

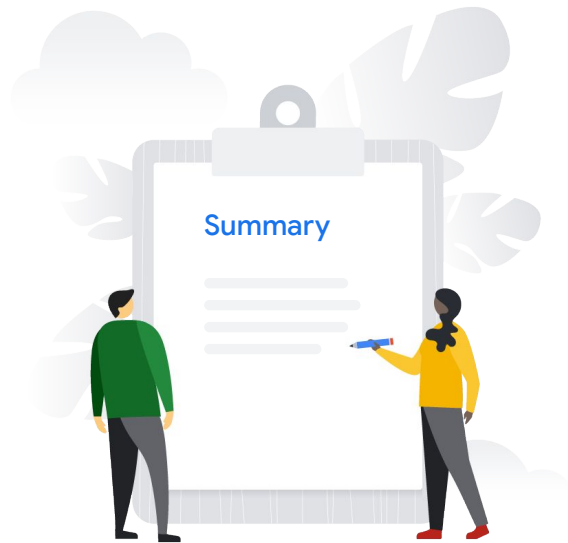


Resources and Next Steps

Tom Stern

In this module we will discuss next steps. This course exposed you to a lot of information. There were many elements designed to help you approach preparing for the exam in different ways.

Summary



First, we described the certification and the exam itself, and you were encouraged to adopt an iterative strategy of answering the questions that you're most confident about first, and bookmarking the others and revisiting them. You also learn that the exam is designed to test practitioners who do the job. It isn't about only knowing certain information or details, it's about being able to reason through cases and make reasoned decisions.

Certification Exam Guide

Professional Data Engineer

[Exam overview](#)

• Exam guide

[Sample questions](#)

[Certification home](#)

[Certification FAQs](#) ➤

[COVID-19 FAQs](#)

[Exam terms & conditions](#)

[Register](#)

[Certified directory](#) ➤

[Google Cloud training](#)

1. Designing data processing systems

1.1 Selecting the appropriate storage technologies. Considerations include:

- Mapping storage systems to business requirements
- Data modeling
- Tradeoffs involving latency, throughput, transactions
- Distributed systems
- Schema design

1.2 Designing data pipelines. Considerations include:

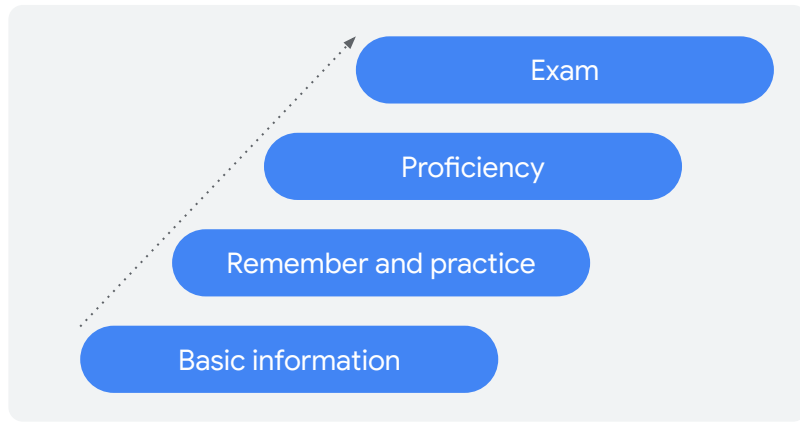
- Data publishing and visualization (e.g., BigQuery)
- Batch and streaming data (e.g., Cloud Dataflow, Cloud Dataproc, Apache Beam, Apache Spark and Hadoop ecosystem, Cloud Pub/Sub, Apache Kafka)
- Online (interactive) vs. batch predictions
- Job automation and orchestration (e.g., Cloud Composer)

1.3 Designing a data processing solution. Considerations include:

- Choice of infrastructure
- System availability and fault tolerance
- Use of distributed systems

