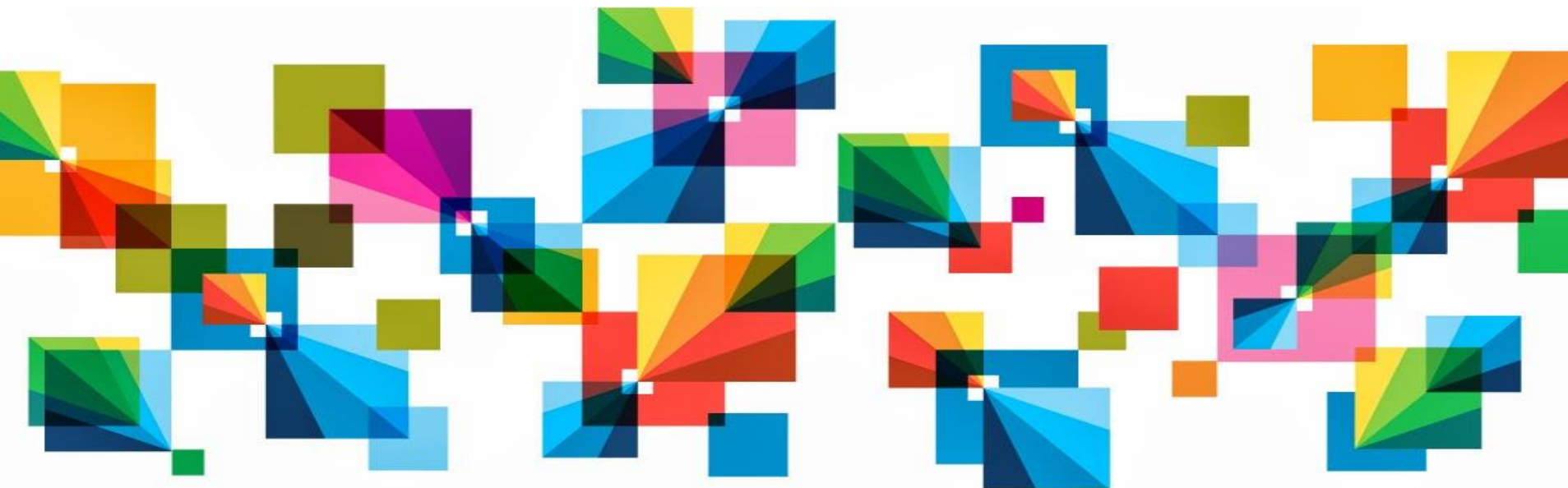


Effectively Integrate All Your Data

InfoSphere Information Server

Highly Scalable Data Movement and Integration



Service Oriented Finance has New Data Integration Challenges

OK, now I understand how to analyze data in motion and data at rest, but...

How do I integrate these with my existing databases?



Service Oriented Finance CTO

Glad you asked!

Let me show you an optimal way to integrate all your big data.



IBM

Big Data Increases Big Integration Challenges



- New data stores
 - Hadoop HDFS
 - NoSQL
- New data types and formats
 - Unstructured, semi-structured
 - JSON, Avro, etc.
 - Video, documents
- Piecemeal open source solutions
 - Lack of standards
- Larger volumes of data
 - Need to move, cleanse, and transform huge amounts of data
 - Big data requires big scalability

Integration Approach 1 – Use Open Source Tools

- Oozie for data flow orchestration
- Flume for bringing in external data sources
- Sqoop for integrating HDFS with relational databases

