DAY 1

Databricks Certification Preparation

Databricks Certified Data Engineer Associate

DAY 1 – KEY MODULES

Daniel Arrizza

HOUSEKEEPING

- This session will be recorded and the slides and notebooks will be shared by the end of this week
- Please use the **Q&A chat** function to ask our team any questions during the training – we will do our best to answer them in the chat or outloud. If we are unable to answer your questions during the session your Databricks account team will follow-up with you
- Find all important links under the **Related Content** widget including the links to the course in Databricks Academy, the Certification Overview, YouTube videos on Delta Live Tables and Data Engineering on Databricks, the notebooks in GitHub, the link to register for the exam, our survey and more!
- Feel free to adjust the widgets on your audience dashboard
- Refresh your page if you have any issues
- Enjoy this training :) and don't forget to **complete our survey**!

THE RAFFLE & CERTIFICATION VOUCHER

Complete the <u>Fundamentals of the Databricks Lakehouse Platform Accreditation</u> by May 27 to be entered into a raffle where 25 lucky winners will be sent a box of cool Databricks swag!
- This is a 30-minute assessment that will test your knowledge about fundamental concepts related to the Databricks Lakehouse Platform. This accreditation is the beginning step in most of the Databricks Academy learning plans – SQL Analysts, Data Scientists, Data Engineers, and Platform Administrators. Business leaders are also welcome to take this assessment.

Complete our short feedback survey, providing your name & company email address and we will send you 1 certification voucher to take on any Databricks certification exam
- Voucher is valid until October 31, 2022

Meet your instructor!

Daniel Arrizza
Sr. Staff Customer Success Engineer
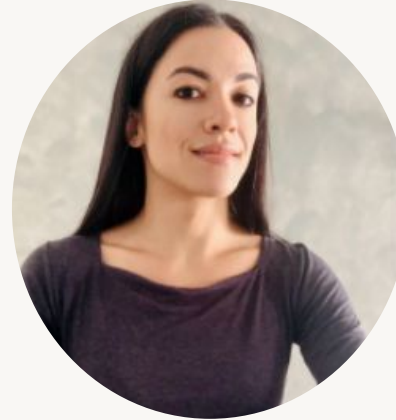
Meet your Teaching Assistants!