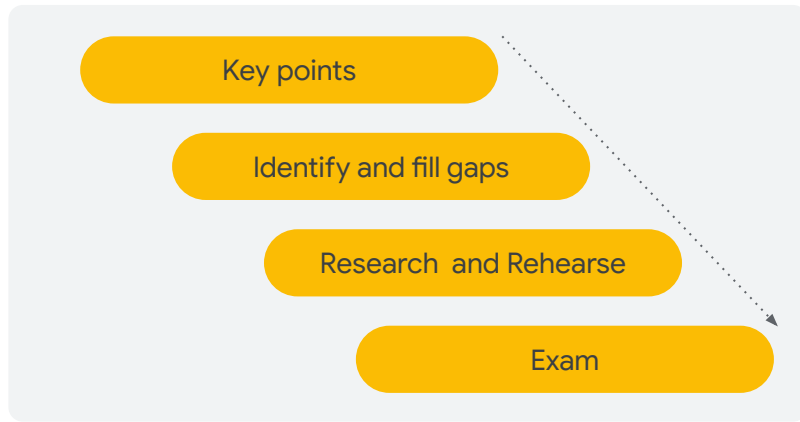
The majority of this course followed the exam guide outline and we explored each of the parts of the exam, the concepts behind them, and map these two important information to know about the technology that's taught in the data engineering courses.

