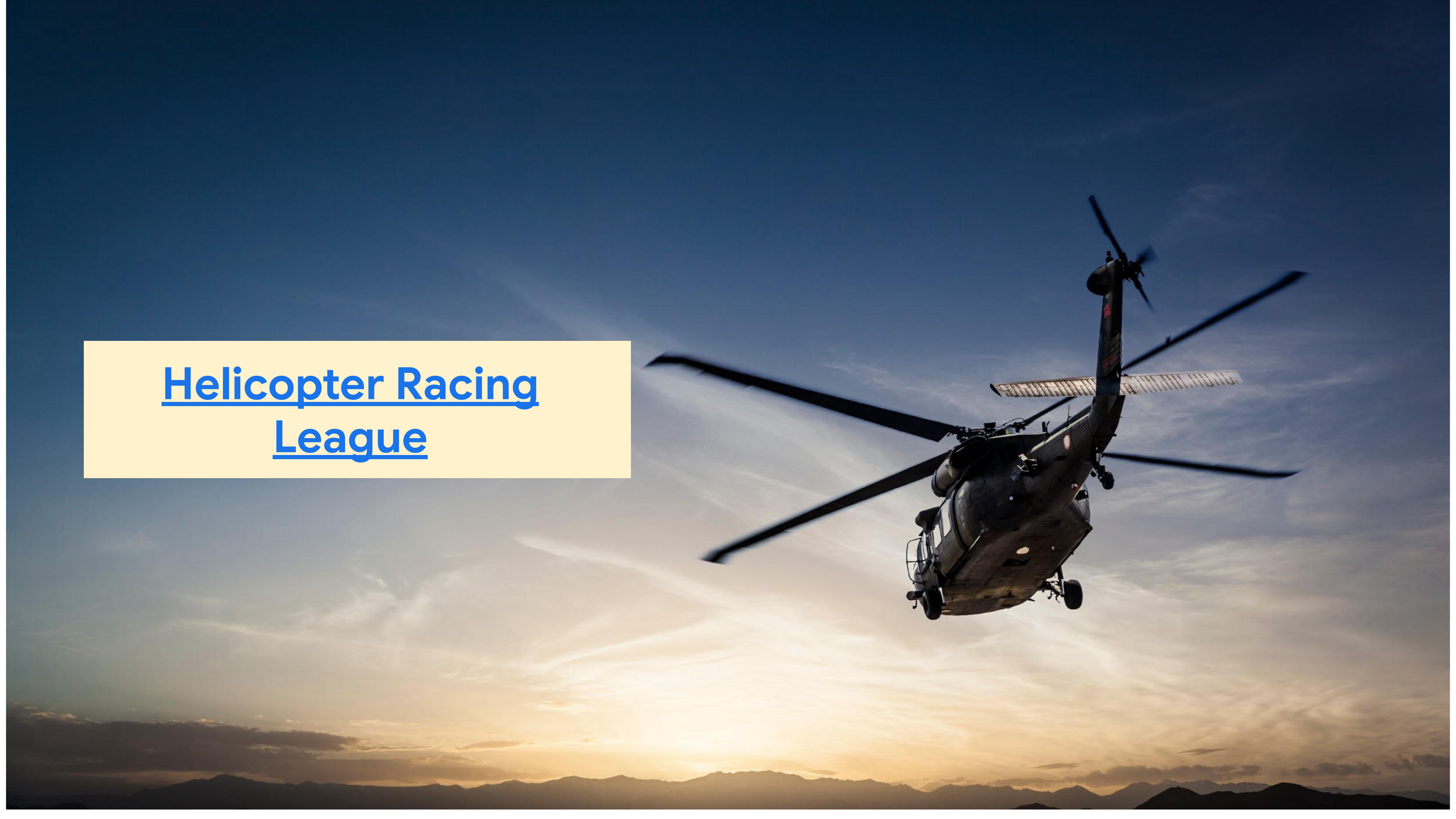
