

DataStage Virtual User Group

November 17th

--

Virginie Grandhaye
Connectivity - Offering Manager
virginie.chassain@fr.ibm.com





Please Note

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM's sole discretion.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.

The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

2020 Connectivity's achievements

2020... many releases to improve connectivity

Information Server

Information server 11.7.1 FP1 (June)

Packs for SAP Apps 8.2.0.3 (September)

Coming next : Packs for SAP BW 4.4.0.2

Cloud Pak for Data

Cloud pak for data 3.0.1 (June)

Coming next : Cloud pak for Data 3.5

1. Better « Cloud » connectivity

New connectors

**Enhancement on
existing connector**

**Datasource
versions - official
certification**

**Technical
connectors**

2. Toward the Data Fabric

Datastage is available in Cloud Pak for Data

SAP Packs are coming in Cloud Pak for Data 3.5

Acceleration of Cloud pak for Data connectivity

**...2021 :
Convergence of Datastage connectivity in the Data fabric**

DataStage Connectivity improvements

2020H1 - 11.7.1 FP1 Connectivity improvements Details

A new Azure Data Lake Storage(ADLS) Connector

- ADLS Gen2 Connector

A new AWS Redshift Connector

Cloud Connectors

- Data type enhancement for Cloud storage connectors
- EoW Implementation for Azure, Cloud Object Store, S3, File Connector (Replication)

Kafka connector

- committed message only

Teradata Connector Enhanced with

- TD Passthrough support by TD Connector

FTP Enhancements

- Support customizable "tmp" directory for FTP Enterprise Stage

SnowFlake Connector enhanced with

- Support Using Snowflake on Azure
- Support for Merge functionality in Snowflake Connector
- Support Using Snowflake on Google

Hive Connector

- Update/Delete support

FileConnector Enhanced

- Add support for High Availability (HA)
- Write to Hive Partitioned Tables
- Support numeric/decimal format
- Rewrite of Parquet Parser for supporting INT96 for Parquet file format

DataStage Distributed Transaction Stage

- Add support for DB2 as a target

XML(Hierarchical) Connector enhanced with

- Hierarchical Stage performance improvement
- XML Connector to handle choice discriminators
- XML Connector to handle references to abstract types in the recurring XSD Structures
- Handling of Large XSDs (enhancement/optimization)

Security

- MQ client-mode connection type needs to support SSL/TLS connectivity

Certifications

Cloudera 6.3.1 & 6.3.2
Oracle Autonomous Data Warehouse Cloud
IBM Performance Server v11
Salesforce API version 47
Oracle 19c
Google BigQuery
Java 1.8 for Java based Connectors

2020H2 DataStage Connectivity

- Seamless job design migration from Teradata to BigQuery via the Rapid Job Update Tool
- **SAP ABAP, BAPI, iDoc, BW:**
 - you can now design and run jobs on Cloud Pak for Data that need to read or write SAP data via ABAP, BAPI, iDoc or BW *
 - Access to these data sources is via the *classic* Windows designer
- Many connector and data source enhancements and certifications (see right side)
- New Connectivity options within the browser-based Flow Designer:
 - Azure Data Lake store (Gen 1 and Gen 2)
 - Amazon Redshift
 - Excel

Connectors enhancements

- Azure Blob/File storage
- DB2 for Z
- Google BigQuery Connector
- Google Cloud Storage connector
- File connector
- Kafka Connector
- Hierarchical Stage
- Hive connector
- Snowflake connector
- S3 connector
- Teradata Connector

Datasources certifications

- JDBC Hive 6.0 driver
- Azure SQL database (JDBC)
- Azure SQL Datawarehouse / Synapse (ODBC/JDBC)
- Salesforce (Progress JDBC)
- Salesforce V48
- Salesforce V49
- SQL Server 2019
- Informix14.10
- HDP 3.1.5
- Teradata 17
- Teradata Vantage
- MariaDB

CP4D DataStage Connectivity improvements+ IS patches

Azure Blob/File storage (JR61221)

- Support of additional formats for MetaData import : ORC, Parquet, Avro)

DB2 for Z (JR62421)

- Support binary file formats, in bulk load
- bulk load : add support for SHRLEVEL REFERENCE, SHRLEVEL and the RESUME NO REPLACE as a properties option to set in

Google BigQuery Connector (JR62847 and JR62850)

- Support of Delete, Update and Upsert operations
- Support encrypted credentials json file
- Support SQL JOIN statements
- Access multiple google projects using single credential file
- Various improvements on SELECT statement
- Support User-defined SQL statements
- Support before/after SQL statements

Google Cloud Storage connector (JR62253)

- Support BigQuery « load » option

File connector (JR62676)

- Improve handling of Date and Timestamps with Hive-exec 3.1

Kafka Connector (JR62782)

- Support of key-value messages
- AVRO support (read, write, schema registry for Cloudera and Hortonworks)
- SASL SSL with SCRAM-SHA-256
- SASL SSL with SCRAM-SHA-512

Hierarchical Stage (JR61345)

- Enhancement of JSON parsing

Hive connector (JR62314)