Bottom-up approach

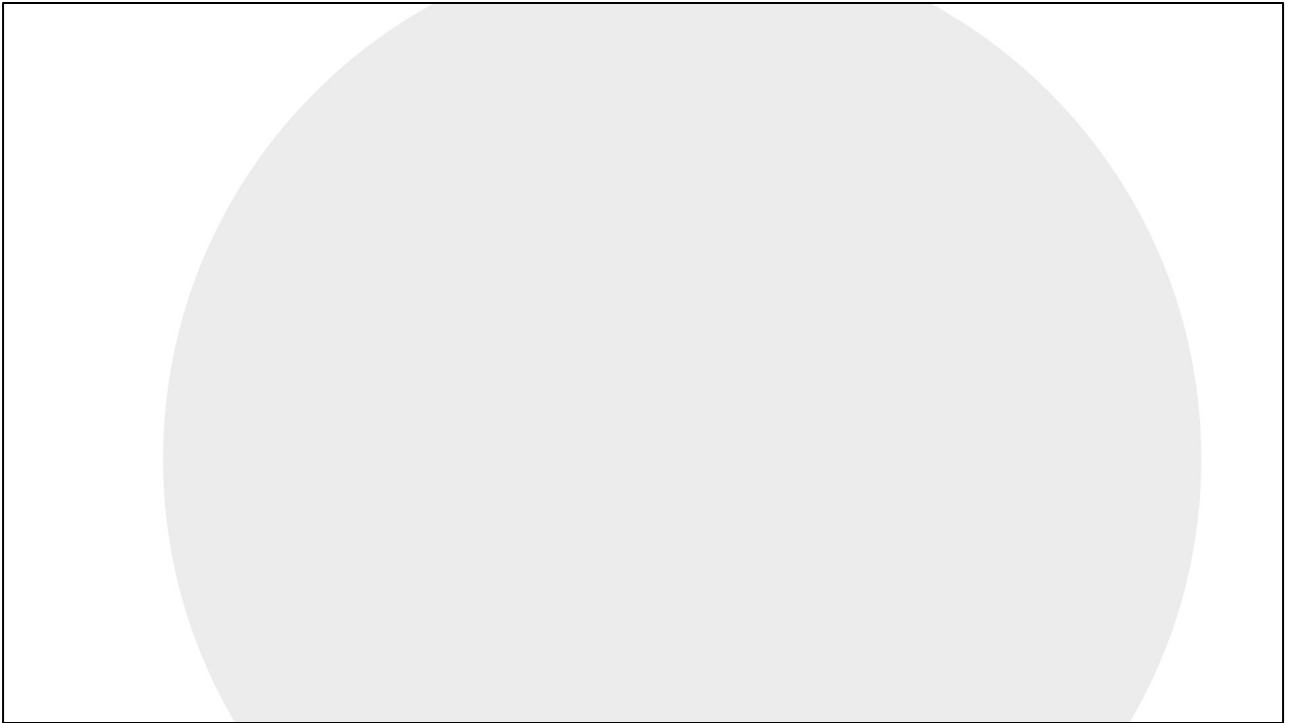


This wasn't a cram session, we were not trying to cover every detail or to build your knowledge from the ground up. In fact, the goal was to remind you of higher-level important concepts that you ought to know and understand, knowledge you probably gained from attending other courses, or from your experience.

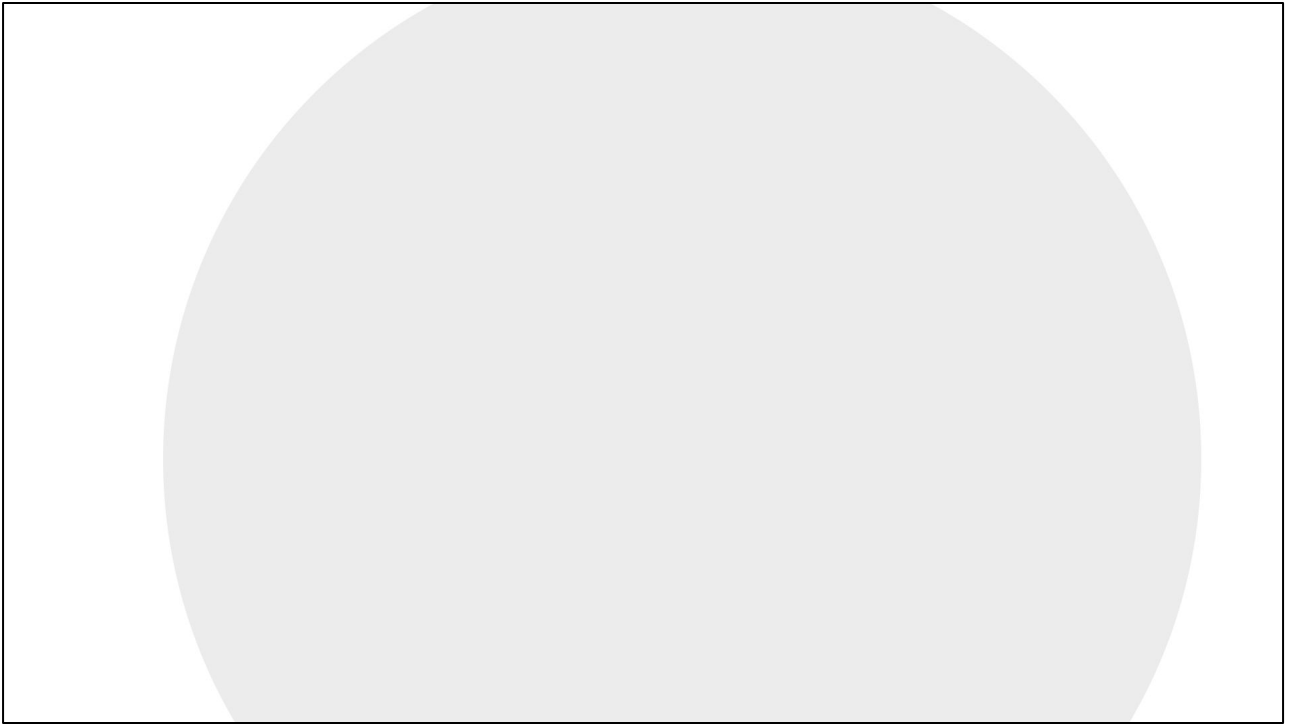
Top-down approach



The idea is, if you understand these touchstone concepts, you'll also need to know all the dependent basic information that they depend on and the reasoning behind them. You can still use a bottom-up mastery approach to help prepare for the exam if you like, the two approaches are not mutually exclusive.