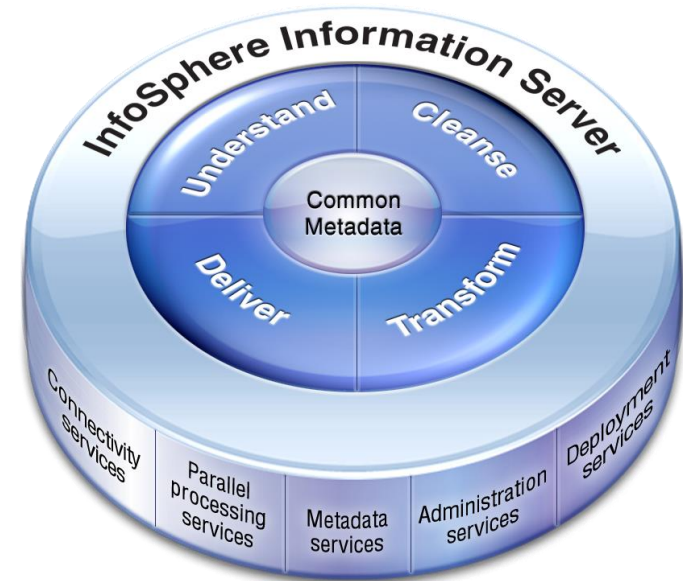


Using three different and immature tools raises concerns

- Productivity and cost
 - ▶ How productive will my team be learning three different tools?
- Risk
 - ▶ The tools are immature. Will they become a huge pain point?
 - ▶ Will these tools scale to meet my big data requirements?
 - ▶ Will I have to add more tools if these can't do everything I need?

Information Server Provides a Complete Enterprise Class Integration Solution

- Same graphical design tools regardless of data source
- Proven tools with years of wide spread usage
- Proven linear scalability
- Speeds productivity, reduces cost and risk
 - Shared meta data fosters collaboration
 - Maximizes reuse – build once and share
 - Hundreds of pre built components
- Connects to every imaginable data source
 - Databases, applications, files, message queues, Hadoop, NoSQL

