Exam Guide

Effective March 29, 2019, the Google Certification organization will be switching the Data Engineering exam and preparation materials to a new version. Persons taking the exam on or after March 29 should prepare with the new Exam Guide.

The Exam Guide outline has a new structure and organization that reflect the changing skills of the Data Engineer job.

The new Exam Guide no longer includes case studies.

Why is the Exam Guide changing?

The Exam Guide is being updated to reflect the changing skills of the Data Engineer job role. This is an incremental and evolutionary process, so the skills you have already learned are still relevant.

Some changes are due to the technology. For example, Cloud Composer (based on Apache Airflow) and Cloud AutoML are new products that did not exist when the previous version of the Exam Guide was published, but are now important for Data Engineers to know.

Some changes are due to job role evolution. For example, the growth in machine learning has resulted in new skills called "Machine Learning Operations," which are the focus of the Data Engineer job role, while the complexity of constructing machine learning models has become the focus of a new Machine Learning Engineer job role.

What do these changes mean for people who are preparing for the exam?

The course has been updated to reflect the new Exam Guide outline. These are the most significant changes:

- All the content that contained tables reviewing parts of the Exam Guide outline has been replaced with content that introduces and describes the new Exam Guide outline.
- The Certification organization no longer publishes case studies as part of the Data Engineer Exam Guide. So the case studies are no longer part of this course.
- The Designing and Building module has been split into Designing Data Processing Systems and Building and Operationalizing Data Processing Systems.
- The title of the Analyzing and Modeling module has been changed to Operationalizing Machine Learning Models.

What has not changed

- "The best way to prepare for the exam is to prepare for the job"
- Design, build, and operationalize
- Data, pipelines, and processing infrastructure

The Data Engineer job has evolved in the following way. DEs are no longer focused mainly on the technology, but are now focused on choosing and using appropriate technologies for specific solutions. So they are now solution-oriented rather than technology-oriented. And the solutions are now focused on common workloads and use cases such as a Data Warehouse, and migrating a Data Warehouse to the cloud or improving a Data Warehouse already in the cloud.

The focus is on solutions rather than technologies. This is not a change in this course, because "Preparing for...." was always focused on defining solutions.

What has changed

Focus on solutions rather than technologies

- Choosing and using technologies for solutions
- Closer association between technologies and solutions, ex: BigQuery = Data Warehouse
- Focus on specific common workloads (migrate Data Warehouse to cloud)

Focus on operations and infrastructure rather than analysis and advanced ML

- ML ...operationalize, pre-built models, improve models (not create models)
- Creating/optimizing a model --> ML on GCP, Machine Learning Engineer's job

Focus on the "-ities". Security, Scalability, Reliability, Flexibility, Portability … and so forth.

This means for content courses like Data Engineering, the focus will shift from technologies to technologies in service of solutions -- how to handle common use cases and workloads.

That is a change in FOCUS not coverage. So mainly, the technical content will remain the same.

Job roles

In addition, there are three adjacent job roles to the Data Engineer. A couple of years ago these were all under the "Data Engineer" job. But now the job roles are growing and becoming more well defined and specialized.

A **Data Scientist** is responsible for understanding the business information and deriving business value from the data. This job is very mathematical and involves abstract thinking skills and strategy skills.

A **Data Analyst** is responsible for performing analysis, looking for patterns, running reports, and providing insights and advice on an ongoing basis.

A **Machine Learning Engineer** is responsible for creating ML models, choosing from among alternative AI and ML tools and approaches, training and developing models, and ensuring they work within required parameters.

Moving these specializations out of the Data Engineer job role helps the Data Engineer focus on selecting and implementing data storage and processing infrastructure. Handling specific workloads. And operationalizing (operating, administering, maintaining, troubleshooting, and securing) data solutions.

Good news -- you don't need to try to encompass all four jobs any longer. You need to know about each area, but mainly to be able to collaborate effectively with people in adjacent roles.

Overall, some skills associated with these job roles will be moving out of the content courses. And some new skills associated with the new technologies and better focus of the DE role will be added to the content courses.