

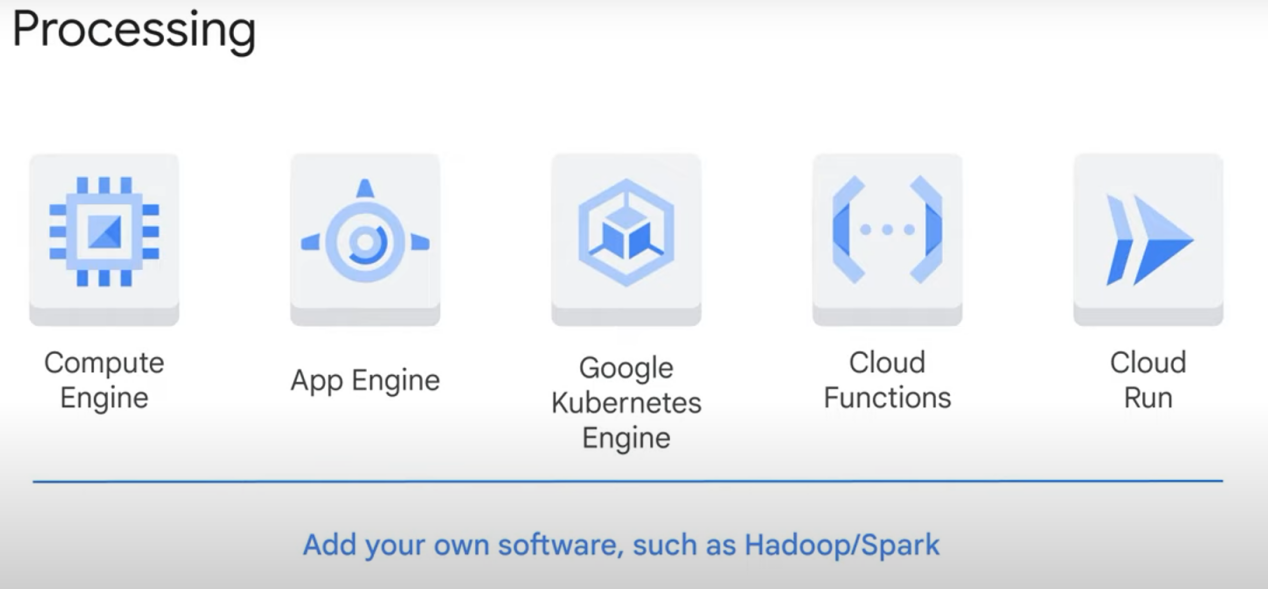
**Managed services** are ones where you can see the individual instance or cluster

VS (opposé de)

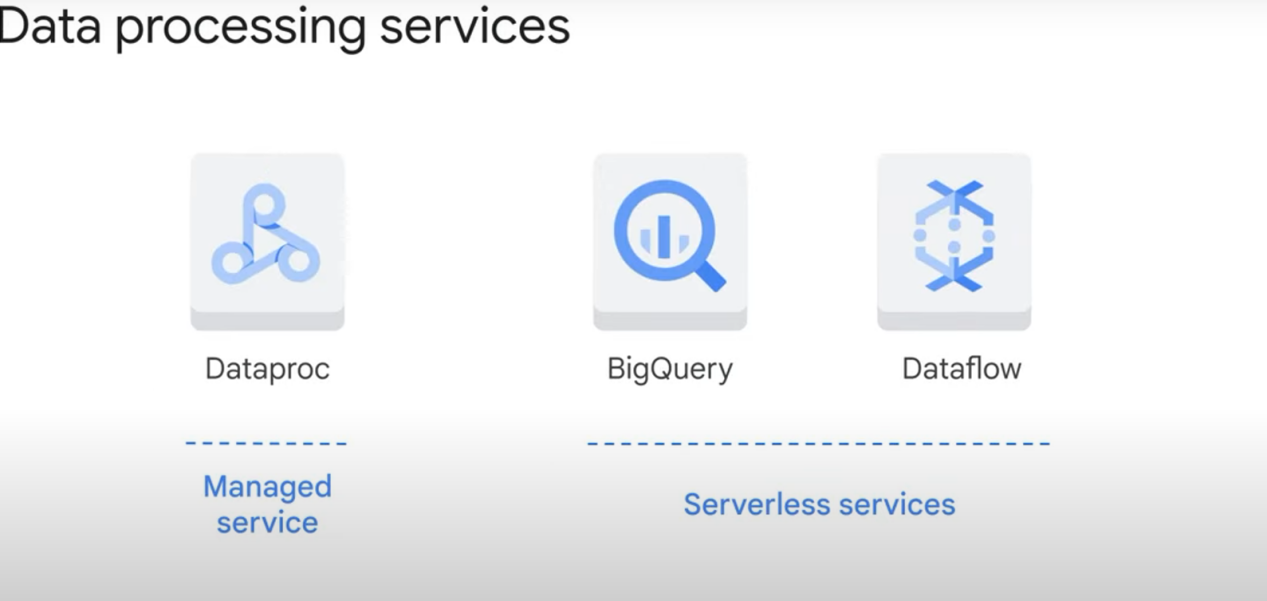
**Serverless services**

exam tip managed services still have some it overhead it doesn't completely eliminate the overhead or manual procedures but it minimizes them compared with on-prem solutions