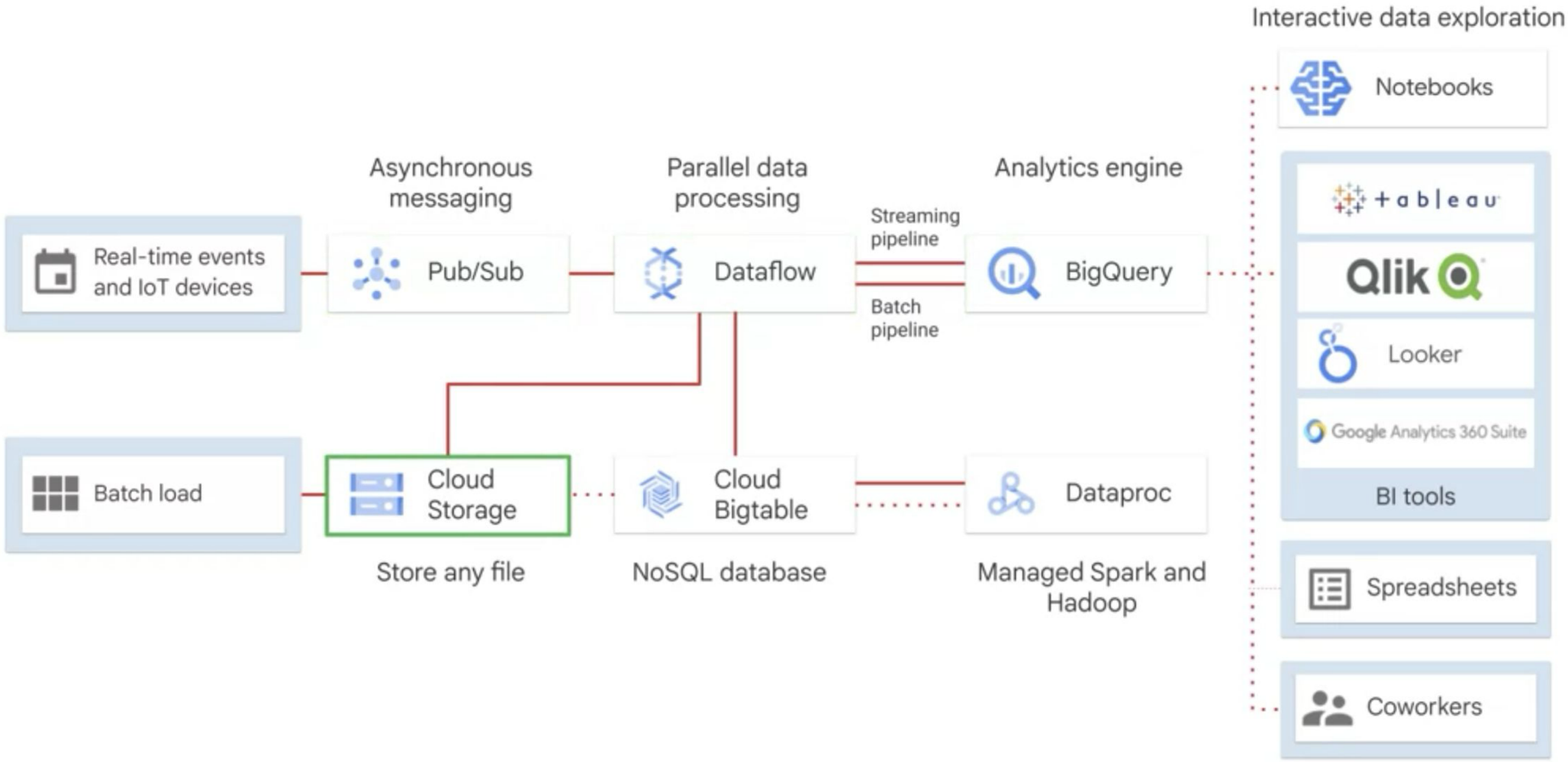