Amar Kadamalakunte

Customer Success Engineer

Debora Reis

Sr. Customer Success Engineer

Lorin Dawson

Sr. Customer Success Engineer

Mark Roepke

Sr. Technical Curriculum Developer

Trevor Bishop

Sr. Customer Success Engineer

Youssef Mrini

Associate Customer Success Engineer

6

Agenda
Topics addressed in this session

1 **Introduction and Overview**
Training series day 1 goals

2 **Why Databricks certification?**
Benefits of certification

3 **Overview of concepts**
Key concepts tested on exam

4 **Next session details**
What to expect next

5 **Available resources**
Helpful resources to help prepare

Agenda

Topics addressed in this session

1 **Introduction and Overview**
Training series day 1 goals

2 **Why Databricks certification?**
Benefits of certification

3 **Overview of concepts**
Key concepts tested on exam

4 **Next session details**
What to expect next

5 **Available resources**
Helpful resources to help prepare

Agenda

Topics addressed in this session

1 **Introduction and Overview**
Training series day 1 goals

2 **Why Databricks certification?**
Benefits of certification

3 **Overview of concepts**
Key concepts tested on exam

4 **Next session details**
What to expect next

5 **Available resources**
Helpful resources to help prepare

Agenda

Topics addressed in this session

1 **Introduction and Overview**
Training series day 1 goals

2 **Why Databricks certification?**
Benefits of certification

3 **Overview of concepts**
Key concepts tested on exam

4 **Next session details**
What to expect next

5 **Available resources**
Helpful resources to help prepare

Agenda

Topics addressed in this session

1 **Introduction and Overview**
Training series day 1 goals

2 **Why Databricks certification?**
Benefits of certification

3 **Overview of concepts**
Key concepts tested on exam

4 **Next session details**
What to expect next

5 **Available resources**
Helpful resources to help prepare

Introduction and Overview

About this training series

Data Engineer

DATABRICKS LAKEHOUSE FUNDAMENTALS ACCREDITATION ★	DATA ENGINEER ASSOCIATE CERTIFICATION ★★	DATA ENGINEER PROFESSIONAL CERTIFICATION ★★★
• What is the Databricks Lakehouse Platform? (SP) • What is Delta Lake? (SP) • What is Databricks SQL? (SP) • What is Databricks Machine Learning? (SP) • **Databricks Lakehouse Fundamentals Accreditation (SP)**	• Data Engineering with Databricks (ILT/SP) • Certification Overview Course for the Databricks Certified Data Engineer Associate (SP) • **Databricks Certified Data Engineer Associate Exam (for $200 USD)**	• Advanced Data Engineering with Databricks *(available now as self-paced or as ILT starting 22-June-2022)* • Certification Overview Course for the Databricks Certified Data Engineer Professional (SP) • **Databricks Certified Data Engineer Professional Exam (for $200 USD)**

SP = self-paced training (free for Databricks customers, available in Data Engineer Learning plan via Databricks Academy)

ILT = instructor-led training course (available for a fee)

Associate Data Engineer Expectations

Therefore, the following is expected of a Associate-level data engineer:

- Understand how to use and the benefits of using the Databricks Lakehouse Platform and its tools

- Build ETL pipelines using Apache Spark SQL and Python

- Incrementally process data

- Build production pipelines for data engineering applications and Databricks SQL queries and dashboards

- Understand and follow best security practices

Helpful background experience

- Work or educational experience as a data or IT professional
- Experience using SQL to query data from enterprise data stores
- Familiarity with basic cloud concepts (virtual machines, object storage, identity management)
- Basic familiarity with Python variables, functions, and control flow (preferred)

What will this training series cover?

2 days to learn from Databricks on how to pass Data Engineering Associate exam

What will this training series cover?

2 days to learn from Databricks on how to pass Data Engineering Associate exam

- Day 1 (today)

  - Certification exam overview

    - Using the Databricks Lakehouse Platform
    - ELT with Spark SQL and Python
    - Incremental Data and Delta Live Tables
    - Managing Data Access and Production Pipelines

What will this training series cover?

2 days to learn from Databricks on how to pass Data Engineering Associate exam

- Day 1 (today)
  - Certification exam overview
    - Using the Databricks Lakehouse Platform
    - ELT with Spark SQL and Python
    - Incremental Data and Delta Live Tables
    - Managing Data Access and Production Pipelines

- Day 2
  - Delta technical dive
  - Certification logistics
  - Testimonials
  - Questions

What this training series will not do

- **Prepare you 100% to take the exam**
- Provide answers to exam questions

Why Databricks certification?
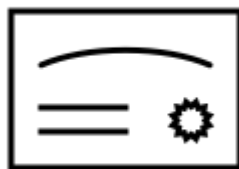
Benefits of certification

Stand out from the crowd

Achieve personal development goals and become more competitive

1

Validate your skills with a respected certification exam in the big data and AI space
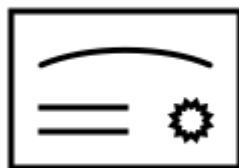
Stand out from the crowd

Achieve personal development goals and become more competitive

1 | **Validate your skills** with a respected certification exam in the big data and AI space
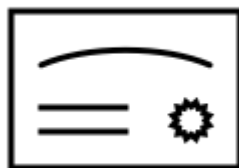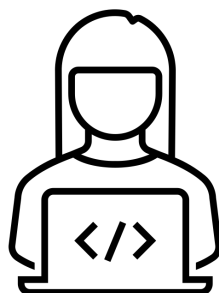
2 | **Increase your efficiency** by learning the things you need to be successful on Databricks
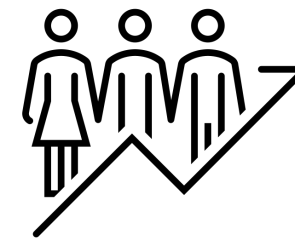
Stand out from the crowd

Achieve personal development goals and become more competitive

1 — **Validate your skills** with a respected certification exam in the big data and AI space

2 — **Increase your efficiency** by learning the things you need to be successful on Databricks

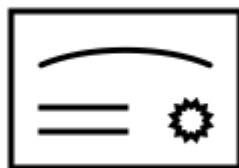**3** — **Turn recruiter and future employer heads** by proudly displaying your certification