You got to practice answering sample exam questions and we shared some actual case studies with you that were from real data engineers explaining how they use the skills on the job. You now have a pretty good sense of the exam and how to prepare for it. You've been exposed to information that will help you decide what to study and how to prepare. Next, you'll be given an opportunity to practice answering sample exam questions and finally, a list of resources for further study and preparation.



In this section, you will have the opportunity to try many more practice exam questions. This will give you practice answering exam-type questions. You can try the questions several times. Use the strategies you've learned to organize your approach.

Preparation resources

Storage and database documentation

Disks:

<https://cloud.google.com/compute/docs/disks/>

<https://cloud.google.com/bigtable/docs/choosing-ssd-hdd>

Cloud Storage: World-wide storage and retrieval of any amount of data at any time

<https://cloud.google.com/storage/docs/>

Cloud Memorystore: Fully managed in-memory data store service.

<https://cloud.google.com/memorystore/docs/redis/>

Cloud SQL: MySQL and PostgreSQL database service.

<https://cloud.google.com/sql/docs/>

Firestore in Datastore Mode: NoSQL document database service.

<https://cloud.google.com/datastore/docs/>

Firestore: Store mobile and web app data at global scale.

<https://cloud.google.com/firestore/docs/>

Firebase Realtime Database: Store and sync data in real Time.

<https://firebase.google.com/docs/database/>

Cloud Bigtable: NoSQL wide-column database service.

<https://cloud.google.com/bigtable/docs/>

Cloud Spanner: Mission-critical, scalable, relational database service.

<https://cloud.google.com/spanner/docs/>

Note: Links can be accessed from the
Course Resources PDF.

Here are links to the storage and database documentation. We recommend that you know the basic information about each service, the key features, performance, reliability, and best practices. Also, review how to control access and security in each.

Preparation resources

Data analytics documentation

BigQuery: A fully managed, highly scalable data warehouse with built-in ML.

<https://cloud.google.com/bigquery/docs/>

Dataproc: Managed Spark and Hadoop service.

<https://cloud.google.com/dataproc/docs/>

Dataflow: Real-time batch and stream data processing.

<https://cloud.google.com/dataflow/docs/>

Datalab: Explore, analyze, and visualize large datasets.

<https://cloud.google.com/datalab/docs/>

Dataprep by Trifacta: Cloud data service to explore, clean, and prepare data for analysis.

<https://cloud.google.com/dataprep/docs/>

Pub/Sub: Ingest event streams from anywhere, at any scale. <https://cloud.google.com/pubsub/docs/>

Google Data Studio: Tell great data stories to support better business decisions.

<https://marketingplatform.google.com/about/data-studio/>

Videos and tutorials are built into the interface of the product.

Cloud Composer: A fully managed workflow orchestration service built on Apache Airflow.

<https://cloud.google.com/composer/docs/>

Analytics and Data Processing documentation. You need to know what each of these services contributes to an overall data engineering solution.

Preparation resources

Machine Learning Documentation

Vertex AI: Build superior models and deploy them into production.

<https://cloud.google.com/vertex-ai/docs/>

Cloud TPU: Train and run ML models faster than ever.

<https://cloud.google.com/tpu/docs/>

AutoML: Easily train high-quality, custom ML models.

<https://cloud.google.com/automl/docs/>

Cloud Natural Language API: Derive insights from unstructured text.

Speech-to-Text: Speech-to-text conversion powered by ML.

Cloud Translation: Dynamically translate between languages.

Text-to-Speech: Text-to-speech conversion powered by ML.

Dialogflow Enterprise Edition: Create conversational experiences across devices and platforms.

Cloud Vision: Derive insight from images powered by ML.

Video Intelligence: Extract metadata from videos.

These are links to the Machine Learning documentation. Same level of detail for the pre-trained models. Did you know that the Vision API doesn't just recognize objects in images, but it can also give you back a color palette and composition information? If you didn't know that, then it is a good indicator of something you need to study. As a Professional Data Engineer, you need to be prepared by knowing what options are available.