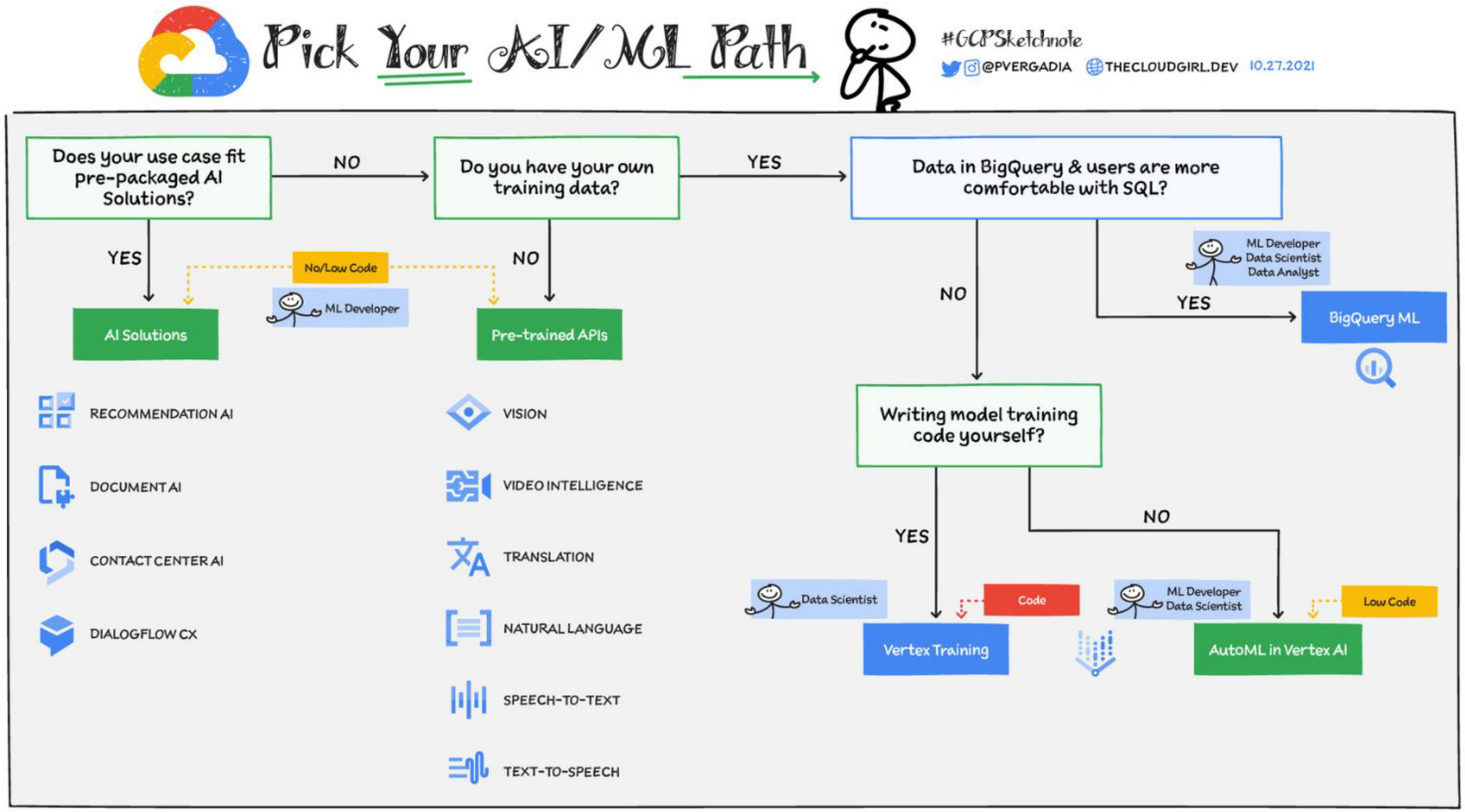
Helicopter Racing League



