

# Google GCP-PDE Study Guide PDF

GRAB THE GOOGLE PROFESSIONAL DATA ENGINEER CERTIFICATION PDF  
QUESTIONS & ANSWERS

---

Exam Summary – Syllabus – Questions

---

## GCP-PDE

Google Cloud Platform - Professional Data Engineer (GCP-PDE)

50 Questions Exam – 70% Cut Score – Duration of 120 minutes

[www.VMExam.com](http://www.VMExam.com)

## Table of Contents

Get an Overview of the GCP-PDE Certification:.....	3
Why Should You Earn the Google GCP-PDE Certification? .....	3
What Is the Google GCP-PDE Professional Data Engineer Certification Exam Structure? .....	3
Enhance Knowledge with GCP-PDE Sample Questions: .....	4
What Study Guide Works Best in Acing the Google GCP-PDE Professional Data Engineer Certification?....	8
Explore the Syllabus Topics and Learn from the Core: .....	8
Make Your Schedule: .....	8
Get Expert Advice from the Training: .....	8
Get Access to the PDF Sample Questions: .....	9
Avoid Dumps and Utilize the Google GCP-PDE Practice Test: .....	9

## Get an Overview of the GCP-PDE Certification:

Who should take the [GCP-PDE exam](#)? This is the first question that comes to a candidate's mind when preparing for the Professional Data Engineer certification. The GCP-PDE certification is suitable for candidates who are keen to earn knowledge on the Cloud and grab their Google Cloud Platform - Professional Data Engineer (GCP-PDE). When it is about starting the preparation, most candidates get confused regarding the study materials and study approach. But GCP-PDE study guide PDF is here to solve the problem. GCP-PDE PDF combines some effective sample questions and offers valuable tips to pass the exam with ease.

## Why Should You Earn the Google GCP-PDE Certification?

There are several reasons why one should grab the GCP-PDE certification.

- The Professional Data Engineer certification proves to be one of the most recognized certifications.
- The certification badge proves the knowledge of the candidate regarding subject matters and makes his resume presentable to potential candidates.
- Thus earning the [Google Cloud Platform - Professional Data Engineer \(GCP-PDE\)](#) is a powerful qualification for a prosperous career.

## What Is the Google GCP-PDE Professional Data Engineer Certification Exam Structure?

<b>Exam Name</b>	Google Professional Data Engineer
<b>Exam Code</b>	GCP-PDE
<b>Exam Price</b>	\$200 USD
<b>Duration</b>	120 minutes
<b>Number of Questions</b>	50
<b>Passing Score</b>	Pass / Fail (Approx 70%)
<b>Recommended Training / Books</b>	<a href="#">Google Cloud training</a> <a href="#">Google Cloud documentation</a> <a href="#">Google Cloud solutions</a>
<b>Schedule Exam</b>	<a href="#">PEARSON VUE</a>
<b>Sample Questions</b>	<a href="#">Google GCP-PDE Sample Questions</a>
<b>Recommended Practice</b>	<a href="#">Google Cloud Platform - Professional Data Engineer (GCP-PDE) Practice Test</a>

# Enhance Knowledge with GCP-PDE Sample Questions:

## Question: 1

Your company is loading comma-separated values (CSV) files into BigQuery. The data is fully imported successfully; however, the imported data is not matching byte-to-byte to the source file.

What is the most likely cause of this problem?

- a) The CSV data loaded in BigQuery is not flagged as CSV.
- b) The CSV data had invalid rows that were skipped on import.
- c) The CSV data has not gone through an ETL phase before loading into BigQuery.
- d) The CSV data loaded in BigQuery is not using BigQuery's default encoding.

**Answer: d**

## Question: 2

You are working on a project with two compliance requirements. The first requirement states that your developers should be able to see the Google Cloud billing charges for only their own projects.

The second requirement states that your finance team members can set budgets and view the current charges for all projects in the organization.

The finance team should not be able to view the project contents. You want to set permissions. What should you do?

- a) Add the finance team members to the Billing Administrator role for each of the billing accounts that they need to manage. Add the developers to the Viewer role for the Project.
- b) Add the finance team members to the default IAM Owner role. Add the developers to a custom role that allows them to see their own spend only.
- c) Add the developers and finance managers to the Viewer role for the Project.
- d) Add the finance team to the Viewer role for the Project. Add the developers to the Security Reviewer role for each of the billing accounts.