Stand out from the crowd

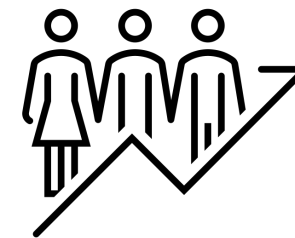## Achieve personal development goals and become more competitive

Validate your skills with a respected certification exam in the big data and AI space
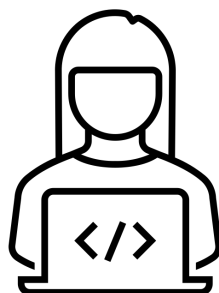
1

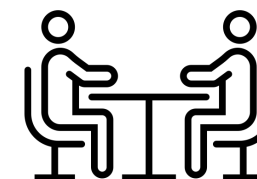**Increase your efficiency** by learning the things you need to be successful on Databricks

2

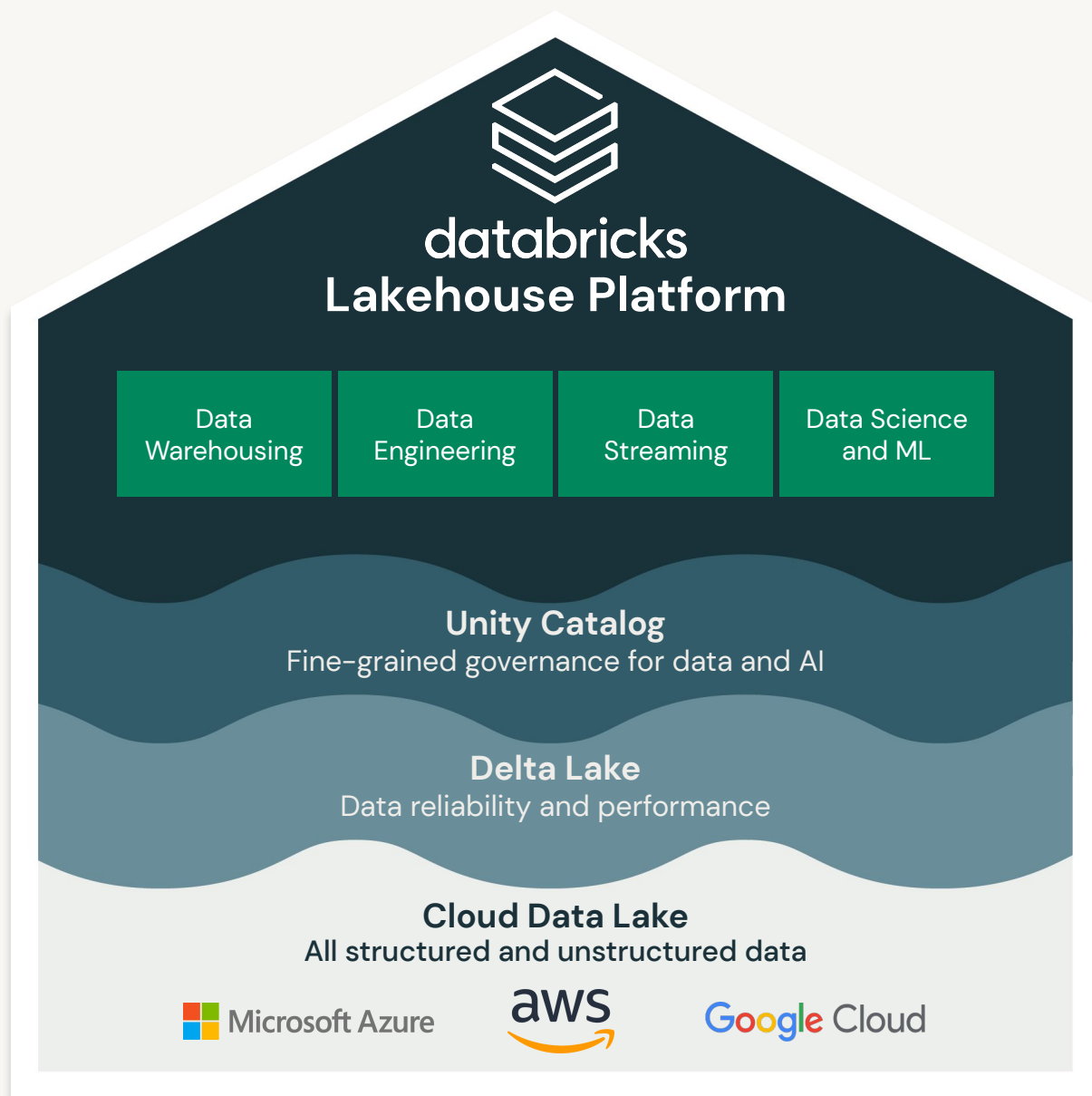**Turn recruiter and future employer heads** by proudly displaying your certification

