

# Google Cloud PDE Quick Summary

Fast 3-5 minute refresh for the core ideas, product choices, and traps that matter most on the exam.

<b>Best for</b> Ultra-fast memory refresh	<b>Focus</b> Definitions, comparisons, and design shortcuts	<b>Use with</b> Use with the cheat sheet for deeper review
--	--	---

## 1. Fast refresh

<b>PDE focus</b>	Design, build, operationalize, secure, and monitor data systems.
<b>Think end-to-end</b>	Ingest, store, transform, serve, govern.
<b>Tool choice matters</b>	BigQuery is not Dataflow, and Dataflow is not Dataproc.
<b>Quality matters</b>	A pipeline that runs but produces bad data is still a failure.

## 2. High-yield comparisons

<b>BigQuery vs Cloud Storage</b>	Warehouse/analytics engine versus raw object storage.
<b>Dataflow vs Dataproc</b>	Managed serverless data processing versus cluster-based Spark/Hadoop control.
<b>Batch vs streaming</b>	Latency and event-time needs drive the choice.
<b>Partitioning vs clustering</b>	Both improve performance, but they solve different query patterns.

## 3. Security shortcuts

<b>Least privilege</b>	Give data access only where justified.
<b>DLP</b>	Useful when sensitive data must be discovered or transformed.
<b>Auditability</b>	Logs and lineage support both trust and compliance.
<b>Service accounts</b>	Treat workload identity carefully.

## 4. Exam traps

<b>Cheapest is not always best</b>	Operational simplicity and reliability matter too.
<b>Scalable is not enough</b>	Governance and quality still matter.
<b>Raw storage is not analytics serving</b>	Do not confuse storage layer with query layer.
<b>Streaming adds complexity</b>	Use it when freshness truly needs it.