# IBM InfoSphere Advanced DataStage - Parallel Framework v11.5

**Course Code: 4499**

This course is designed to introduce advanced parallel job development techniques in IBM DataStage v11.5. In this course you will develop a deeper understanding of the DataStage architecture, including a strong foundation of the DataStage development and runtime environments. This will enable you to design parallel jobs that are robust, less subject to errors, reusable and optimized for better performance.

## What You'll Learn

- Design a job that creates robust test data
- Describe the job execution process
- Generate sequences of numbers (surrogate keys) in a partitioned, parallel environment
- Describe sort key and partitioner key logic in the parallel framework
- Work with complex data
- Describe the Balanced Optimization workflow

## Who Needs to Attend

Experienced DataStage developers seeking training in more advanced DataStage job techniques and who seek an understanding of the parallel framework architecture.

## Prerequisites

IBM InfoSphere DataStage Essentials course or equivalent and at least one year of experience developing parallel jobs using DataStage.

# IBM InfoSphere Advanced DataStage - Parallel Framework v11.5

**Course Code: 4499**

CLASSROOM LIVE

$2,595 USD

3 days

## Classroom Live Outline

**Course Outline:**

**1. Introduction to the Parallel Framework Architecture**

- Describe the parallel processing architecture
- Describe pipeline and partition parallelism
- Describe the role of the configuration file

**2. Compiling and Executing Jobs**

- Describe the main parts of the configuration file
- Describe the compile process and the OSH that the compilation process generates
- Describe the role and the main parts of the Score

**3. Partitioning and Collecting Data**

- Understand how partitioning works in the Framework
- Viewing partitioners in the Score
- Selecting partitioning algorithms

**4. Sorting Data**

- Sort data in the parallel framework
- Find inserted sorts in the Score
- Reduce the number of inserted sorts
- Optimize Fork-Join jobs
- Use Sort stages to determine the last row in a group

**5. Buffering in Parallel Jobs**

- Describe how buffering works in parallel jobs
- Tune buffers in parallel jobs

- Avoid buffer contentions

**6. Parallel Framework Data Types**

- Describe virtual data sets
- Describe schemas
- Describe data type mappings and conversions
- Describe how external data is processed
- Handle nulls

**7. Reusable Components**

- Create a schema file
- Read a sequential file using a schema
- Describe Runtime Column Propagation (RCP)
- Enable and disable RCP
- Create and use shared containers

**8. Balanced Optimization**

- Enable Balanced Optimization functionality in Designer
- List the different Balanced Optimization options.
- Push stage processing to a data source
- Push stage processing to a data target
- Optimize a job accessing Hadoop HDFS file system
- Understand the limitations of Balanced Optimizations

## Classroom Live Labs

**Labs:**

You'll participate in hands-on labs.

# IBM InfoSphere Advanced DataStage - Parallel Framework v11.5

**Course Code: 4499**

VIRTUAL CLASSROOM LIVE

$2,595 USD

3 days

## Virtual Classroom Live Outline

**Course Outline:**

**1. Introduction to the Parallel Framework Architecture**

- Describe the parallel processing architecture
- Describe pipeline and partition parallelism
- Describe the role of the configuration file

**2. Compiling and Executing Jobs**

- Describe the main parts of the configuration file
- Describe the compile process and the OSH that the compilation process generates
- Describe the role and the main parts of the Score

**3. Partitioning and Collecting Data**

- Understand how partitioning works in the Framework
- Viewing partitioners in the Score
- Selecting partitioning algorithms

**4. Sorting Data**

- Sort data in the parallel framework
- Find inserted sorts in the Score
- Reduce the number of inserted sorts
- Optimize Fork-Join jobs
- Use Sort stages to determine the last row in a group

**5. Buffering in Parallel Jobs**

- Describe how buffering works in parallel jobs
- Tune buffers in parallel jobs

- Avoid buffer contentions

**6. Parallel Framework Data Types**

- Describe virtual data sets
- Describe schemas
- Describe data type mappings and conversions
- Describe how external data is processed
- Handle nulls

**7. Reusable Components**

- Create a schema file
- Read a sequential file using a schema
- Describe Runtime Column Propagation (RCP)
- Enable and disable RCP
- Create and use shared containers

**8. Balanced Optimization**

- Enable Balanced Optimization functionality in Designer
- List the different Balanced Optimization options.
- Push stage processing to a data source
- Push stage processing to a data target
- Optimize a job accessing Hadoop HDFS file system
- Understand the limitations of Balanced Optimizations

## Virtual Classroom Live Labs

**Labs:**

You'll participate in hands-on labs.

Dec 18 - 20, 2017 | 11:30 AM - 7:30 PM EST

Jan 3 - 5, 2018 | 8:30 AM - 4:30 PM EST

Feb 26 - 28, 2018 | 8:30 AM - 4:30 PM EST

Apr 30 - May 2, 2018 | 8:30 AM - 4:30 PM EST

Jun 18 - 20, 2018 | 11:30 AM - 7:30 PM EST

# IBM InfoSphere Advanced DataStage - Parallel Framework v11.5

**Course Code: 4499**

SELF-PACED

$1,080 USD

1 session

## Self-Paced Outline

**Course Outline:**

**1. Introduction to the Parallel Framework Architecture**

- Describe the parallel processing architecture
- Describe pipeline and partition parallelism
- Describe the role of the configuration file

**2. Compiling and Executing Jobs**

- Describe the main parts of the configuration file
- Describe the compile process and the OSH that the compilation process generates
- Describe the role and the main parts of the Score

**3. Partitioning and Collecting Data**

- Understand how partitioning works in the Framework
- Viewing partitioners in the Score
- Selecting partitioning algorithms

**4. Sorting Data**

- Sort data in the parallel framework
- Find inserted sorts in the Score
- Reduce the number of inserted sorts
- Optimize Fork-Join jobs
- Use Sort stages to determine the last row in a group

**5. Buffering in Parallel Jobs**

- Describe how buffering works in parallel jobs
- Tune buffers in parallel jobs

- Avoid buffer contentions

**6. Parallel Framework Data Types**

- Describe virtual data sets
- Describe schemas
- Describe data type mappings and conversions
- Describe how external data is processed
- Handle nulls

**7. Reusable Components**

- Create a schema file
- Read a sequential file using a schema
- Describe Runtime Column Propagation (RCP)
- Enable and disable RCP
- Create and use shared containers

**8. Balanced Optimization**

- Enable Balanced Optimization functionality in Designer
- List the different Balanced Optimization options.
- Push stage processing to a data source
- Push stage processing to a data target
- Optimize a job accessing Hadoop HDFS file system
- Understand the limitations of Balanced Optimizations

## Self-Paced Labs

**Labs:**

You'll participate in hands-on labs.

# IBM InfoSphere Advanced DataStage - Parallel Framework v11.5

**Course Code: 4499**

PRIVATE GROUP TRAINING

3 days

Visit us at www.globalknowledge.com or call us at 1-866-716-6688.

Date created: 10/24/2017 1:50:27 PM