

The modern data warehouse How IBM DataStage is an essential ingredient to deliver real-time analytics

—
July 16th
Virginie Grandhaye – Senior Offering Manager

IBM Data And AI

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.

Community Webinar series: DataStage-Data Integration community

Title: Cost optimization on Multi cloud environments using IBM DataStage – available on demand

Overview: In today's business landscape, businesses are looking to maximize their IT budgets and reduce costs by moving their on-prem data in data warehouses to multi-cloud environments.

Title: Tech-talk: Operationalizing containers within your organization and speed up data pipelines using IBM DataStage– available on demand

Overview: Join this webinar to understand why Kubernetes is a key requirement to automated deployment of containers and IBM DataStage on Cloud Pak for Data delivers improved resource utilization, reduced TCO and increased scalability.

Title: The modern data warehouse: How IBM DataStage is an essential ingredient to deliver real-time analytics– 16 July 2020 – 9:00 AM EDT

Overview: The explosive growth of data has forced organizations to use their enterprise data warehouse (EDW) for purposes that it was never intended for – including running extract, transform, load (ETL) workloads and storing large volumes of unused data.

Title: Tips and tricks to get more of IBM DataStage – 21 July 2020 – 11:00 AM EDT

Overview: Learn about new features in IBM DataStage and DataStage Flow Designer to speed up the time to connect to new sources, targets while building ETL jobs and improve your productivity.

Speakers



Beate Porst
Program Director Offering Management, DataOps - Data Integration worldwide offering leader
IBM Data and AI

Speakers



Scott Brokaw
Lead Offering Manager, DataOps - Data Integration
IBM Data and AI

Speakers



Virginie Grandhay
DataOps Connectivity Offering Manager
IBM Data and AI

Speakers



Kevin Wei
Offering Manager, DataOps - Data Integration
IBM Data and AI

[Community webinars](#)

Datasphere

storage capacity
worldwide will **grow**
to 6.8 zettabytes (ZB)
this year

increase of 16.6%
over 2019.

2019-2024
forecast :

17.8% CAGR

IDC - May 13, 2020:

<https://www.idc.com/getdoc.jsp?containerId=prUS46303920>

Coronavirus global online traffic impact as of June 2020

Published by [J. Clement](#), Jun 19, 2020



The global coronavirus pandemic is affecting consumer behavior worldwide. In the week ending June 14, online traffic across 20 different industries increased by 10.2 percent compared to the reference period in January and February 2020. Online transactions increased by 32.9 percent compared to the index period, and decreased 3.1 percent versus prior week.

For further information about the coronavirus (COVID-19) pandemic, please visit our dedicated [Facts and Figures](#) page.

Data ...

...~~where, how~~

Data...

... why,

Business value

Define your Strategy

89%

of businesses struggle with managing data.¹ Understanding an organization's business goals is critical to developing an effective data strategy for analytics and AI.

Deliver business ready data fast with DataOps

<https://www.ibm.com/downloads/cas/GLNAZBXN>

IBM DataStage

Deliver business-ready data in real time for AI with IBM Cloud Pak for Data DataStage

<https://www.ibm.com/downloads/cas/M0E2YKB6>

Why you can't have AI without 'information architecture'

<https://siliconangle.com/2019/02/13/cant-ai-without-information-architecture-think2019/>

Poll

In my current job role, this is more relevant to my areas of interest

- A. Data Governance, Data Masking, Optimized infrastructure, Information systems, ETL
- B. Decision intelligence, AI, Business ROI, Machine learning
- C. I care equally for both

Data Governance

Data masking

Optimized
infrastructure

Enterprise compliance

Cost reduction

Self service data access

Automation

Cloud

IT

« I have an IA strategy »

Business

« I have an AI strategy »

Decision intelligence

Machine learning

Data science

Business intelligence

Digital transformation

AI

Trust in data

5 Steps to Build the Right AI Infrastructure in Your Organization

Gartner (2020 July 6th)

<https://www.gartner.com/document/3987153?ref=algobottomrec&refval=3986597>

What is a Modern Datawarehouse?