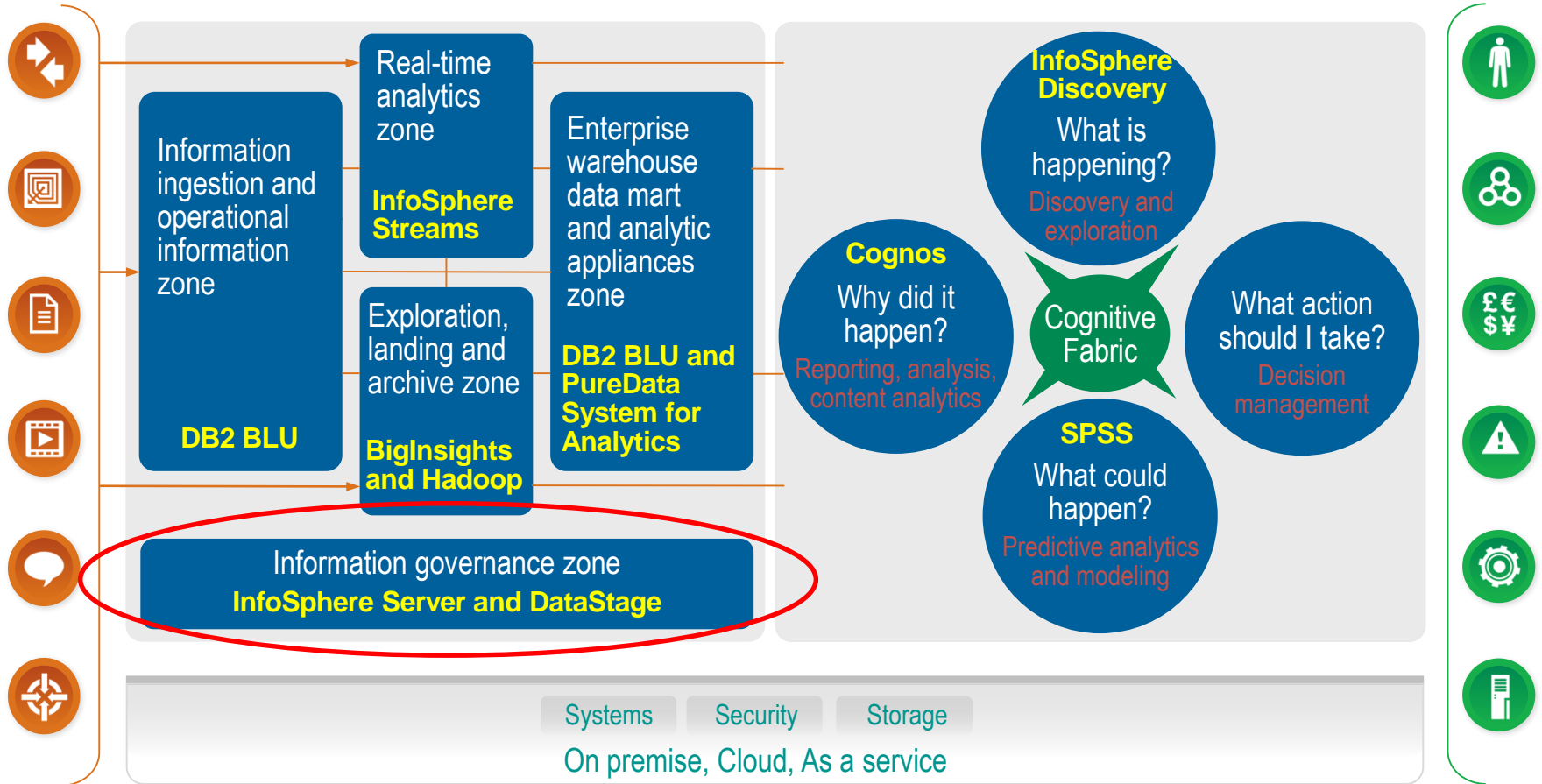


Make Sense out of BigData with IBM Watson Foundations

All Data

IBM Watson Foundations

New/Enhanced Applications



Analyze all data, from any source, with the right technology

Service Oriented Finance Wants to Build the 360 View of Their Customers

How do I integrate the negative sentiment Twitter data that we analyzed with our relational database?



Service Oriented Finance CTO

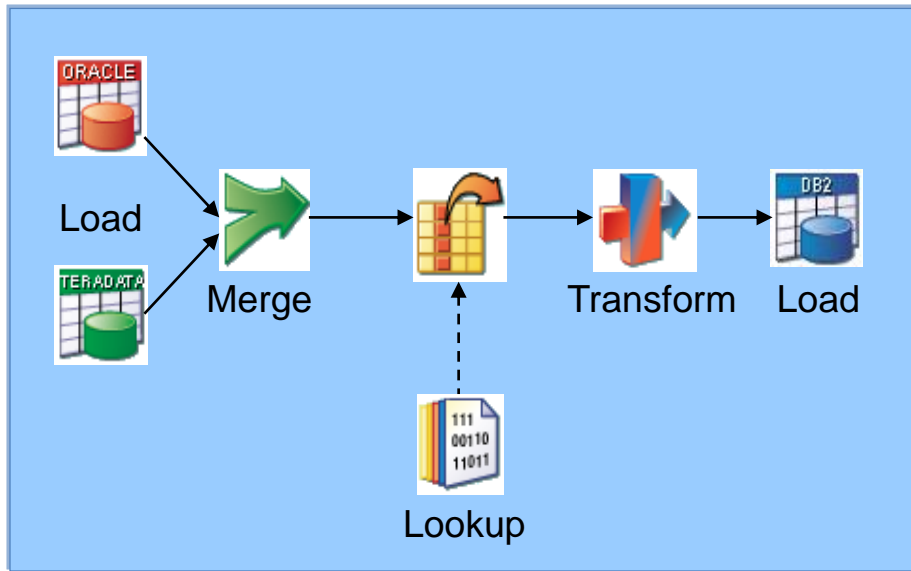
We'll use one of the key tools in the Information Server suite – DataStage.