- Load balancing across Hive servers

Snowflake connector (JR62815)

- Support of Runtime Column Propagation (RCP)
- Performance improvements

S3 connector enhancement: (JR62461)

- Read credentials from Amazon profiles

Teradata Connector : (JR62886)

- Allow « Delete then Insert » when RCP enabled
- Certification of Teradata 17

RJUT:

- From Teradata to BigQuery

Datasource Certifications :

- JDBC Hive 6.0 driver (on 11.7 and 11.5)
- Sybase IQ 16.1 and Sybase ASE 16 (on 11.3.1.2)
- Azure SQL database (JDBC)
- Azure SQL Datawarehouse / Synapse (ODBC/JDBC)
- Salesforce (Progress JDBC) (JR61159)
- Salesforce V48 (JR62456)
- Salesforce V49 (JR62659)
- SQL Server 2019 (JR61936)
- Informix14.10 (Progress ODBC on 11.7.1.1)
- HDP 3.1.5
- CDP 7.0
- Teradata 17 (also applies to Teradata Vantage)
- MariaDB (ODBC, JDBC)

SAP Packs :

- [SAP packs 8.2.0.3](#) in DataStage Cartridge
- Load into SAP S4/HANA systems (Odata stage)
- RHEL 8 support
- C4/Hana support

SAP Packs coming in Cloud pak for Data 3.5

SAP architectures and IBM's response to it

SAP Architecture	SAP Apps Pack Link to Detailed System requirements				SAP BW Pack Detailed System requirements		No additional charge
	ABAP*	BAPI	IDOC	Delta Extract*	BW extract*	BW load	OData
CRM	X	X	X	X			SAP Sales Cloud and SAP services Cloud
ECC/ERP	X	X	X	X			X
S4Hana	X	X	X	X	X	X	X
S4Hana on Cloud							X
BW application					X	X	X*
BW 4 Hana					X		X*

* Extract only

Why should I use the SAP packs, instead of the Odata connector available for free? (or what is offered by other stages, that OData doesn't allow)

Run your jobs as backend / asynchronous tasks (ABAP, IDOC)

Odata cannot allow 'partial extract of new records only' based on the last extract (Delta stage)

Other stages are more suitable for on-prem systems (ABAP, BAPI, IDoc and Delta)

OData relies on REST API standards : lower performances in a real-time/near real time use case. IDoc Stages can provide near to real time integration

DB migration use case: ABAP is for high volume data extraction from table through custom built SQL queries using Easy to use Wizard

Odata offers only a limited set of operations (GET, POST, PUT, PATCH, DELETE). BAPI provides an ability to extract or load data using the SAP as well as custom defined function modules

BW extract allows extracting the objects that are available in BW/4 HANA. With ODATA only what is defined in the ODATA service

ABAP Extract, BAPI, IDoc and Delta Extract stages supports parallelism and can be typically used for the high volume data transfer

Delta Extract can fetch full as well as delta data using DataSources or through ODP exposed on DataSources

Acceleration of Cloud Pak for Data connectivity...

...**2021** : Convergence of
Datastage in the Data fabric

CPD 3.5 More Datasources, better

Storage datasources

- Amazon RDS for MySQL
- Amazon RDS for PostgreSQL
- Apache Cassandra
- Apache Derby
- Azure MariaDB
- Box
- CosmosDB
- ElasticSearch
- HTTP
- Microsoft Azure Blob
- MongoDB (read only)
- Storage volumes
- SAP Hana (no driver provided)
- OData / SAP OData write support

Services datasources

- Data Virtualization Manager (z/OS)
- DB2 EventStore
- SPSS Analytics Server
- Amazon GovCloud (through S3 connector)

SSL/TLS v1.2 wider support

- Apache Derby
- Apache HDFS
- Data virtualization
- Data Virtualization for z/OS
- DB2 family
 - DB2,
 - DB2 BigSQL,
 - DB2 Event store,
 - DB2 for i,
 - DB2 for z/OS,
 - DB2 Warehouse on Cloud

Not everything is available in Datastage yet...
they will come in a continuous delivery mode.

[CPD 3.5 Documentation](#)

Let's build stronger products together...

[Connectivity Aha Ideas Portal](#)

1. Register / Login to the portal
2. Vote for existing Ideas
3. Create your own

More votes... higher likelihood to happen sooner

IBM DataOps Community

Connect with experts and peers to elevate technical expertise, solve problems and share insights.

[Join / Log in](#)

DataStage - Data Integration

[Group Home](#)

[Discussion](#) 34

[Library](#) 13

[Blogs](#) 4

[Events](#) 5

[Members](#) 158

[Datastage Community](#)



IBM Data and AI

[Welcome to the IBM Data and AI Ideas Portal for Clients!](#)

We welcome and appreciate your feedback on IBM Data and AI Products to help make them even better than they are today!

Before you submit an idea, please perform a search first as a similar idea may have already been reported in the portal. If a related idea is not yet listed, please create a new idea and include with it a description which includes expected behavior as well as why having this feature would improve the service and how it would address your use case.

Thank you

Virginie Grandhaye

virginie.chassain@fr.ibm.com