Machine Learning -- clearly that is going to be a component of many or most data engineering solutions in the future. So you need to know quite a bit about it.

Preparation resources

Infrastructure documentation

Google Cloud's operations suite: Monitoring and management for services, containers, applications, and infrastructure.

- **Cloud Monitoring:** Monitoring for applications on Google Cloud and AWS.
- **Cloud Logging:** Logging for applications on Google Cloud and AWS.
- **Error Reporting:** Identifies and helps you understand application errors.
- **Cloud Trace:** Find performance bottlenecks in production.
- **Cloud Debugger:** Investigate code behavior in production.
- **Cloud Profiler:** Continuous CPU and heap profiling to improve performance and reduce costs.

Stackdriver is now
Google Cloud's
operations suite.

Transparent Service Level Indicators: Monitor Google Cloud services and their effects on your workloads.

Terraform: Manage cloud resources with simple templates.

Cloud Console: Google Cloud's integrated management console.

Cloud Shell: Command-line management from any browser.

Here are some of the important infrastructure services. You really need to know Google Cloud's operations suite. Because every data engineering solution is going to need to be monitored and maintained.

Training resources

<https://cloud.google.com/training/>

<https://cloud.google.com/training/data-ml>

<https://qwiklabs.com/>

<https://qwiklabs.com/catalog>

<https://www.coursera.org/programs/google-specialization?browseProductType=COURSE>

<https://www.coursera.org/courses?query=google%20cloud>

Links to our training resources. That is Instructor Led Training, lab-based training on Qwiklabs, and on-demand video based training on Coursera.

All sections of the exam are covered in the Data Engineering and Machine Learning curriculum

Instructor Led Training
[Data Engineer Track](#)

coursera for **Google**
[Data Engineering on
Google Cloud Specialization](#)

Classes

- Google Cloud Fundamentals: Big Data and Machine Learning
- Data Engineering on Google Cloud

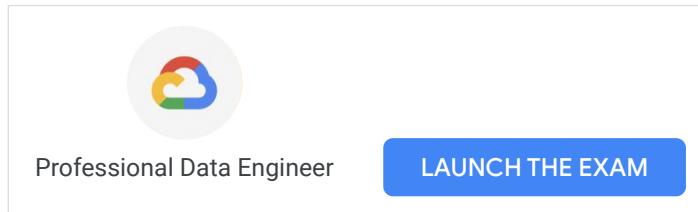
Courses

- Google Cloud Big Data and Machine Learning Fundamentals
- Leveraging Unstructured Data with Dataproc on Google Cloud
- Serverless Data Analysis with BigQuery and Dataflow
- Serverless Machine Learning with TensorFlow on Google Cloud
- Building Resilient Streaming Systems on Google Cloud

The Professional Data Engineer exam is specifically related to the Data Engineer track.

For example, in the Data Analyst track there is a great class called "Data to Insights" that covers in detail how to use BigQuery. And that class is not part of the Data Engineering track, so you would not need that class for this exam.

Data Engineer Practice Exam



<https://cloud.google.com/certification/practice-exam/data-engineer>

The Data Engineer practice exam will familiarize you with types of questions you may encounter on the certification exam and help you determine your readiness or if you need more preparation and/or experience.

Successful completion of the practice exam does not guarantee you will pass the certification exam as the actual exam is longer and covers a wider range of topics.

For a full list of the topics you could be tested on, see the Exam Guide.

