

## Migration of Data to GCP (ETL/Pipeline) - Documentation

### ***Overview:***

The data migration journey involves the seamless transfer of information from local databases to Google Cloud Platform (GCP) using the Extract, Transform, Load (ETL) method. This method is chosen for its efficiency, compatibility with GCP services, and robust data security measures, ensuring a smooth and secure transition of the data assets.

### **Step-by-Step Instructions:**

#### ***Get the Data***

- Initiate the migration process by meticulously extracting all necessary data from local databases, ensuring comprehensive coverage and accurate capture of required information.
- Validate the extracted data to ensure integrity and completeness before proceeding.

#### ***Prepare the Data:***

- Focus on refining the extracted data to enhance usability, accuracy, and security.
- Create a normalized database schema, built out from the LLP, to include at least 3 separate dimension tables containing relevant attributes useful for data analysis, following a star schema structure.
- Transform categorical data into numerical formats for efficient analysis.
- Safeguard personal information by hashing social security numbers and pseudonymizing phone numbers to maintain privacy while retaining analytical utility.

#### ***Put the Data in GCP:***

- Securely store the refined data in Google Cloud Storage, organizing it for easy access.
- Leverage GCP's data transfer services to seamlessly move the data from Cloud Storage to its designated destination within GCP.
- Explore alternative options for larger datasets, considering factors such as bandwidth constraints and cost optimization.

#### ***Keep it Safe:***

- Implement robust measures to safeguard the confidentiality, integrity, and availability of data assets.
- Apply encryption techniques to protect data during transit and at rest.
- Configure access controls and permissions to restrict access to sensitive information.
- Adhere to Google Cloud's security best practices and guidelines, including multi-factor authentication and audit logging.

- Establish an audit strategy to monitor data access and track changes, ensuring transparency and accountability.

***Watch and Check:***

- Continuously monitor the migration process to detect and address issues promptly.
- Configure automated alerts and notifications to notify stakeholders of deviations from expected behaviors.
- Perform comprehensive checks and validations post-migration to ensure data integrity and consistency.
- Monitor performance metrics and key performance indicators to evaluate the efficiency of the migration process and identify areas for improvement.

**Conclusion:**

The meticulous approach to data migration ensures the secure and seamless transfer of data to Google Cloud Platform, preserving data integrity, confidentiality, and compliance with regulatory requirements. By following clear guidelines and best practices, risks are mitigated, ensuring a successful migration process and laying the foundation for effective data management and analysis in the cloud environment.