**Cleaning Data with PySpark**

### Course Description

Working with data is tricky - working with millions or even billions of rows is worse. Did you receive some data processing code written on a laptop with fairly pristine data? Chances are you’ve probably been put in charge of moving a basic data process from prototype to production. You may have worked with real world datasets, with missing fields, bizarre formatting, and orders of magnitude more data. Even if this is all new to you, this course helps you learn what’s needed to prepare data processes using Python with Apache Spark. You’ll learn terminology, methods, and some best practices to create a performant, maintainable, and understandable data processing platform.