GDG Cloud Chennai

Google Cloud
**Community Day
2022 Chennai**

📅 July 30th, 2022
📍 Infosys Limited,
Sholinganallur, Chennai

Experts
Google Cloud

Data Governance
on GCP

Nirav Kothari
Technical Architect

Agenda

- Introduction

- Features

- GCP Services

GDG Cloud Chennai

What is Data Governance

Data governance is everything you do to ensure data is **secure**, **private**, **accurate**, **available**, and **usable**. It includes the actions **people** must take, the **processes** they must follow, and the **technology** that supports them throughout the data life cycle

CIA Triad

A model designed to guide policies for information security within an organization

Data Governance Features

- Data Discoverability

- Data Security

- Data Quality

- Data Retention / Disposal

GDG Cloud Chennai

Data Discoverability / Metadata Management

- Visibility of the data available for use
- Technical metadata
  - Column names
  - Field types
  - Record count
  - Last refreshed time
- Business metadata
  - Data domain
  - Data classification based on sensitivity
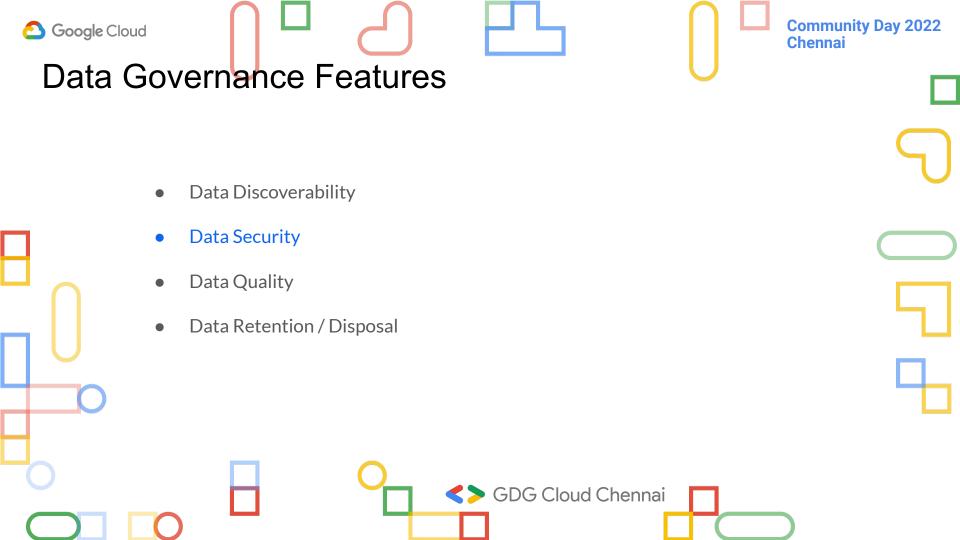 - PII or Non-PII
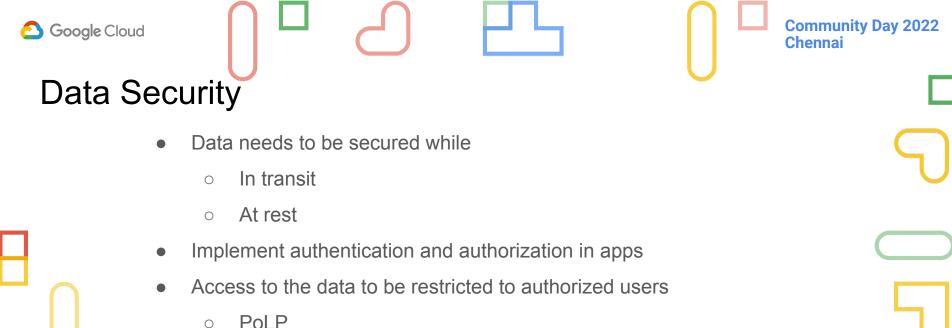- Data Lineage
  - Source of data
  - What transformation applied

