

Databricks - ETL Part 3 - Production

Code: ETL3-PROD-SELF

URL: [View Online](#)

In this course data engineers optimize and automate Extract, Transform, Load (ETL) workloads using stream processing, job recovery strategies, and automation strategies like REST API integration. By the end of this course you will schedule highly optimized and robust ETL jobs, debugging problems along the way.

Skills Gained

- Perform an ETL job on a streaming data source
- Parameterize a code base and manage task dependencies
- Submit and monitor jobs using the REST API or Command Line Interface
- Design and implement a job failure recovery strategy using the principle of idempotence
- Optimize ETL queries using compression and caching best practices with optimal hardware choices

Prerequisites

- ETL Part 1 self-paced course (optional, but strongly encouraged)
- ETL Part 2 self-paced course (optional, but strongly encouraged)

Course Details

Course Outline

- Course Overview and Setup
- Streaming ETL
- Runnable Notebooks
- Scheduling Jobs
- Job Failure
- ETL Optimizations
- Capstone Project

Platforms

Supported platforms include Azure Databricks and AWS Databricks. Note: This course will not run on Databricks Community Edition.

- If you're planning to use the course on Azure Databricks, select the "Azure Databricks" Platform option.
- If you're planning to use the course on Databricks Community Edition or on a non-Azure version of Databricks, select the "Other Databricks" Platform option.

Format

The course is a series of six self-paced lessons available in both Scala and Python. A final capstone project involves refactoring a batch ETL job to a streaming pipeline. In the process, students run the workload as a job and monitor it. Each lesson includes hands-on exercises.

ExitCertified® Corporation and iMVP® are registered trademarks of ExitCertified ULC and ExitCertified Corporation and Tech Data Corporation, respectively
Copyright ©2019 Tech Data Corporation and ExitCertified ULC & ExitCertified Corporation.
All Rights Reserved.

Generated 11