Course goals

Hi, welcome to Apache Spark SQL for data analysts. My name is Kate Sullivan and along with my colleagues at Databricks, we're going to guide you through how to leverage your existing SQL skills to start working with Apache Spark immediately.

If you're not too familiar with SQL, that's okay. We're going to give you all the resources and information you need to get you started and get you through this course.

>> So a question I'm sure some of you are wondering is, why Spark? And that's a really good question. The short answer is Spark is awesome. And there's no better way to take massive amounts of data, coming in in all sorts of formats, and creating actionable business insights from that data. Now, even though Spark isn't super popular with data analysts right now, this course will totally change that.

>> By the end of this course, you'll be ready to start using Spark for your data analysis projects. In addition to lots of hands-on practice to leverage Spark to adjust, transform, query and visualize data, we'll also be answering questions that we typically get from a lot of data analysts. Questions like, what is Spark SQL and how does Spark work with SQL? Also, what tools do I need to leverage Spark? Plus, we'll also introduce you to and show you how to work with some pretty powerful echnology in addition to Spark, which will make you stand out in your current or future roles.

And you'll also be getting a certificate of completion that you can share with your networks as testimony for your new Spark SQL skills.

>> Once again, welcome to the course. We're thrilled to have you.

# Before you begin

Please review this lesson for some of the most frequently asked questions about this course.

**Is this course part of a specialization on Coursera?**

Yes! This course is part of a three-course Coursera specialization, "Data Science with Databricks for Data Analysts". The courses in that specialization include:

1. Apache Spark SQL for Data Analysts (this course!)
2. Data Science Fundamentals for Data Analysts
3. Applied Data Science for Data Analysts

**How much time will it take me to complete this course?**

We have prepared this material to take you approximately eight hours to complete. This includes the hands-on labs that help you apply the concepts you’ll learn throughout the course.

**Where can I find a course agenda?**

The course agenda can be found below. We recommend completing one lesson per week, but you’re able to move at the pace that works best for you.

You will notice that the lessons are organized into modules – this is based on the topics that they cover. At the end of the final lesson within each module, you will need to complete and pass a graded quiz or activity in order to move on to the next module.

| **Week** | **Module** | **Lesson** | **Estimated Time** |
| --- | --- | --- | --- |
| 1 | 1 | Welcome to Apache Spark SQL for Data Analysts | 15 min |
| 1 | 2 | Spark Makes Big Data Easy | 30 min |
| 2 | 3 | Using Spark SQL on Databricks | 60 min |
| 2 | 4 | Spark Under the Hood | 60 min |
| 3 | 5 | Complex Queries | 60 min |
| 3 | 6 | Applied Spark SQL | 60 min |
| 4 | 7 | Data Storage and Query Optimization | 30 min |
| 4 | 8 | Delta with Spark SQL | 60 min |
| 5 | 9 | Completing Coding Challenges | 120 |

**What tools do I need to complete this course?**

If you are planning on completing the hands-on labs in this course, you will need to create a free account on Databricks Community Edition. Don’t worry - you don’t have to do it now. We’ll guide you through how to perform this simple step later on in the course. Other than that, you just need a computer or laptop with an internet connection, and a desire to learn!

**What if I don't know SQL?**

We will provided guided examples to help you learn the SQL commands and clauses we use in this course, and often, we'll point you to documentation you can use to learn more. If you're completely unfamiliar with SQL, it may help to work through an online tutorial like [SQL Zoo](https://sqlzoo.net/wiki/SQL_Tutorial). Lessons 1 - 8 should provide all the background you need.

**How do I receive my accreditation?**

In order to receive the accreditation (Associate SQL Analyst) provided by completing this course, you must complete all course content and pass a graded exam at the end of the last module of the course.

**How do I provide feedback?**

You will have the opportunity to provide feedback in a survey at the end of the course. We take your feedback very seriously -- it is used to help improve our offerings on Coursera. Please send thoughts, concerns, suggestions, etc. our way so that we can keep them in mind for future course offerings.

**Where is the course content?**

Most of the course content will be available through Coursera, and you’ll also be completing work in Databricks Community Edition. An overview of Databricks Community Edition is available later in this lesson.

The content for the Databricks Community Edition portion of the course can be downloaded from <https://files.training.databricks.com/courses/moocs/SQLDA/Lessons.dbc>. This link is a Databricks Archive file (DBC). DBC files can be loaded directly into Databricks Community Edition. An overview of this process is available in Module 3. Please note that DBC files cannot be opened directly on your computer.

**Who are you?**

Good question! Your hosts for this course and this entire specialization if you choose to complete it (you should!) all work at Databricks. You’ll learn more about Databricks (as well as your hosts) as we progress through this course. Fun fact: The founders of Databricks were the original creators of Apache Spark (pretty, cool, right?). In this MOOC, we promise to bring you the latest and greatest news about Apache Spark - specifically, Apache Spark and SQL.