



DATA ENGINEERING Master

Delivery mode: self-paced, onsite, live online

Course length: 2 days

Target Audience: Data Engineer, Data Scientist

Prerequisite courses: Data Engineering Professional



Data Engineer



Business Analyst



Domain Expert



Data Scientist



Executive



Administrator

Practice building advanced data preparation processes to handle complex requirements in RapidMiner.

Overview

This course provides the opportunity to directly work on creating and modifying RapidMiner Processes to perform data engineering. It covers techniques needed to handle complex data engineering requirements that exceed common data preparation steps.

In addition to *Data Engineering Professional*, consider taking both *Applications & Use Cases Professional* as well as *Machine Learning Professional* before this course

Course Objectives

- Build robust processes
- Perform the most common data cleansing techniques
- Process text data for data mining
- Access web-based data
- Use databases and Hadoop
- Leverage scripts



Data Engineer



Business Analyst



Domain Expert



Data Scientist



Executive



Administrator



Data Engineering Master Topic Guide:

Practice creating and using Regular Expressions.

This includes creation of Regular Expressions because they frequently play a key role in complex RapidMiner Processes.

Review the purpose and uses of Macros.

View processes that leverage Scripts.

Although details of the languages are not covered, the ability to use *R*, *Python*, or *Groovy scripts* is covered.

Practice a variety of techniques for flow control and error handling including:

Loop Attributes, Loop Examples, Loop Collection, Branch and Select Subprocess, Handle Exceptions, and Throw Exception

Review how to implement logging for various types of information.

Practice implementing common advanced cleansing operations including:

Sample data, Replace Missing Values or Impute Missing Values, Normalize, Bin or discretize numerical data, Handle Columns with Many Values, Remove Duplicate examples, Outlier Detection, Dimensionality Reduction.

Learn how to use the Text Processing extension for common tasks.

Learn how the Process Documents operators work, and some of the common document operators including Tokenize, Filter Stopwords, Stemming, and Generate n-Grams.

Learn how to use the Web Mining extension for common tasks including:

Process Documents from Web, Read RSS Feed, and Enrich Data by Webservice.



DATA ENGINEERING Master



Delivery mode: self-paced, onsite, live online
Course length: 2 days



Data Engineer



Business Analyst



Domain Expert



Data Scientist



Executive



Administrator

We designed our program with the most common user personas and the required areas of expertise for applied data science in mind:

