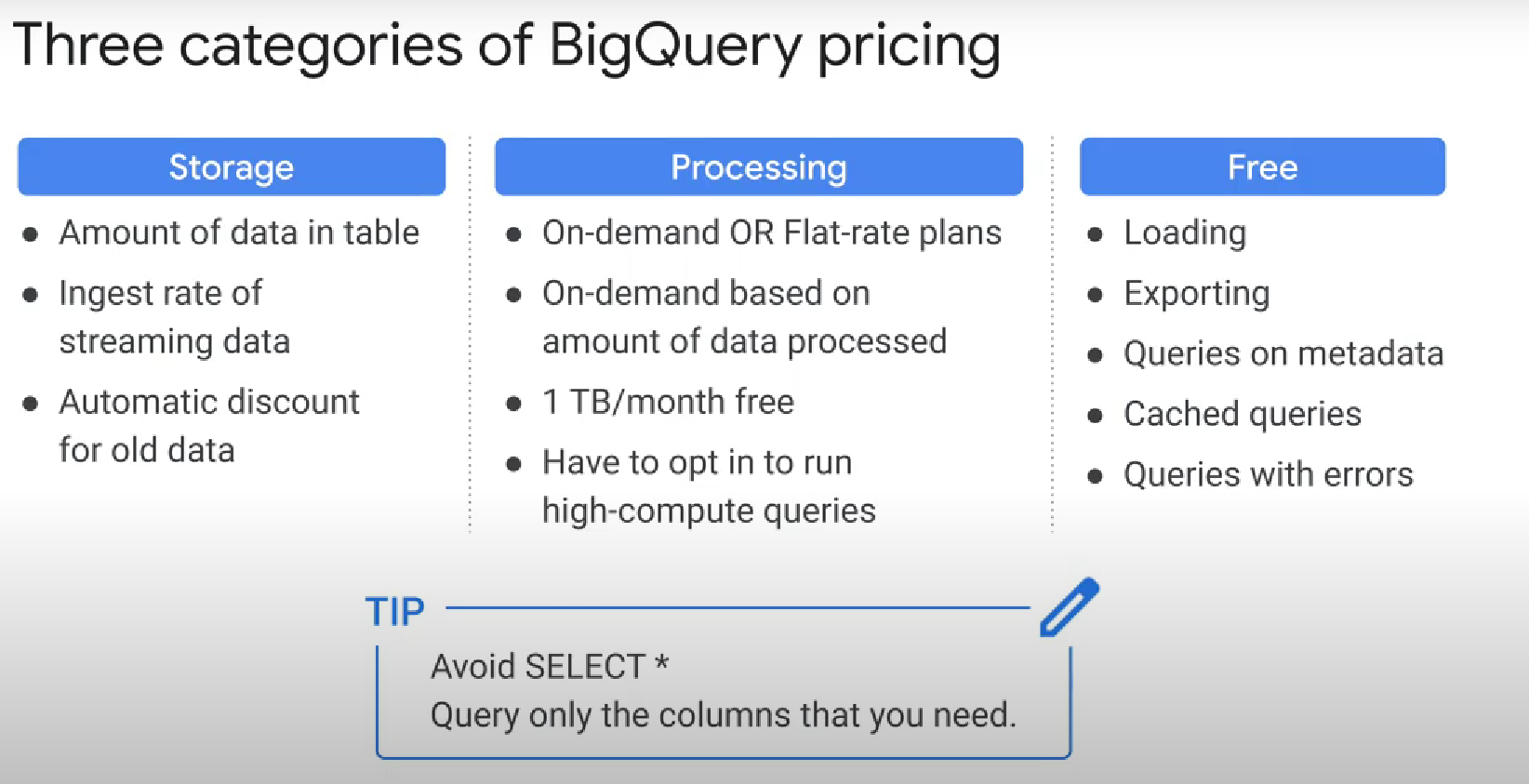


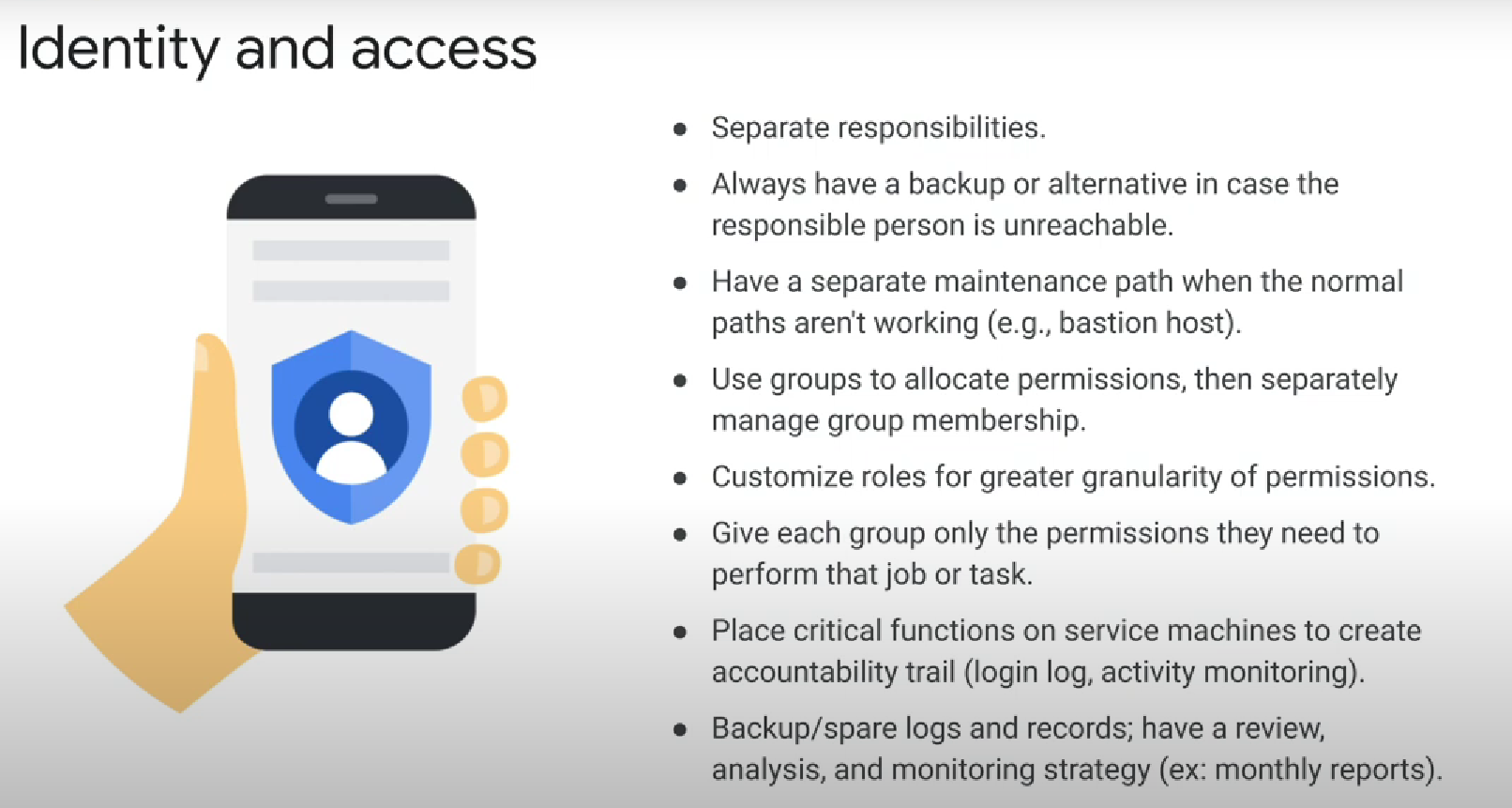
Google cloud pricing calculator can be used with BigQuery to estimate the cost of a query.

