IBM DataStage Flow Designer:
Tips & Best Practices

Kevin Wei
Offering Manager
kevinw@ibm.com

IBM

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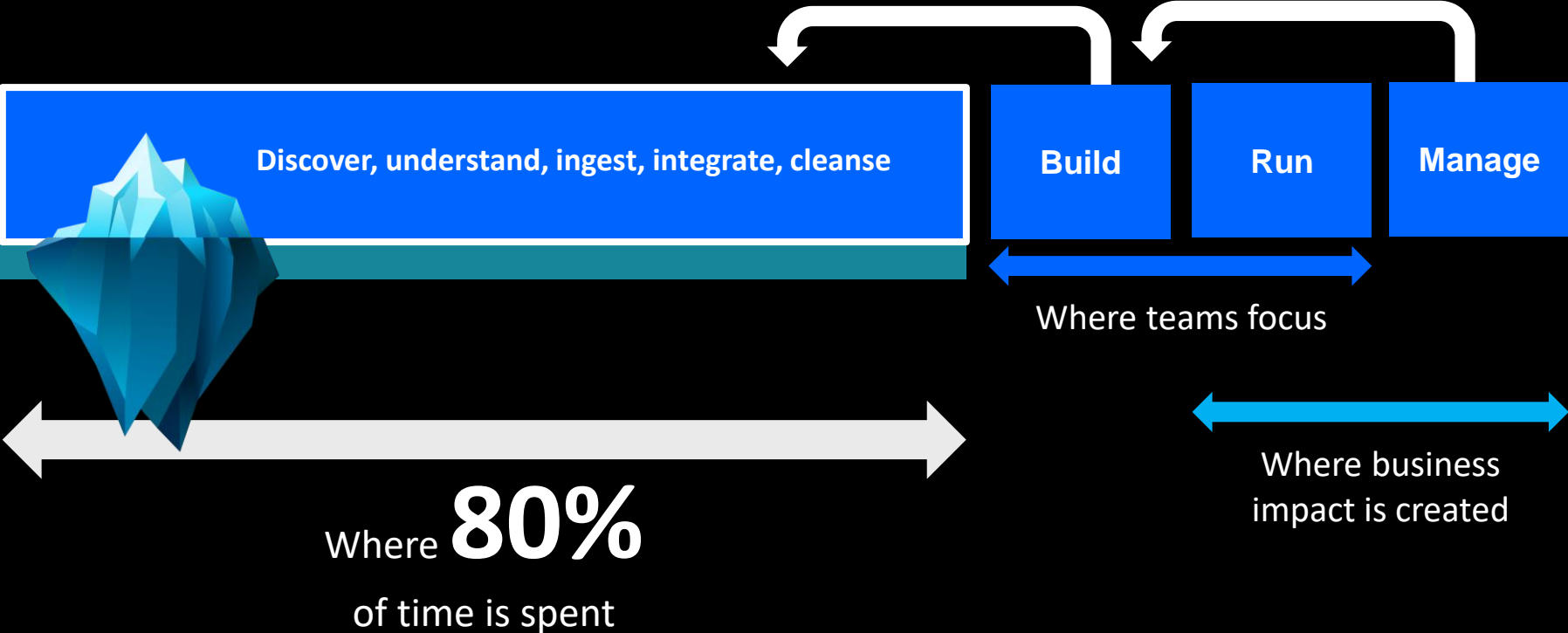
Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.

