University of Rochester

databricks

University Alliance

Data Science at Scale

DSCC 202/402

Spring 2026

Your Instructors & TA

Lloyd Palum

CTO at a data intensive application co.

UofR Alum… long ago!

Ajay Anand

Deputy Director, GIDS

Teaches Capstone and Time Series Courses

Brendan Mort

Director, CIRC

Teaches Data Science Tools Course

Mai Pham

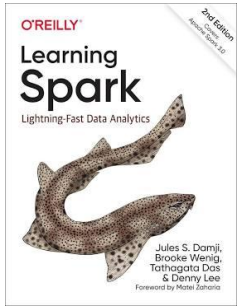
Our Teaching Assistant

Prerequisites

- Python proficiency consistent with Intro to CS and Python
- SQL experience e.g how to select and transform data using SQL
- Familiarity with training and validating models.
- Familiarity with Git version control. Github.
- Familiarity with Unix command line
- Familiarity with Jupyter Notebooks is helpful

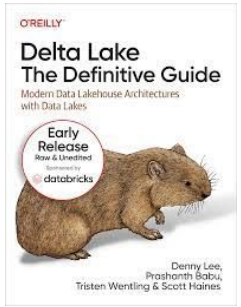
GitHub

DSaS Course Materials



Learning Spark, 2nd Edition

by Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee
Released July 2020
Publisher(s): O'Reilly Media, Inc.
ISBN: 9781492050049
[student copy link]