Consumable by various teams

Business leaders

Data scientists

Data Analysts

Data Engineers

IT

Cloud based

Scalable

Cost efficient

Optimized

Business ready data fast access

High Performances

Trusted

Integration ready

Provides business value

Datastage Datawarehouse connectivity

- IBM Netezza Performance server
- Oracle Autonomous Data Warehouse Cloud
- Google BigQuery
- Amazon Redshift
- Teradata / Teradata Vantage
- Snowflake
- Azure Synapse (coming 2020H2)
- Yellowbrick (coming 2020H2)

Now Tech : Cloud Data Warehouse, Q2 2020

—

Forrester, July 1st

<https://www.forrester.com/report/Now+Tech+Cloud+Data+Warehouse+Q2+2020/-/E-RES160679>

Enterprise data warehouse optimization

—

IBM

<https://www.ibm.com/account/reg/us-en/signup?formid=urx-35177>

Influence the roadmap

<https://ibm.biz/IBM-Data-and-AI-Ideas>

DataStage : Broader, Faster, Safer Connectivity

IS 11.7.1+ / CPD 3.0

Hadoop

- **HBase** connector
- **Hadoop File** Connector
- **Kafka** Connector [enhanced](#)
- **Hive** Connector [enhanced](#) ([Write to Hive Partitioned Tables](#))
- MongoDB support
- **Cassandra** connector (incl. Data Lineage and metadata import)
- BDFS Kerberos improvements for non Hadoop environments
- Apache Sequential File support for File Connector
- HA support for HDFS/File Connector
- Presto
- Cloudera [up to CDH 6.3.3 / CDP 7.0](#)
- HDP up to 3.1

File

- XML connector [enhanced](#)
- [INT96](#) for Parquet file

Cloud

- Amazon EMR/Hive
- Amazon Redshift
- Amazon S3 KMS Support
- Amazon S3 Parquet and ORC support
- AWS Aurora PostgreSQL
- AWS Dynamo DB (limited)
- **Snowflake** connector [enhanced](#)
- **Azure Cloud Storage** connector
- Azure CosmosDB support via Cassandra connector
- **Azure Data Lake Storage** Connector (Gen 1 and [Gen 2](#))
- **Salesforce** (PK Chunking) [API 48 support](#)
- **IBM Cloud Object Storage** connector
- **Google BigQuery** Connector
- **Google Cloud Storage** Connector
- SAP Odata support
- [Oracle Autonomous Data Warehouse Cloud](#)
- EoW Implementation for Azure, Cloud Object Store, S3, File Connector (Replication)