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.
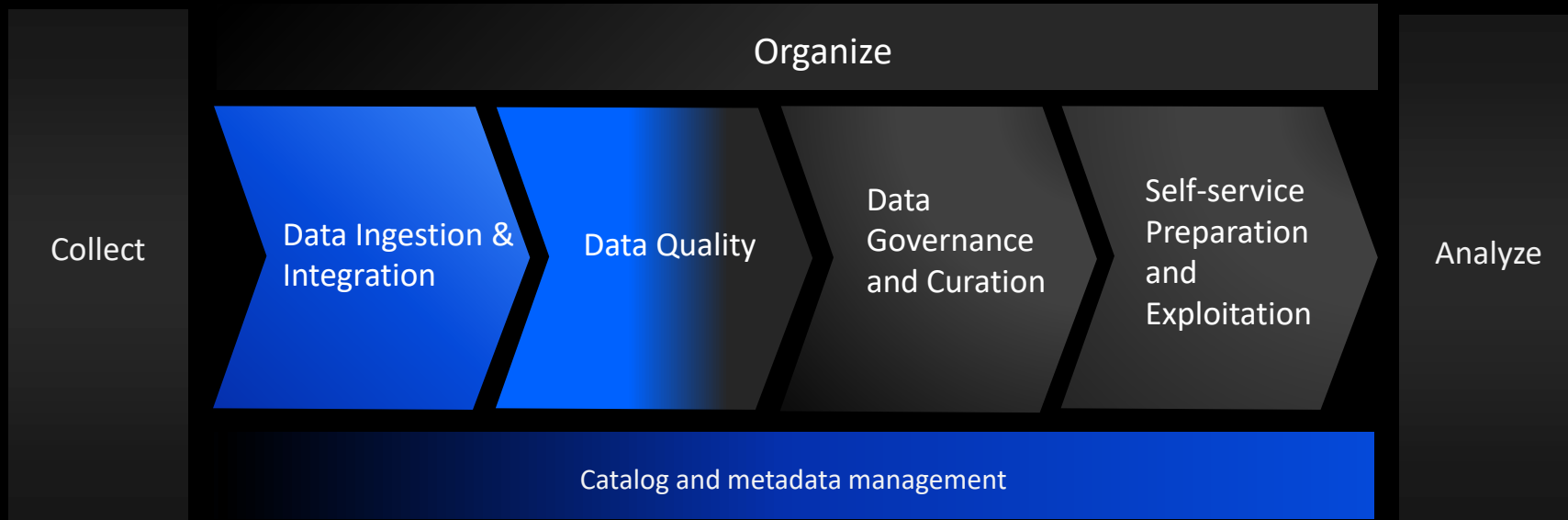
Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user's job stream, 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.
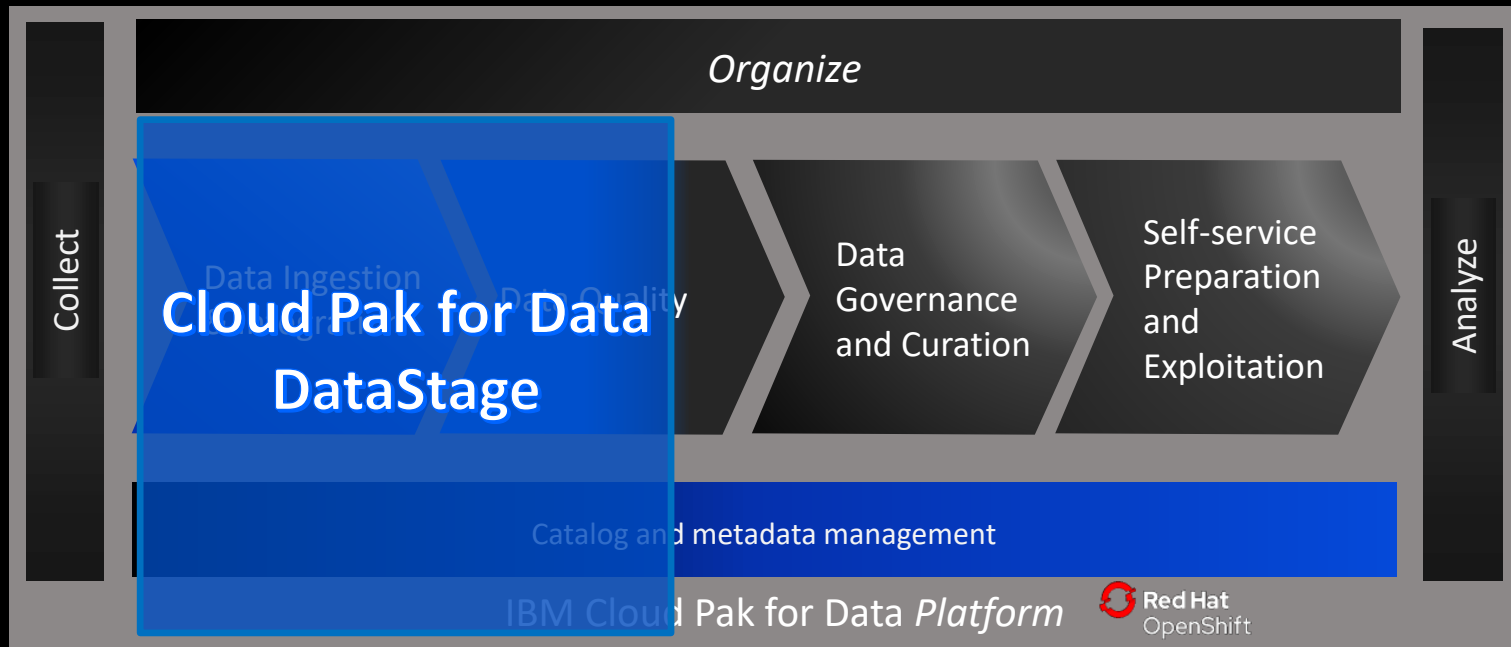
Getting Data to your AI Initiatives is Hard

