Analyzing Big Data
With Microsoft R

The main purpose of the course is to give participants the ability to use Microsoft R Server to create and run an analysis on a large dataset, and show how to utilize it in Big Data environments, such as a Hadoop or Spark cluster, or a SQL Server database.

Who Should Attend
The primary audience for this course is people who wish to analyze large datasets within a big data environment. The secondary audience are developers who need to integrate R analyses into their solutions.

Course Prerequisites
Before attending this course, participants should have:
> Programming experience using R, and familiarity with common R packages
> Knowledge of common statistical methods and data analysis best practices
> Basic knowledge of the Microsoft Windows operating system and its core functionality
> Working knowledge of relational databases

Course Objectives
After completing this course, participants will be able to:
> Explain how Microsoft R Server and Microsoft R Client work
> Use R Client with R Server to explore big data held in different data stores
> Visualise data by using graphs and plots
> Transform and clean big data sets
> Implement options for splitting analysis jobs into parallel tasks
> Build and evaluate regression models generated from big data
> Create, score, and deploy partitioning models generated from big data
> Use R in the SQL Server and Hadoop environments

Course Code: 20773a
Course Duration: 5 Days
SATV: Yes

Course Summary
Module 1: Microsoft R Server and R Client
Module 2: Exploring Big Data
Module 3: Visualising Big Data
Module 4: Processing Big Data
Module 5: Parallelizing Analysis Operations
Module 6: Creating and Evaluating Regression Models
Module 7: Creating and Evaluating Partitioning Models
Module 8: Processing Big Data in SQL Server and Hadoop

contact@wardyit.com
1300 927 394

register for this course today as places are strictly limited
Analyzing Big Data
With Microsoft R

Course Outline

Module 1: Microsoft R Server and Microsoft R Client
This module explains how Microsoft R Server and Microsoft R Client work

- What is Microsoft R Server?
- Using Microsoft R Client
- The ScaleR Functions

Module 2: Exploring Big Data
At the end of this module, the participant will be able to use R Client with R Server to explore big data held in different data stores

- Understanding ScaleR data sources
- Reading data into an XDF object
- Summarizing data in an XDF object

Module 3: Visualising Big Data
This module explains how to visualize data by using graphs and plots

- Visualising in-memory data
- Visualising big data

Module 4: Processing Big Data
This module explains how to transform and clean big data sets

- Transforming Big Data
- Managing datasets

Module 5: Parallelizing Analysis Operations
This module explains how to implement options for splitting analysis jobs into parallel tasks

- Using the RxLocalParallel compute context with rxExec
- Using the revoPemaR package

Module 6: Creating and Evaluating Regression Models
This module explains how to build and evaluate regression models generated from big data

- Clustering Big Data
- Generating regression models and making predictions

Module 7: Creating and Evaluating Partitioning Models
This module explains how to create and score partitioning models generated from big data

- Creating partitioning models based on decision trees
- Test partitioning models by making and comparing predictions

Module 8: Processing Big Data in SQL Server and Hadoop
This module explains how to transform and clean big data

- Using R in SQL Server
- Using Hadoop Map/Reduce
- Using Hadoop Spark

register for this course today as places are strictly limited

1300 927 394
contact@wardyit.com