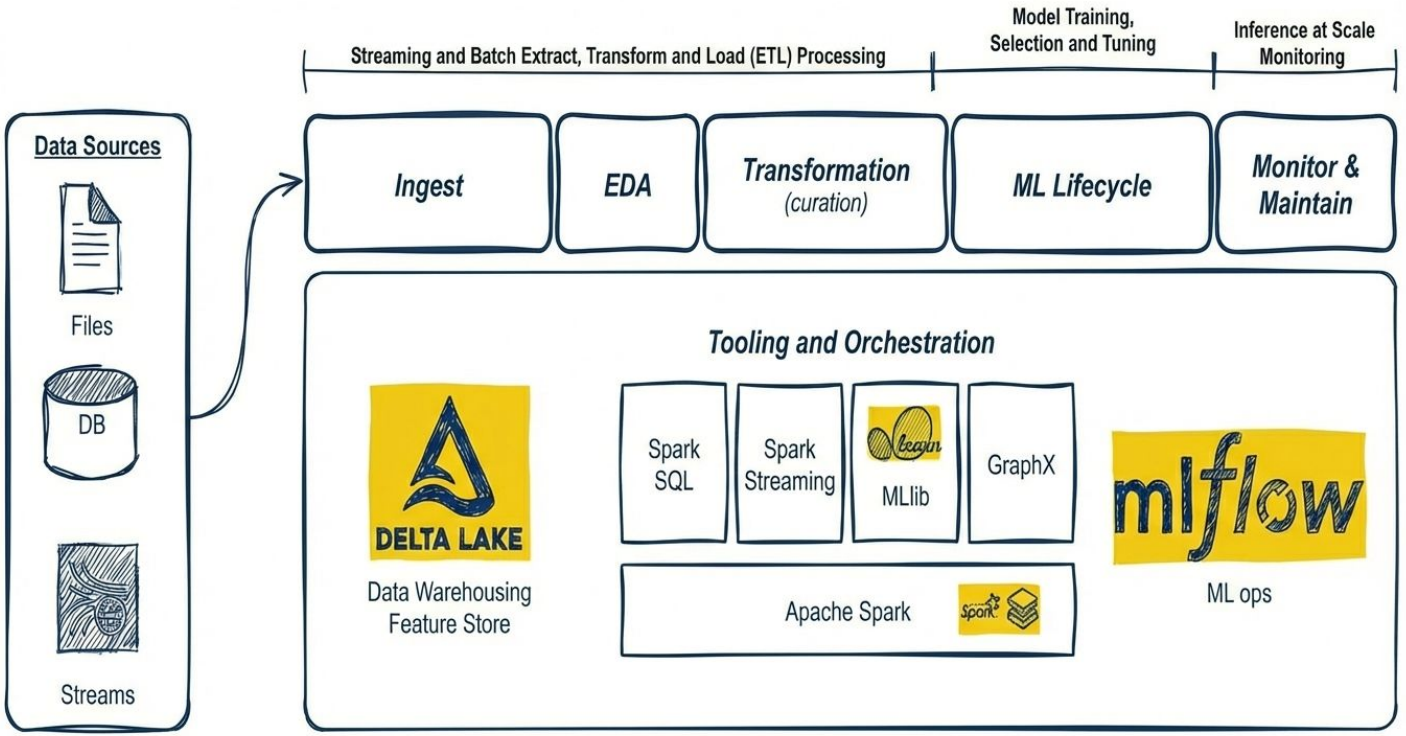


Delta Lake: the Definitive Guide

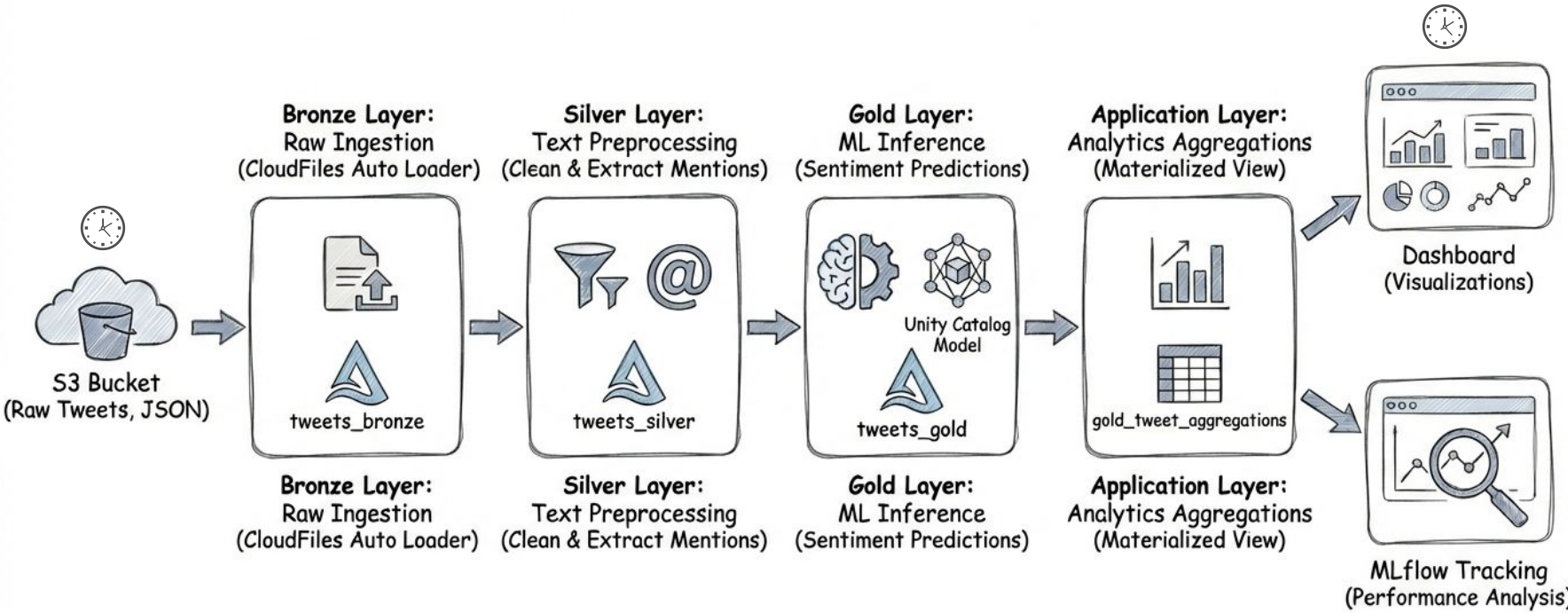
by Denny Lee, Tathagata Das, Vini Jaiswal
Released April 2022
Publisher(s): O'Reilly Media, Inc
ISBN: 9781098104528
[student copy link]

- **Two platforms of interest…**
  - **Databricks Free edition** with serverless compute
  - **Github codespaces** - cloud based development environment
- **How we teach…**
  - Lecture content (presentation)
  - Notebook coding walkthroughs
  - Quizzes
  - Lab Notebooks
  - Student projects