\***firestore** is a nosql document database built for automatic scaling. It can be run in its native mode (firestore mode) or in datastore mode (compatible with the old technology named «datastore»).

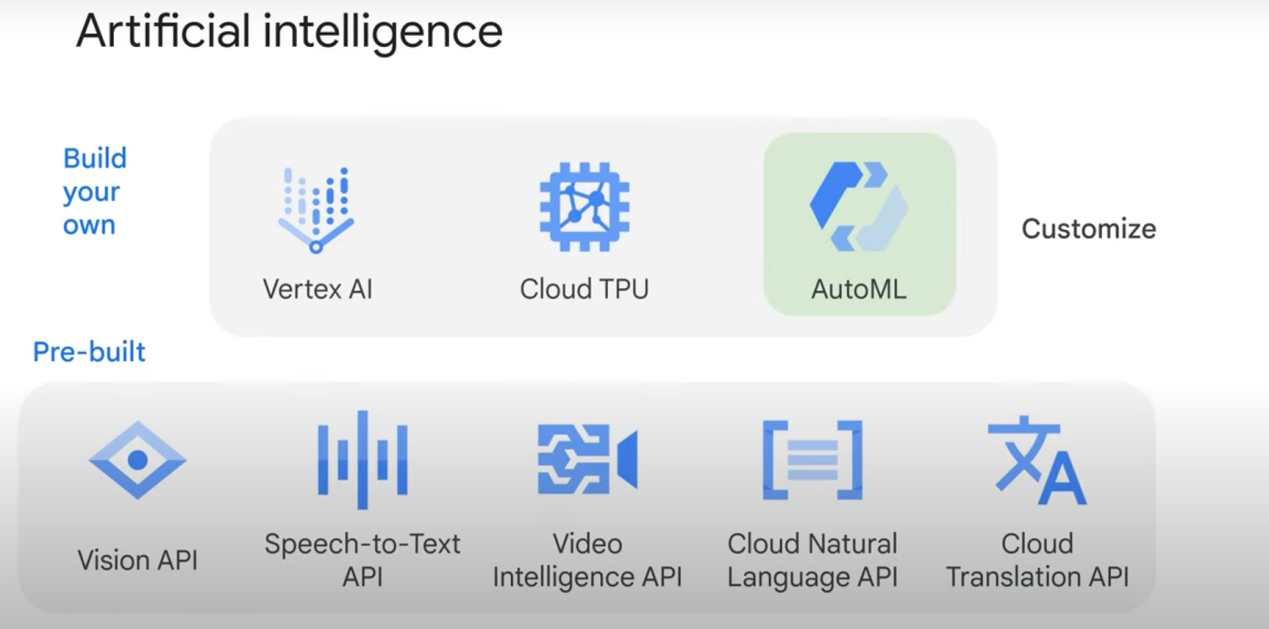


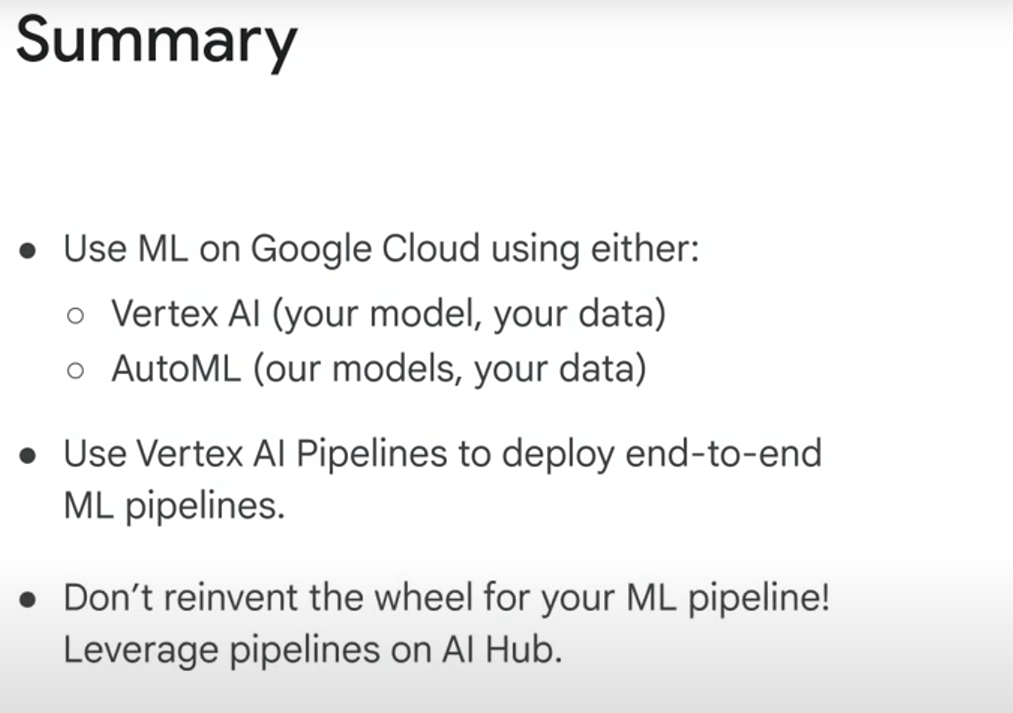
\* On compute engine (VMs ?), you can install any open-source tool you want and use it to process your data. This takes more time for IT devs than using cloud services.