Discover, understand, ingest, integrate, cleanse

Build | **Run** | **Manage**

Where teams focus

Where **80%** of time is spent

Where business impact is created

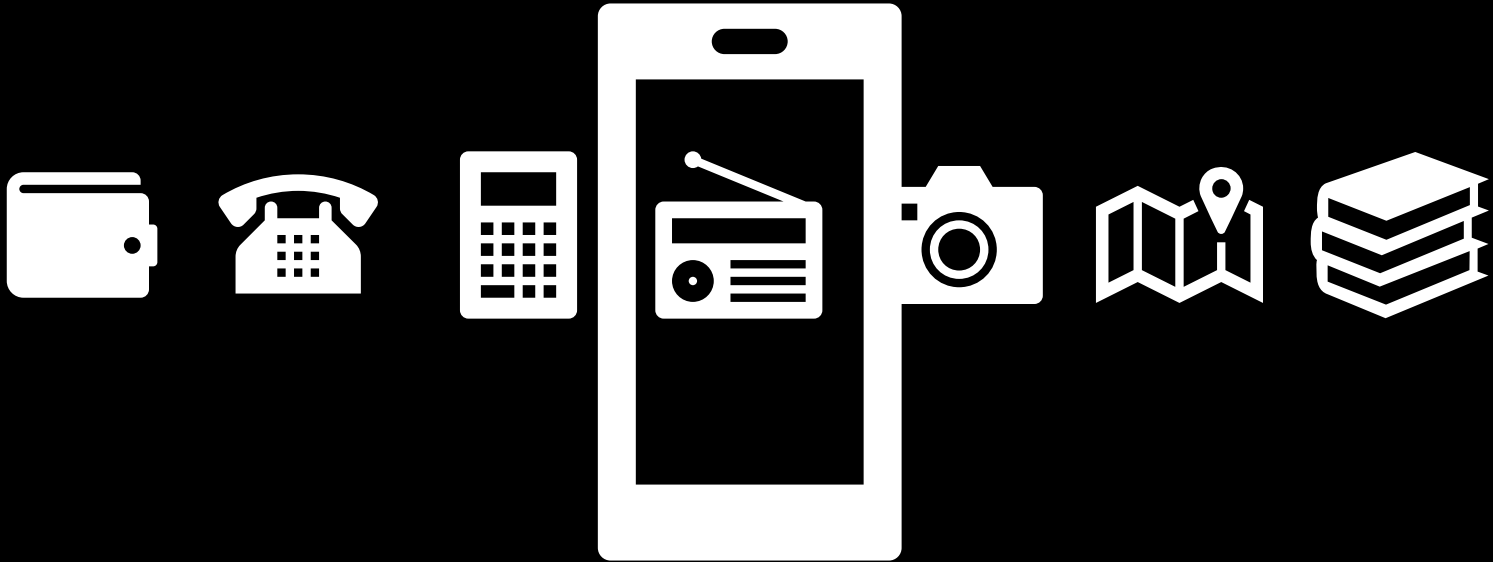# DataOps aims to increase efficiency to operationalize Data Delivery
Data Integration & Quality are at the heart of your DataOps value chain

