20773: Analyzing Big Data with Microsoft R

Course Details

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
   - Using R client in VSTR and RStudio
   - Exploring ScaleR functions
   - Connecting to a remote server

2. **Exploring Big Data**
   - Understanding ScaleR data sources
   - Reading data into an XDF object
   - Summarizing data in an XDF object
   - Lab : Exploring Big Data
   - Reading a local CSV file into an XDF file
   - Transforming data on input
   - Reading data from SQL Server into an XDF file
   - Generating summaries over the XDF data

3. **Visualizing Big Data**
   - Visualizing In-memory data
   - Visualizing big data
   - Lab : Visualizing data
   - Using ggplot to create a faceted plot with overlays
   - Using rxlinePlot and rxHistogram

4. **Processing Big Data**
   - Transforming Big Data
   - Managing datasets
   - Lab : Processing big data
   - Transforming big data
   - Sorting and merging big data
   - Connecting to a remote server

5. **Parallelizing Analysis Operations**
    - Using the RxLocalParallel compute context with rxExec
    - Using the revoPemaR package
    - Lab : Using rxExec and RevoPemaR to parallelize operations
    - Using rxExec to maximize resource use
    - Creating and using a PEMA class

6. **Creating and Evaluating Regression Models**
    - Clustering Big Data
    - Generating regression models and making predictions
    - Lab : Creating a linear regression model
    - Creating a cluster
    - Creating a regression model
    - Generate data for making predictions
    - Use the models to make predictions and compare the results

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
    - Splitting the dataset
    - Building models
    - Running predictions and testing the results
    - Comparing results

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
    - Creating a model and predicting outcomes in SQL Server
    - Performing an analysis and plotting the results using Hadoop Map/Reduce
    - Integrating a sparklyr script into a ScaleR workflow