, orchestration, and monitoring layers cutting across the pipeline.

Although this research is focused on native Google Cloud analytics offerings, some organizations may prefer to use non-native products for several reasons:

- Best-of-breed options may provide deeper functionality or certain domain-specific features that general-purpose tools may lack.
- Organizations with footprints in other public clouds or on-premises that they want to reuse may want to maintain cloud vendor neutrality.

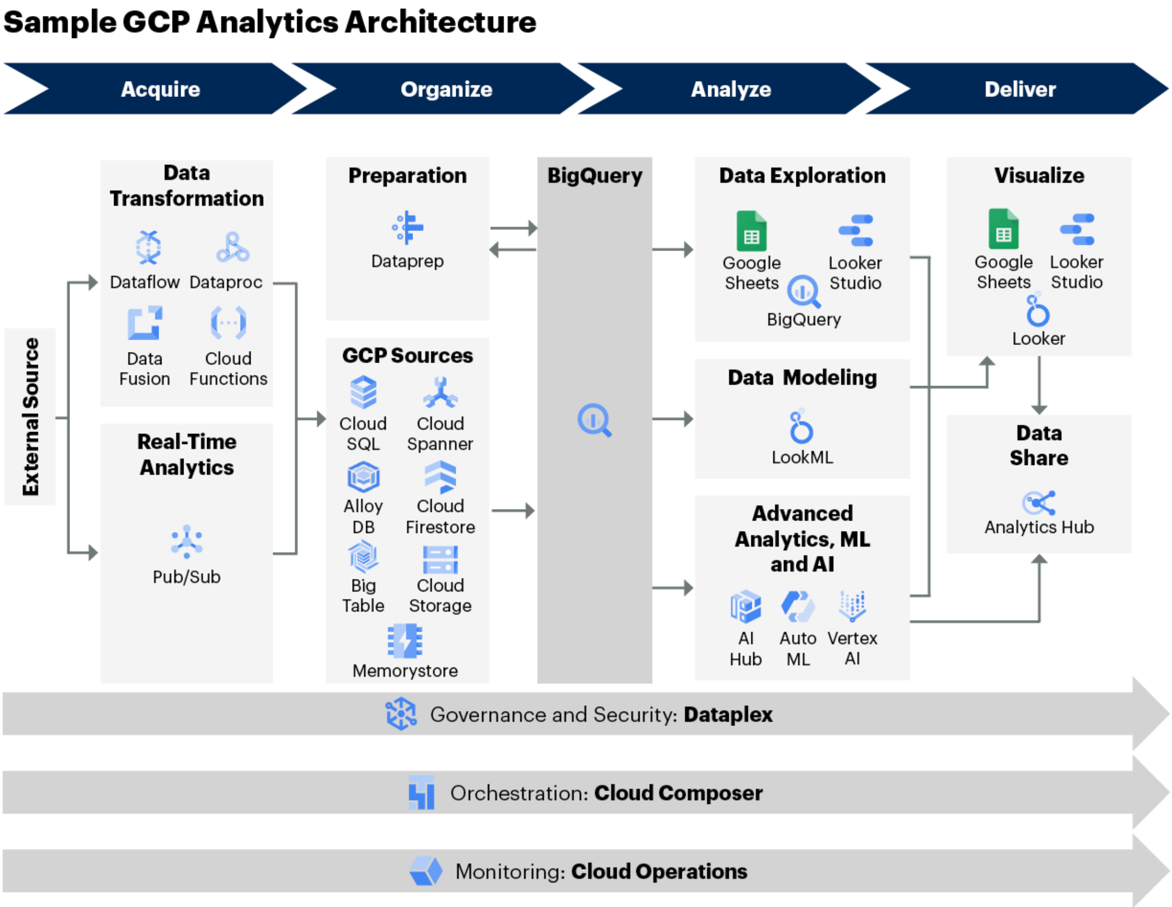
- Some organizations may prefer to take advantage of incumbent tools with existing skills, licensing and vendor relations. D&A technical professionals look to lift and shift existing on-premises niche and business-centric tools when they move to one or more public cloud vendors to reduce training and hiring costs and complexity.

Additionally, in 2022, Google joined multiple data and analytics providers to launch the Data Cloud Alliance. The purpose of this alliance is to make data more transferable and accessible across business systems, platforms and environments. This research does not compare Google's features or functionality with non-native products or services (including impacts from the Data Cloud Alliance), but this type of comparison will often be necessary to select the right tool for your organization's needs.

Each of the core components in Figure 1 represents a complex subecosystem across data and analytics. This research explores the corresponding native product and service offerings within Google as they relate to the analytics life cycle.

Figure 2 shows a curated list of popular components of Google Cloud analytics offerings that can be chosen to deliver an analytical solution. It is not a representation of Google's complete breadth of offerings. This research explains the analytics capabilities of Google's analytics products in terms of data preparation, analysis, delivery and exploration, governance, orchestration, and monitoring, and will provide insight to when tools may be used within your organization's analytics architecture.

Figure 2. Sample GCP Analytics Architecture



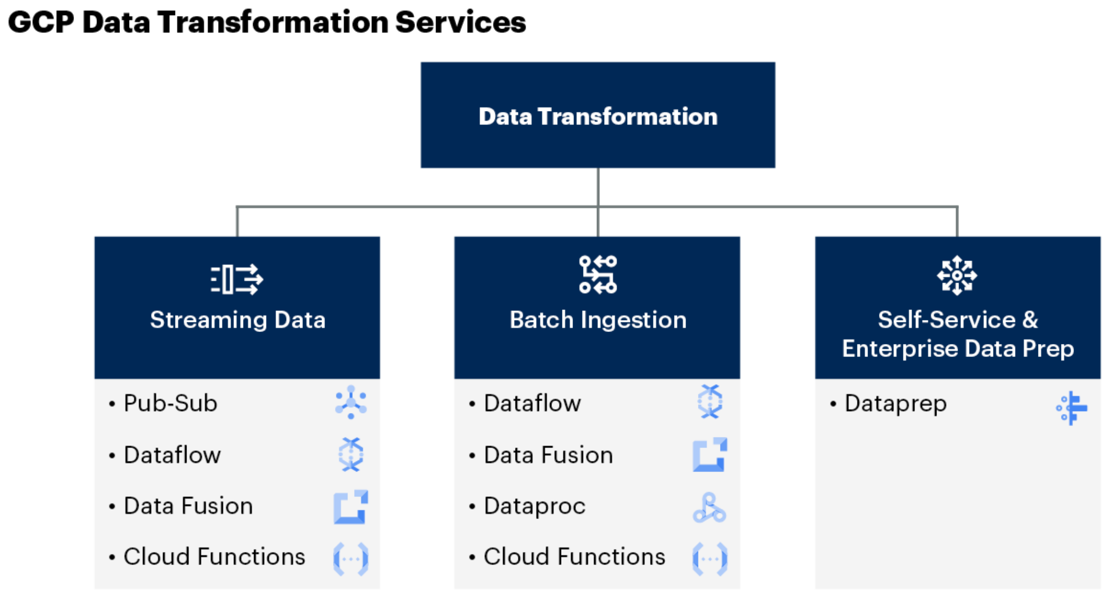
Source: Gartner
765723_C

Acquire and Organize

Google's analytics products and services can connect to a wide range of data sources. This research focuses on preparation and analysis of data from Google-based services. The related research, Building a Data Management Architecture in Google Cloud Platform, identifies the tools and services used to acquire, ingest and persist data within Google Cloud.

Figure 3 outlines the tools available in Google Cloud to ingest and transform data for persistence.

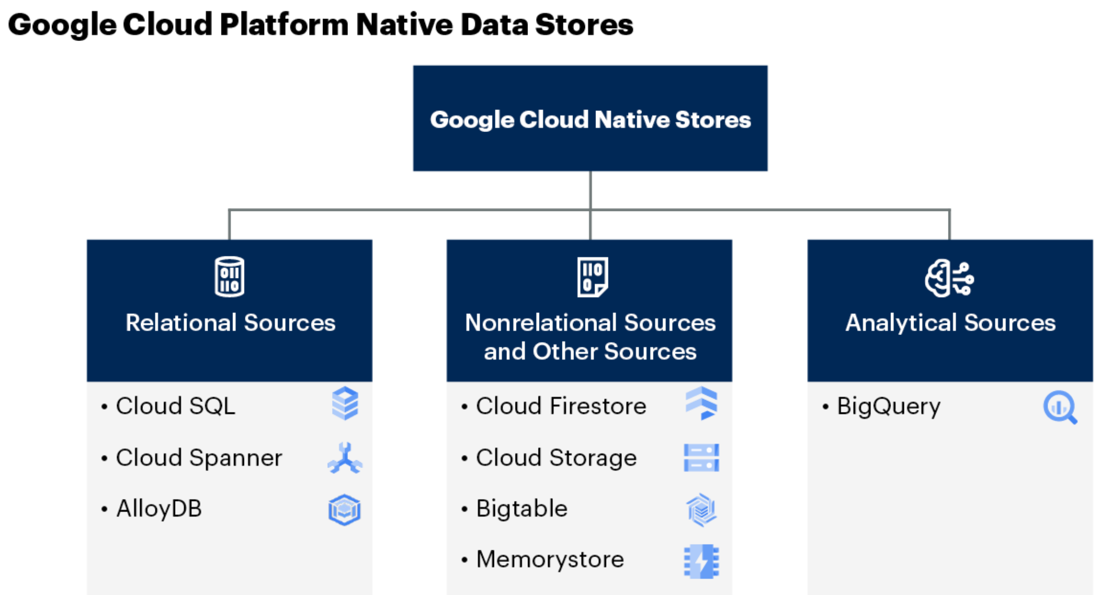
Figure 3. GCP Data Transformation Services



Source: Gartner
765723_C

Figure 4 outlines the native Google Cloud data stores.

Figure 4. Google Cloud Platform Native Data Stores



Source: Gartner
765723_C

This research does not detail the differences or use cases across these transformation and persistence services, but it does note them throughout. See the companion research, Building a Data Management Architecture in Google Cloud Platform, for further details on these services.

Analyze

Google provides multiple ways to analyze data in the Google Cloud Platform. Each of these services serve specific data analysis needs. At a high level, the analytics options include:

- **Analytical data modeling:** Dataprep by Trifacta, BigQuery and Looker Modeling Service
- **Machine learning and AI:** BigQuery ML, Google Cloud ML and Vertex AI
- **Data exploration:** BigQuery search indexes and Dataplex (Explore)

Data Preparation

Dataprep

Business analysts and data scientists spend a majority of their time in preparing the data, commonly known as "data wrangling," rather than performing the necessary analytics. This has led to the creation of self-service data preparation solutions.

Dataprep, based on Trifacta Wrangler Enterprise software, is a serverless solution that provides an intelligent data service to visually explore, clean and prepare structured and unstructured data for analysis and machine learning. It leverages Dataflow as its execution engine and scales on demand to meet data preparation needs.

The Dataprep service automatically detects schemas, data types and possible joins along with identifying anomalies (such as missing values, type mismatches, outliers and duplicates). The inclusion of AI-driven functions throughout Dataprep provides data engineers and self-service users with suggested transformations. These suggested transformation examples provide the capability for users to quickly identify, analyze for purpose, and incorporate appropriate transformation steps. Tools like RapidTarget also provide automated mapping of data to existing schema in BigQuery or Google Cloud Storage (see Trifacta for Data Engineers: Introducing RapidTarget).

By automating and suggesting common transformation tasks, analytics developers can reduce the time-consuming work of cleaning up messy data and move on to the data analysis faster. Further, the flows created within Dataprep may be orchestrated using Google Cloud Composer to schedule data wrangling pipelines.

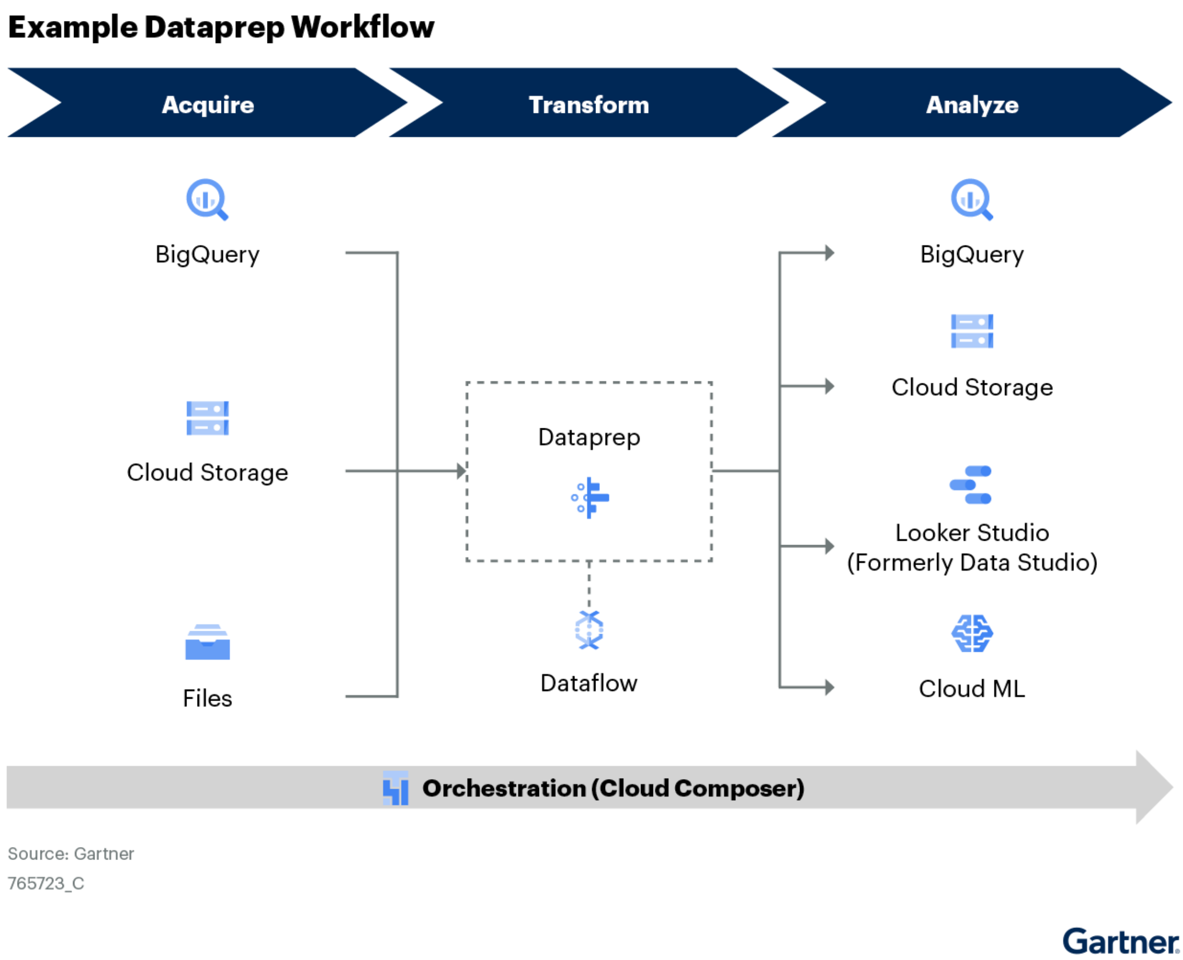
Based on the version of Dataprep by Trifacta implemented, users may find certain features are unsupported, including:

- Sort transformations
- User-defined functions
- Custom data types
- Collaborative suggestions
- Limited access to APIs

Developers looking to identify both the version of Dataprep used and related limitations should review Product Limitations.

Figure 5 demonstrates an example workflow of Dataprep used to transform and prepare data for analytical use.

Figure 5. Example Dataprep Workflow



In February 2022, analytics data preparation company Alteryx completed its acquisition of Trifacta. Although Google continues to include Dataprep by Trifacta as part of its ecosystem, the long-term future of this service and support as a native GCP capability is unknown. Additionally, Google does not have (at time of publication) a planned solution to replace Dataprep with a new business-user-focused analytics data preparation tool. Therefore, being the primary user-focused, self-service data preparation tool available natively on GCP, a loss of this product will open a gap in the end-to-end analytics delivery capability.

With this in mind, the following tools are available today for analytics data preparation (including for self-service):

- **Dataprep by Trifacta:** This is Google's primary service for low-code, self-service, visual data preparation. However, as noted, consumers should monitor Google's roadmap for support and be prepared to find alternatives.

- **Connected Sheets:** The use of spreadsheets to prepare data is common across industries for business analysts. However, while Google Sheets may fill this common task through Connected Sheets, it is not scalable or operationalizable at an enterprise level.
- **Cloud Data Fusion:** While this is a low-code data transformation/preparation tool, it is not designed for business users. The primary audience for this tool remains data engineers and may not represent a self-service alternative to Dataprep.

Data Querying

BigQuery

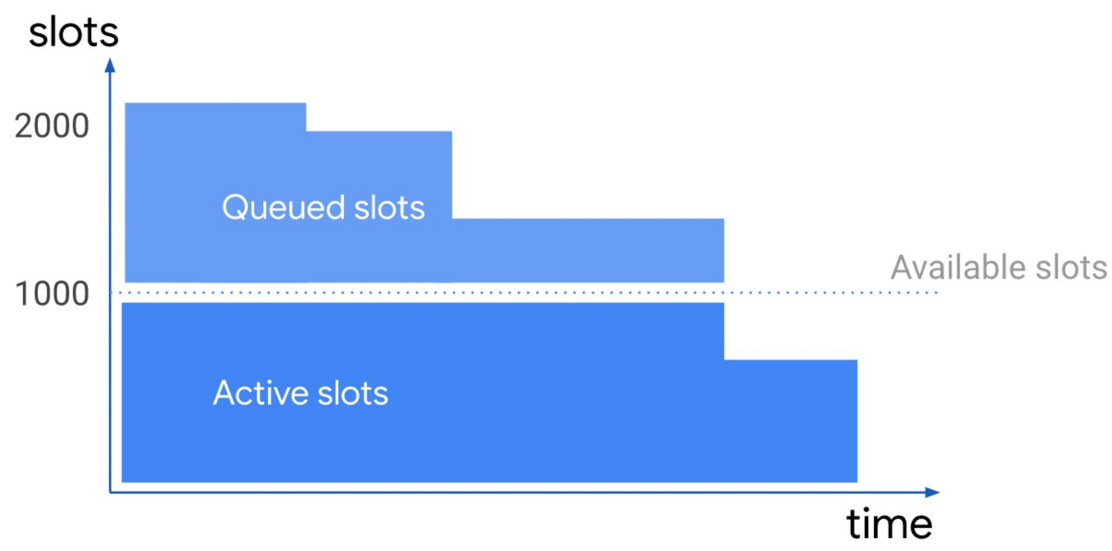
BigQuery is Google's fully managed, serverless data warehouse and analytical engine. This service uses SQL to execute queries via a distributed analysis engine. The core of BigQuery's strength comes from its separation of storage and compute. By separating, these capabilities may be managed and scaled independently, based on demand.

The performance of query jobs in BigQuery and concurrency of processing are dependent on the reservation of "slots" (that is, virtual CPUs). A mix of on-demand and flat-rate pricing slot reservation models provides the means to allocate capacity across the organization. Based on the model used, slot resources are allocated to submitted queries.

Where slot demand exceeds current availability, additional slots are queued and held for processing once capacity is available. This processing model allows for continued processing of concurrent large query workloads.

Figure 6 shows slots over time and queued for processing.

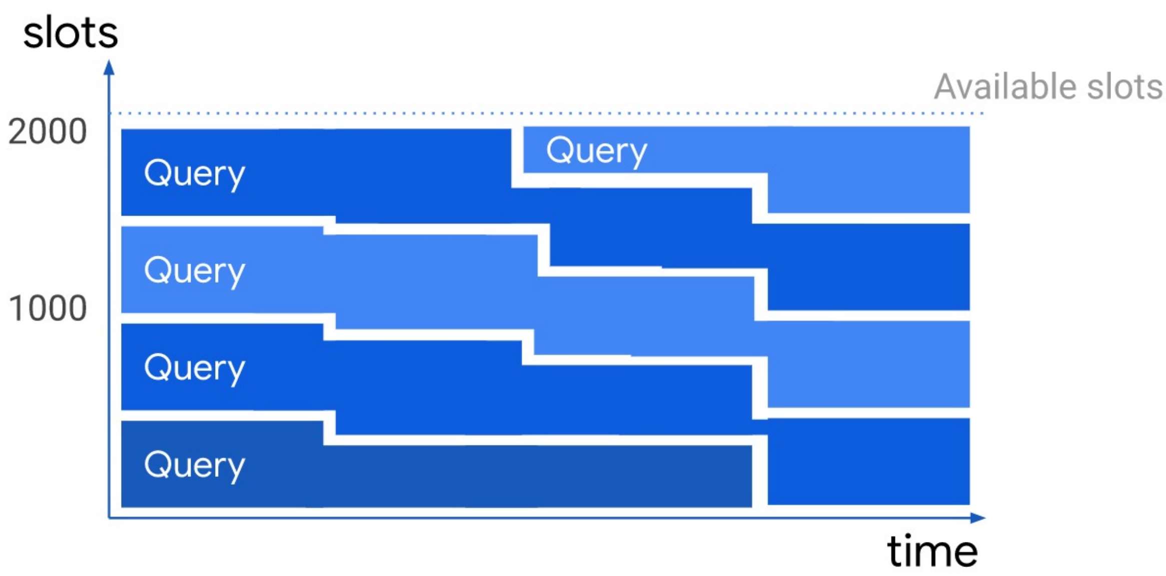
Figure 6. Google BigQuery Slot Reservations



Source: Google

Google's query engine dynamically distributes slots fairly across jobs within projects. Figure 7 demonstrates slots as allocated across multiple query workloads over time.

Figure 7. Google Slot Scheduling



Source: Google

Several analytics features in BigQuery include:

- **Ad hoc analysis:** BigQuery supports ad hoc analysis using a SQL dialect known as Google Standard SQL.
- **Geospatial analysis:** Analyze and visualize geospatial data in BigQuery by using geography data types and Google Standard SQL geography functions.
- **Machine learning:** BigQuery ML lets you create and execute machine learning models in BigQuery by using Google Standard SQL queries.
- **Business intelligence:** BigQuery BI Engine is a fast, in-memory analysis service.

BigQuery BI Engine

In addition to the distributed processing of queries in BigQuery, Google offers the BigQuery BI Engine as an add-on service that provides accelerated, in-memory data analysis. The BigQuery BI Engine uses cached tables and automated memory management to deliver subsecond response times to processes where the BI Engine is engaged. The BI Engine SQL interface also allows for integration with third-party BI tools such as Tableau Software and Microsoft Power BI. However, note that this enhanced query service may include additional costs and time to spin up prior to query execution.

So while processing may be accelerated once running, engaging the BigQuery BI Engine is not immediate. Planning is required when this feature is desired.

For further details on BigQuery as a persistent data warehouse, see the companion document, Building a Data Management Architecture in Google Cloud Platform.

Analytical Data Modeling

Business users want self-service analytics by accessing a semantic layer to visualize and report on the data. The purpose of analytical data models (semantic layer) is to represent corporate data using simple business terms like “customer” or “account” in place of cryptic column names.