3

Improve stakeholder and peers perceptions when it comes to reputation, credibility, and confidence

4

Overview of concepts

Key concepts tested on exam

Databricks Lakehouse Platform

Simple
Unify your data warehousing and AI use cases on a single platform

Open
Built on open source and open standards

Multicloud
One consistent data platform across clouds

Next session details

What to expect next

Join us next time for...

Deep dive of Delta notebooks
a technical dive into an important piece of the certification exam

Join us next time for...

Deep dive of Delta notebooks

a technical dive into an important piece of the certification exam

Certification exam logistics

learn about the exam itself and the logistics behind registering for it

Join us next time for...

Deep dive of Delta notebooks

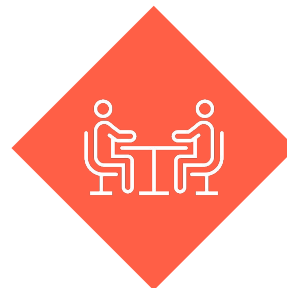a technical dive into an important piece of the certification exam

Certification exam logistics

learn about the exam itself and the logistics behind registering for it

Testimonials from exam takers

hear tips and tricks from folks who have taken the exam

Join us next time for...

Deep dive of Delta notebooks
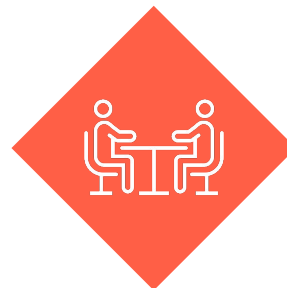a technical dive into an important piece of the certification exam

Certification exam logistics
learn about the exam itself and the logistics behind registering for it

Testimonials from exam takers
hear tips and tricks from folks who have taken the exam

Ask us anything
get your questions about Databricks certification answered

Available resources

Helpful resources to help prepare

Resources to help you prepare starting now!

- Databricks Academy
  - Data Engineering with Databricks (free, self-paced course/$1500 ILT course)
  - Certification Overview for the Databricks Certified Associate Data Engineer Exam (free, self-paced course)

DAY 2

databricks

Databricks Certification Preparation

Databricks Certified Data Engineer Associate

DAY 2 – EXAM PREP

Daniel Arrizza

Meet today's team!

Daniel Arrizza,
Sr. Staff Customer
Success Engineer

Debora Reis,
Sr. Customer
Success Engineer

Amar Kadamalakunte,
Customer
Success Engineer

Agenda
Topics addressed in this session

1 **Introduction and Overview**
Training series day 2 goals

2 **Delta Notebooks**
Review of Delta topics on exam

3 **Certification Exam Logistics**
Information about the exam

4 **Tips and Tricks**
Hear from individuals who took the exam

5 **Q&A**
Ask us anything about the exam

Agenda

Topics addressed in this session

1 **Introduction and Overview**
Training series day 2 goals

2 **Delta Notebooks**
Review of Delta topics on exam

3 **Certification Exam Logistics**
Information about the exam

4 **Tips and Tricks**
Hear from individuals who took the exam

5 **Q&A**
Ask us anything about the exam

Agenda
Topics addressed in this session

1 **Introduction and Overview**
Training series day 2 goals

2 **Delta Notebooks**
Review of Delta topics on exam

3 **Certification Exam Logistics**
Information about the exam

4 **Tips and Tricks**
Hear from individuals who took the exam

5 **Q&A**
Ask us anything about the exam

Agenda

Topics addressed in this session

1 **Introduction and Overview**
Training series day 2 goals

2 **Delta Notebooks**
Review of Delta topics on exam

3 **Certification Exam Logistics**
Information about the exam

4 **Tips and Tricks**
Hear from individuals who took the exam

5 **Q&A**
Ask us anything about the exam

Agenda

Topics addressed in this session

1 **Introduction and Overview**
Training series day 2 goals

2 **Delta Notebooks**
Review of Delta topics on exam

3 **Certification Exam Logistics**
Information about the exam

4 **Tips and Tricks**
Learn from individuals who took the exam

5 **Q&A**
Ask us anything about the exam

Introduction and Overview

Training series goals

Before we get started... who is this for?

