



Overview

- Week 1
- Week 2
- Week 3
- Week 4
- Grades
- Notes
- Discussion Forums
- Messages
- Course Info

Week 1

Scalable Machine Learning on Big Data using Apache Spark

Week 1

Discuss this week's modules here.
317 threads · Last post 2 days ago

Go to forum

Week 1: Introduction

Romeo Kienzler

This is an introduction to Apache Spark. You'll learn how Apache Spark internally works and how to use it for data processing. RDD, the low level API is introduced in conjunction with parallel programming / functional programming. Then, different typ...

More

Key Concepts

- Understanding key concepts of distributed computing for big data - Understanding key concepts on Apache Spark and how it achieves

More

Course Introduction

Video: Introduction to Apache Spark for Machine Learning on BigData

6 min

Resume

Video: What is Big Data?

11 min

Reading: Course Syllabus

10 min

Reading: Setup of the grading and exercise environment

10 min

Understanding how Apache Spark works

Video: Data storage solutions

5 min

Video: Parallel data processing strategies of Apache Spark

7 min

Reading: Exercise 1 - working with RDD

10 min

Video: Functional programming basics

6 min

Reading: Exercise 2 - functional programming basics with RDDs

10 min

Video: Resilient Distributed Dataset and DataFrames - ApacheSparkSQL

6 min

Reading: Exercise 3 - working with DataFrames

10 min

Practice Quiz: Practice Quiz (Ungraded) - Apache Spark concepts

4 questions

Quiz: Apache Spark and parallel data processing

7 questions

Due Sep 14, 1:59 AM CDT

Reading: Programming Lanuage Options for Apache Spark (optional)

10 min

