

## Analyzing Big Data with Microsoft R (20773)

- **Formato do curso:** Presencial
- **Localidade:** Lisboa
- **Data:** 07 Jan. 2019 a 18 Jan. 2019
- **Preço:** 1170€
- **Promoção:** -290€
- **Horário:** Pós-laboral - 2ª, 4ª e 6ª, das 18h45 às 22h15
- **Duração:** 21 horas

The main purpose of the course is to give students 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.

### Destinatários

- 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.

### Pré-requisitos

- 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.

### Objectivos

- 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
- Visualize 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
- 

## Programa

### **Microsoft R Server and R Client**

- What is Microsoft R server
- Using Microsoft R client
- The ScaleR functions

### **Exploring Big**

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

### **Visualizing Big Data**

- Visualizing In-memory data
- Visualizing big data

### **Processing Big Data**

- Transforming Big Data
- Managing datasets

### **Parallelizing Analysis Operations**

- Using the RxLocalParallel compute context with rxExec
- Using the revoPemaR package
- Creating and Evaluating Regression
- Clustering Big Data
- Generating regression models and making predictions

### **Creating and Evaluating Partitioning Models**

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

### **Processing Big Data in SQL Server and Hadoop**

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