- Data Engineer

- Beginner-level certification

- Assess candidates at a level equivalent to **six months of experience** with data engineering with Databricks

Professional

6 months

Associate

Associate Data Engineer Expectations

Therefore, the following is expected of a Associate-level data engineer:

- Understand how to use and the benefits of using the Databricks Lakehouse Platform and its tools

- Build ETL pipelines using Apache Spark SQL and Python

- Incrementally process data

- Build production pipelines for data engineering applications and Databricks SQL queries and dashboards

- Understand and follow best security practices

Helpful background experience

- Work or educational experience as a data or IT professional
- Experience using SQL to query data from enterprise data stores
- Familiarity with basic cloud concepts (virtual machines, object storage, identity management)
- Basic familiarity with Python variables, functions, and control flow (preferred)

What will this training series cover?

2 days to learn from Databricks on how to pass Data Engineering Associate exam

What will this training series cover?

2 days to learn from Databricks on how to pass Data Engineering Associate exam

- Day 1 (recap)

  - Certification exam overview

    - Using the Databricks Lakehouse Platform

    - ELT with Spark SQL and Python

    - Incremental Data and Delta Live Tables

    - Managing Data Access and Production Pipelines

What will this training series cover?

2 days to learn from Databricks on how to pass Data Engineering Associate exam

- ## Day 1 (recap)

  - Certification exam overview
    - Using the Databricks Lakehouse Platform
    - ELT with Spark SQL and Python
    - Incremental Data and Delta Live Tables
    - Managing Data Access and Production Pipelines

- ## Day 2

  - Delta technical dive
  - Certification logistics
  - Testimonials
  - Questions

What this training series will not do

- **Prepare you 100% to take the exam**
- Provide answers to exam questions.

Delta Notebooks

Review of Delta topics on exam

Notebook walk-through and demo
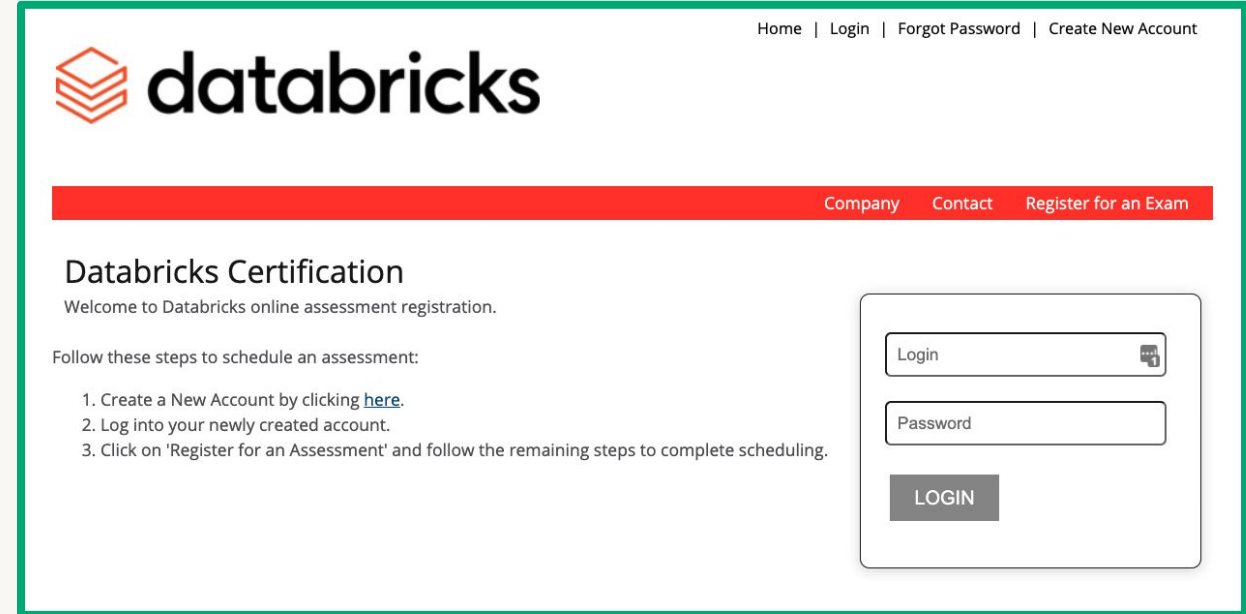
- Daniel to add any needed slides here

Certification Exam Logistics

Information about the exam / taking the exam

Exam Platform

- Databricks Academy certifications are offered through Kryterion's Webassessor platform.
- Webassessor is a simple, scalable assessment solution resulting in an easy test-taking experience.