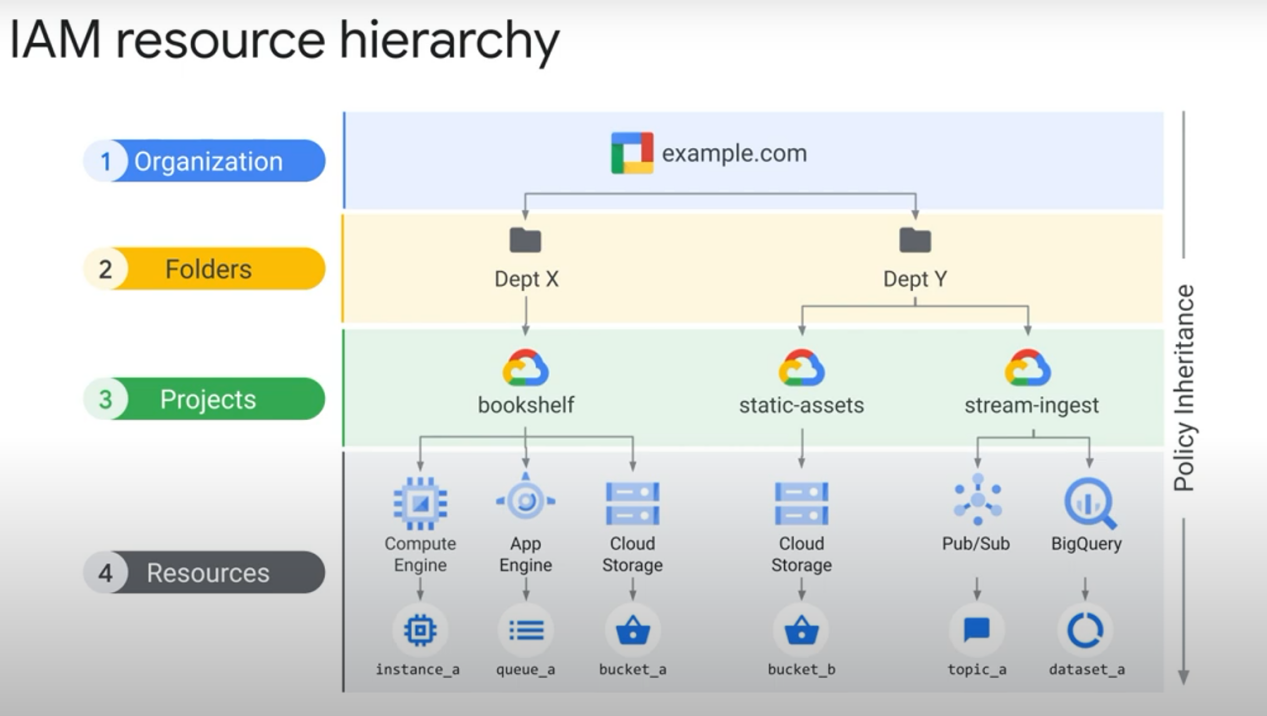
+

Query validator estimates about how much data will be used by the query. Use this data on the pricing calculator to get an idea of the cost



When you enroll in flat-rate pricing, you purchase dedicated query processing capacity, measured in BigQuery slots (cpu, memory, & network).

