

MSDA-1 - MOC 20773 - ANALYZING BIG DATA WITH MICROSOFT R

Categoria: **Data and Analytics**

INFORMAZIONI SUL CORSO



Durata:
3 Giorni



Categoria:
Data and Analytics



Qualifica Istruttore:
Microsoft Certified
Trainer



Dedicato a:
Professionista IT



Produttore:
Microsoft

OBIETTIVI

- Spiegare come funzionano Microsoft R Server e Microsoft R Client
- Utilizzare il client R con il server R per esplorare i Big Data contenuti nei diversi store di dati
- Visualizzare i dati utilizzando grafici e diagrammi
- Trasformare e pulire grandi set di dati
- Implementare le opzioni per dividere i lavori di analisi in attività parallele
- Costruire e valutare i modelli di regressione generati dai Big Data
- Creare, contrassegnare e distribuire i modelli di partizionamento generati dai Big Data
- Utilizzare R nell'ambiente SQL Server e Hadoop

PREREQUISITI

- Esperienza di programmazione usando R, e familiarità con i comuni pacchetti R.
- Conoscenza dei metodi statistici comuni e delle migliori pratiche di analisi dei dati.
- Conoscenza di base del sistema operativo Microsoft Windows e delle sue funzionalità principali.

CONTENUTI

Module 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
- After completing this module, students will be able to:
 - Explain the purpose of R server.
 - Connect to R server from R client
 - Explain the purpose of the ScaleR functions.

Module 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
After completing this module, students will be able to:
Explain ScaleR data sources
Describe how to import XDF data
Describe how to summarize data held in XCF format

Module 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
After completing this module, students will be able to:
Use ggplot2 to visualize in-memory data
Use rxLinePlot and rxHistogram to visualize big data

Module 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
After completing this module, students will be able to:
Transform big data using rxDataStep
Perform sort and merge operations over big data sets

Module 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
After completing this module, students will be able to:
Use the rxLocalParallel compute context with rxExec
Use the RevoPemaR package to write customized scalable and distributable analytics.

Module 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

After completing this module, students will be able to:

Cluster big data to reduce the size of a dataset.

Create linear and logit regression models and use them to make predictions.

Module 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

After completing this module, students will be able to:

Create partitioning models using the rxDTree, rxDForest, and rxBTree algorithms.

Test partitioning models by making and comparing predictions.

Module 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

After completing this module, students will be able to:

Use R in the SQL Server and Hadoop environments.

Use ScaleR functions with Hadoop on a Map/Reduce cluster to analyze big data.

INFO

Esame: 70-773 - Analyzing Big Data with Microsoft R

Manuale: Il Materiale Didattico Ufficiale per tutti i corsi Microsoft MOC può essere richiesto, se disponibile, in forma elettronica (DMOC) invece che cartacea e lo studente iscritto potrà scaricarlo dal sito Microsoft. Chi acquista un DMOC ha diritto a consultare tutte le versioni del manuale, sia quelle precedenti a quella che acquista sia quelle che usciranno successivamente, dove troverà corretti eventuali errori e/o le novità del prodotto.

Prezzo manuale: 170 € incluso nel prezzo del corso

Natura del corso: Operativo (previsti lab su PC)