

Building and Operationalizing Data Processing Systems

Exam Guide

Building and operationalizing storage systems.	<ul style="list-style-type: none">• Effective use of managed services.• Storage costs and performance• Data lifecycle management
---	--

Tip: Data management is often influenced by business requirements. After the data has been used for the "live" application, is it collected for reporting, for backup and recovery, for audits, or for legal compliance? What are the changing business purposes of the data in different time frames?

Tip: "Effective use of managed services": Choose the right service and the correct settings/features for specific use cases.

Consider costs, performance, and effective use cases (key features) for these:

- Cloud Bigtable
- Cloud Spanner
- Cloud SQL
- BigQuery
- Cloud Storage
- Cloud Datastore
- Cloud Memorystore

Exam Guide

Building and operationalizing pipelines.	<ul style="list-style-type: none">• Data cleansing• Batch and streaming• Transformation• Data acquisition and import• Integrating with new data sources
---	---

Tip: What is Data cleansing? Data cleansing is improving the data quality through consistency. You could use Cloud Dataprep to Extract, Transform, or Load (ETL). You could run a data transformation job on Cloud Dataproc.

Tip: Batch and streaming together? You should already be thinking "Cloud Dataflow."

Tip: Integrating with new data sources? You should be familiar with the connectors available between services in the cloud and common import/acquisition configurations.

Exam Guide

Building and operationalizing processing infrastructure.	<ul style="list-style-type: none">• Provisioning resources• Monitoring pipelines• Adjusting pipelines• Testing and quality control
---	---

Tip: Testing and quality control and monitoring: You should be familiar with the common approaches to testing that are used in production environments, such as A/B testing, and other rollout scenarios. Similarly, there are operational and administrative monitoring elements for most services in the GCP Console, and statistical and log monitoring in Stackdriver. Do you know how to enable and use Stackdriver with common services?