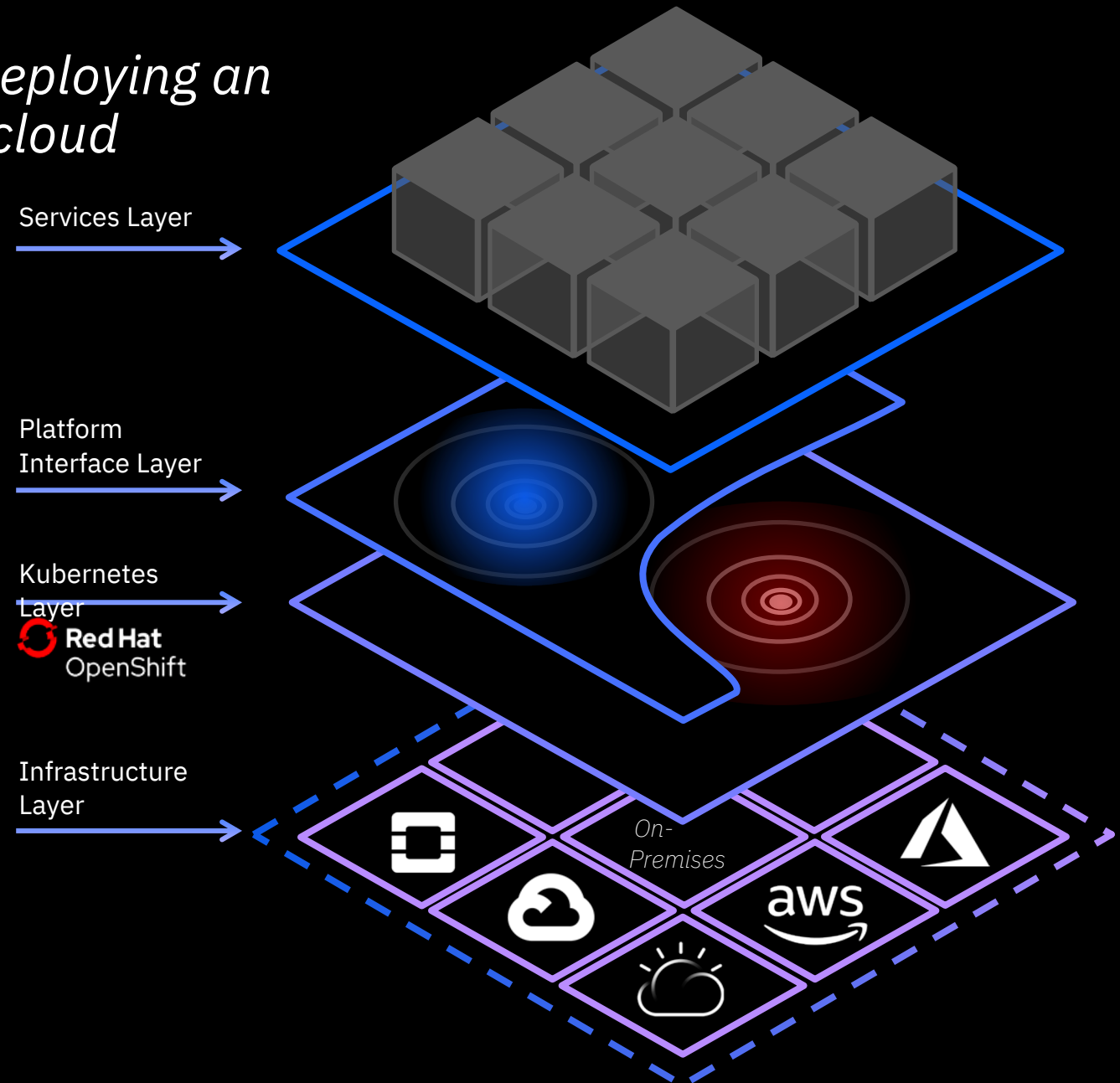
Enterprise

- Oracle [19c](#) (incl. CBD and PDB)
- Siebel 8.2.2.4 certification
- Sybase datatype enhancement & IQ 16.1 support
- New SAP BW & ERP Ppacks
- Data Masking ODPP v11.3 support and expanded Data masking policy support
- DTS Connector: MQ Client mode
- MQ Connector version update
- ILOG Connector Decision Engine
- Db2 v12 z/OS certification
- Greenplum v5.4 certification
- Teradata Connector V16.2 (Big Buffer Support, [passthrough support](#))
- **SAP** ERP Pack V8.1 ([Delta extract stage](#) , [contenerized delivery](#))
- Db2 connector support for External Tables
- RJUT usability improvements for easy PDA→IIAS migration
- Filter condition push-down
- FTP support for customizable /tmp and FTPS
- Teradata Connector [enhanced](#)
- [Netezza Performance Server V11](#)
- [Security enhancement](#)

IBM Cloud Pak for Data

Delivers the foundational platform for deploying an information architecture for AI, on any cloud

- Eliminate data silos, connect all data
- Automate and govern the data and AI lifecycle
- Operationalize AI with trust and transparency
- Avoid lock-in, run anywhere with agility



IBM Cloud Pak for Data

Collect

Db2 Warehouse
Db2 Event Store
Virtualization
Db2^s
Performance
Server^s

Organize

Watson Knowledge
Catalog
DataStage^s
MDM^s

Analyze

Streaming Analytics
Watson Studio
Cognos Dashboards
Watson Assistant^s
Cognos Analytics^s