Improving Data Efficiency needs a more holistic approach
From DataStage to Cloud Pak for Data DataStage

Organize

Collect

Data Ingestion

Cloud Pak for Data DataStage

Data Quality

Data Governance and Curation

Self-service Preparation and Exploitation

Analyze

Catalog and metadata management

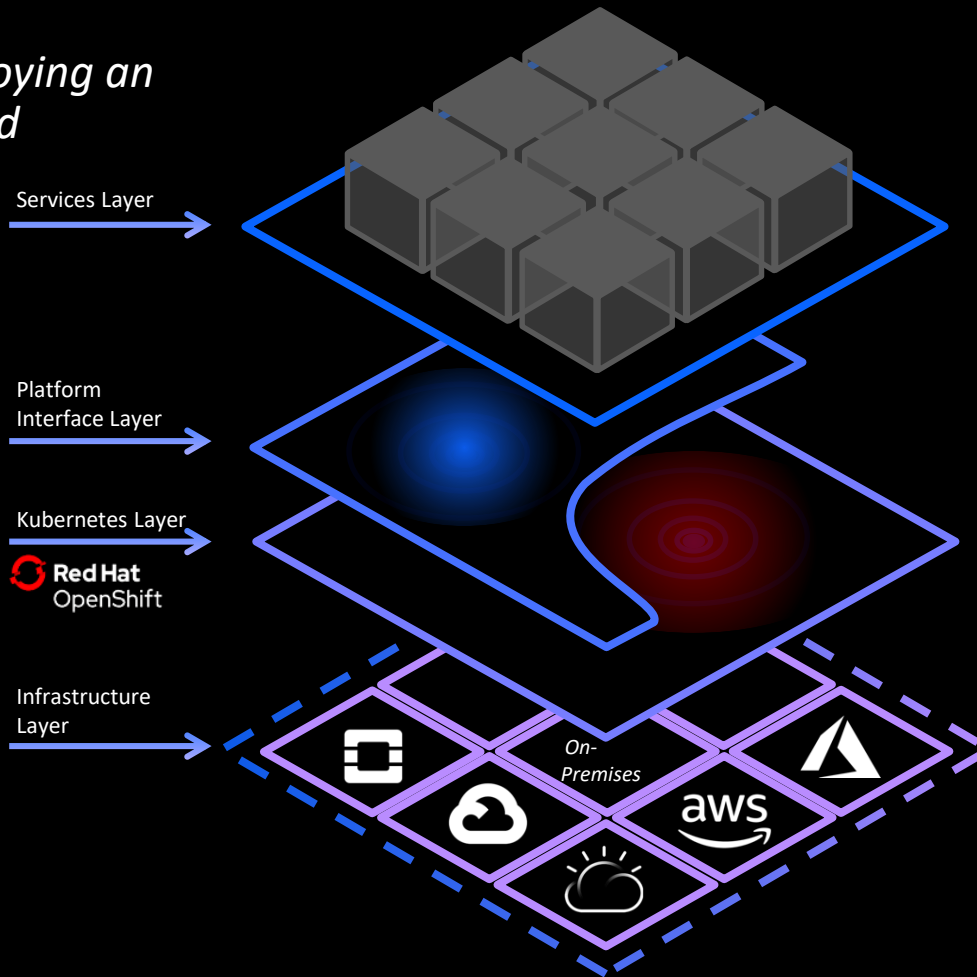IBM Cloud Pak for Data *Platform*

Red Hat OpenShift

The Smartphone revolutionized how we communicate and obsoleted the need to use dedicated tools

IBM Cloud Pak for Data is the *smartphone* for your Data and Analytics apps
