

IBM DataStage for Practitioners

Advanced DataStage training

VIQTOR
DAVIS®

DATA CRAFTSMANSHIP

viqtor-davis.com

DataStage for Practitioners

1 Introduction

During this training, an experienced DataStage consultant will teach how to use the IBM ETL tool DataStage. DataStage is meant for batch- or near real-time processing of large amounts of data.

This training provides a further exploration of themes introduced during the Foundation training and includes additional subjects to further deepen and extend the DataStage knowledge.

2 Content

This training introduces parallel job development techniques in DataStage. During this training the student will gain more extensive knowledge of the DataStage architecture, and also deepened understanding of the DataStage develop and runtime environments. This enables developers to develop parallel jobs that are robust and re-usable. The jobs will also be less sensitive to mistakes and optimized for better performance.

Next to the theory lessons, the student will also work with DataStage during the training. There will be fun exercises to solve during which the instructor will share do's and don'ts within the development process. For the faster students, additional exercises are available.

We will use a Linux environment that is set-up to be used with a few standard commands. Extensive Linux knowledge will thus not be necessary.

The IIS-products can be installed through various methods. Single-, multiserver or grid. For this reason, we will discuss these different topologies of the IIS installation options.

During the training we will follow a Business Case to work on different hands-on exercises. In these exercises many of the discussed topics are combined and you can determine if you can apply the newly acquired knowledge.

3 Program

Additional to theory, applying knowledge to practical situations will also be widely covered. The training takes 3 days, during which different topics will be discussed. For each topic, when relevant, the material consists of theory demo's and hands-on exercises:

Topics*:

- Parallel Framework Architecture
 - o Parallel processing architecture
 - o Pipeline and partition parallelism
 - o Configuration file structure, nodes
 - o Orchestrate job score

- Partitioning, Collecting and Sorting Data
 - o Different partitioning algorithms
 - o Sorting data methods
 - o Sort stage vs inline sort

- Reusable Components
 - o Shared containers
 - o Run time column propagation
 - o Schema files (sequential files)
 - o Dynamic parameter set/value files

- Advanced Data Processing
 - o Different connector stages and their properties
 - o Datasets and their use
 - o Hierarchical Data processing

- Balanced Optimization

- Data Masking

- Performance Tuning of DataStage Jobs
 - o Tunable parameters
 - o Buffering in Jobs

- Creating restartable ETL processes
 - o Error handling
 - o Exception stages
 - o Sequence snapshots

- Tools and Utilities

4 Location

During Covid-19 lockdown, these courses will be held online. Contact us for exclusive on-site training.

VIQTOR DAVIS
Parijsboulevard 143 A
3541 CS Utrecht
The Netherlands

T +31 30 221 90 50
info.nl@viqtordavis.com
viqtordavis.com

VIQTOR
DAVIS®

DATA CRAFTSMANSHIP