GDG Cloud Chennai

Data Discoverability - GCP

- Metadata as a service

- Central Catalog

- Search and Discovery

- Fully Managed, Serverless, Highly

  Scalable

- Cloud IAM integration

- Cloud DLP integration

- On-prem connectors

GDG Cloud Chennai

Cloud Data Catalog - Demo

Data Governance Features

- Data Discoverability

- Data Security

- Data Quality

- Data Retention / Disposal

GDG Cloud Chennai

Data Security

- Data needs to be secured while
  - In transit
  - At rest
- Implement authentication and authorization in apps
- Access to the data to be restricted to authorized users
  - PoLP
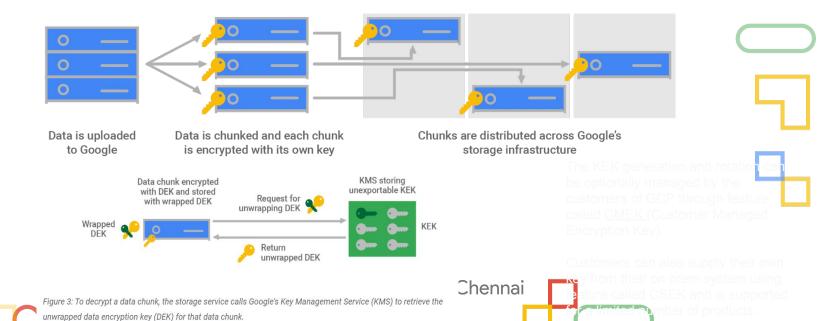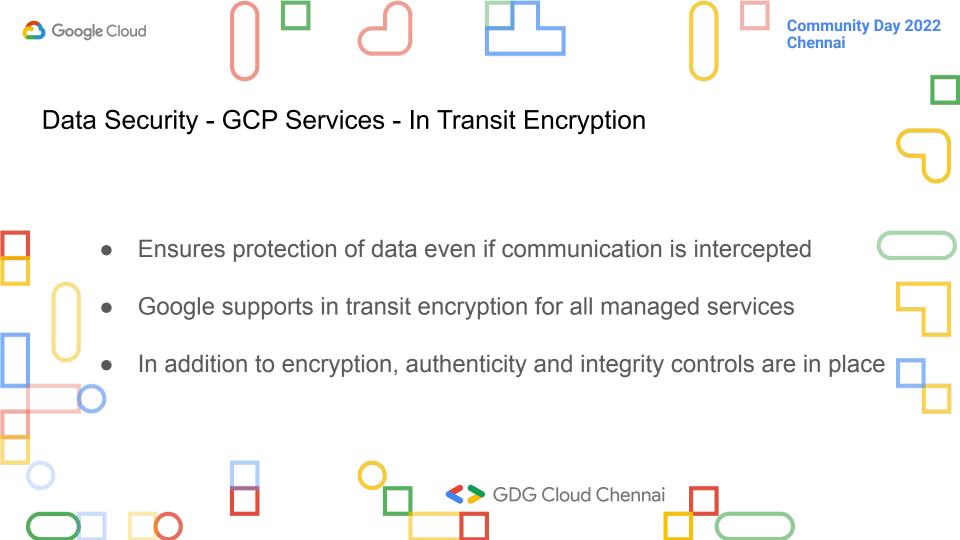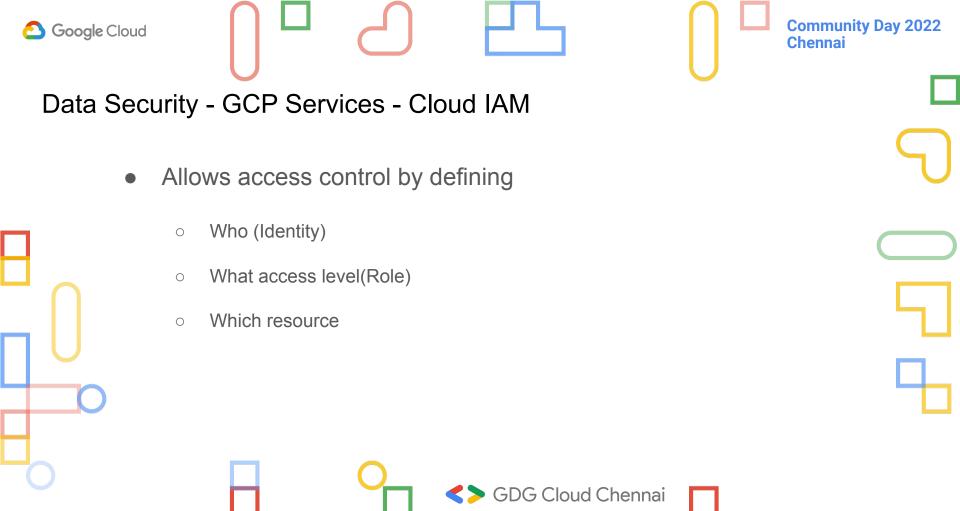- Personally identifiable data should be protected and kept hidden