Modern serverless data management architecture



The Google Cloud machine learning spectrum



Exam Tip: *4 high-level ways to approach ML in GCP. All depends on particular use-case, but the approach should be like with compute: start from more managed / easier approaches and choose more advanced ones only if needed.*

Proposed Technical Solutions

- Processing environment:
  - GCE (VMs) for lift and shift, add managed instance groups for automatic scaling.
  - If they'd like to optimize for costs and scalability: containerize the app and deploy on GKE with multiple clusters in different regions.
    i. Use GKE multi-cluster capabilities or Anthos to manage/orchestrate workloads across clusters.
  - Transcoding API to perform video transcoding (from raw streams to different quality of ready to consume video files).
  - **Video Intelligence Streaming API with HLS Protocol for live streaming.**
- Connectivity between GCP and other hyperscalers: standard patterns.
- Multi-regional GCS buckets for increased availability and reduced latency to end users.
  - Use on-line Storage Transfer Service to migrate the data from other cloud provider to GCP.
- Global HTTPS Load Balancer + Cloud CDN (serving data from GCS buckets or from a different hyperscaler) + Cloud Armor.
- Preemptible VMs (maybe with GPUs or even TPUs); Spot VMs are “2nd generation of Preemptible VMs”.
- Telemetry
  - Native services (Cloud Operations Suite), maybe stronger focus on Cloud Trace (importance of latency) and VPC Flow Logs.
  - Network Intelligence Center for more insights into networking topology / connectivity / performance.
- **BigQuery for real-time analytics and data mart**
  - (if ingested data is structured) GCS bucket => BigQuery
  - (if ingested data needs transformations) GCS bucktes => (optionally if streaming) Pub/Sub => Dataflow => BigQuery
  - Data Studio or Looker for visualization and analytics
- AI/ML:
  - Enriching video (tags, labels, object detection etc): Video Intelligence API.
  - **Predicting race results, experimental forecasting: Vertex AI** with Jupyter Notebooks and options to deploy and expose the model to partners.
  - Viewers sentiment analysis: Natural Language API.
- **Expose APIs / ML models to partners: Apigee (monetization, rate limiting, merchandising revenue stream etc)**

[HRL case study] Diagnostic Question #1

For this question, refer to the Helicopter Racing League (HRL) case study. HRL is looking for a cost-effective approach for storing their race data such as telemetry. They want to keep all historical records, train models using only the previous season's data, and plan for data growth in terms of volume and information collected. You need to propose a data solution.

Considering HRL business requirements and the goals expressed by CEO S. Hawke, what should you do?

A. Use Firestore for its scalable and flexible document-based database. Use collections to aggregate race data by season and event.

B. Use Cloud Spanner for its scalability and ability to version schemas with zero downtime. Split race data using season as a primary key.

C. Use BigQuery for its scalability and ability to add columns to a schema. Partition race data based on season.

D. Use Cloud SQL for its ability to automatically manage storage increases and compatibility with MySQL. Use separate database instances for each season.

[HRL case study] Diagnostic Question #1

For this question, refer to the Helicopter Racing League (HRL) case study. HRL is looking for a cost-effective approach for storing their race data such as telemetry. They want to keep all historical records, train models using only the previous season's data, and plan for data growth in terms of volume and information collected. You need to propose a data solution.

Considering HRL business requirements and the goals expressed by CEO S. Hawke, what should you do?

A. Use Firestore for its scalable and flexible document-based database. Use collections to aggregate race data by season and event.

B. Use Cloud Spanner for its scalability and ability to version schemas with zero downtime. Split race data using season as a primary key.

C. Use BigQuery for its scalability and ability to add columns to a schema. Partition race data based on season.

D. Use Cloud SQL for its ability to automatically manage storage increases and compatibility with MySQL. Use separate database instances for each season.

[HRL case study] Diagnostic Question #2

For this question, refer to the Helicopter Racing League (HRL) case study. A recent finance audit of cloud infrastructure noted an exceptionally high number of Compute Engine instances are allocated to do video encoding and transcoding. You suspect that these Virtual Machines are zombie machines that were not deleted after their workloads completed. You need to quickly get a list of which VM instances are idle.