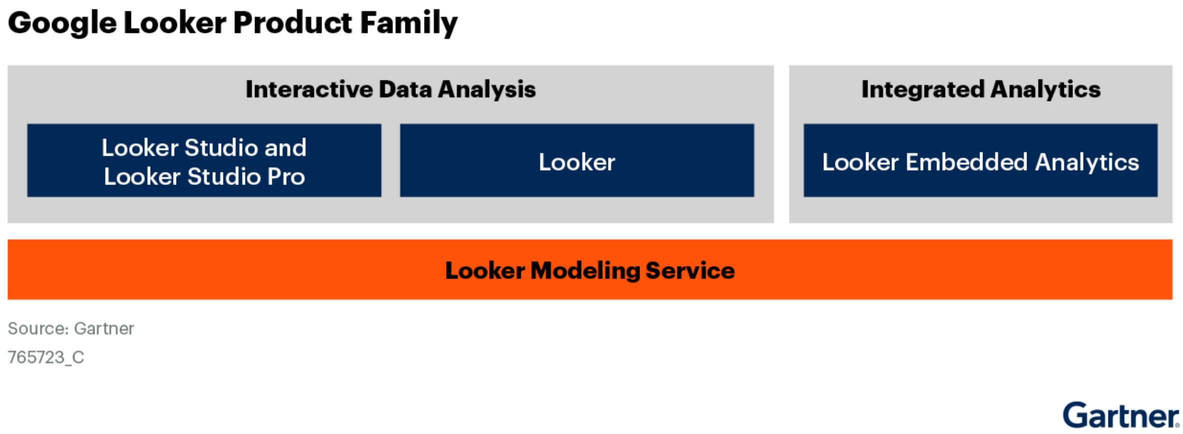
As detailed in Demystifying Semantic Layers for Self-Service Analytics, semantic layers perform the following core functions:

- Translate the underlying database structures into business-user-oriented terms and constructs, which are then intuitive to business users.
- Provide a mechanism to define, and store calculations and business rules.

Looker Product Family

In October 2022, Google announced the Looker product family, which is a rebranded combination of Looker (originally acquired in February 2020), LookML and Data Studio. Figure 8 shows these products combined for analytical delivery, now sharing the same underlying modeling platform.

Figure 8. Google Looker Product Family



With this announcement, Data Studio is now Looker Studio. Looker Studio still offers the free user version and now also offers Looker Studio Pro, a paid tier that includes enhanced enterprise capabilities and technical support. The interactive reporting tool capability of Looker remains, as do the features of Looker Embedded Analytics. These analytical delivery tools are now underpinned by the Looker Modeling Service. Further details on these data analysis tools can be found in the Analytics Delivery and Exploration section of this document.

Looker Modeling Service

The Looker Modeling Service underpins Google's analytics delivery tools. This service delivers a semantic layer written in LookML, which is Google's declarative language for describing dimensions, aggregates, calculations and relationships in a SQL database. The LookML modeling layer is made up of a collection of projects that include models, views and Explores. Each of these objects has a role in defining the dimensions, aggregates, calculations and relationships of the underlying database.

Table 1 below describes objects included in the LookML modeling layer.

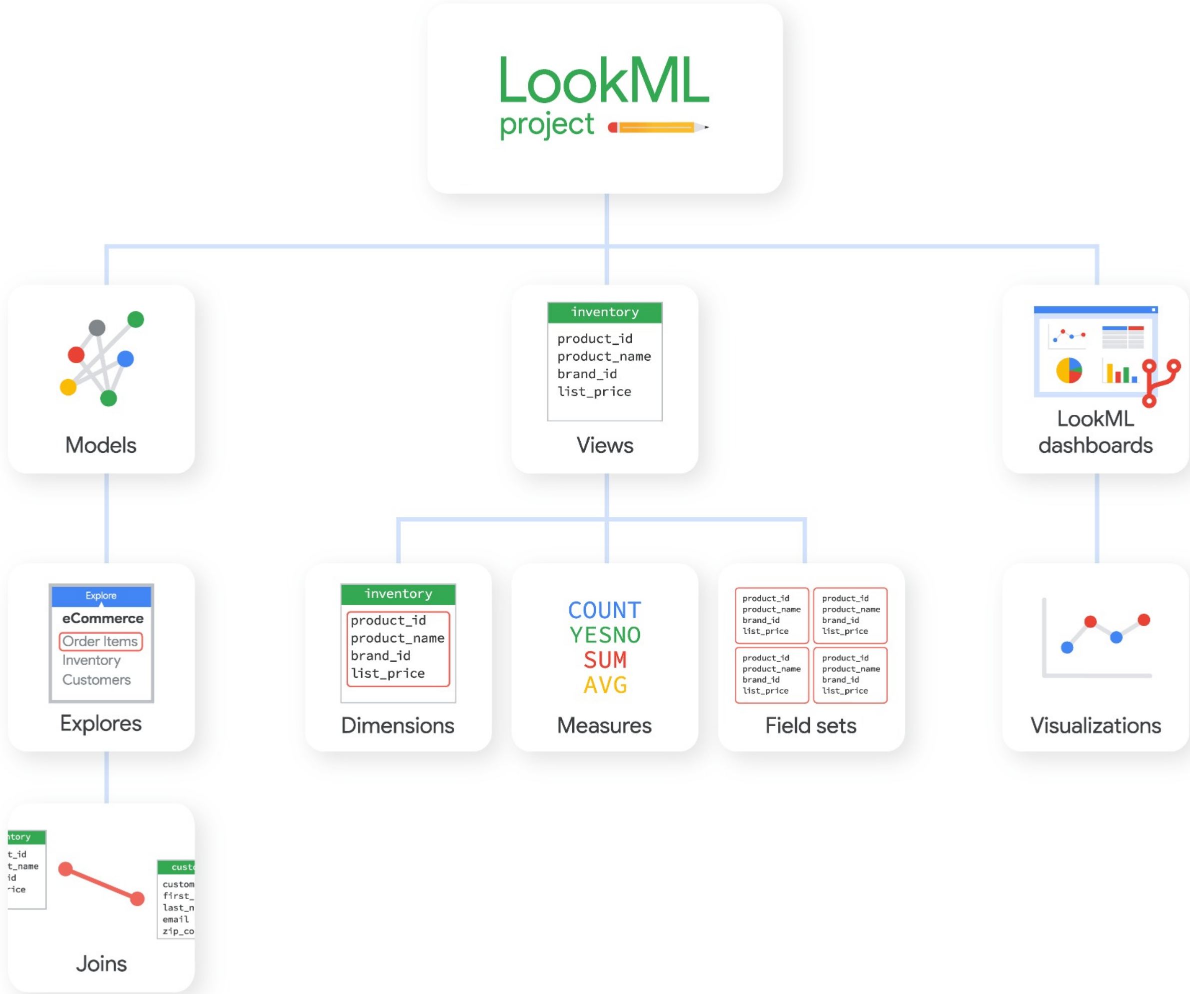
Table 1: Primary Looker Modeling Service Components

Object ↓	*Description* ↓
Project	Projects are the containers that collect your files that describe tables and relationships and that define how Looker controls the user interface's behavior. These files principally include models and views.
Model	Models hold information about which tables are included and their relationships. Connections defined in the model also define Explores for users to query on.
View	Views define the fields (measures and dimensions) and their connection to underlying tables. Views are typically defined as part of the project Explores.
Explore	Explores are the starting point for user queries. For consumers familiar with SQL, this is comparable to the "FROM" SQL statement

Source: Gartner (January 2023)

To put all these elements together, Figure 9 provides an overall view of a LookML project.

Figure 9. LookML Core Elements



Source: Google

These objects ultimately separate out data structure from the SQL query, and when combined through the Looker query builder, Looker's SQL generator translates and sends the resulting query to the data store. One advantage of this approach is that organizations can create a common data model using the LookML modeling layer, which abstracts individual data models. Additionally, the Looker family of tools (see Figure 8) now share this semantic platform to deliver consistent data across use cases.

Gartner research and client inquiry has also identified the following concerns to the Looker modeling layer:

- LookML adds a new language to be learned and can be challenging to find when staffing for development.
- LookML is less focused on self-service business users.

- LookML models primarily source structured SQL-type data.

This semantic layer can evolve and be versioned inside Looker. Objects within these collections are generally version-controlled via Git and stored as files in GitHub. However, Looker may also be configured to use other Git providers, including GitLab, Atlassian (Bitbucket) or other Git servers utilizing SSH keys for authentication.

Overall, the separation of the Looker Modeling Service from the delivery tools positions LookML to be used as a headless analytics governance tool. As a headless BI layer (or a metrics store), LookML can be accessible via APIs from third-party data visualization products. Google previously announced (in October 2021) its partnership with Salesforce (Tableau) to provide native connectivity from LookML to Tableau, combining the development and governance of complex data with the visual explorative capabilities within Tableau. Google is also planning to expand this capability to other third-party analytic and BI platforms, including Power BI (in preview at time of publication). For additional details on how headless BI or metrics stores impact analytics architectures, see Gartner's Video: Demystifying the Metrics Store.

Machine Learning and AI

The data science and machine learning (DSML) market is fast-growing, and businesses continue to find new and expanded needs for the analyses delivered by data scientists and ML professionals. Google Cloud offers multiple options for incorporating artificial intelligence (AI) and machine learning (ML) into your analytics and application workflows, including pretrained models, APIs and intelligent AI agents to deliver business value.

Table 2 identifies pretrained models/APIs and AI agents that can be used without significant ML expertise.

Table 2: Google Cloud AI Agents and Pretrained Models/AI Capabilities

(Enlarged table in Appendix)

AI Agents		*Example Use Case*
Contact Center AI (powered by Dialogflow CX)		Prebuilt customer service AI solution
Document AI		Document image quality detection, classification, labeling and text extraction
Discovery AI (includes Retail AI and Recommendations AI)		Deliver tailored recommendations to specific customer preferences
Dialogflow CX		Build agents to maintain conversations through text and audio with end users
Healthcare Insights AI (includes Google Cloud Healthcare Data Engine, Cloud Healthcare API and Cloud Natural Language API)		Deliver forecasting models to understand the progression of infection across populations
Pretrained Models and APIs		*Example Use Case*
Vision	Vision AI	Detect and extract text from images or properties for further classification
	Video AI	Moderate and automate captioning to streaming videos
Language	Translation AI	Provide text translation in application workflow for service agents receiving requests in multiple languages
	Natural Language AI	Identify entities (phrases) in text and perform sentiment analysis from social media streams
Conversation	Speech-to-Text AI	Convert voicemail recordings to text files for documentation
	Text-to-Speech AI	Provide audio response to application users upon task completion
Structured Data	Timeseries Insights API, Vertex AI Forecast, AutoML Tables, TabNet and Cloud Fleet Routing API	Develop predictions for daily demand on products or services

Source: Gartner (January 2023)

Vertex AI

In addition to Google's pretrained AI capabilities, Vertex AI provides a managed machine learning platform that supports the development and deployment of artificial intelligence models, whether delivered from low-code AutoML or code-based ML models. This unified platform serves to simplify the challenges faced by data scientists and developers to develop, train, deploy and maintain models.