**Answer: a**

**Question: 3**

Your company is streaming real-time sensor data from their factory floor into Bigtable and they have noticed extremely poor performance.

How should the row key be redesigned to improve Bigtable performance on queries that populate real-time dashboards?

- a) Use a row key of the form <timestamp>.
- b) Use a row key of the form <sensorid>.
- c) Use a row key of the form <timestamp>#<sensorid>.
- d) Use a row key of the form <sensorid>#<timestamp>.

**Answer: d**

**Question: 4**

You are building storage for files for a data pipeline on Google Cloud. You want to support JSON files. The schema of these files will occasionally change.

Your analyst teams will use running aggregate ANSI SQL queries on this data. What should you do?

- a) Use BigQuery for storage. Provide format files for data load. Update the format files as needed.
- b) Use BigQuery for storage. Select "Automatically detect" in the Schema section.
- c) Use Cloud Storage for storage. Link data as temporary tables in BigQuery and turn on the "Automatically detect" option in the Schema section of BigQuery.
- d) Use Cloud Storage for storage. Link data as permanent tables in BigQuery and turn on the "Automatically detect" option in the Schema section of BigQuery.

**Answer: b**

**Question: 5**

You have 250,000 devices which produce a JSON device status event every 10 seconds. You want to capture this event data for outlier time series analysis. What should you do?

- a) Ship the data into BigQuery. Develop a custom application that uses the BigQuery API to query the dataset and displays device outlier data based on your business requirements.
- b) Ship the data into BigQuery. Use the BigQuery console to query the dataset and display device outlier data based on your business requirements.
- c) Ship the data into Cloud Bigtable. Use the Cloud Bigtable cbt tool to display device outlier data based on your business requirements.
- d) Ship the data into Cloud Bigtable. Install and use the HBase shell for Cloud Bigtable to query the table for device outlier data based on your business requirements.

**Answer: c**

**Question: 6**

You are designing a relational data repository on Google Cloud to grow as needed. The data will be transactionally consistent and added from any location in the world.

You want to monitor and adjust node count for input traffic, which can spike unpredictably. What should you do?

- a) Use Cloud Spanner for storage. Monitor CPU utilization and increase node count if more than 70% utilized for your time span.
- b) Use Cloud Spanner for storage. Monitor storage usage and increase node count if more than 70% utilized.
- c) Use Cloud Bigtable for storage. Monitor data stored and increase node count if more than 70% utilized.
- d) Use Cloud Bigtable for storage. Monitor CPU utilization and increase node count if more than 70% utilized for your time span.

**Answer: a**

**Question: 7**

You are using Pub/Sub to stream inventory updates from many point-of-sale (POS) terminals into BigQuery.

Each update event has the following information: product identifier "prodSku", change increment "quantityDelta", POS identification "termId", and "messageId" which is created for each push attempt from the terminal.

During a network outage, you discovered that duplicated messages were sent, causing the inventory system to over-count the changes. You determine that the terminal application has design problems and may send the same event more than once during push retries.

You want to ensure that the inventory update is accurate. What should you do?

- a) Add another attribute orderId to the message payload to mark the unique check-out order across all terminals. Make sure that messages whose "orderId" and "prodSku" values match corresponding rows in the BigQuery table are discarded.
- b) Inspect the "messageId" of each message. Make sure that any messages whose "messageId" values match corresponding rows in the BigQuery table are discarded.
- c) Instead of specifying a change increment for "quantityDelta", always use the derived inventory value after the increment has been applied. Name the new attribute "adjustedQuantity".
- d) Inspect the "publishTime" of each message. Make sure that messages whose "publishTime" values match rows in the BigQuery table are discarded.

**Answer: a**

**Question: 8**

You need to stream time-series data in Avro format, and then write this to both BigQuery and Cloud Bigtable simultaneously using Dataflow. You want to achieve minimal end-to-end latency.

Your business requirements state this needs to be completed as quickly as possible. What should you do?