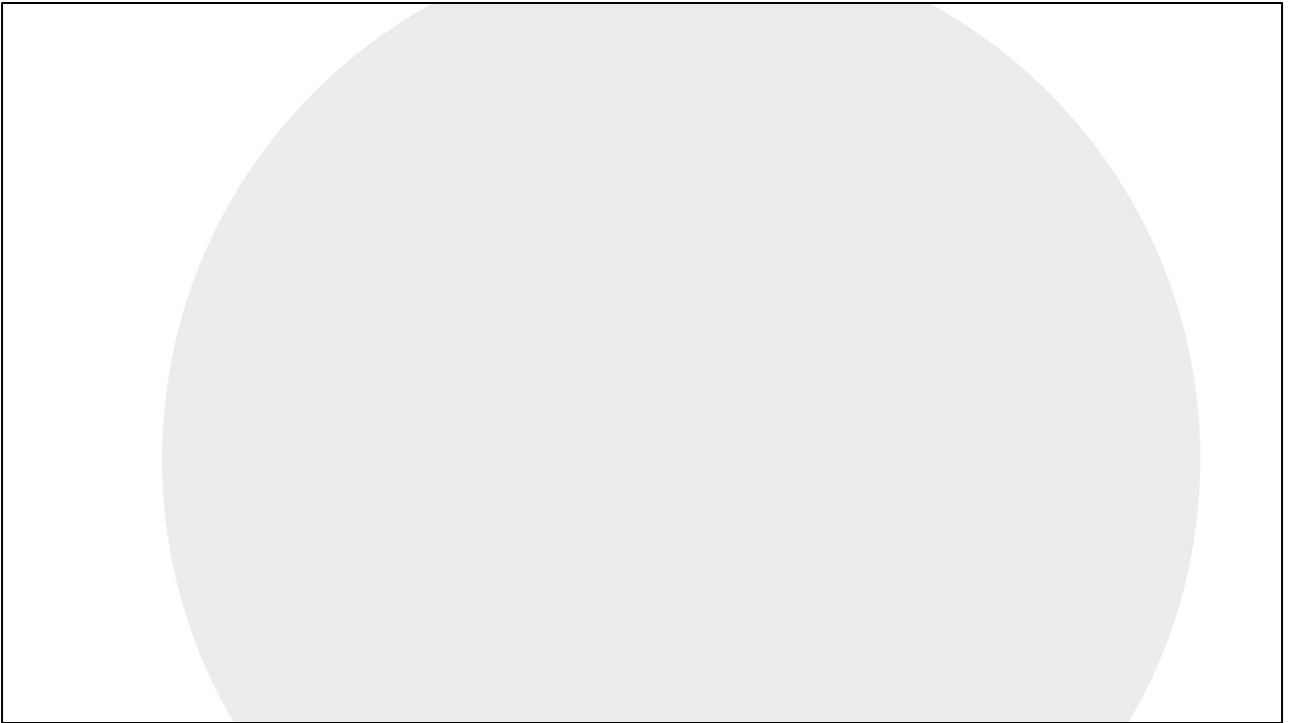
- There is no limit to the number of times you can take this practice exam.
- You can't save your progress. If you close the practice exam window, you must start from the beginning.
- There is no time limit for the practice exam, but we recommend completion in 45 minutes or less.
- This practice exam is available in English and Japanese.

Data Engineer - Quest



- Creating a Data Transformation Pipeline with Dataprep
- Building an IoT Analytics Pipeline on Google Cloud
- ETL Processing on Google Cloud Using Dataflow and BigQuery
- Predict Visitor Purchases with a Classification Model in BigQuery ML
- Cloud Composer: Copying BigQuery Tables Across Different Locations

This advanced-level quest is unique amongst the other Qwiklabs offerings. The labs have been curated to give IT professionals hands-on practice with topics and services that appear in the Google Cloud Certified Professional Data Engineer Certification. From BigQuery, to Dataproc, to TensorFlow, this quest is composed of specific labs that will put your Google Cloud data engineering knowledge to the test. Be aware that while practice with these labs will increase your skills and abilities, you will need other preparation too. The exam is quite challenging and external studying, experience, and/or background in cloud data engineering is recommended.



Thanks for taking the course "Preparing for the Professional Data Engineer Examination".

We hope you enjoyed the course and feel better prepared to attempt the exam.

Please fill out the evaluation and let us know what we got right and what we can improve.

And best of luck on the exam!