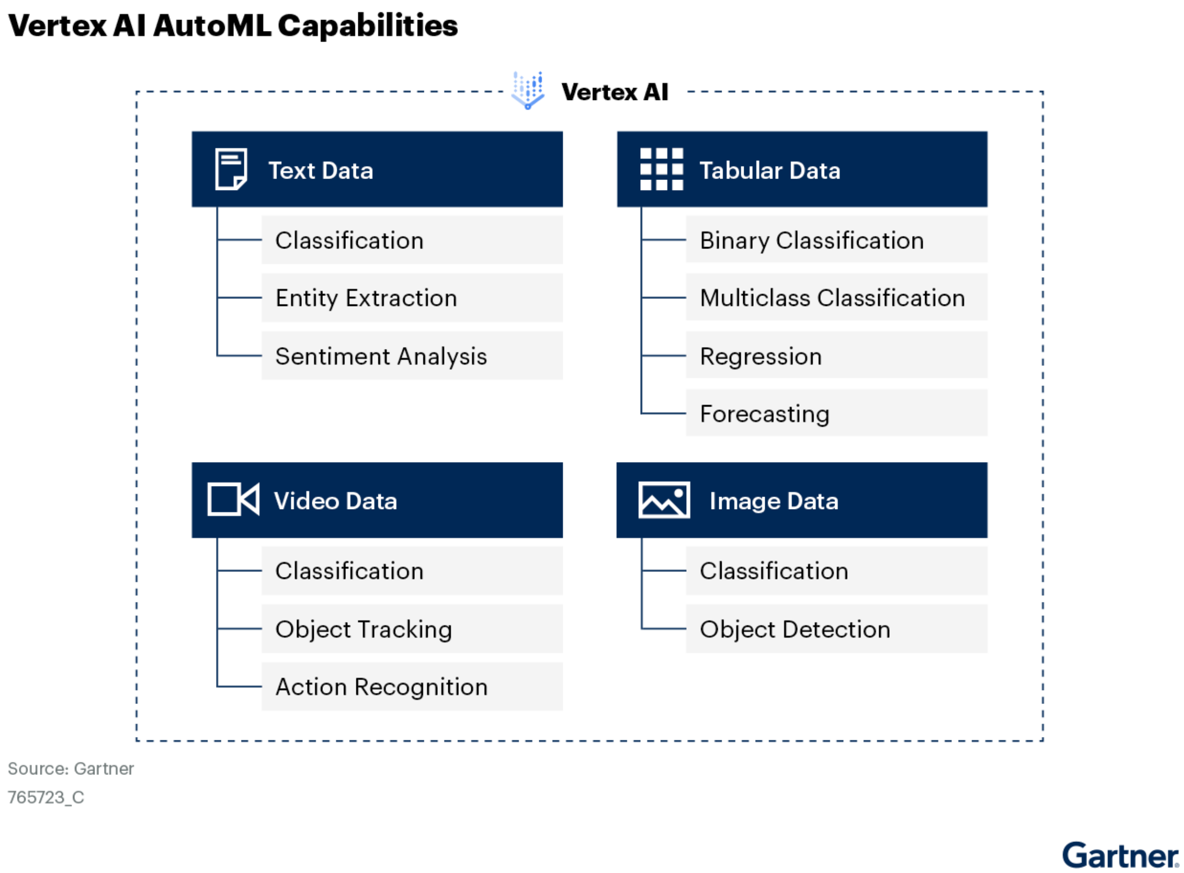
Along with training models through AutoML or code-based training, there are several notable features of Vertex AI:

- **Vertex AI Data Labeling:** Request mechanism for human labeling in a dataset to train custom machine learning models

- **Vertex AI Feature Store:** Fully managed repository to ingest, serve and share ML features
- **Vertex AI Workbench:** A Project Jupyter notebook-based development environment with capabilities encapsulating the full data science workflow

The use of AutoML in Vertex AI allows developers to incorporate machine learning without data science expertise. Figure 10 Identifies Vertex AI's AutoML available capabilities.

Figure 10. Vertex AI AutoML Capabilities



The Vertex AI AutoML capabilities outlined above may fit many common use cases. However, there are limitations to the control data developers have over the training environment and no manual customization to hyperparameters is possible. If these are concerns in your use case, consider code-based training in Vertex AI. Additionally, users considering the Vertex AI AutoML capabilities should note the following limitations when preparing data in Table 3.

Table 3: Vertex AI AutoML Data Limitations

Text Data	*Video Data*	*Image Data*
Minimum training docs: 20 Max training docs: 1 million Min categorization labels: 2 Max categorization labels: 5,000	File Types: MOV, MPEG4, MP4, AVI Max size: 50 gigabytes (3 hours) Max categorization labels: 1,000	Training — max file size 30MB (JPEG, GIF, PNG, BMP, ICO) Prediction — max file size 1.5MB (JPEG, GIF, PNG, WebP, BMP, TIFF, ICO)
Note that these limitations may change as Vertex AI evolves. Users should check Google's documentation for updated information.		

Source: Gartner (January 2023)

For a more detailed assessment of Vertex AI, see Solution Comparison for Cloud Data Science and Machine Learning Platforms.

Additionally, any organization seeking to leverage AI in its analytics stack should consider the exposure to algorithmic risks, regulatory scrutiny, bias and lack of transparency. Explainable AI is the concept of providing explanations and transparency into why ML models are making certain predictions. Vertex Explainable AI provides insight into how much each feature in the data contributed to the predicted result. This transparency helps users verify that the model is behaving as expected.

Table 4 outlines the types of models Vertex Explainable AI is available for and how the feature is enabled.

Table 4: Vertex Explainable AI Use Cases

Model Type	Data Type	Enabling
AutoML	Image (classification only)	Must explicitly enable during training
AutoML	Tabular (classification and regression)	Enabled by default
Custom-trained	Image	Configure explanations when building or deploying model
Custom-trained	Tabular	Configure explanations when building or deploying model

Source: Gartner (January 2023)

Consumers implementing Vertex Explainable AI should consider each data type, the attribution method applied and how they may impact your prediction results. Table 5 outlines the three methods used (*sampled Shapley, integrated gradients and XRAI*), recommended model types and use cases.

Table 5: Vertex Explainable AI Feature Attribution Methods

(Enlarged table in Appendix)

Method	Basic Explanation	Recommended Model Types	Example Use Cases	Compatible Vertex AI Model Resources
Sampled Shapley	Assigns credit for the outcome to each feature and considers different permutations of the features; provides a sampling approximation of exact Shapley values	Nondifferentiable models, such as ensembles of trees and neural networks	Classification and regression on tabular data	Custom-trained models (any prediction container) AutoML tabular models
Integrated gradients	A gradients-based method to efficiently compute feature attributions with the same axiomatic properties as the Shapley value	Differentiable models, such as neural networks; recommended especially for models with large feature spaces Recommended for low-contrast images, such as X-rays	Classification and regression on tabular data Classification on image data	Custom-trained TensorFlow models that use a TensorFlow prebuilt container to serve predictions AutoML image models
XRAI (eXplanation with Ranked Area Integrals)	Based on the integrated gradients method; assesses overlapping regions of the image to create a saliency map, which highlights relevant regions of the image rather than pixels	Models that accept image inputs; recommended especially for natural images, which are any real-world scenes that contain multiple objects	Classification on image data	Custom-trained TensorFlow models that use a TensorFlow prebuilt container to serve predictions AutoML image models

Source: Compare Feature Attribution Method, Google

Those looking to understand more about explainable AI should see Incorporate Explainability and Fairness Within the AI Platform.

BigQuery ML

Complimenting Vertex AI is BigQuery ML. This additional ML capability enables data analysts and data scientists to build and run machine learning models in BigQuery using SQL on structured data. By building models using existing SQL tools and skills, implementation of ML models is broadened to a range of users outside of traditional data scientists, such as citizen data scientists. BigQuery ML also recently announced support for unstructured data, which will enable users to perform inference using Google intellectual property (IP) in speech recognition, vision, translation and text processing using SQL.

Since BigQuery ML's release in 2018, Google has expanded its supported model types to include:

- Regression (linear, binary logistic and multiclass logistic)
- K-means clustering
- Matrix factorization
- Time series
- Boosted tree
- Random forests
- Deep neural network (DNN)
- Wide-and-deep models
- AutoML Tables
- TensorFlow model importing
- Autoencoder
- Principal component analysis

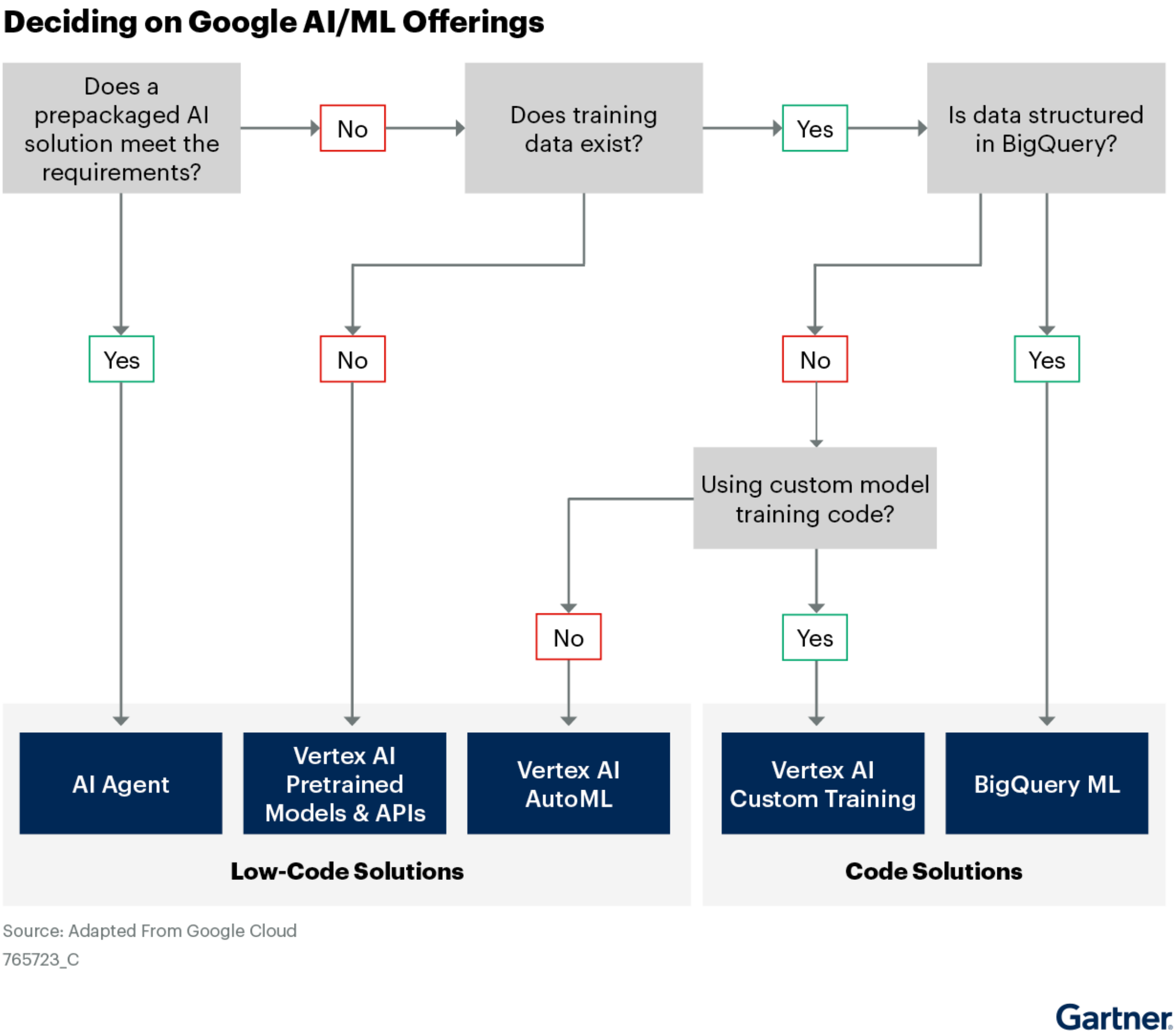
BigQuery ML has notable benefits for those looking to implement ML capabilities, including:

- **Use of existing skill sets:** BigQuery ML does not require knowledge of programming languages like Python and allows for traditional SQL developers to bring ML into their analyses.