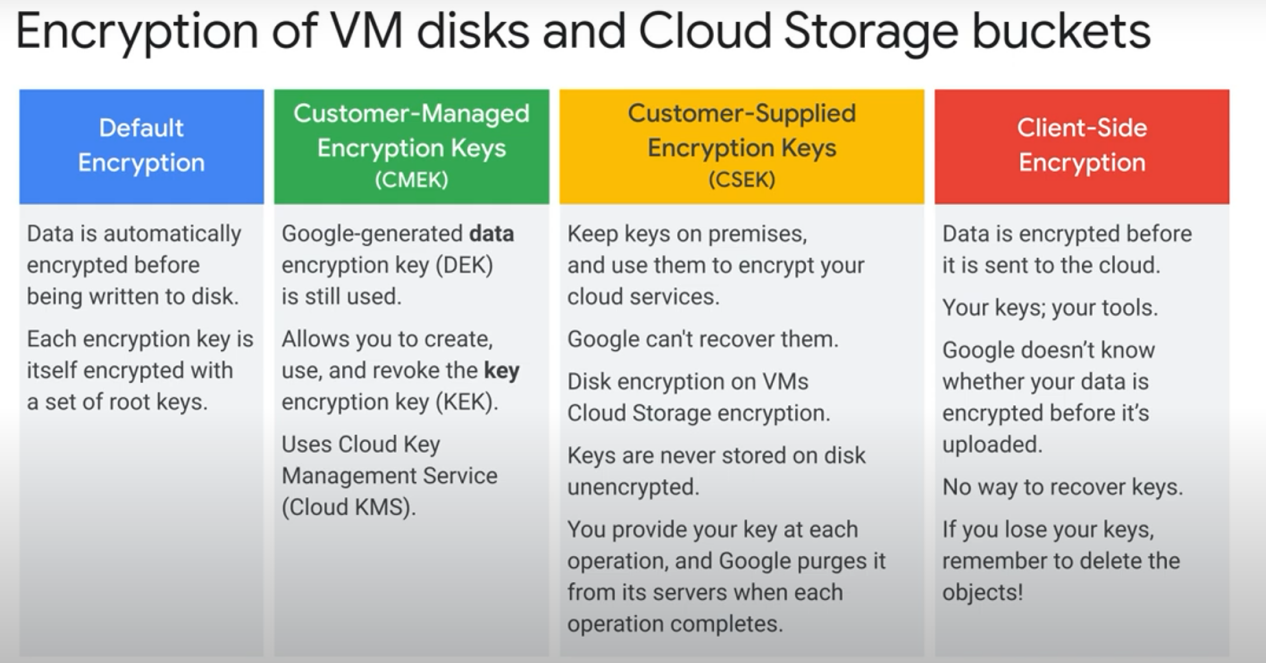




the final policy is the union of the parent policy and the resource policy what happens when these two policies are in conflict

what if the policy on the resource only gives access to a singlecloud storage bucket and restricts access to all other buckets however at the project level a rule exists that grants access to all buckets in the project

Answer: if the **parent policy is less restrictive** it **overrides** a more restrictive resource policy



key concepts

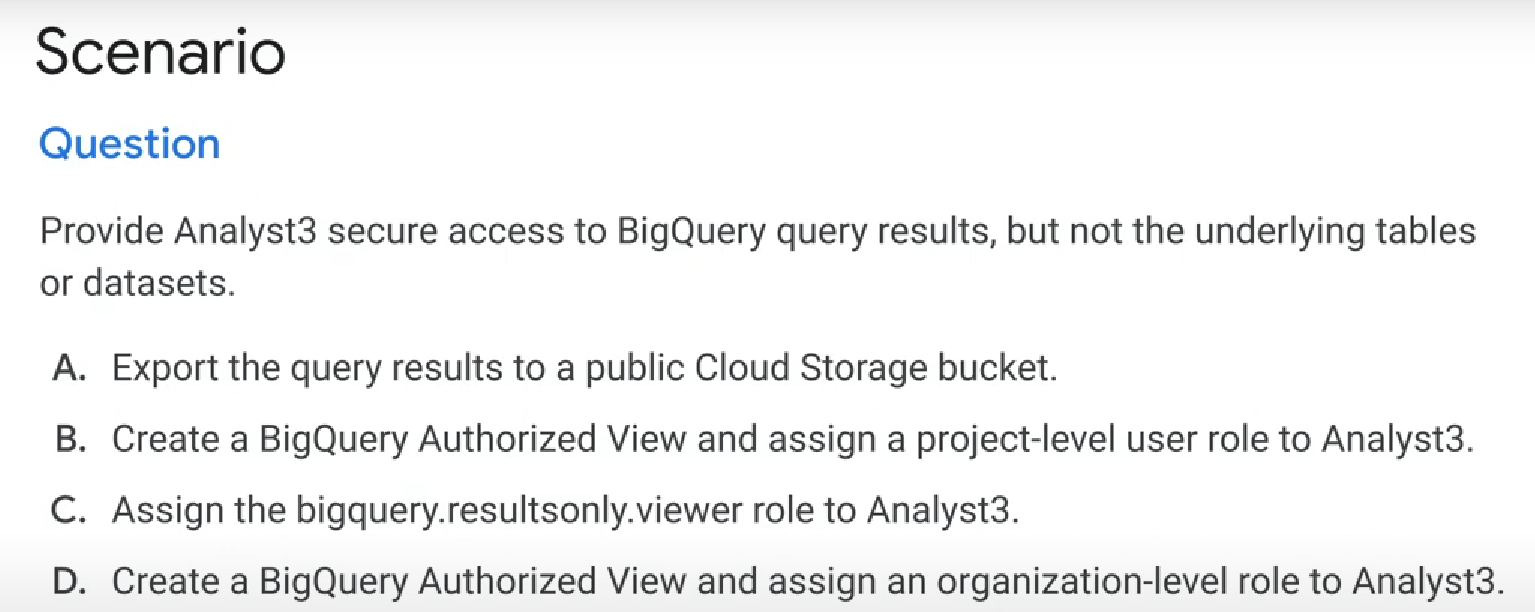
\* cloud armor: a service that protects from Ddos