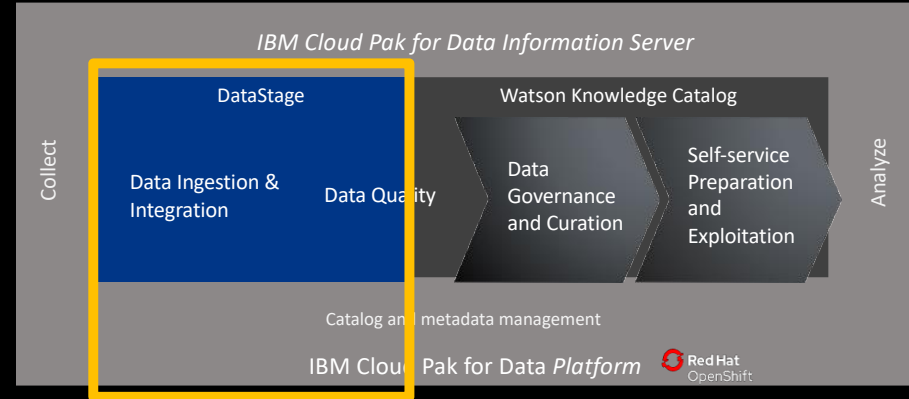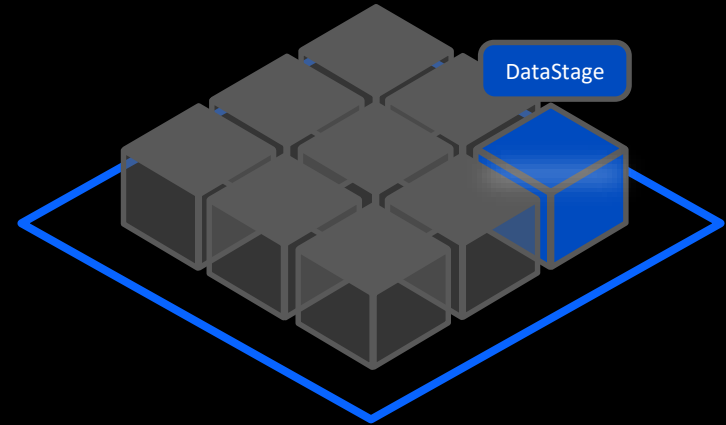
IBM Cloud Pak for Data

*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

Services Layer

Platform Interface Layer

Kubernetes Layer

Red Hat OpenShift

Infrastructure Layer

On-Premises

aws

IBM Cloud Pak for Data DataStage

- Provides state of the art *Data Integration* and *Quality* capabilities ready to power your AI initiative

- Underpinned by an enterprise container platform for enhanced security, reliability and scalability

- Flexibility to runs across any cloud for easily integrate their data to speed innovation.