Proctoring Details

- During the exam, you will be:
  - Monitored via webcam by a Webassessor proctor.
  - Asked to provide valid, photo-based identification.
- The proctor will:
  - Monitor you during the exam.
  - Answer any exam delivery questions you might have.
  - Provide technical support.
- The proctor will not provide assistance on the content of the exam.
- No test aids will be available during the exam.

Exam Grading

- Certification exams are automatically graded.
- Following the exam, the proctor's session notes and the recorded grade will be reviewed by Databricks Academy
- It will take about one week for you to find out whether or not you passed the exam.

Certificate Awarding Process

- If it's been determined that you've passed the exam, your certificate will be awarded via Databricks Academy.

Basic Exam Details

- Time allotted to complete exam = **1.5 hours**

- Passing scores = At least **70%**

- Number of Questions = **45**

- More info. on the Databricks Academy FAQ:

  http://files.training.databricks.com/lms/docebo/databricks-academy-faq.pdf

Code Examples

- Data manipulation code will be provided in **SQL** when possible

- In all other cases, code will be in **Python**

- Runtime Version = **DBR 9.1**

Exam Question and Topic Distribution

- Exam questions are distributed into five categories:
  - Databricks Lakehouse Platform – 24%
  - ELT with Spark SQL and Python – 29%
  - Incremental Data Processing – 22%
  - Production Pipelines – 16%
  - Data Governance – 9%

- All of the questions in all categories are **multiple-choice questions** – this means there's only one correct answer for each question.

Conceptual Question Types

- **Definitions:** what something is or does
- **Relationships:** how something compares to or is related to something else
- **Results:** If _____ occurs, ...
- **Classification:** in which category does something belong
- **Configurations:** based on this cluster configuration, ...

Tips and Tricks

Learn from individuals who took the exam

Debora Reis,
Sr. Customer
Success Engineer
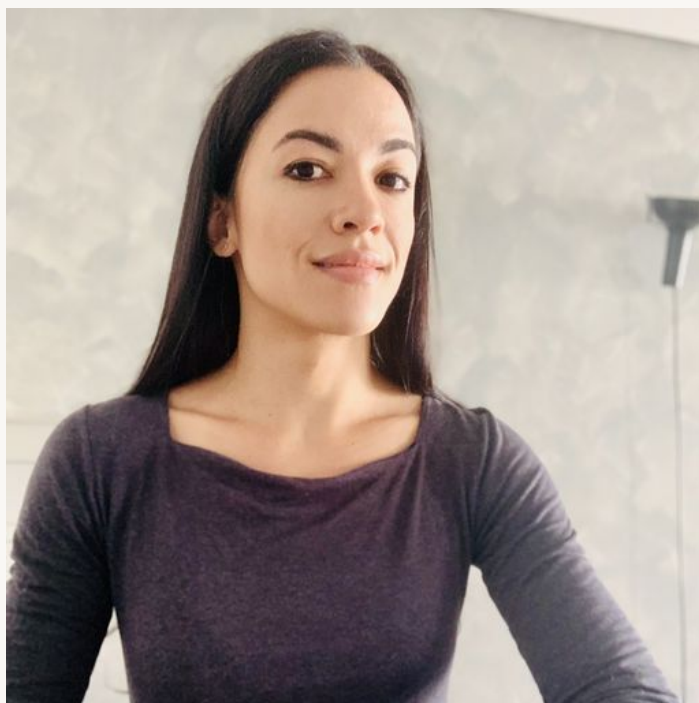
Amar Kadamalakunte,
Customer
Success Engineer

Available resources

Helpful resources to help prepare

Resources to help you prepare starting now!

- Databricks Academy
    - Data Engineering with Databricks (free, self-paced course/$1500 ILT course)
    - Certification Overview for the Databricks Certified Associate Data Engineer Exam (free, self-paced course)

Q&A

Ask us anything about the exam