- **Data stays in place:** BigQuery ML does not require data to be exported from the data warehouse to be analyzed. This can speed up time to insights by bypassing steps to move data into other ML frameworks.
- **Vertex AI integration:** Models trained in BigQuery ML can be registered and deployed in Vertex AI, allowing for a scalable MLOps experience.
- **Built-in explainable AI:** Native support of Explainable AI in both time series and non-time series models helps customers understand the results that predictive models generate for classification and regression tasks.

Analytics developers have multiple options when implementing machine learning and AI. Figure 11 outlines a decision path for identifying which capability may fit your use case, whether through prepackaged AI solutions, pretrained API models, Vertex AI (via AutoML or code-based training) or SQL-based BigQuery ML.

Figure 11. Deciding on Google AI/ML Offerings



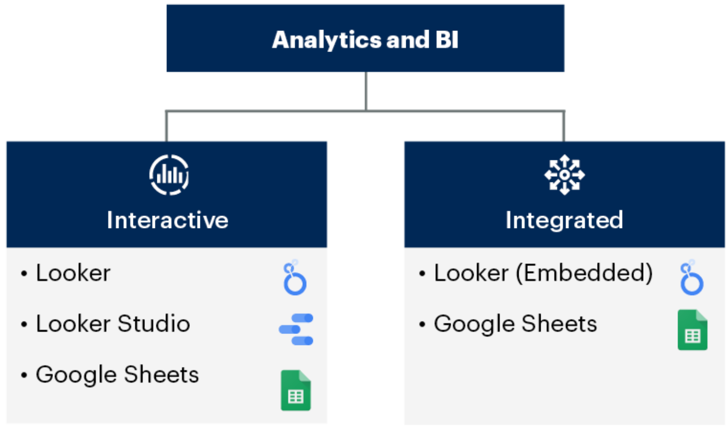
Analytics Delivery and Exploration

With Google's latest announcement (detailed in Looker Product Family), the Looker name now encapsulates Google's full suite of interactive and integrated analytics offerings. With this advancement, Looker, Looker Studio and Looker Embedded share access to the same governed base with the Looker Modeling Service.

Figure 12 displays the tools available for delivering analytics and BI to the enterprise.

Figure 12. Google Analytics Delivery Offerings

Google Analytics Delivery Offerings



Source: Gartner
765723_C

Gartner

Interactive Analytics and BI

Looker

Looker is Google's tool for delivering enterprise analytics to the organization through a browser-based UI. Analytics are delivered to users through three primary means:

- Web interface
- Scheduled delivery
- Looker Embedded Analytics

Web Interface

Looker's web interface exposes data to end users under the Looker label by sharing any of the following artifacts:

- **Explores:** The starting point for queries, this artifact exposes the fields incorporated into the resulting view. These may be used as building blocks for Looks.
- **Looks:** A stand-alone report or visualization, these may be used as components to build a dashboard.
- **Dashboards:** These are a collection of queries combined as visualizations or text tiles on a page

Scheduled Delivery

Looker Scheduler allows analysts and developers to send data to a large number of productivity and SaaS tools, including Microsoft Excel or Google Sheets, either on-demand or on a scheduled or one-time basis. Standard delivery targets include email, webhook, Amazon Simple Storage Service (Amazon S3) buckets, and SFTP servers.

Looker Embedded Analytics

Looker Embedded Analytics offers APIs that application developers use to embed visual analytics and reports (Explores, Looks and dashboards) in their business applications and dashboards. Embedding may be used to deliver insights to your organization's customers or to internal users.

Google provides three primary ways to embed Looker into business processes:

- **iFrame embedding:** Use of standard HTML iframes to insert Looker dashboards into organization's web applications
- **Looker extensibility:** Extending Looker through scripting hosted in Looker, thus creating mini applications
- **Application APIs:** Full-stack application development utilizing API code to interact with and expose analytics from Looker

Additional assessment of the Looker platform can also be found in Gartner's Critical Capabilities for Analytics and Business Intelligence Platforms.

Looker Studio (Formerly Data Studio)

Looker Studio initially started as Data Studio as part of the Google Marketing Platform. It has connectivity to Google Cloud Platform's Looker Modeling Service, BigQuery, Cloud Storage, Cloud Spanner and Cloud SQL, MySQL, and over 800-plus sources via partner-built data connectors. Looker Studio complements Looker as a self-service business intelligence and analytics tool designed to support ad hoc reporting and analysis across your data landscape.

Although there are overlapping capabilities between Looker and Looker Studio, these tools may be used in concert to provide for various enterprise and self-service analytics use cases. Figure 13 highlights the capability differences between these products.

Figure 13. Google Looker vs. Looker Studio Use Cases

Google Looker vs. Looker Studio Use Cases

Google Looker	Feature	Looker Studio
✓	Basic Data Modeling	✓
✓	Advanced Data Modeling	✗
✓	Predictive Analytics	✗
✓	Embedded Analytics	✓[a]
✓	API Access	✓[a]
✓	Custom Visualizations	✓
✗	Ad Hoc Data Acquisition	✓

Source: Gartner
[a] Indicates limited capabilities
765723_C

As Figure 13 indicates, there are subtle differences between Looker and Looker Studio. A principle difference between these is Looker Studio's capability for ad hoc data acquisition and subsequent analysis. Of Google's offerings for analytics delivery, Looker Studio is the tool most preferred by self-service analysts looking to connect to and explore data.

Connected Sheets

Connected Sheets allows analysts to leverage the familiar spreadsheet UI of Google Sheets. This capability enables these users by providing no-code capabilities to access Looker data models and BigQuery to explore data and develop new insights using standard worksheet functions like pivot tables, charts and functions.

Table 6 summarizes Google Cloud's analytics and BI offerings and example use cases.

Table 6: Google Interactive Analytics and BI Use Cases

Product	Function	Example
Looker	Deliver consistent, governed models and metrics to end users for enterprise-level consumption	Deploying a centralized regulatory compliance dashboard Providing enterprise-focused interactive KPI dashboards including time series forecasting
Looker Studio	Enable business analysts and citizen data scientists capabilities to connect enterprise data, local domain data, and external data to provide detailed, contextual insight	Developing targeted LOB interactive dashboards for performance monitoring
Looker Embedded Analytics	Integrating analytics delivery into end user applications for reduced time to insight and response workflow	Incorporating analytics views into client service consoles to provide agents with comprehensive views of customer behavior
Connected Sheets	Enable business and data analysts to connect to and prepare large amounts of data without SQL knowledge, leveraging common spreadsheet formulas and pivot tables	Aggregating large amounts of finance data for internal sales and revenue reporting

Source: Gartner (January 2023)

Data Exploration

Users of the Google Cloud Platform may explore data and discover insights through multiple lenses. Commonly, the tools found to interact with and view data may be with:

- **BigQuery SQL workspace:** A code-first, data analyst view to connect with and query BigQuery resources
- **Connected Sheets:** A low-code, self-service lens for business analysts
- **Vertex AI Workbench:** Jupyter Notebook service in GCP for use by data science professionals

In addition to using BigQuery's SQL workspace, Connected Sheets and the Vertex AI Workbench to query and explore data, Google provides discoverability capabilities through BigQuery search indexes and the Dataplex data exploration workbench.

BigQuery Search Indexes

In Search, BigQuery combines columnar store and full-text search in a single platform. This, coupled with Google Standard SQL, allows users to discover insights hidden in unstructured text and semistructured JSON data.

Identified common use cases for Search include:

- Searching application and system logs stored in BigQuery tables
- Identifying and acting on data to meet compliance needs
- Performing security audits

Dataplex (Explore)

The data exploration workbench in Dataplex offers an interactive, serverless, Apache Spark-powered data exploration capability for users to query data in BigQuery and Cloud Storage data using Apache Software Foundation (ASF) Spark SQL and Jupyter notebooks. Use of Explore is dependent on linking your Dataplex lake with a Google Dataproc Metastore (DPMS). The related DPMS is a fully managed Apache Hive metastore running on Google Cloud (note that the details of DPMS are outside the scope of this research; see Design).

The data exploration workbench is principally useful for those users leveraging Spark with the following limitations identified:

- Explore is limited to lakes in the us-central1 and europe-west2 regions.

- Spark SQL queries can query data only within a specific lake. To explore data in another lake, analysts must switch environments to another lake.
- Although saving queries and notebooks is supported, sharing of Spark SQL queries and notebooks is not yet supported.

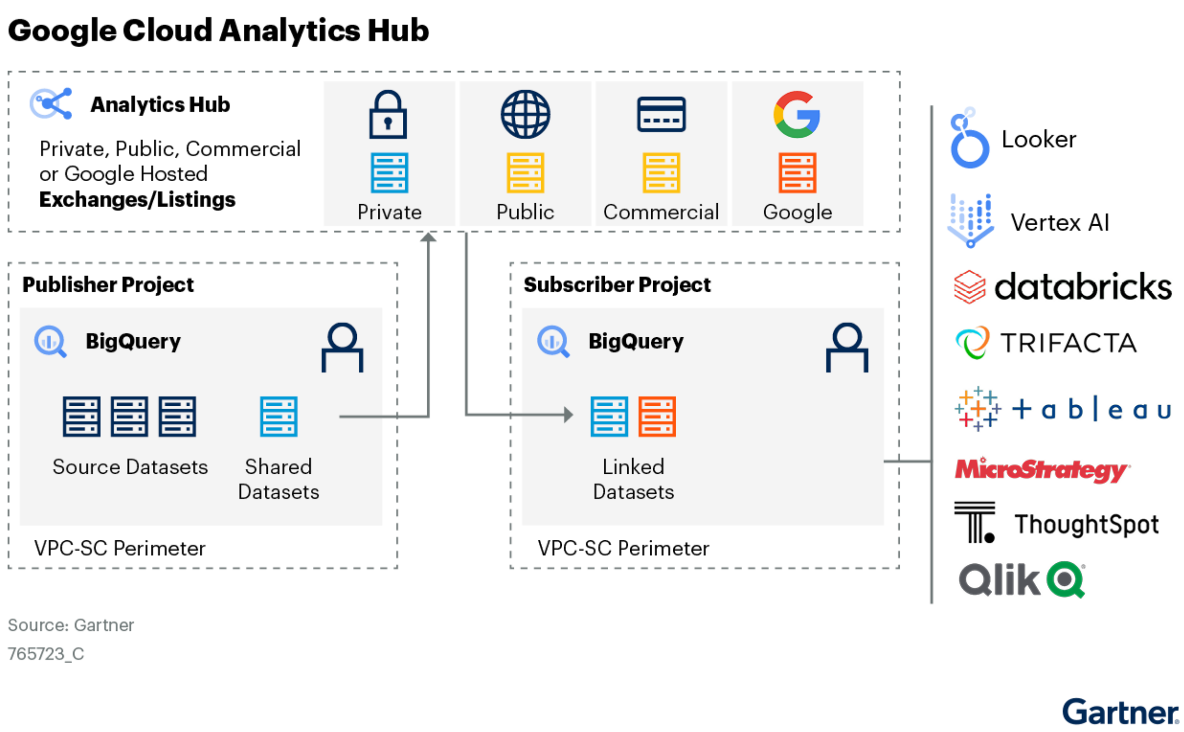
Data Sharing

Analytics Hub

With Google Analytics Hub, data providers create data exchanges to share data and analytics assets across the organization internally and with external partners/consumers. As a container, the data exchange provides a self-service interface for subscribers to access data. These exchanges are either private (default), where access is managed by the provider through Google's Identity and Access Management (IAM) policy, or public, where the assets may be discovered and subscribed to by Google Cloud users.