- a) Create a pipeline and use ParDo transform.
- b) Create a pipeline that groups the data into a PCollection and uses the Combine transform.
- c) Create a pipeline that groups data using a PCollection, and then use Avro I/O transform to write to Cloud Storage. After the data is written, load the data from Cloud Storage into BigQuery and Bigtable.
- d) Create a pipeline that groups data using a PCollection and then uses Bigtable and BigQueryIO transforms.

**Answer: d**

**Question: 9**

You want to publish system metrics to Google Cloud from a large number of on-prem hypervisors and VMs for analysis and creation of dashboards.

You have an existing custom monitoring agent deployed to all the hypervisors and your on-prem metrics system is unable to handle the load. You want to design a system that can collect and store metrics at scale. You don't want to manage your own time series database.

Metrics from all agents should be written to the same table but agents must not have permission to modify or read data written by other agents. What should you do?

- a) Modify the monitoring agent to write protobuf messages directly to BigTable.
- b) Modify the monitoring agent to publish protobuf messages to Pub/Sub. Use a Dataproc cluster or Dataflow job to consume messages from Pub/Sub and write to BigTable.
- c) Modify the monitoring agent to write protobuf messages to HBase deployed on Compute Engine VM Instances
- d) Modify the monitoring agent to write protobuf messages to Pub/Sub. Use a Dataproc cluster or Dataflow job to consume messages from Pub/Sub and write to Cassandra deployed on Compute Engine VM Instances.

**Answer: b**

**Question: 10**

You are designing storage for CSV files and using an I/O-intensive custom Apache Spark transform as part of deploying a data pipeline on Google Cloud. You intend to use ANSI SQL to run queries for your analysts.

How should you transform the input data?

- a) Use BigQuery for storage. Use Dataflow to run the transformations.
- b) Use BigQuery for storage. Use Dataproc to run the transformations.
- c) Use Cloud Storage for storage. Use Dataflow to run the transformations.
- d) Use Cloud Storage for storage. Use Dataproc to run the transformations.

**Answer: b**

## What Study Guide Works Best in Acing the Google GCP-PDE Professional Data Engineer Certification?

The GCP-PDE study guide is a combination of some proven study tips and the combination of all valuable study materials like sample questions, syllabus and practice tests in one place.

### Explore the Syllabus Topics and Learn from the Core:

If you are determined to earn success in the Professional Data Engineer exam, getting in full touch of the [syllabus](#) is mandatory. During preparation, you might not like all syllabus sections or topics, but try to get at least the fundamental knowledge from the sections you don't like. The more you possess knowledge on all syllabus sections, the more is the chance to attempt maximum number of questions during the actual exam.

### Make Your Schedule:

Studying and completing the syllabus becomes easier, if you work on the syllabus topics after making a schedule. Your syllabus must mention what areas you want to cover and within what time. Once you make a schedule and follow it regularly, syllabus completion becomes easier and preparation becomes smoother.

### Get Expert Advice from the Training:

Do not forget to join the Google GCP-PDE training if it is providing any. Training enhances the practical knowledge of a candidate, which helps them to work well in the practical field during projects.



## Get Access to the PDF Sample Questions:

If your study material is in a [PDF format](#) or the materials are mobile-friendly, what could be better than that? Get access to the free sample questions and keep enhancing your knowledge beyond the syllabus.

## Avoid Dumps and Utilize the Google GCP-PDE Practice Test:

Why should you rely on practice tests? The reason is simple: you must get familiar with the exam pattern before reaching the exam hall. An aspirant aware of the exam structure and time management during the exam preparation can perform well in the actual exam and attempt the maximum number of questions during the exam.

Many aspirants prefer to read from dumps, but they miss out on the self assessment method. Therefore, GCP-PDE practice tests always stand out to be the better choice than dumps PDF.

### Avail the Proven GCP-PDE Practice Test for Success!!!

Do you want to pass the GCP-PDE exam on your first attempt? Stop worrying; we, VMExam.com are here to provide you the best experience during your Google Professional Data Engineer preparation. Try out our free mock tests to get a glimpse of our quality study materials, and build your confidence with the premium [GCP-PDE practice tests](#). Our expert-designed questions help you to improve performance and pass the exam on your first attempt.