GDG Cloud Chennai

Data Security - GCP Services - At Rest Encryption

- GCP, by default, encrypts data at rest using AES256
- Malicious access to disk does not result in access to data
- Google encrypts data prior to being written

Data is uploaded to Google

Data is chunked and each chunk is encrypted with its own key

Chunks are distributed across Google's storage infrastructure

Data chunk encrypted with DEK and stored with wrapped DEK

Request for unwrapping DEK

KMS storing unexportable KEK

Wrapped DEK

KEK

Return unwrapped DEK

The KEK generation and rotation can be optionally managed by the customers of GCP through feature called CMEK (Customer Managed Encryption Key).

Customers can also supply their own key from their on-prem system using feature called CSEK and is supported for a limited number of products.

Chennai

Figure 3: To decrypt a data chunk, the storage service calls Google's Key Management Service (KMS) to retrieve the unwrapped data encryption key (DEK) for that data chunk.

Data Security - GCP Services - In Transit Encryption

- Ensures protection of data even if communication is intercepted

- Google supports in transit encryption for all managed services

- In addition to encryption, authenticity and integrity controls are in place

GDG Cloud Chennai

Data Security - GCP Services - Cloud IAM

- Allows access control by defining

  - Who (Identity)

  - What access level(Role)

  - Which resource

GDG Cloud Chennai

Data Security - Data Privacy

- Relates to proper handling of data
  - How you collect it
  - How you use it
  - Is the data hosted within country / region permitted by governing body
  - Maintain compliance

- Example data privacy regulations
  - GDPR
  - HIPAA
  - CCPA

GDG Cloud Chennai

Data Security - GCP Services - Cloud DLP

Managed service designed to help you protect sensitive data -

- Discover
- Classify
- Protect

Features -

- Data discovery and Classification
- Data masking / De-identification of PII
- Integrate with other GCP services

GDG Cloud Chennai

Data Governance Features

- Data Discoverability

- Data Security

- Data Quality

- Data Retention / Disposal

GDG Cloud Chennai

Data Quality

Degree to which the data is available at the time it is needed

Degree to which necessary data is available for use

Degree to which the data represents the reality

Degree to which data is unique and cannot be mistaken for other entries

Degree to which the data is within defined requirements like formate, type and range

Degree to which the data is equal within and between datasets

Timeliness

Completeness

Accuracy

Uniqueness

Validity

Consistency

6 1
5 2
4 3

GDG Cloud Chennai

Data Quality - GCP Services

- Checks can be implemented at multiple layers

    - Web application

    - Schema validation at API level

    - Database schema validation

- Data Quality Monitoring solution

    - With evolving set of rules

- GCP - Cloud Dataprep

GDG Cloud Chennai

Data Governance Features

- Data Discoverability

- Data Security

- Data Quality

- Data Retention / Disposal

GDG Cloud Chennai

Data Retention

- Data needs to be retained for

  - Operational requirements

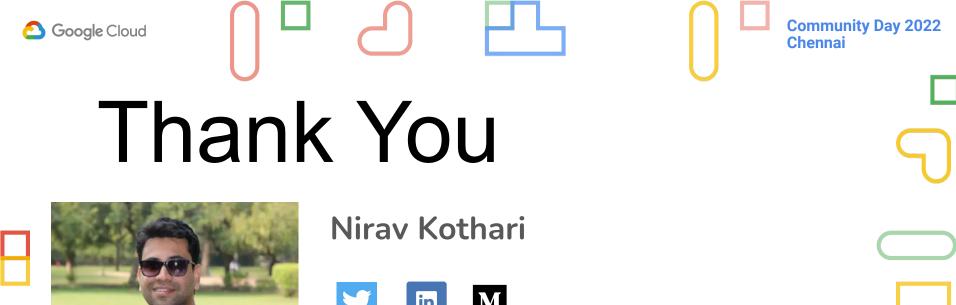 - Meet regulatory requirements

- Should not be retained beyond retention period

  - Regulatory requirements

- Approaches for disposing data

  - Physically remove data

  - Make data inaccessible

    - Anonymisation

    - Delete decryption key

GDG Cloud Chennai