Figure 14 shows how the Analytics Hub connects to BigQuery and delivers data to subscribers.

Figure 14. Google Cloud Analytics Hub



Currently, the following BigQuery objects may be shared through the Analytics Hub:

- Tables (including external and snapshot)
- Views (including materialized and authorized views)
- Authorized datasets
- BigQuery ML models

For those consumers looking to provide data externally or monetize their data, the data exchange in Analytics Hub delivers the capability to publish data for subscribers. These exchanges may be public or private and subscribers (existing and potential) can search for new data to consume.

BigQuery Export

In cases where analytical data needs to be shared, users may export tables from BigQuery in the following forms and compression types (see Table 7).

Table 7: BigQuery Export Formats

File Format	Compression Type
CSV	GNU Gzip
JSON	Gzip
Apache Avro	DEFLATE, Snappy
Apache Parquet	Snappy, Gzip

Source: Gartner (January 2023)

Users looking to share data via export should note the following limitations:

- Resulting files may only be exported to Cloud Storage.
- Resulting files are limited to 1 gigabyte (GB) in size. Data exceeding this size will be segmented into multiple files.
- A single export job cannot include data from multiple tables.

Governance

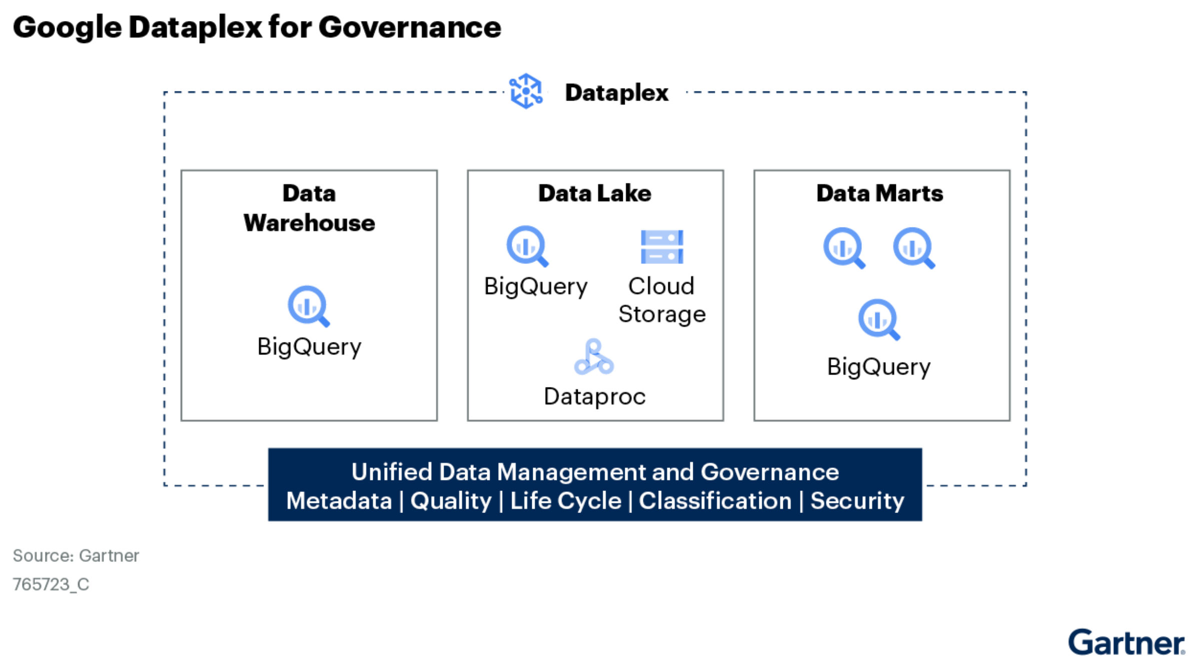
D&A governance is increasingly important as data is collected, transformed and analyzed from various sources. While a comprehensive governance program to maintain security and manage sensitive data is more extensive than just technology, Google does provide several GCP-based tools to support these efforts. For more details on developing a governance framework in your organization, see Building a Comprehensive Governance Framework for Data and Analytics.

Dataplex

Google positions Dataplex as a data fabric to centrally manage, monitor and govern data across your landscape. Through the use of external tables in BigQuery, information can be ingested into Dataplex to provide insight beyond your Google Cloud landscape. Dataplex automatically detects schemas, extracts metadata, and performs quality checks without relocating data to build out a data fabric. Broadly, Dataplex follows the principles of data fabric design. However, a full analysis of data fabric is outside the scope of this research. For additional information on this topic, see Quick Answer: What Is Data Fabric Design?

As part of Dataplex, this service also incorporates the legacy tool (Google Cloud Data Catalog) under a single user interface. For additional details on Google Cloud Platform data governance, privacy and security, see Building a Data Management Architecture in Google Cloud Platform. Figure 15 shows Dataplex as a unified platform to manage and govern enterprise data across the landscape.

Figure 15. Google Dataplex for Governance



Vertex AI Model Registry

With the launch of the Vertex AI Model Registry, Google makes additional steps to model governance in the Google Cloud Platform. The registry is a central repository for ML life cycle management and supports the following:

- Custom models
- AutoML (text, tabular, image and video)
- BigQuery ML

Through the registry, users can deploy models to endpoints using a simplified UI. BigQuery ML models do not need to be exported and deployed manually to Vertex AI. Metadata and lineage of BigQuery ML models is now retained.

Additional features of the Vertex AI Model Registry include:

- **Version control**: Versions of existing models may be created directly in the Vertex AI Model Registry or imported as a new version.

- **Use of aliases:** Named references can be used to identify particular model versions. Consuming models by reference ensures a specific version is utilized without the user needing a specific model ID.
- **Metadata labels:** Use of custom labels allows developers to insert custom fields for logical organization and classification. Labels are available at the model and version levels.

Vertex AI Feature Store

Feature engineering is the iterative process data scientists undergo to source, acquire, clean and transform the raw data they need to create features for ML development projects. The goal for all enterprises seeking to mature their ML operations must be to implement a feature management practice and related tooling to ensure the reproducibility, reusability and reliability of features for ML. This increases the efficiency of ML model development and productionization.

The Vertex AI Feature Store provides a central repository for organizing, storing and serving ML features. The feature store enables sharing across the organization rather than reengineering features for each project and use case. Feature Store offers search and filter capabilities and provides lineage for data used in models.

For further details on feature stores for machine learning, see the following Gartner research:

- Feature Stores for Machine Learning (Part 1): The Promise of Feature Stores
- Feature Stores for Machine Learning (Part 2): Current State and Future Directions

Orchestration

This section outlines the tools in GCP that data and analytics professionals may use to author, monitor and execute workflows in a scheduled and automated manner. In order to deliver a complete end-to-end data and analytics solution, GCP provides the essential networking, security, DevOps and infrastructure components. Data and analytics professionals must work with their infrastructure teams to ensure that all the key pieces are in place before the analytical pipeline can be successfully deployed in production.

Detailed discussion or analysis on these topics is beyond the scope of this assessment.

Cloud Composer

Cloud Composer is Google's fully managed workflow orchestration tool based on Apache Airflow. Workflows in Cloud Composer are representations of tasks created using directed acyclic graphs (DAGs).

Tasks in each DAG may perform a variety of tasks including:

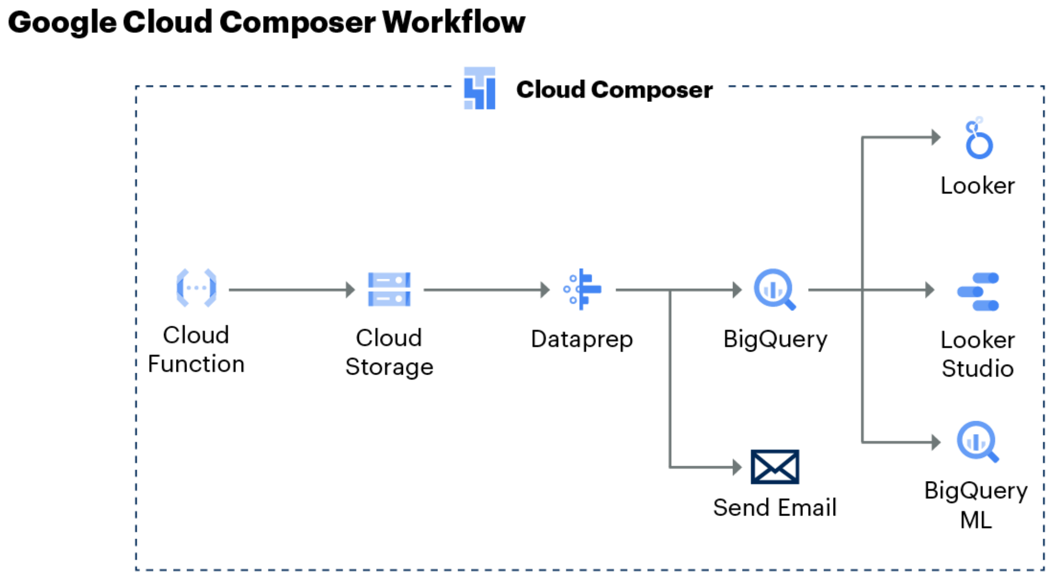
- Data preparation
- Monitoring
- Sending communications
- Running pipelines

Cloud Composer currently offers two versions; scaling capabilities are the principal difference. Cloud Composer 1 offers manual scaling, while Cloud Composer 2 includes autoscaling environments. For a full list of the differences between versions, see Google's product documentation, Major Versions of Cloud Composer.

In addition to Cloud Composer's tight integration with the Google Cloud Platform, it includes functionality to allow for orchestration of workflows across multiple cloud providers and on-premises solutions.

Figure 16 displays an end-to-end workflow supported by Cloud Composer.

Figure 16. Google Cloud Composer Workflow

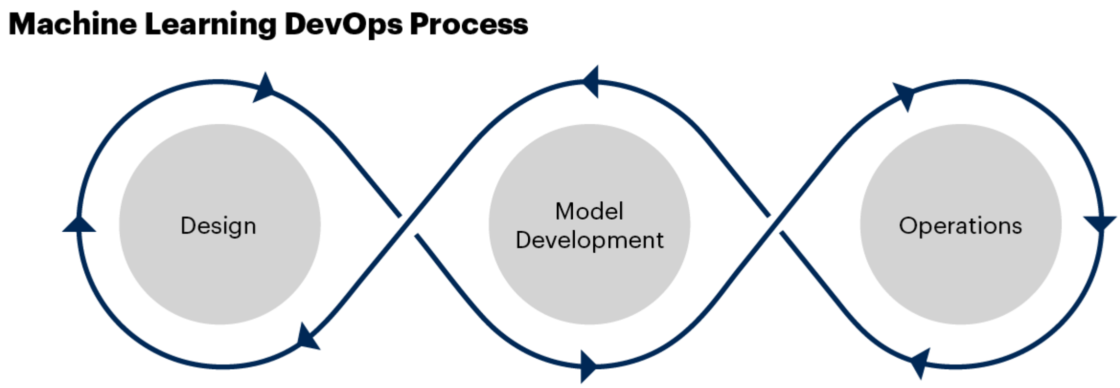


Source: Gartner
765723_C

Machine Learning DevOps

Machine learning DevOps (MLOps) differ from traditional DevOps projects. ML models need to be regularly evaluated to reduce potential bias and drift. Figure 17 shows this cyclical pattern of data exploration, development and operationalization of ML models.