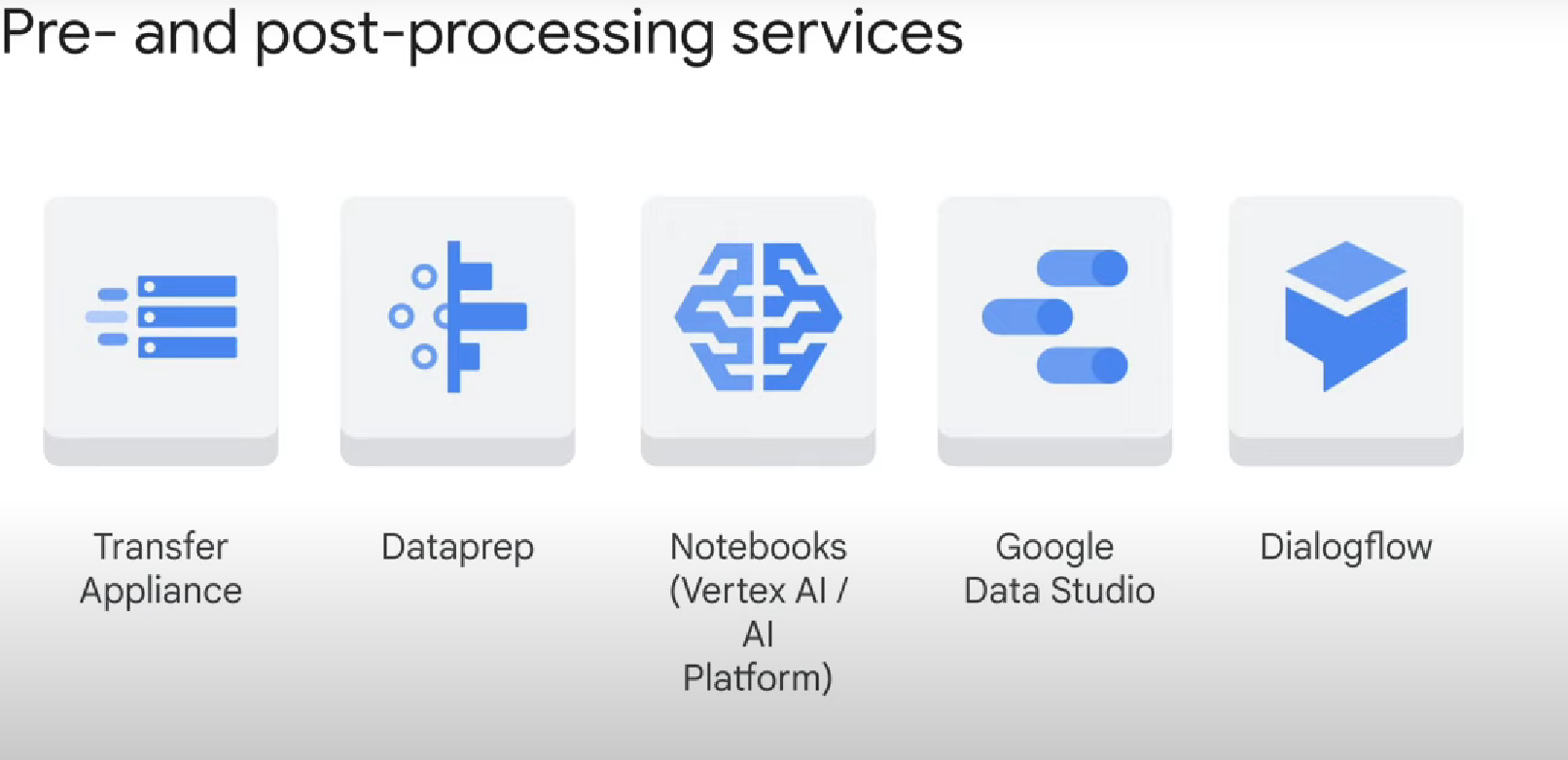


\* Data processing services offer

1. processing abstraction (E.g: in dataproc, directed acyclic graph DAGs)
2. Storage abstraction (E.g: in dataproc too, resilient distributed dataset or rdd . **A Resilient Distributed Dataset (RDD)** is a low-level API and **Spark's** underlying data abstraction. An RDD is a **static set of items distributed across clusters to allow parallel processing**. The data structure stores **any Python, Java, Scala, or user-created object**.)







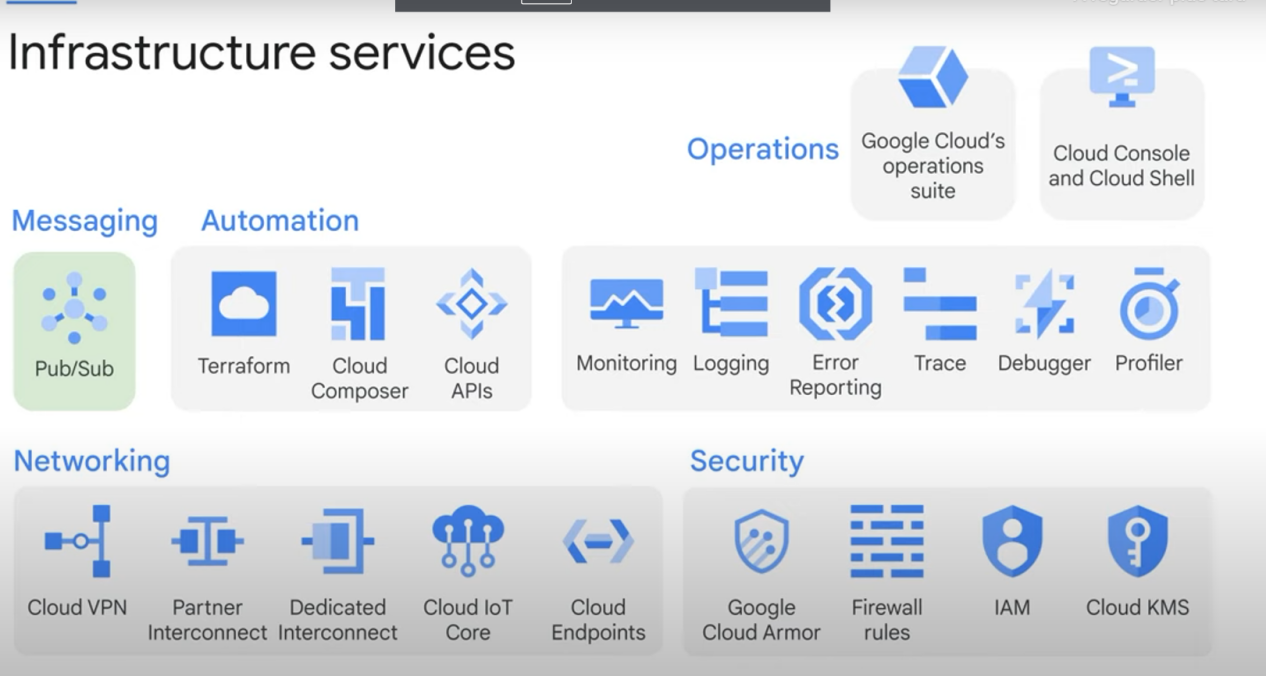
\* **Data transfer appliance** is a shippable device that's used for synchronizing data in the cloud with an **external source cloud**

\* **Data studio** is used for **visualization of data** after it has been processed

\* **Cloud data prep** is used to **prepare or condition data** and to prepare pipelines before processing

\* **Cloud data lab is a notebook** that is a **self-contained workspace** that holds code, executes the code, and displays results

\* **dialogflow** is a service for creating chat bots. It uses ai to provide a method for direct human interaction with data



Note: pub/sub can hold data for 7 days == resilience

\* **Cloud vpn, Partner interconnect, or Dedicated interconnect**, play a role **whenever there's data on premise that must be transmitted to services in the cloud**

Iops = input/output per second

An n4 machine has more iops, and costs more, than an n1 machine

**Designing flexible data representations:**

Which storage to use?

\* File system (hadoop, HDFS): easy to find elements by name,

\* Storing in a database = easier to find data by logic (like sql queries)

\* storing in a processing system = easier to process and transform the data