Infuse AI

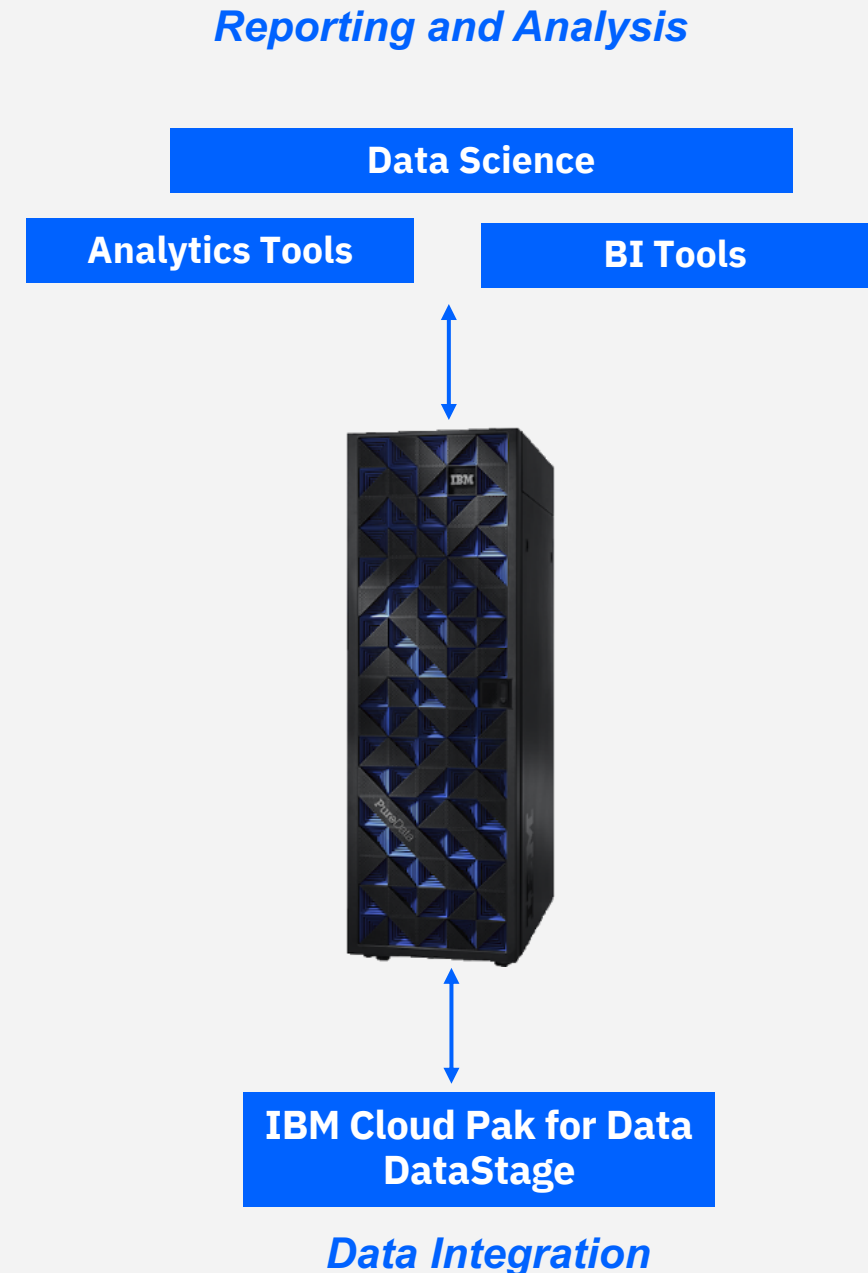
Auto AI
Watson Machine
Learning
OpenScale

Common operation and management



Modernize your on-prem Data Warehouse or move to the cloud

- Cloud-ready data integration architecture for multi-cloud deployments with elastic scaling and balancing to support high data volumes
- Optimized load times with co-located IBM Netezza on IBM Cloud Pak for Data System
- Trusted data delivery with Data Lineage and in-line quality for governance initiatives
- Reduced total cost of ownership by up to 50% with automated administration of backup, recovery, and patch management
- Run your workloads more efficiently on any cloud with built-in workload balancing and parallel engine



Customer success stories

Business Challenge

Decision makers at NYGH must monitor key metrics that influence the hospital's care processes and funding reimbursement. The hospital needed an analytics architecture that could provide dynamic insight into large volumes of data.

Tech Solution

Buisness Intelligence (Cognos)
DataWarehouse (Netezza)
ETL (DataStage)

Results

“With our new IBM Analytics platform, we now have a fast, powerful operational decision support solution...”

BI and Data Architect at NYGH

Solution: **IBM Hybrid Cloud** Industry: **Healthcare**

North York General Hospital

Continuously improving the quality of care with powerful, flexible analytics

For Canadian hospitals, monitoring quality of care and operational performance isn't just important for day-to-day management—it's also the key to securing government funding reimbursement. North York General Hospital (NYGH) is using IBM® Analytics solutions to help ensure quality of care by calculating daily performance metrics, identifying trends, and fine-tuning its processes.

Share this    

[Learn about Cognos Analytics](#)



North York General Hospital: Improving patient care with IBM Analytics

Success Story :

<https://www.ibm.com/case-studies/north-york-general-hospital-hybrid-cloud-cognos-analytics>



Business Challenge

As competition in the energy industry heated up, ENTEGA wanted to gain an advantage over its rivals by sharpening its ability to win new customers, retain existing ones and secure their loyalty.

Tech Solution

Business Intelligence (Cognos)
DataWarehouse (Netezza)
ETL (DataStage)
Analytic tools (SPSS)

Results

« Now, we can achieve four times as many sales by writing to just half of the customers »

Head of Marketing at Entega

Solution: **IBM Hybrid Cloud** Industry: **Energy & Utilities**

ENTEKA

Dazzling customers with compelling, relevant offers powered by IBM Analytics

ENTEKA harnessed IBM Analytics solutions to dig deep into customer and operational data, using the new insight to send targeted offers to the right customer at the right time, via the right channel.