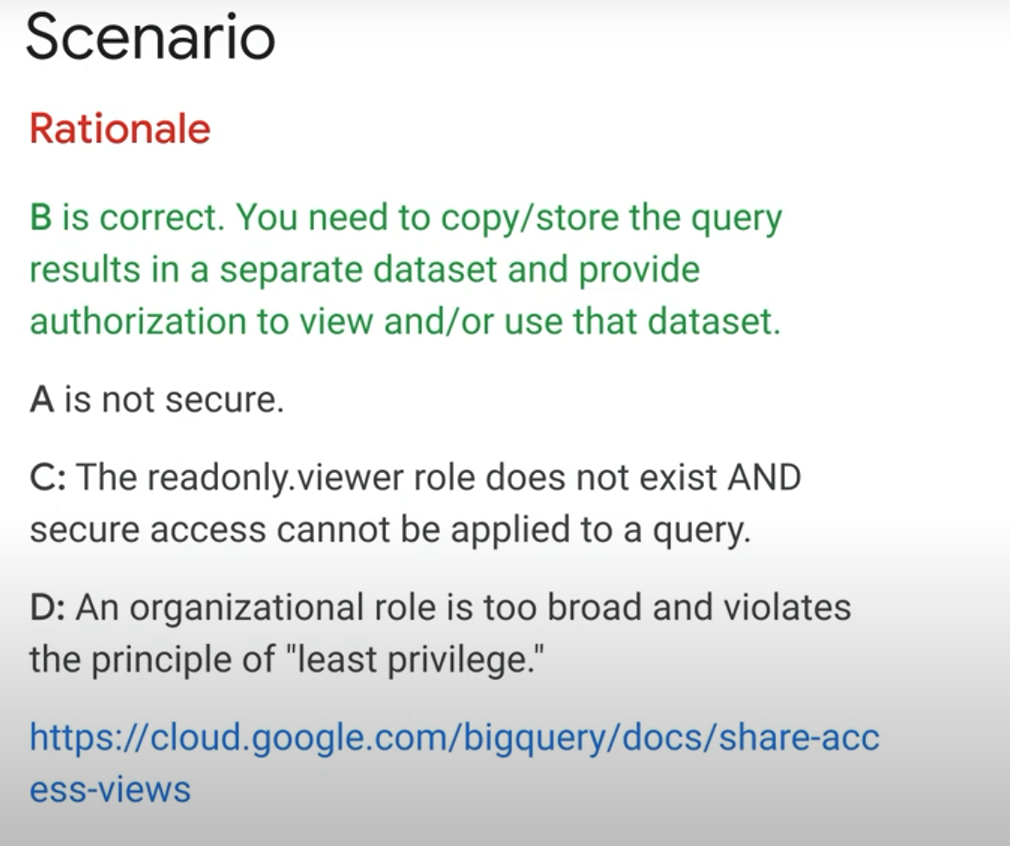
\* cloud load balancing

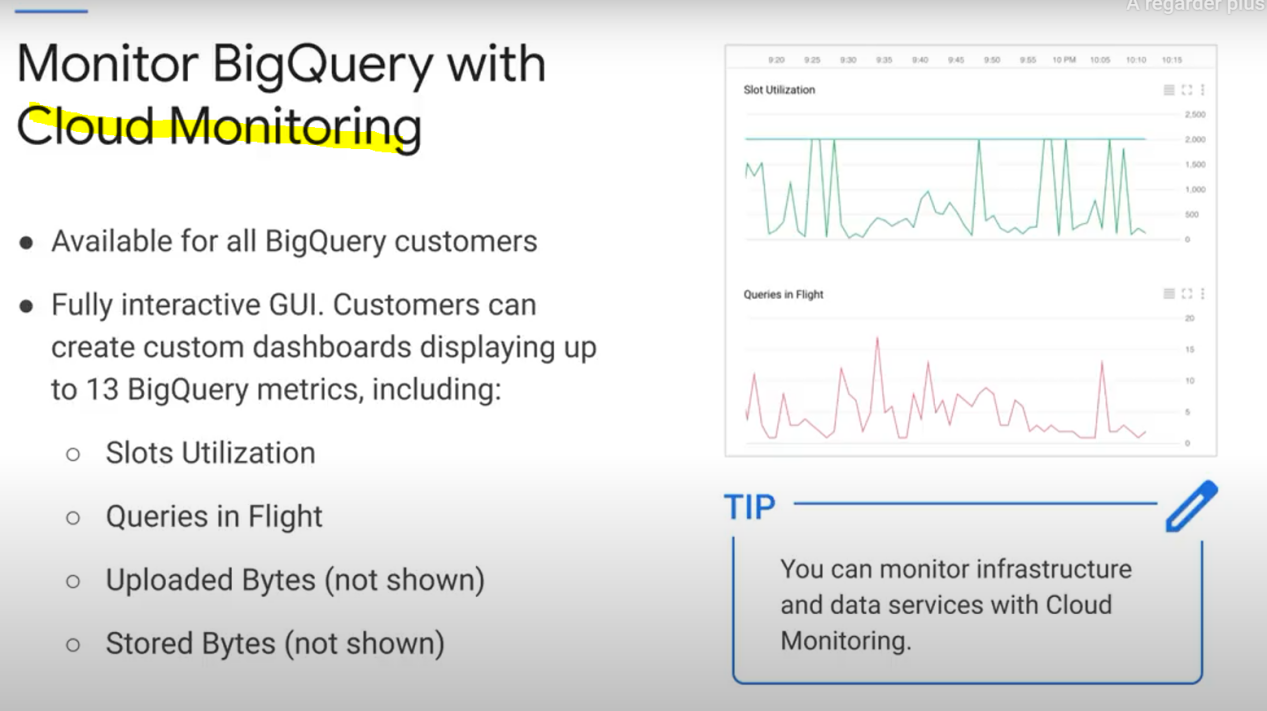
\* cloud firewall rules

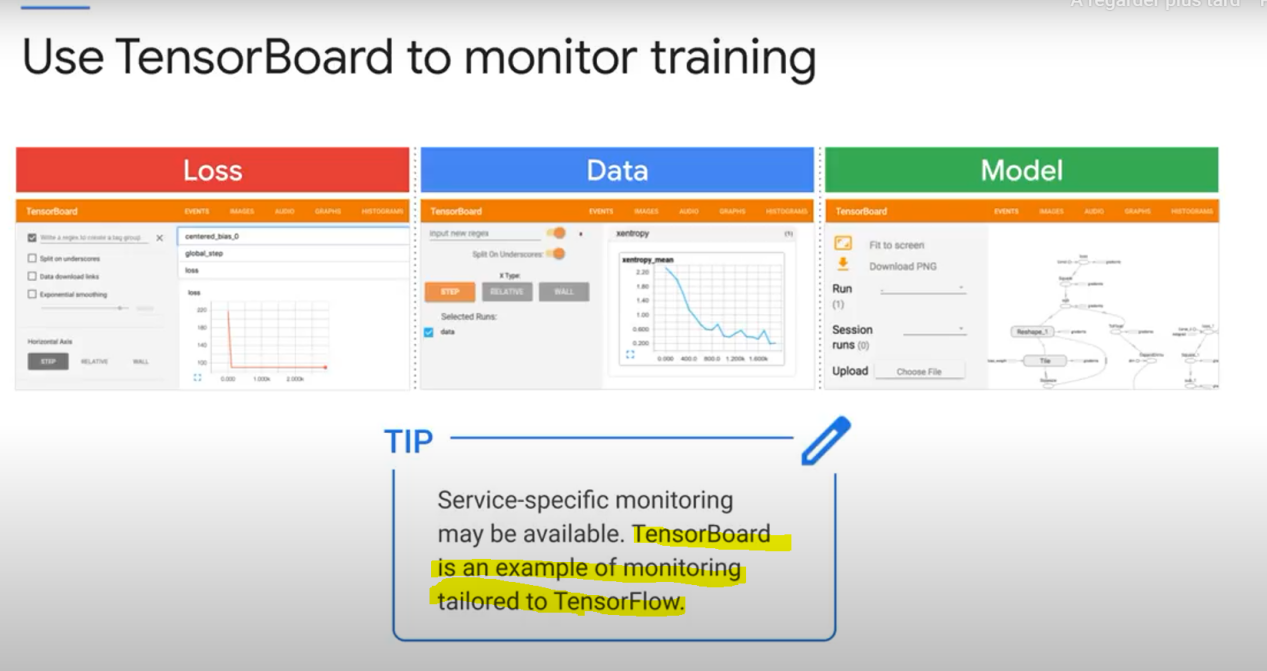
\* service accounts separation into front end and back end