IBM

DataStage Makes it Easy to Integrate All Data Sources

Drag, drop, configure

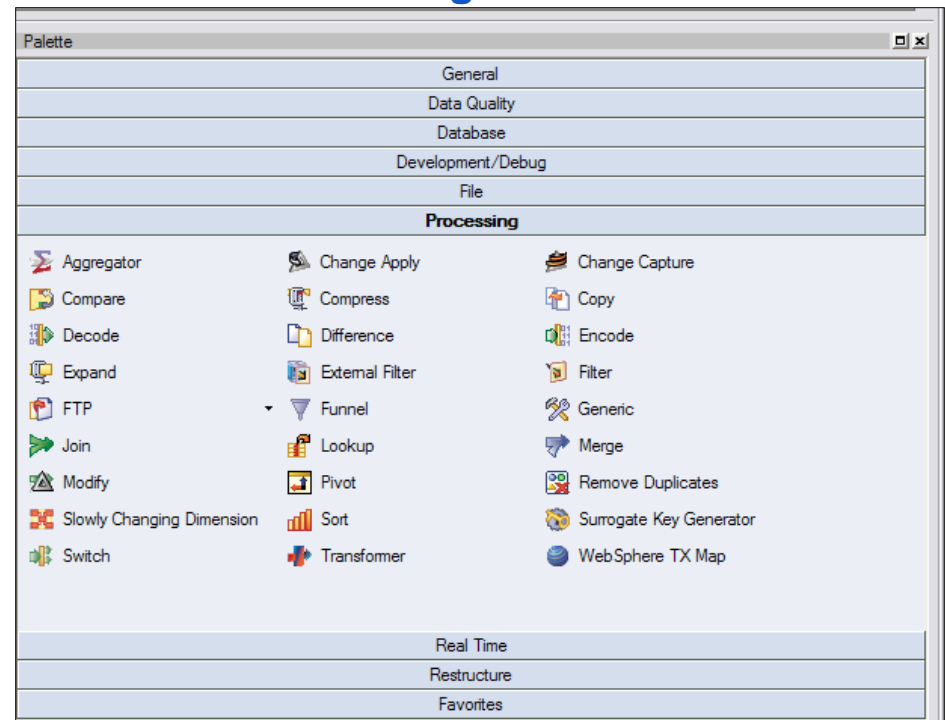


- **Productive**
 - ▶ Graphical design of data flows
 - Focus on flow rather than low level implementation
 - ▶ Numerous pre-built components
- **Scalable**
 - ▶ Leverages parallel processing
- **Reduces risk, reduces cost**
 - ▶ Modular approach maximizes reusability of components
 - ▶ Shared metadata improves collaboration
- **One tool integrates all data sources**

Numerous Pre-Built Transformation Components Speed Productivity

- 50+ pre-built transformation components for fast job generation and increased productivity
- Many built-in functions for easy to write advanced transformation logic
- Easy to use editors allow for quick configuration