Share this [f](#) [t](#) [G+](#) [in](#)

5x higher

response rates from accurately targeted communications

4x sales

from contacting 50 percent fewer customers with targeted offers

Cuts

cost per campaign

Success Story :

<https://www.ibm.com/case-studies/entega>



“One of the great things about the Cloud Pak for Data System is the speed with which we'll be able to launch and scale our analytics platform. The integrated stack contains what we need to improve data quality, catalog our data assets, enable data collaboration, and build/ operationalize data sciences.”

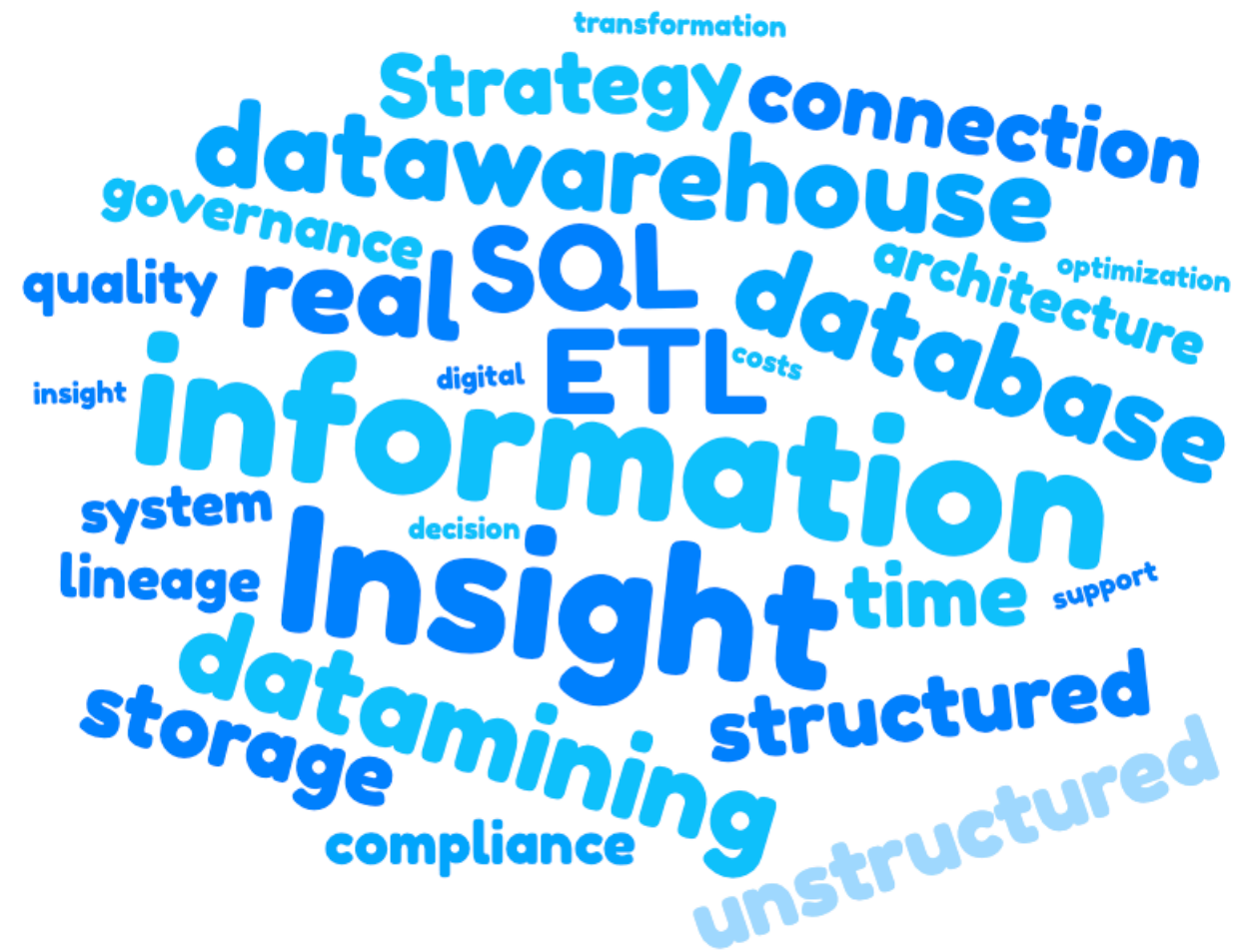
— Steve Lueck, Vice President, Data Management Associated Bank

Video :

<https://www.youtube.com/watch?v=oO9cnbOlVCQ>



What is your data strategy?



Let us know

<https://www.ibm.com/products/infosphere-datastage?schedulerform>



Thank you

-

Virginie Grandhaye
virginie.chassain@fr.ibm.com

Join me on LinkedIn