What should you do?

A. Log into each Compute Engine instance and collect disk, CPU, memory, and network usage statistics for analysis.

B. Use the gcloud compute instances list to list the virtual machine instances that have the idle: true label set.

C. Use the gcloud recommender command to list the idle virtual machine instances.

D. From the Google Console, identify which Compute Engine instances in the managed instance groups are no longer responding to health check probes.

[HRL case study] Diagnostic Question #2

For this question, refer to the Helicopter Racing League (HRL) case study. A recent finance audit of cloud infrastructure noted an exceptionally high number of Compute Engine instances are allocated to do video encoding and transcoding. You suspect that these Virtual Machines are zombie machines that were not deleted after their workloads completed. You need to quickly get a list of which VM instances are idle.

What should you do?

A. Log into each Compute Engine instance and collect disk, CPU, memory, and network usage statistics for analysis.

B. Use the gcloud compute instances list to list the virtual machine instances that have the idle: true label set.

C. Use the gcloud recommender command to list the idle virtual machine instances.

D. From the Google Console, identify which Compute Engine instances in the managed instance groups are no longer responding to health check probes.

[HRL case study] Diagnostic Question #3

For this question, refer to the Helicopter Racing League (HRL) case study. Your team is in charge of creating a payment card data vault for card numbers used to bill tens of thousands of viewers, merchandise consumers, and season ticket holders. You need to implement a custom card tokenization service that meets the following requirements:

- It must provide low latency at minimal cost.
- It must be able to identify duplicate credit cards and must not store plaintext card numbers.
- It should support annual key rotation.

Which storage approach should you adopt for your tokenization service?

A. Store the card data in Secret Manager after running a query to identify duplicates.

B. Encrypt the card data with a deterministic algorithm stored in Firestore using Datastore mode.

C. Encrypt the card data with a deterministic algorithm and shard it across multiple Memorystore instances.

D. Use column-level encryption to store the data in Cloud SQL.

[HRL case study] Diagnostic Question #3

For this question, refer to the Helicopter Racing League (HRL) case study. Your team is in charge of creating a payment card data vault for card numbers used to bill tens of thousands of viewers, merchandise consumers, and season ticket holders. You need to implement a custom card tokenization service that meets the following requirements:

- It must provide low latency at minimal cost.
- It must be able to identify duplicate credit cards and must not store plaintext card numbers.
- It should support annual key rotation.

Which storage approach should you adopt for your tokenization service?

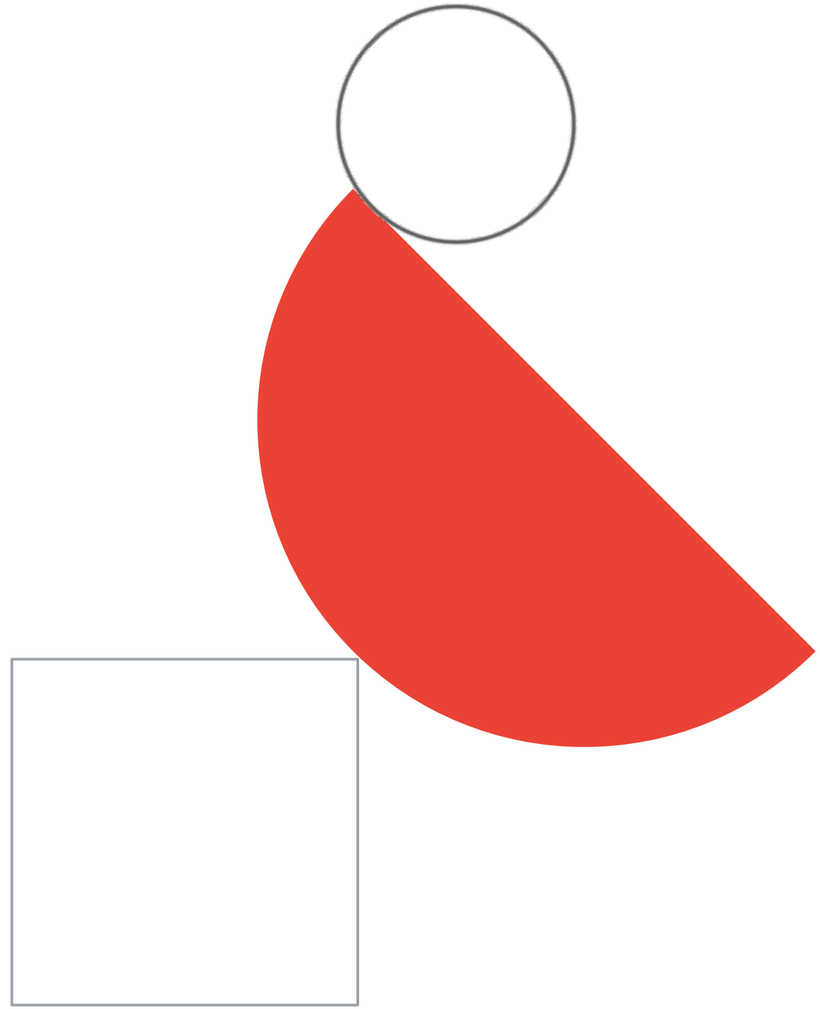
A. Store the card data in Secret Manager after running a query to identify duplicates.

