# Ensuring Solution Quality: Exam Guide Review

## Security

| | |
|---|---|
| **Designing for security and compliance.** | **Identity and access management**<br>**Data security**<br>Ensuring privacy<br>Legal compliance |

**Tip:** IAM -- Understand permissions and custom roles. Under what conditions are custom roles preferred over standard predefined roles?

**Tip:** 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.

Be familiar with all these:

- Cloud IAM
- Encryption, Key Management
- Data Loss Prevention API
- HIPPA, COPPA, FedRAMP, GDPR

## Efficiency

| | |
|---|---|
| **Ensuring scalability and efficiency.** | Building and running test suites<br>Pipeline monitoring<br>**Assessing, troubleshooting, and improving data representations and data processing infrastructure**<br>Resizing and autoscaling resources |

**Tip:** A lot of administration over resources is presented in the console. But a lot of runtime information such as logs, performance, and so forth is presented and reported in Cloud Monitoring and Cloud Logging. The tools within Google Cloud's operations suite provide information for troubleshooting both functional and performance issues.

- Google Cloud's operations suite

# Reliability

| | |
|---|---|
| **Ensuring reliability and fidelity.** | **Performing data preparation and quality control**<br>Verification and monitoring<br>**Planning, executing, and stress testing data recovery**<br>Choosing between ACID, idempotent, eventual consistency requirements |

**Tip:** Establishing standard data quality at ingress using Dataprep by Trifacta or by running an ETL pipeline can prevent many problems later in processing that would be difficult to troubleshoot.

**Tip:** Keep in mind the business purpose of the data processing. How resilient does the application need to be? For example, financial transactions usually cannot be dropped and must not be duplicated. But a statistical analysis might be equally valid if a small amount of data is lost. These assumptions influence the approach to rerunning failed jobs.

Study these:

- Dataprep by Trifacta
- Fault-tolerance
- Rerunning failed jobs
- Performing retrospective re-analysis

# Portability

| | |
|---|---|
| **Ensuring flexibility and portability.** | Mapping to current and future business requirements<br>Designing for data and application portability |

**Tip:** Where is the official authoritative data (sometimes called the source of truth) and where are the replicas? How frequently does data need to be shared or updated? Can smaller parts of the data be synchronized to reduce costs?

**Tip:** Where is the data stored? Where is the data going to be processed? Can data storage and data processing be in locations near one another?

**Tip:** When will the data need to be exported? How difficult and expensive will it be to make this happen? For example, you might want to store data in a different location or in a different type of storage to meet business requirements for portability.

- Multi-cloud data residency requirements