Can Spark or BigQuery be used as base for the platform?

# Spark (Free to use)

https://spark.apache.org/

Apache Spark is an **open-source data processing and analytics platform** that provides a range of tools and **libraries for working with large-scale data sets**. It is designed to be fast, scalable, and easy to use, and it provides support for a variety of data processing and analytics tasks, including batch processing, streaming data analysis, **machine learning**, and graph processing.

Spark provides a unified programming model that allows users to write applications using a **single API**, and it provides support for a variety of programming languages, including Java, **Python**, R, and Scala. It also integrates with a variety of other tools and frameworks, such as Hadoop and other big data technologies, allowing users to build and deploy complex data pipelines and applications.

**Spark provides a range of capabilities for working with data, including tools for data ingestion, data exploration and visualization, data preparation, and data modeling. It also provides support for a variety of machine learning algorithms and techniques, including supervised and unsupervised learning, deep learning, and natural language processing.**

Overall, Apache Spark provides a fast and scalable platform for data processing and analytics, with a unified programming model and a range of tools and libraries for working with large-scale data sets. It is designed to be easy to use and to integrate with other tools and technologies, and it provides support for a variety of data processing and analytics tasks, including machine learning.

# BigQuery

BigQuery is a cloud-based data warehousing and analytics platform that is offered by Google Cloud. It provides a scalable and performant platform for storing and querying large-scale data sets, and it allows users to run complex SQL queries over petabytes of data in seconds.

BigQuery is designed to be easy to use and to integrate with other Google Cloud services, such as Cloud Storage, Cloud Pub/Sub, and Cloud Dataproc. It also integrates with a variety of other tools and technologies, such as data visualization tools, machine learning frameworks, and other big data technologies.

BigQuery provides a range of capabilities for working with data, including tools for data ingestion, data preparation, data modeling, and data visualization. It also provides support for a variety of data formats, including structured and semi-structured data, as well as support for both batch and streaming data.

Overall, BigQuery is a cloud-based data warehousing and analytics platform that provides a scalable and performant platform for storing and querying large-scale data sets. It is designed to be easy to use and to integrate with other tools and technologies, and it provides a range of capabilities for working with data.

## BigQuery cheaper alternatives (cloud-based data warehousing and analytics platforms)

Amazon Redshift

Snowflake

What