DSaS Course Context - Practice and Platforms



Tweet Sentiment Analysis Pipeline Architecture



DSaS Course Outline - what we teach

Modules

1. Spark Introduction - Prof. Palum
2. Spark Optimization & Streaming - Prof. Anand
3. Data Lakes & MLops - Prof. Mort
4. End 2 End Application Development - Prof. Palum

Topic	Date	5 Quizzes (In Class)	5 Notebooks & 1 Project (BB upload/GitHub)
Spark Introduction Ch. 3, 4, & 5	**Wednesday, February 4th, 2026**	**10** points	**5** points labs/0.1 - Spark Core.py
Spark Optimization Ch. 7, 12	**Wednesday, February 18th, 2026**	**10** Points	**5** Points labs/0.2 - Spark Optimization.py
Spark Streaming Ch. 8	**Wednesday, March 4th, 2026**	**10** Points	**5** Points labs/0.3 - Spark Streaming.py
Delta Lakes Ch. 9	**Wednesday, March 25th, 2026**	**10** Points	**5** Points labs/0.4 - Delta Lake.py
MLops Ch. 11	**Wednesday, April 8th, 2026**	**10** Points	**5** Points labs/0.5 - MLops.py
Final Project	**Friday, May 8th, 2026**	N/A	**25** Points final_project

100 Total Points

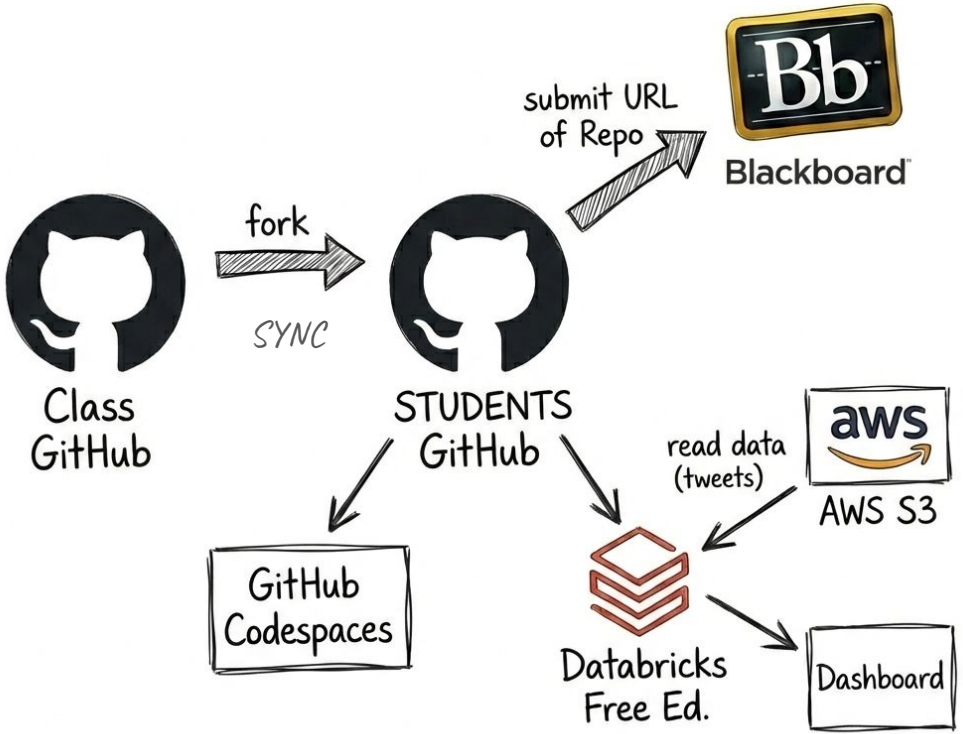
How to engage - our course materials

- Lectures on M & W 16:50 EST Gavett Hall 202
- Blackboard class references & notices
- Office hours by appointment via Zoom
- Teaching Assistant: Mai Pham(mpham8@u.rochester.edu)

You will get out what you put in. Engage. Come to lectures, Ask questions. Respond. There are no stupid questions only missed opportunities to learn.

What to do now...

- Login to Blackboard
- Read and understand the Course Overview - Getting Started
  - *Syllabus*
  - Schedule
- Establish a Github account if you do not already have one.
- Fork the class repo to your GitHub Account
- Follow the instructions in the README.md in the repo



How we use GitHub, Databricks, and Blackboard