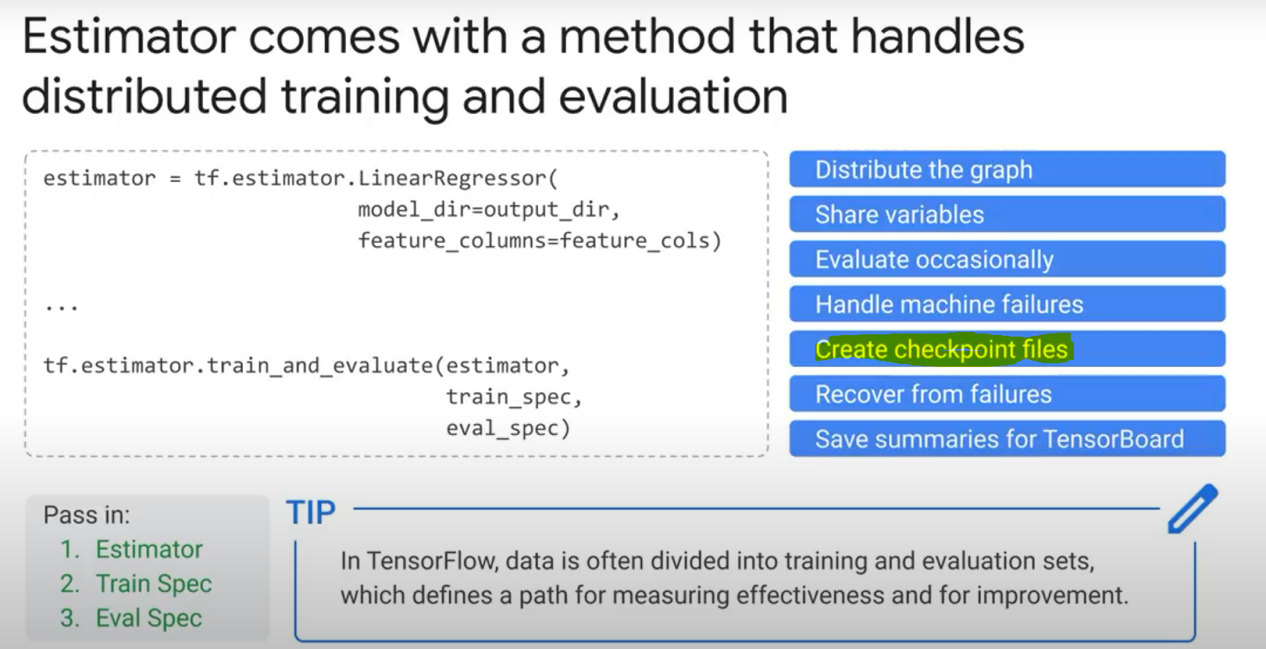
\* isolation of resources using separate service accounts between services











**Data protection (advocating policies):**

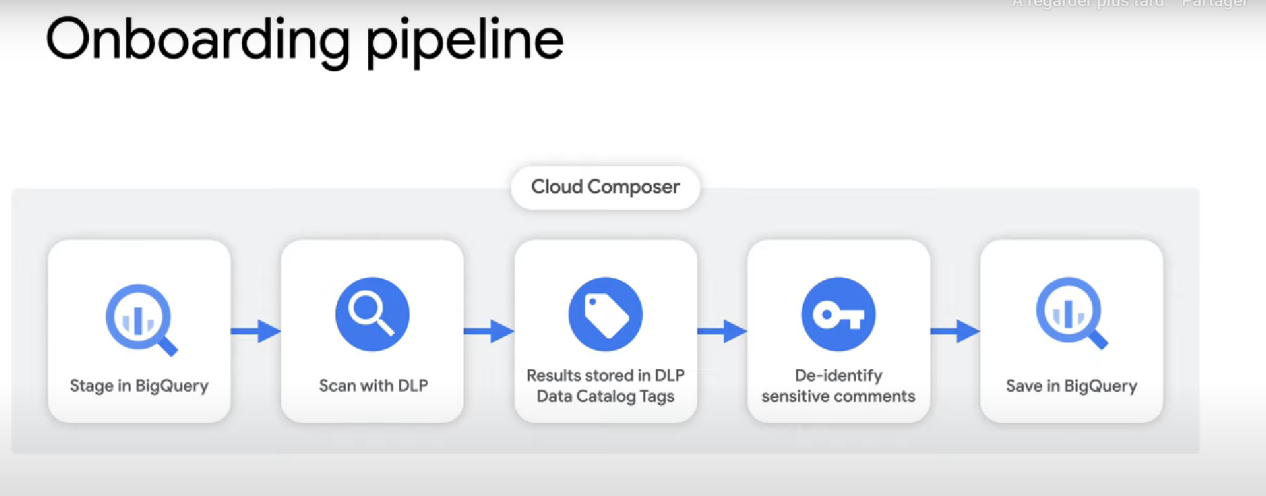
\* We can use **Data Catalog** to create a **Tag** that we will assign to certain column(s) containing personal information that we need to protect (pii)

Note: We can’t directly add a tag in bigquery to a specific column, as native bigquery tags are added to the entire table

You can use **«DLP» (cloud data loss prevention)** helps you discover sensitive data in datasets/tables.