B. Encrypt the card data with a deterministic algorithm stored in Firestore using Datastore mode.

C. Encrypt the card data with a deterministic algorithm and shard it across multiple Memorystore instances.

D. Use column-level encryption to store the data in Cloud SQL.

[optional] Links to useful materials



Optional materials 1

[READING]

- Boto configuration file | Cloud Storage
- Use customer-supplied encryption keys | Cloud Storage
- Object change notification | Cloud Storage
- Read about Database Migration Service.
- What are the options for connecting to Cloud SQL instance:
  a. https://cloud.google.com/sql/docs/mysql/connect-overview
  b. https://cloud.google.com/sql/docs/postgres/external-connection-methods
- How to choose optimal AI/ML path in GCP?

[VIDEOS]

- Cloud Networking 103 (Securing Network): Cloud OnAir: CE TV: Google Cloud Networking 103 - Securing your Network
- Google Cloud Storage options: Difference between object store, block store and file store | Google Cloud Storage options
- GCS Offline Transfer Appliance: Introducing Google Cloud's Transfer Appliance
- How to transfer data to GCS: How to transfer data to Google Cloud? #GCPSketchnote
- Authentication controls for Cloud Storage
- What's new with Cloud SQL

Optional materials 2

- [IMPORTANT TO KNOW] Different patterns for connecting to Cloud SQL: Cloud SQL: Concepts of Networking
- Great demo of how to centralize network management and set up Shared VPC in GCP: Level Up From Zero Episode 4: Shared VPC
- Accelerating cloud migrations with managed databases: Accelerating cloud migration with managed databases
- [Highly recommended] Choose your database on Google Cloud: Choose your database on Google Cloud
- Introducing Database Migration Service: Introducing Database Migration Service
- How to achieve high resiliency and availability with GCP: How to achieve high resiliency and availability with Google Cloud infrastructure
- Deploying MongoDB via GCP Marketplace: Deploying MongoDB from Google Cloud Marketplace
- Infrastructure as code with Terraform and Cloud Run: Infrastructure as code with Terraform and Cloud Run
- Build ETL Pipelines using Cloud Dataflow: Build ETL Pipelines using Cloud Dataflow

Optional materials 3

[PODCASTS]

- Cloud SQL
- Database Migration Service
- Beam and Spark
- Cloud Dataflow

[DEEP DIVES]

- The battle of relational and non-relational databases | SQL vs NoSQL Explained: The battle of relational and non-relational databases | SQL vs NoSQL Explained
- [video] How to accelerate migration to GCP: Tools and services to accelerate your migration to Google Cloud
- 5 ways Google can help you succeed in the multicloud world.