Google Cloud

Data Engineering on Google Cloud v2.5.x

Instructor Materials

Last updated: March 25, 2022

Data Engineering on Google Cloud 4-Day ILT courseware:

Day 1: Modernizing Data Lakes and Data Warehouses with Google Cloud

Day 2: Building Batch Data Pipelines on Google Cloud

Day 3: Building Resilient Streaming Analytics Systems on Google Cloud

Day 4: Smart Analytics, Machine Learning and AI on Google Cloud

Timing Guide

Logging Issues

Changelog

Day 1: Modernizing Data Lakes and Data Warehouses with Google Cloud					
#	**Module**	**Trainer**	**Student**	**Lab**	**Demo**
1.0	Introduction	PDF	PDF	-	-
1.1	Introduction to Data Engineering	PDF	PDF	Using BigQuery to do Analysis	Finding PII in your dataset with the DLP API
1.2	Building a Data Lake	PDF	PDF	Loading Taxi Data into Google Cloud SQL	-

1.3	Building a Data Warehouse	PDF	PDF	Loading Data into BigQuery Working with JSON and Array Data in BigQuery	Query TB+ of data in seconds Exploring BigQuery Public Datasets with SQL using INFORMATION_SCHEMA Nested and repeated fields in BigQuery
1.4	Summary	PDF	PDF	-	-

Day 2: Building Batch Data Pipelines on Google Cloud					
#	**Module**	**Trainer**	**Student**	**Lab**	**Demo**
2.0	Introduction	PDF	PDF	-	-
2.1	Introduction to Building Batch Data Pipelines	PDF	PDF	-	ELT to improve data quality in BigQuery
2.2	Executing Spark on Dataproc	PDF	PDF	Running Apache Spark jobs on Dataproc	-
2.3	Serverless Data Processing with Dataflow	PDF	PDF	Serverless Data Analysis with Dataflow: A Simple Dataflow Pipeline (Python/Java) Serverless Data Analysis with Dataflow: MapReduce in Dataflow (Python/Java) Serverless Data	-

				Analysis with Dataflow: Side Inputs (Python/Java)	
2.4	Manage Data Pipelines with Cloud Data Fusion and Cloud Composer	PDF	PDF	Building and Executing a Pipeline Graph in Cloud Data Fusion [HOMEWORK IF NECESSARY] An Introduction to Cloud Composer	[OPTIONAL] Event-triggered Loading of data with Cloud Composer, Cloud Functions, Cloud Storage, and BigQuery
2.5	Summary	PDF	PDF	-	-

Day 3: Building Resilient Streaming Analytics Systems on Google Cloud					
#	**Module**	**Trainer**	**Student**	**Lab**	**Demo**
3.1	Introduction	PDF	PDF	-	-
3.2	Serverless Messaging with Pub/Sub	PDF	PDF	Streaming Data Processing: Publish Streaming Data into Pub/Sub	-
3.3	Dataflow Streaming Features	PDF	PDF	Streaming Data Processing: Streaming Data Pipelines	-
3.4	High-Throughput BigQuery and Bigtable Streaming Features	PDF	PDF	Streaming Data Processing: Streaming Analytics and Dashboards Streaming Data Processing: Streaming Data Pipelines into	-

				Bigtable	
3.5	Advanced BigQuery Functionality and Performance	PDF	PDF	Optimizing your BigQuery Queries for Performance [OPTIONAL] Partitioned Tables in BigQuery	Mapping Fastest Growing Zip Codes with BigQuery GeoViz
3.6	Summary	PDF	PDF	-	-

Day 4: Smart Analytics, Machine Learning and AI on Google Cloud					
#	**Module**	**Trainer**	**Student**	**Lab**	**Demo**
4.0	Introduction	PDF	PDF	-	-
4.1	Analytics and AI: Introduction	PDF	PDF	-	-
4.2	Prebuilt ML Model APIs for Unstructured Data	PDF	PDF	Using the Natural Language API to Classify Unstructured Text	-
4.3	Big Data Analytics with Notebooks	PDF	PDF	BigQuery in JupyterLab on Vertex AI	-
4.4	Production ML Pipelines with Kubeflow	PDF	PDF	Running ML Pipelines on Kubeflow	-
4.5	Custom Model building with SQL in BigQuery ML	PDF	PDF	Predict Bike Trip Duration with a Regression Model in BigQuery ML Movie Recommendations	Train a model with BigQuery ML to predict NYC taxi fares

				in BigQuery ML	
4.6	Custom Model Building with AutoML	PDF	PDF	-	-
4.7	Summary	PDF	PDF	-	-

Timing Guide

Access the recommended Timing Guide here.

Logging Issues

We are committed to ensuring a positive experience for both yourself and your students. You are encouraged to use the channels available to log issues as follows:

- **Customer Issues** allows you to log issues that compromise the learning experience. This can relate to functionality or content. Please use the **Priority** settings provided to allow us to prioritize remedial action.
    - Internal CI link
    - External ATP CI link
- **Feature Requests** allows you to log suggestions for the improvement of the overall learning experience.
    - Internal FR link
    - External ATP FR link

    NOTE: When logging issues as an ATP, please do not delete the issue title prefix that is generated by the template. Kindly add a descriptor after the prefix.

Changelog

Access the Changelog for this course here.