Figure 17. Machine Learning DevOps Process

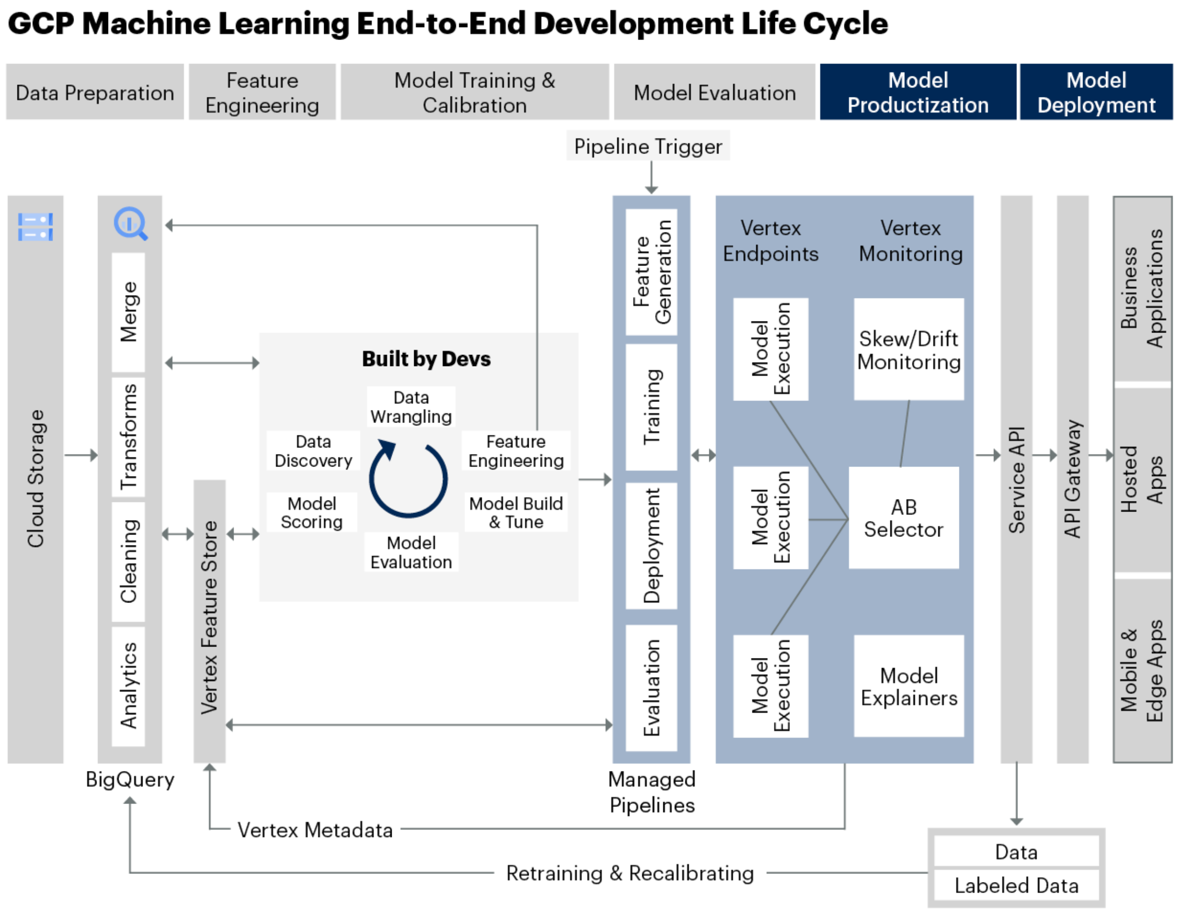


Source: Gartner
763681_C

In a seminal research document called Hidden Technical Debt in Machine Learning Systems, Google researchers demonstrated that creating an ML pipeline is more than simply writing the code. In fact, writing the code is actually a very small component of the pipeline. Much bigger tasks include the process of ingesting and preparing the data, which, in the ML parlance, are termed "data collection" and "feature engineering." Once the model is trained, it must be continually tested, monitored and scored against continuously changing input data. This requires a sophisticated model management process that tracks model versions and deploys the right version.

Figure 18 expands on the ML model development life cycle and how various GCP products are utilized to achieve the desired solution.

Figure 18. GCP Machine Learning End-to-End Development Life Cycle



Source: Gartner
765723_C

Vertex AI Pipelines

Vertex AI Pipelines facilitate the ML development life cycle by orchestrating ML workflows. Once your ML workflow is outlined as a pipeline, Vertex AI Pipelines is used to build portable containers for orchestration. Each container stores the self-contained code to perform that step's specific task (such as data preprocessing, transformation or training).

Vertex AI Pipelines can also be used to run ML pipelines built using Kubeflow or TensorFlow Extended (TFX). If you have already built pipelines using TFX, it is recommended to continue using that software development kit (SDK) in Vertex AI Pipelines. If not using TFX, the Kubeflow SDK allows for the use of Google Cloud Pipeline Components, which can make it simpler to incorporate AutoML into your pipeline.

For additional information regarding XOps, see Demystifying XOps: DataOps, MLOps, ModelOps, AIOps and Platform Ops for AI.

Monitoring

Looker Monitoring

To manage activity and performance of your Looker service, Google provides a series of system activity dashboards to monitor the health and performance of analytics:

- **User activity:** Information on users and their usages on your Looker instance
- **Content activity:** Information on which dashboards, Looks and Explores are viewed and scheduled
- **Database performance:** Performance of content and persistent derived tables (PDTs)
- **Instance performance:** Information about the load and performance of the scheduler and performance-intensive content
- **Performance recommendations:** Insight and opportunities to improve performance on your Looker content
- **Errors and broken content:** Identification of dashboards, Looks, schedules and PDTs producing errors

Vertex AI Model Monitoring

Once machine learning models are deployed to production, they must be monitored for performance degradation, including feature skew and drift. Vertex AI Model Monitoring enables analytics professionals to monitor tabular AutoML and tabular custom-trained models for:

- **Training-service skew:** Difference between the model performance during training and serving
- **Prediction drift:** Change in distribution of the predictions on evaluation over time

Model Monitoring uses TensorFlow Data Validation (TFDV) to calculate distributions and distance scores on skew and drift for categorical and numeric features. ML developers set related thresholds and alerts that are delivered when calculations are breached. Through the Google Cloud console, ML developers can view feature distribution results to evaluate for potential retraining activities.

For more details about monitoring ML models, see Getting Started With Machine Learning Monitoring in Production.

Google Cloud Operations

In addition to the monitoring capabilities specific to analytics through Looker, Google provides Google Cloud operations suite (formerly Stackdriver). Cloud operations include the following six primary modules for monitoring activity across your Google Cloud Platform instance:

- **Cloud Monitoring:** Provides visibility on metrics for CPU usage, disk input/output (I/O), memory, traffic and custom metrics
- **Cloud Logging:** Provides real-time log data for events across Google Cloud
- **Cloud Error Reporting:** Provides a centralized management interface to aggregate and display errors from your cloud services
- **Cloud Debugger:** Provides code inspection capabilities to view the state of applications without runtime interruption
- **Cloud Trace:** Provides near real-time reporting of latency data from applications
- **Cloud Profiler:** Provides insight to application performance by gathering and profiling CPU usage and memory allocation data from production applications

Further details of Google Cloud's operations is beyond the scope of this research.

Strengths

Google Cloud Platform's technical strengths in analytics include:

- **Accessibility and availability of ML/AI integration:** Through AI agents, pretrained AutoML and BigQuery ML, Google Cloud offers a wide range of ML capabilities to integrate with your analytics and applications. These options open the gateway to implementing ML to an increasingly growing base of developers with differing skill sets.
- **Unified analytics governance:** Users can create scalable, enterprise-focused data models in the Looker Modeling Service using LookML, consumable by all Looker analytics tools (Looker, Looker Studio and Looker Embedded Analytics). These enterprise models may also be consumed through third-party BI tools, extending their reach and reusability across diverse organizations.
- **Multicloud openness:** BigQuery Omni (see Note 1) provides multicloud query capabilities without requiring copying or replication of data from other cloud environments. This provides flexibility for organizations leveraging multiple cloud service providers across their technical landscape.

Weaknesses

Google Cloud Platform's technical weaknesses in analytics include:

- **Available developer pool:** Although the LookML modeling layer is a strength for Looker, its steep learning curve (relative to other platforms) can make Looker less appealing. Without knowledge of the LookML language, organizations could experience limited growth through this platform.
- **Some data processing tools are not "Google" owned, but are part of the larger Google Cloud ecosystem:** Dataprep by Trifacta (acquired by Alteryx) is a data preparation tool used to create curated datasets as part of the data pipeline. Its acquisition by Alteryx raises questions on long-term support as part of the Google Cloud Platform.
- **Vertex AI has limited capabilities for collaboration:** It does not have capabilities to enable real-time multiuser collaboration, and it does not provide native integration with project management tools from within Vertex AI.

Guidance

With a full range of analytics offerings — with capabilities for data engineers, data scientists, modelers and analysts — GCP may provide compelling options for organizations looking to derive competitive advantage through advanced analytics, including ML and AI.

The accessibility of AI and ML services is a key strength of the Google Cloud Platform. Google's offerings from code-first to low-code capabilities include:

- Prepackaged AI agents with code solutions for analyzing documents, providing customer recommendations, and customer interaction
- Pretrained ML APIs providing common-use models for developers and analysts working with text, tabular, video and image data
- Vertex AI service that provides a unified experience to build, train and deploy both AutoML and code-based ML models
- BigQuery ML that brings machine learning capabilities to data engineers using their familiarity with SQL.

Analysts looking to utilize these AI/ML capabilities find themselves working to decide whether to build and custom-train their ML models or use AutoML. This decision is not too far from application implementation decisions on build vs. buy. Prepackaged AI and pretrained ML models may provide ease of implementation, but could lack discrete nuances to your use case. Therefore, the pros and cons of each must be weighed carefully.

See Table 8 for guidance on choosing AutoML models or code-based model training for your use cases in Google Cloud.

Table 8: AutoML vs. Code-Based Training

(Enlarged table in Appendix)

	AutoML	Code-Based Training
Data science expertise needed?	No	Yes Expertise needed for developing the training application and data preparation, including feature engineering.
Programming ability needed?	No	Yes
Time it takes to train a model?	Lower Less data preparation is required, and no development is required.	Higher More data preparation is required, and training application development is needed.
Limits on machine learning objectives?	Yes AutoML includes predefined data types and objectives.	No
Can manually optimize model performance with hyperparameter tuning?	No AutoML does some automated hyperparameter tuning. It cannot be modified manually.	Yes
Can control aspects of the training environment?	Limited Image and tabular datasets allow for node hours and early stopping configurations.	Yes
Limits on data size?	Yes AutoML uses managed datasets; data size limitations vary depending on the type of dataset.	For unmanaged datasets, no. Managed datasets have the same limits as Vertex AI datasets that are used to train AutoML models.

Source: Choose a Training Method, Google

Finally, continued inquiries with Gartner clients show that organizations struggle with analytical data silos resulting from advanced business intelligence (BI) tools supporting self-service analytics by an increased number of citizen developers. Organizations looking to provide governed, modeled data to self-service analysts and citizen data scientists will find that the Google Cloud Looker Modeling Service (based on LookML) directly addresses this. It positions itself as a governed analytical modeling layer able to deliver analytics across various use cases and tools.