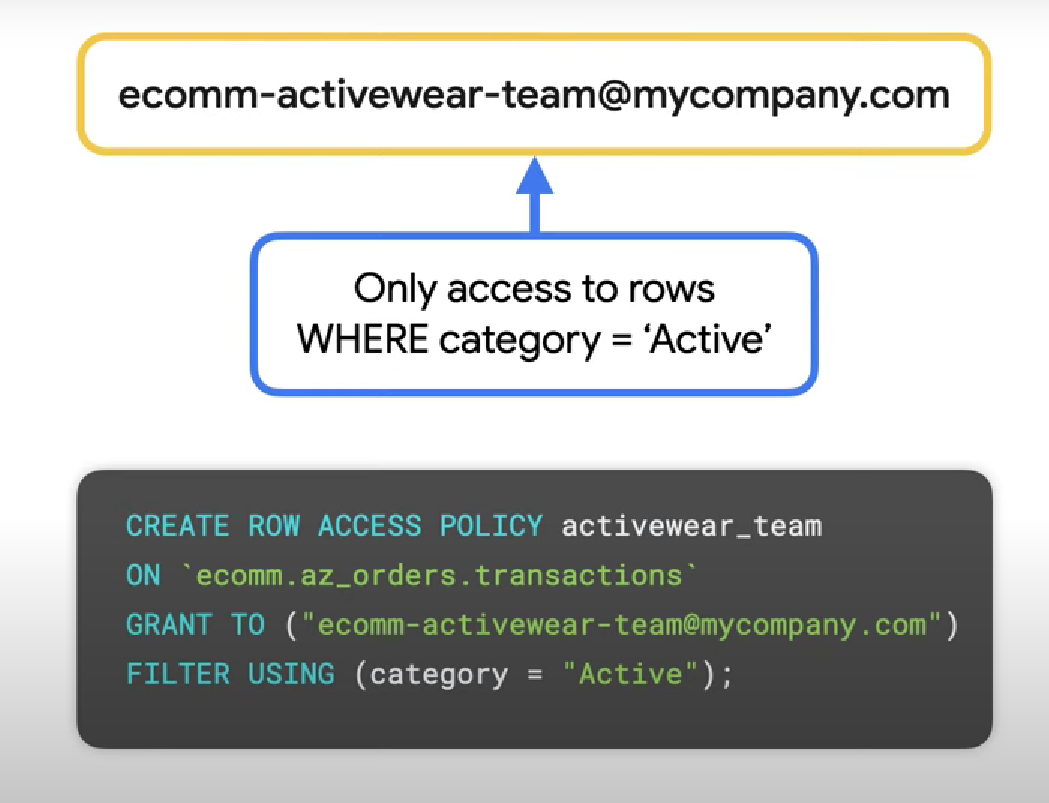
Official definition: Data security -- DLP -- Cloud DLP allows you to minimize what you collect, store, expose, or copy. Classify or automatically redact sensitive data from text streams before you write to disk, generate logs or perform analysis.

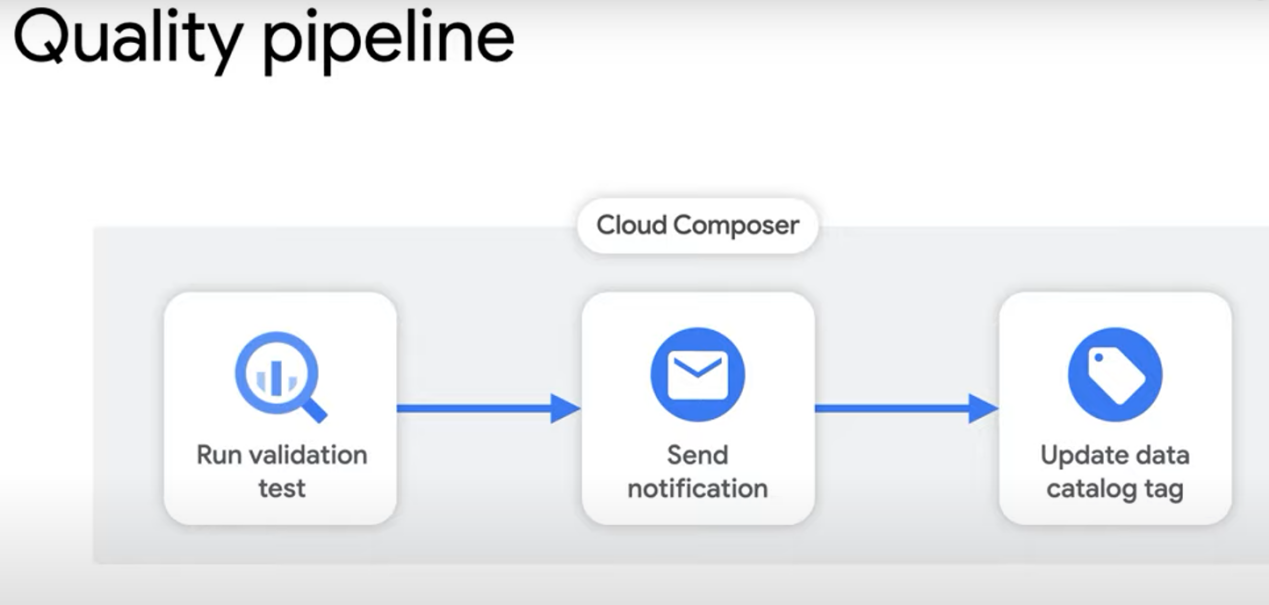
**De-identify** data: replace sensitive data with asterixes



