

**IBM InfoSphere
Information
Server**

Lesson 8: New Features



Lesson Objectives

- On completion of this lesson, you will be able to understand some of the new features:
- Big Data File Stage
- Hierarchical Data Stage
- Greenplum Connector Stage
- InfoSphere Master Data Management Connector Stage
- Amazon S3 Connector Stage
- Unstructured Data Stage – Microsoft Excel (.xls and .xlsx)
- Sort Stage Optimization
- Improved Flexibility in Record Delimiting
- Operations Console/Workload Management

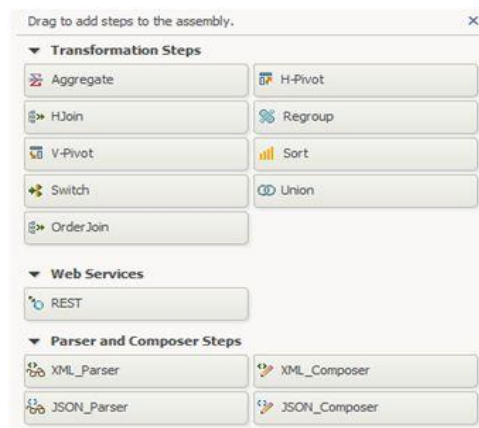


Hierarchical Data Stage



- Ability to process all types of Hierarchical Data (JSON) as opposed to strictly being limited to XML.
- This stage also has some additional functionality which wasn't previously available. If you are familiar with this stage (Hierarchical Data/XML) you will know it has various steps which are added in the Assembly Editor, for a sequence of processing events.
- There are now three new steps:
 - REST – Invokes a RESTful web service
 - JSON_Parser – Parse JSON content with a selected type
 - JSON_Composer – Compose JSON content with a selected type

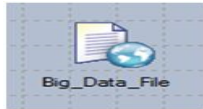
Hierarchical Data Stage



Big Data File Stage



- The Big Data File stage is used to read and write to files on Hadoop (HDFS).
- The Big Data File stage is now compatible with Hortonworks 2.1, Cloudera 4.5, and InfoSphere BigInsights 3.0.



Greenplum ConnectorStage



- You can now use the Greenplum Connector stage for a native connection for accessing data which is located in a Greenplum database.
- You can now also import Table Definitions using the Greenplum Connector framework.



InfoSphere Master Data Management Connector Stage



- The Master Data Management Connector stage can be used to read and write data from the IBM master data management solution – InfoSphere MDM.
- This stage can be configured for Member read and Member write interactions from the MDM server.



Amazon S3 Connector Stage



- Amazon S3 (Simple Storage Service) is a cheap cloud file storage system which offers availability through web services (REST, SOAP, and BitTorrent).
- It offers scalability, high availability, and low latency at extremely competitive prices. The Amazon S3 Connector stage can be used to read and write data residing in Amazon S3.



Unstructured Data Stage – Microsoft Excel (.xls and .xlsx)

- The Unstructured Data stage was first introduced in DataStage v9.1 and was used to read Excel files through a native interface.
- Previously, Excel data was staged as a .csv file or accessed through ODBC. The stage can also now be used to write data to Excel files.



Sort Stage Optimization



- The Sort stage now tries to optimize your DataStage sort operations by converting length bounded columns to variable length before the sort and then converts it back to a length bounded column after the sort.
- When a record's actual size of data is smaller than the defined upper bound, the optimization will result in reduced disk I/O.

Improved Flexibility in Record Delimiting



- The Sequential File stage now gives developers more flexibility with how a source flat file has to be delimited.
- A new environment variable, `APT_IMPORT_HANDLE_SHORT`, can be set to enable the import operator the ability the read in records which do not contain all of the fields defined in the import schema. Previously, these records were rejected by the stage.
- The values assigned to any missing field depends on the data type and nullability.

Operations Console/Workload Management



- IBM lists the Operations Console and Workload Management as new features of the 11.3 release documentation, even though these components have already been introduced in previous releases.
- Both components are now part of the base Information Server installation and Workload Management is now by default enabled.

Operations Console/Workload Management



➤ Sustainable Data Quality

- Governance Dashboard
- Cognos now entitled to support rollout of data quality program

➤ Performance Optimizations

- Parallelism for column analysis
 - Faster algorithms for multicolumn
 - primary key analysis
 - Expanded native connector support for faster read/write

➤ Productivity Enhancements

- Data connections & metadata from IMAM for profiling & rules
- Additional Data Governance Dashboard reports

• Engine Optimizations

- Sort compress reduces I/O size for temp files improving run time
- Tuning +
- 20% performance improvement in some cases