

20773 Analyzing Big Data with Microsoft R

Course Outline

1 - MICROSOFT R SERVER AND R CLIENT

- What is Microsoft R server
- Using Microsoft R client
- The ScaleR functions
- Lab : Exploring Microsoft R Server and Microsoft R Client

2 - EXPLORING BIG DATA

- Understanding ScaleR data sources
- Reading data into an XDF object
- Summarizing data in an XDF object
- Lab : Exploring Big Data

3 - VISUALIZING BIG DATA

- Visualizing In-memory data
- Visualizing big data
- Lab : Visualizing data

4 - PROCESSING BIG DATA

- Transforming Big Data
- Managing datasets
- Lab : Processing big data

5 - PARALLELIZING ANALYSIS OPERATIONS

- Using the RxLocalParallel compute context with rxExec
- Using the revoPemaR package
- Lab : Using rxExec and RevoPemaR to parallelize operations

6 - CREATING AND EVALUATING REGRESSION MODELS

- Clustering Big Data
- Generating regression models and making predictions
- Lab : Creating a linear regression model

7 - CREATING AND EVALUATING PARTITIONING MODELS

- Creating partitioning models based on decision trees.
- Test partitioning models by making and comparing predictions
- Lab : Creating and evaluating partitioning models

8 - PROCESSING BIG DATA IN SQL SERVER AND HADOOP

- Using R in SQL Server
- Using Hadoop Map/Reduce
- Using Hadoop Spark
- Lab : Processing big data in SQL Server and Hadoop