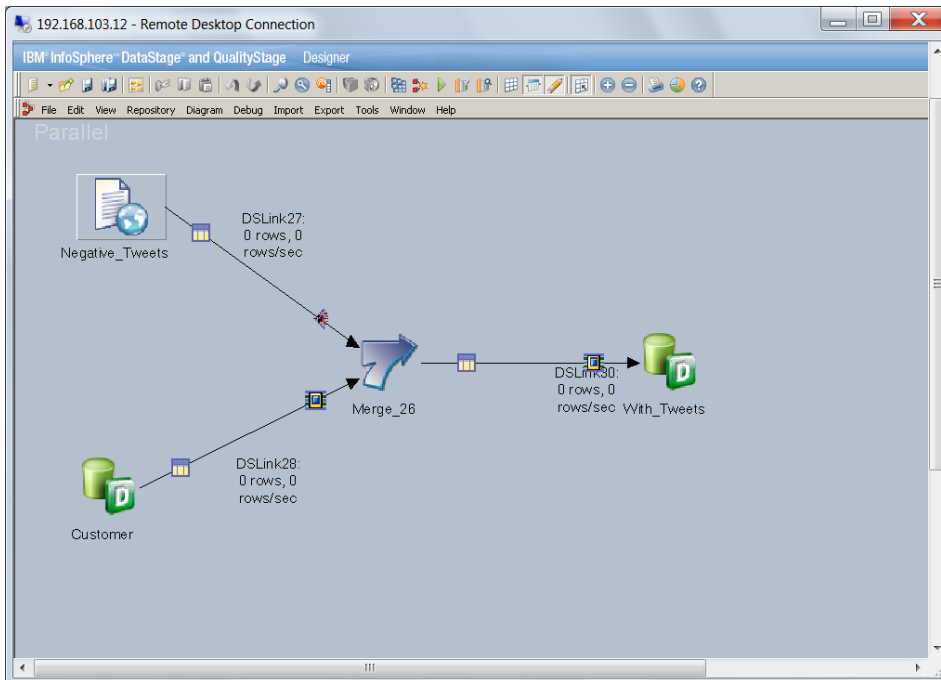
DataStage Palette



DEMO: DataStage Makes it Easy to Complete the 360 Degree View of the Customer

Goal: Combine Twitter Tweets With Relational Customer Data

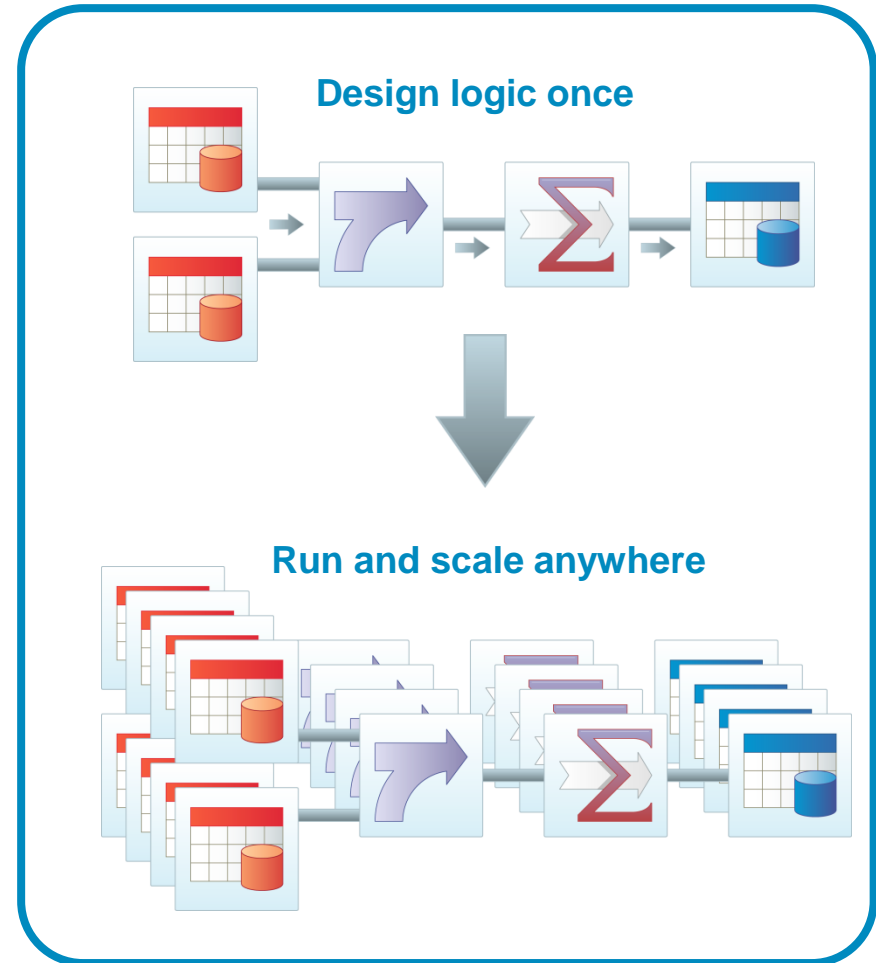
DataStage Editor



- One tool for all data integration needs
- Drag and drop simple
- Easy to see what processes use what data