Evidence

2020 Gartner Cloud Data and Analytics Survey: This survey was conducted online from 17 June through 29 June 2020 with 86 members of Gartner's Research Circle — a Gartner-managed panel. Fifty-six percent of qualified respondents were involved in setting the cloud data and analytics management strategy of their organizations. Secondary research delivered by Gartner Secondary Research Service (SRS) included:

- Articles
- Blogs
- Charts
- News journals
- Press releases
- Published materials
- Reports
- Subscribed databases
- Surveys
- Websites

Note 1: BigQuery Omni

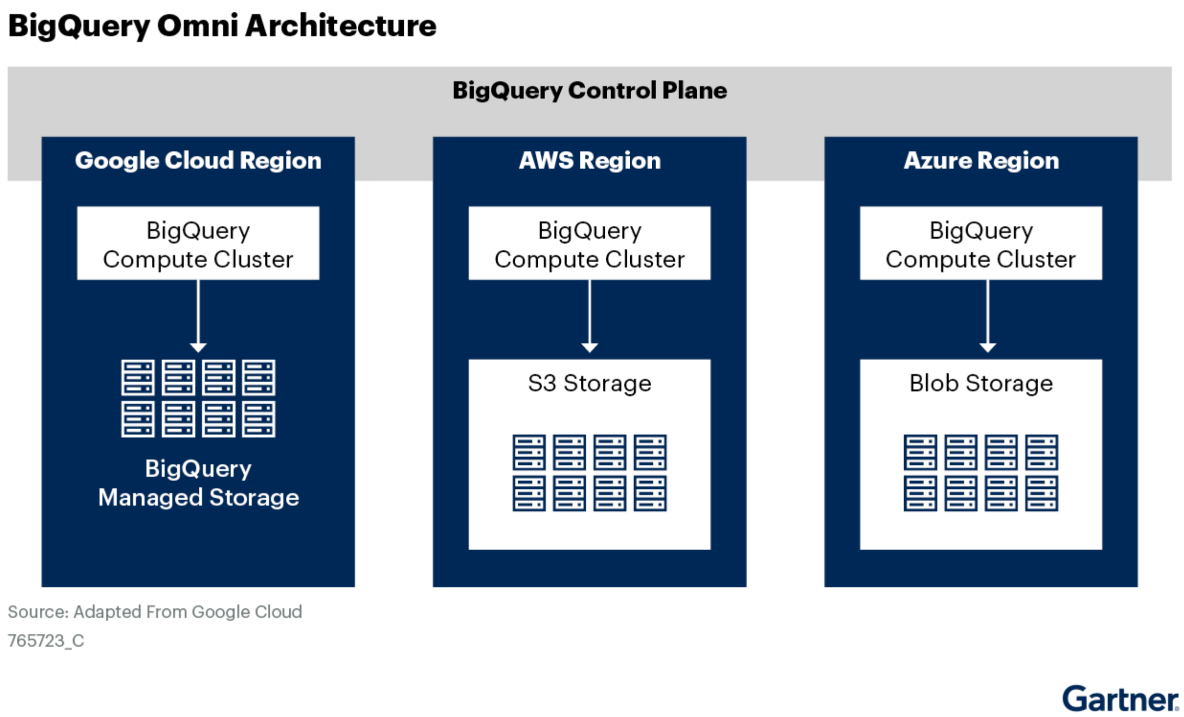
Many more organizations today store data in multiple clouds, whether by design, residency restrictions or organically from acquisitions. This presents a challenge to analysts looking to combine this data and deliver insights.

By separating query compute from storage, Google offers a unique capability through BigQuery Omni to query and retrieve data from multiple clouds without traditional efforts to move or copy any data. This presents a flexible opportunity to scale across large query workloads.

To accomplish this, a BigQuery compute cluster is required in the cloud where data is stored (e.g., Amazon Web Services [AWS] or Microsoft Azure). This compute cluster may be used in the local cloud to prevent cross-cloud movement and related egress fees or return its results through to Google Cloud. Supported data formats for BigQuery Omni include CSV, JSON, Apache Parquet, ORC and Apache Avro.

Figure 19 demonstrates the architecture to connect the BigQuery control plane with storage in multiple external clouds.

Figure 19. BigQuery Omni Architecture



While BigQuery Omni offers a flexible method for accessing data across public clouds, there are known limitations to this feature. Limitations include:

- Only external tables are supported, and all related limitations of external tables apply.
- Materialized views are not supported.
- BigQuery ML statements are not supported.
- Maximum result set is limited to 10GB (in preview at time of publication).

- Connectivity to AWS and Azure regions is limited to aws-us-east-1 and azure-eastus2, respectively.

Note 2: BigQuery GIS

More organizations today revolve their business around location and mobility data, and still more are looking to derive insights from this same type of data. BigQuery GIS supports geographic data types and functions as first-class citizens, so geospatial analyses can be performed directly, using SQL, where your location-based data persists.

Google provides multiple ways for analysts to visualize this data, including:

- **Looker Studio (formerly Data Studio):** Studio provides users with capabilities to query data with geographic field types and visual interactions similar to Google Maps.
- **BigQuery Geo Viz:** This is a web-based companion app with queries performed in SQL. The underlying computational libraries are the same as those used by Google Earth Engine and Google Maps.
- **Google Earth Engine:** To utilize the Google Earth Engine, BigQuery data must be exported to Cloud Storage, then imported into Earth Engine.
- **Jupyter Notebook:** Through the GeoJSON extension, analysts may visualize data (GeoJSON format only) in Jupyter notebooks.

Recommended by the Authors

Some documents may not be available as part of your current Gartner subscription.

Solution Comparison for Cloud Data Science and Machine Learning Platforms

Solution Scorecard for Google Cloud Vertex AI

Solution Scorecard for Google Cloud Analytical Data Stores

Critical Capabilities for Analytics and Business Intelligence Platforms

Solution Path for Building Modern Analytics and BI Architectures

---

Table 1: Primary Looker Modeling Service Components

Object	Description
Project	Projects are the containers that collect your files that describe tables and relationships and that define how Looker controls the user interface's behavior. These files principally include models and views.
Model	Models hold information about which tables are included and their relationships. Connections defined in the model also define Explores for users to query on.
View	Views define the fields (measures and dimensions) and their connection to underlying tables. Views are typically defined as part of the project Explores.
Explore	Explores are the starting point for user queries. For consumers familiar with SQL, this is comparable to the "FROM" SQL statement

Source: Gartner (January 2023)

Table 2: Google Cloud AI Agents and Pretrained Models/AI Capabilities

AI Agents		*Example Use Case*
Contact Center AI (powered by Dialogflow CX)		Prebuilt customer service AI solution
Document AI		Document image quality detection, classification, labeling and text extraction
Discovery AI (includes Retail AI and Recommendations AI)		Deliver tailored recommendations to specific customer preferences
Dialogflow CX		Build agents to maintain conversations through text and audio with end users
Healthcare Insights AI (includes Google Cloud Healthcare Data Engine, Cloud Healthcare API and Cloud Natural Language API)		Deliver forecasting models to understand the progression of infection across populations
Pretrained Models and APIs ↓		*Example Use Case* ↓
Vision	Vision AI	Detect and extract text from images or properties for further classification
	Video AI	Moderate and automate captioning to streaming videos

Language	Translation AI	Provide text translation in application workflow for service agents receiving requests in multiple languages
	Natural Language AI	Identify entities (phrases) in text and perform sentiment analysis from social media streams
Conversation	Speech-to-Text AI	Convert voicemail recordings to text files for documentation
	Text-to-Speech AI	Provide audio response to application users upon task completion
Structured Data	Timeseries Insights API, Vertex AI Forecast, AutoML Tables, TabNet and Cloud Fleet Routing API	Develop predictions for daily demand on products or services

Source: Gartner (January 2023)

Table 3: Vertex AI AutoML Data Limitations

Text Data	*Video Data*	*Image Data*
Minimum training docs: 20 Max training docs: 1 million Min categorization labels: 2 Max categorization labels: 5,000	File Types: MOV, MPEG4, MP4, AVI Max size: 50 gigabytes (3 hours) Max categorization labels: 1,000	Training — max file size 30MB (JPEG, GIF, PNG, BMP, ICO) Prediction — max file size 1.5MB (JPEG, GIF, PNG, WebP, BMP, TIFF, ICO)
Note that these limitations may change as Vertex AI evolves. Users should check Google's documentation for updated information.		

Source: Gartner (January 2023)

Table 4: Vertex Explainable AI Use Cases

Model Type	*Data Type*	*Enabling*
AutoML	Image (classification only)	Must explicitly enable during training
AutoML	Tabular (classification and regression)	Enabled by default
Custom-trained	Image	Configure explanations when building or deploying model
Custom-trained	Tabular	Configure explanations when building or deploying model

Source: Gartner (January 2023)

Table 5: Vertex Explainable AI Feature Attribution Methods

Method	Basic Explanation	Recommended Model Types	Example Use Cases	Compatible Vertex AI Model Resources
Sampled Shapley	Assigns credit for the outcome to each feature and considers different permutations of the features; provides a sampling approximation of exact Shapley values	Nondifferentiable models, such as ensembles of trees and neural networks	Classification and regression on tabular data	Custom-trained models (any prediction container) AutoML tabular models
Integrated gradients	A gradients-based method to efficiently compute feature attributions with the same axiomatic properties as the Shapley value	Differentiable models, such as neural networks; recommended especially for models with large feature spaces Recommended for low-contrast images, such as X-rays	Classification and regression on tabular data Classification on image data	Custom-trained TensorFlow models that use a TensorFlow prebuilt container to serve predictions AutoML image models

Method	Basic Explanation	Recommended Model Types	Example Use Cases	Compatible Vertex AI Model Resources
XRAI (eXplanation with Ranked Area Integrals)	Based on the integrated gradients method; assesses overlapping regions of the image to create a saliency map, which highlights relevant regions of the image rather than pixels	Models that accept image inputs; recommended especially for natural images, which are any real-world scenes that contain multiple objects	Classification on image data	Custom-trained TensorFlow models that use a TensorFlow prebuilt container to serve predictions AutoML image models

Source: Compare Feature Attribution Method, Google

Table 6: Google Interactive Analytics and BI Use Cases

Product	Function	Example
Looker	Deliver consistent, governed models and metrics to end users for enterprise-level consumption	Deploying a centralized regulatory compliance dashboard Providing enterprise-focused interactive KPI dashboards including time series forecasting
Looker Studio	Enable business analysts and citizen data scientists capabilities to connect enterprise data, local domain data, and external data to provide detailed, contextual insight	Developing targeted LOB interactive dashboards for performance monitoring
Looker Embedded Analytics	Integrating analytics delivery into end user applications for reduced time to insight and response workflow	Incorporating analytics views into client service consoles to provide agents with comprehensive views of customer behavior
Connected Sheets	Enable business and data analysts to connect to and prepare large amounts of data without SQL knowledge, leveraging common spreadsheet formulas and pivot tables	Aggregating large amounts of finance data for internal sales and revenue reporting

Source: Gartner (January 2023)

Table 7: BigQuery Export Formats

File Format	Compression Type
CSV	GNU Gzip
JSON	Gzip
Apache Avro	DEFLATE, Snappy
Apache Parquet	Snappy, Gzip

Source: Gartner (January 2023)

Table 8: AutoML vs. Code-Based Training

	AutoML	*Code-Based Training*
Data science expertise needed?	No	Yes Expertise needed for developing the training application and data preparation, including feature engineering.
Programming ability needed?	No	Yes
Time it takes to train a model?	Lower Less data preparation is required, and no development is required.	Higher More data preparation is required, and training application development is needed.
Limits on machine learning objectives?	Yes AutoML includes predefined data types and objectives.	No
Can manually optimize model performance with hyperparameter tuning?	No AutoML does some automated hyperparameter tuning. It cannot be modified manually.	Yes
Can control aspects of the training environment?	Limited Image and tabular datasets allow for node hours and early stopping configurations.	Yes

	AutoML	*Code-Based Training*
Limits on data size?	Yes AutoML uses managed datasets; data size limitations vary depending on the type of dataset.	For unmanaged datasets, no. Managed datasets have the same limits as Vertex AI datasets that are used to train AutoML models.

Source: Choose a Training Method, Google