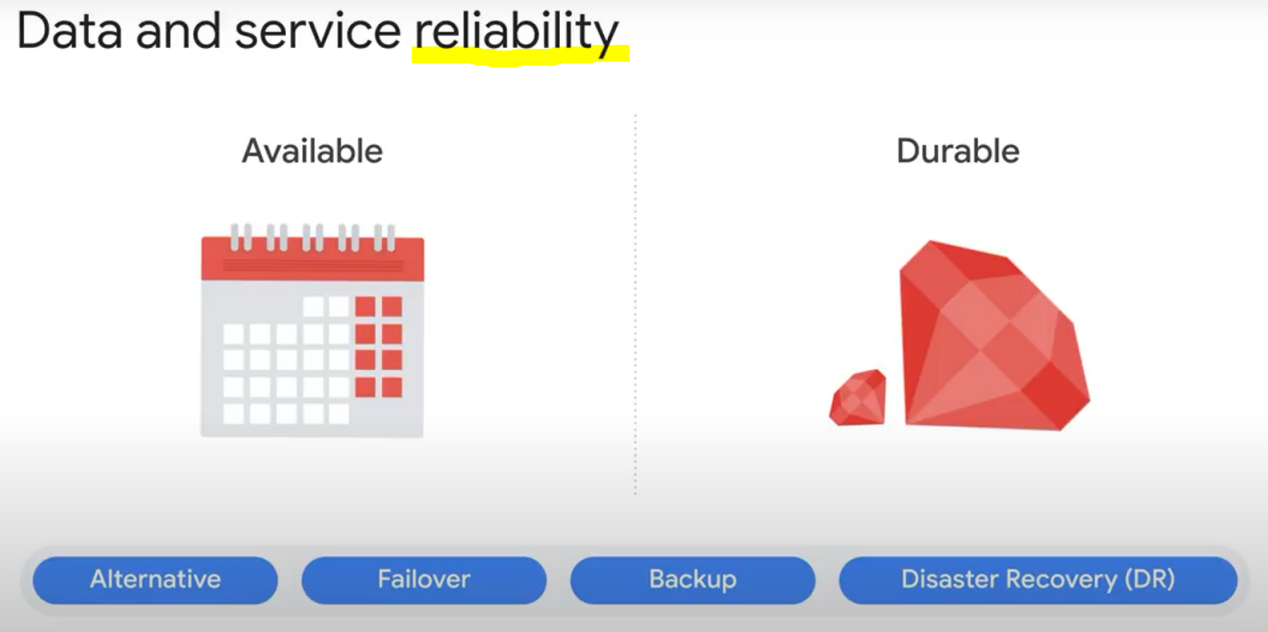
With data catalog, we can add a «clearance level» policy tag on specific column. It can have many values (E.g: low, medium, high). Then, any person who needs to read a column with a certain clearance (E.g: high), will need to be assigned the «fine-grained reader» role to the «High» ressource in data catalog.

You can also create a **row policy** that only allows certain readers to read rows with certain column/value pairs. E.g:





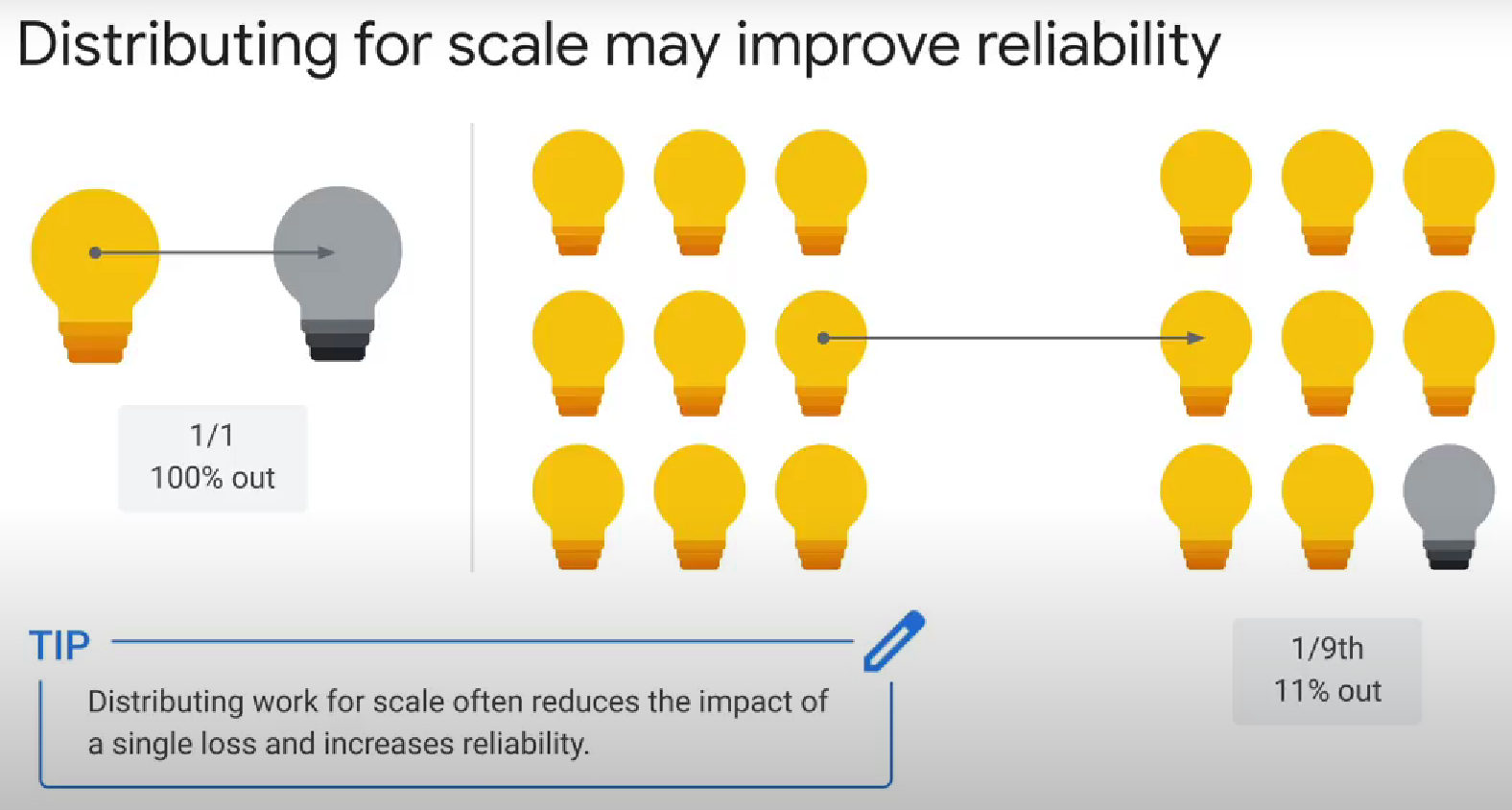
(data catalog is used here to add tags about data quality/validation to the processed data)



\* Automatic failover consists of automatically moving data or applications to the standby server if the primary system fails.

\* The alternative is a passive system where the process happens manually. Most failover processes operate automatically to reduce downtime.