- Seamlessly combine with other Cloud Pak for Data capabilities such as Governance, Virtualization, Data Science and more

DataStage

IBM Cloud Pak for Data Information Server

Collect

DataStage

Data Ingestion & Integration | Data Quality

Watson Knowledge Catalog

Data Governance and Curation

Self-service Preparation and Exploitation

Analyze

Catalog and metadata management

IBM Cloud Pak for Data *Platform*

Red Hat OpenShift

2019 Release History for IBM Data Integration

Q1 2019	Q2 2019	Q3 2019	Q4 2019

New Release of IBM InfoSphere Information Server

11.7.1

New: IBM Data Edition for Cloud Pak for Data

New: IBM Advanced Data Preparation

New: IBM InfoSphere Information Server V11.7.1 SR1

New: Major update to the DataStage Flow Designer

New: Update to the SAP Pack delivering OData support

New: IBM InfoSphere Information Server V11.7.1 SR2
→ 90+% stage completion for DFD
→ 1.Phase of multi cloud invocation

New: DataStage Real Time Capture Capabilities Tech-Preview

Update: New Version for IBM Advanced Data Prep

New: Cloud Pak for Data Information Server and DataStage offering

Empower the User

DataStage Flow Designer
A New Integration Experience

Intuitive, browser-based
(no-install) experience

– Reducing total cost of ownership

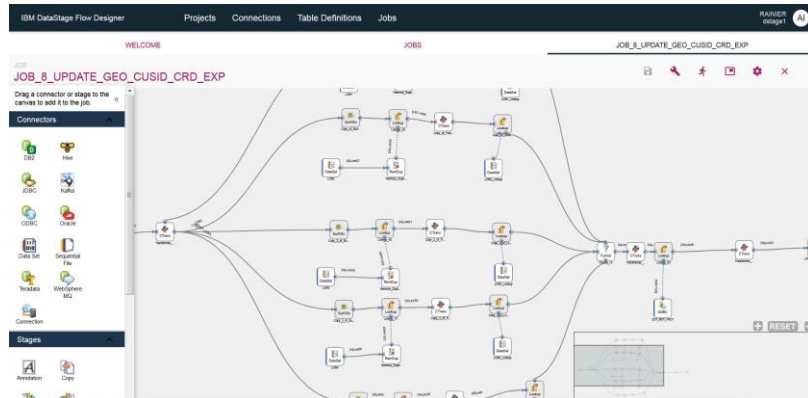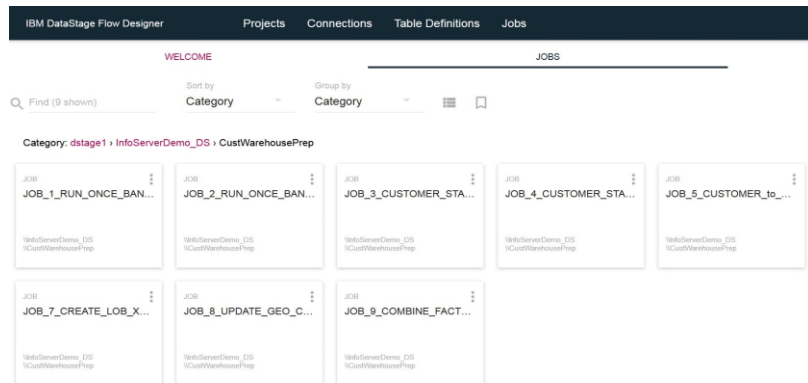Full backwards compatibility to DS/QS Thick client

Accelerated productivity through:
– Automatic schema propagation
– Git Source code control Integration
– Highlighted design errors
– Powerful type-ahead search
– Server-side compilation
– M/L-based design acceleration

Supports 3 Canvas types and 95+ stages
– continuously adding new/more stages

IBM Confidential

IBM DataStage Flow Designer:
Tips & Best Practices

Agenda

1. Job Clustering
2. Smart Palette
3. Stage Suggestions
4. One-click Stage Linking
5. Search Support
6. Horizontal Layout
7. Automatic Propagation
8. Auto-Scaling
9. Importing Jobs (ISX)

Thank you

Kevin Wei
Offering Manager
IBM Data and AI
—
kevinw@ibm.com
http://ibm.com/datastage