**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. All the content that contained tables of parts of the Exam Guide outline has been replaced with content that introduces and describes the new Exam Guide outline.

Although the Certification organization no longer publishes case studies as part of the Data Engineer Exam Guide, the Training organization still provides them. We think that using case studies is a great way to rehearse and integrate the problem-solving skills of the Data Engineer. This is exactly the kind of thinking that is needed to understand and answer the exam questions, so the case studies are now part of this course.

You need to know how the job role and skills needed of the Data Engineer have changed so that you can make sure you are current in these areas as part of your preparation.

**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

**What has changed**

Focus on solutions instead of technologies

- Choosing and using technologies for solutions
- Closer association between technologies and solutions: e.g., BigQuery = data warehouse
- Focus on specific common workloads (migrate data warehouse to cloud)

Focus on operations and infrastructure instead of 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, etc.

**The job role is changing**

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 instead of technology-oriented. And the solutions are now focused on common workloads and use cases, such as data warehouses, migrating a data warehouse to the cloud, or improving a data warehouse already in the cloud.

**The curriculum is changing**

The focus on solutions instead of technologies is a change in the general DE curriculum. 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.

However, the "Preparing for…" class was always focused on solutions, because thinking through problems to solutions is a key job skill needed for the exam.

**Job roles**

In addition, there are three job roles adjacent 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 that 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. Some new skills associated with the new technologies and better focus of the DE role will be added to the content courses.