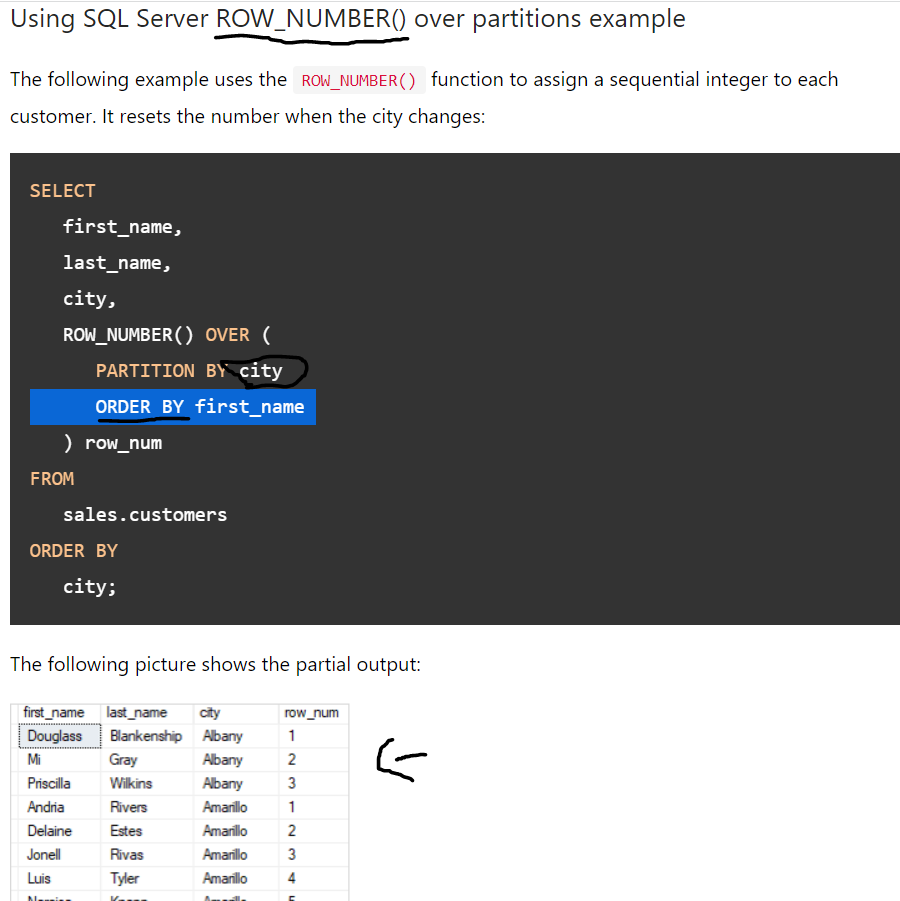
\* Backup and DR (disaster recovery) are the slowest. If we’re in a system that requires atomic transactions, we cannot accept these!



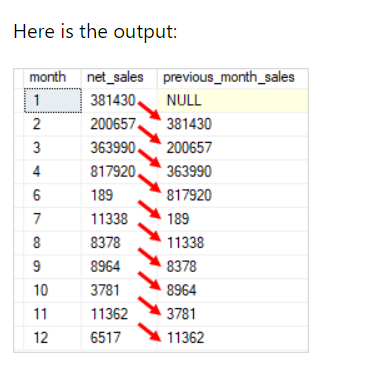
Google Cloud's operations suite = Cloud monitoring + Cloud Logging

Getting rid of duplicates when each row has an ID column, using a SQL query:

Use ROW\_NUMBER(Partition by ID, ORDER BY ID) as rowNumber, then add WHERE rowNumber = 1 (it starts at 1 not 0)



Lag function



(get 1 column from the nTH previous row (offset = n))

LEAD is the opposite of LAG

**Need-to-study if I have time:**

- HIPPA, COPPA, FedRAMP, GDPR policies (data protection)

- Partitionning VS clustering in bigquery?

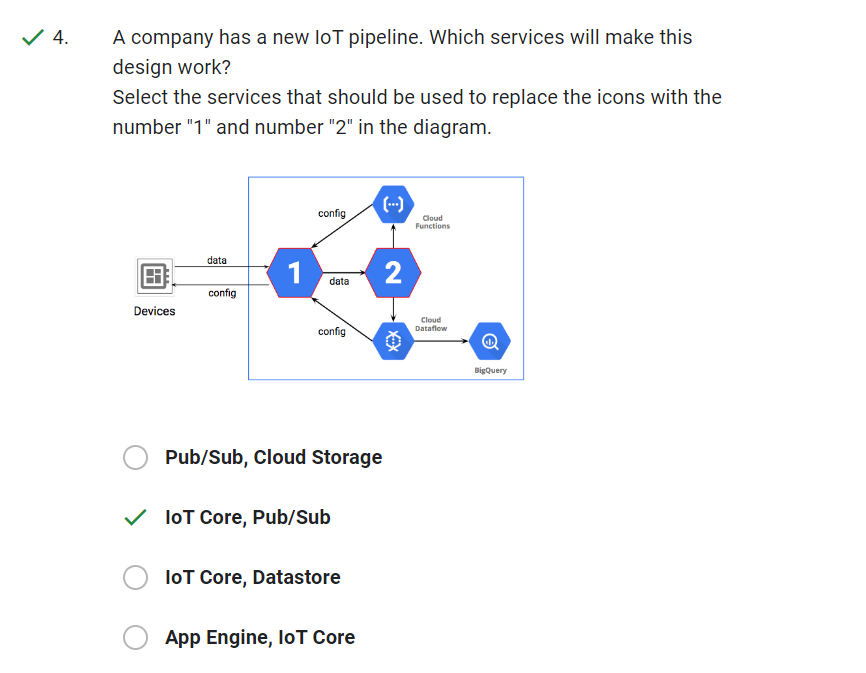
- partition expiration dates?

When you create a table partitioned by ingestion time or time-unit column, you can specify a partition expiration. This setting specifies how long BigQuery keeps the data in each partition.

- TODO: read about SIGNED URLS

- TODO: LEARN ABOUT THE IAM ROLES

- IOT Core :



- Cloud Router with VPN

- Partner VS direct interconnect:

**Dedicated Interconnect** provides a direct physical connection between your on-premises network and Google's network. **Partner Interconnect** provides connectivity between your on-premises and VPC networks through a supported service provider.

Note: If your traffic doesn't require a 10-Gbps or 100-Gbps circuit, consider **Cloud VPN** or **Partner Interconnect** instead of a dedicated interconnect.

Note 2: For partner interconnect: more points of connectivity through one of our supported service providers. Traffic flows between networks through a service provider, **not through the public internet.**