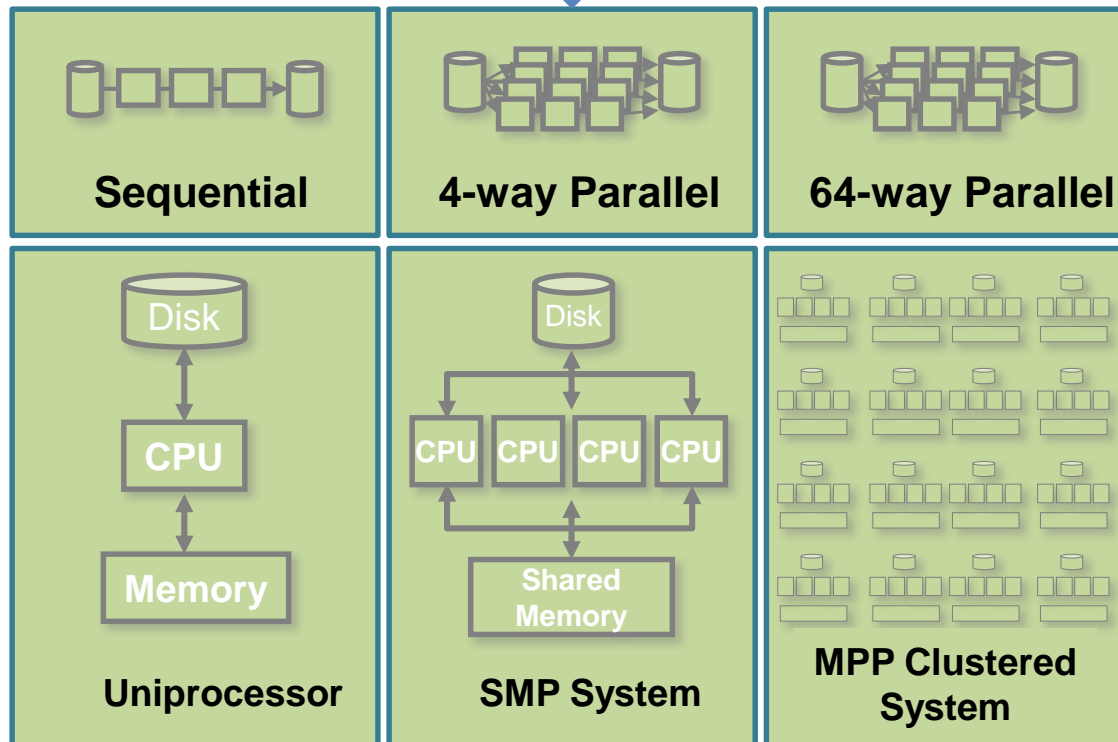
Design Once – Run and Scale Anywhere

- Job design is the same regardless of how it will be deployed/scaled
- Scaling is configured by simple text file
- Just like Hadoop, data and code are partitioned out to a cluster of nodes for parallel processing
- Instantly get better performance as hardware resources are added
- Extract, transform and load any volume of information



Unprecedented Scalability!

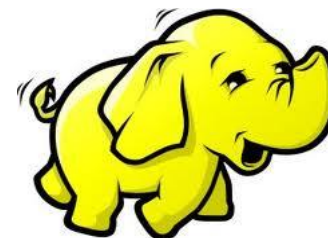
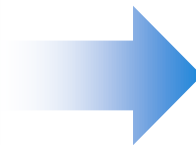
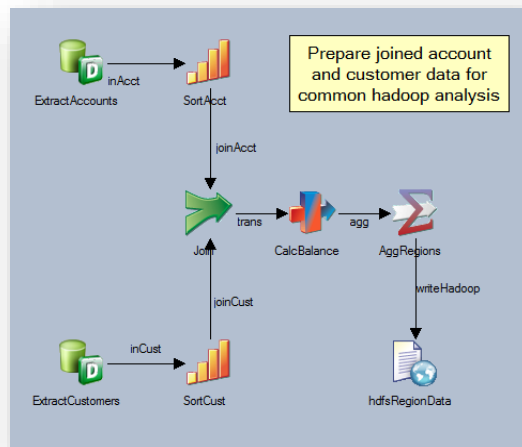
InfoSphere DataStage is Big Data Integration



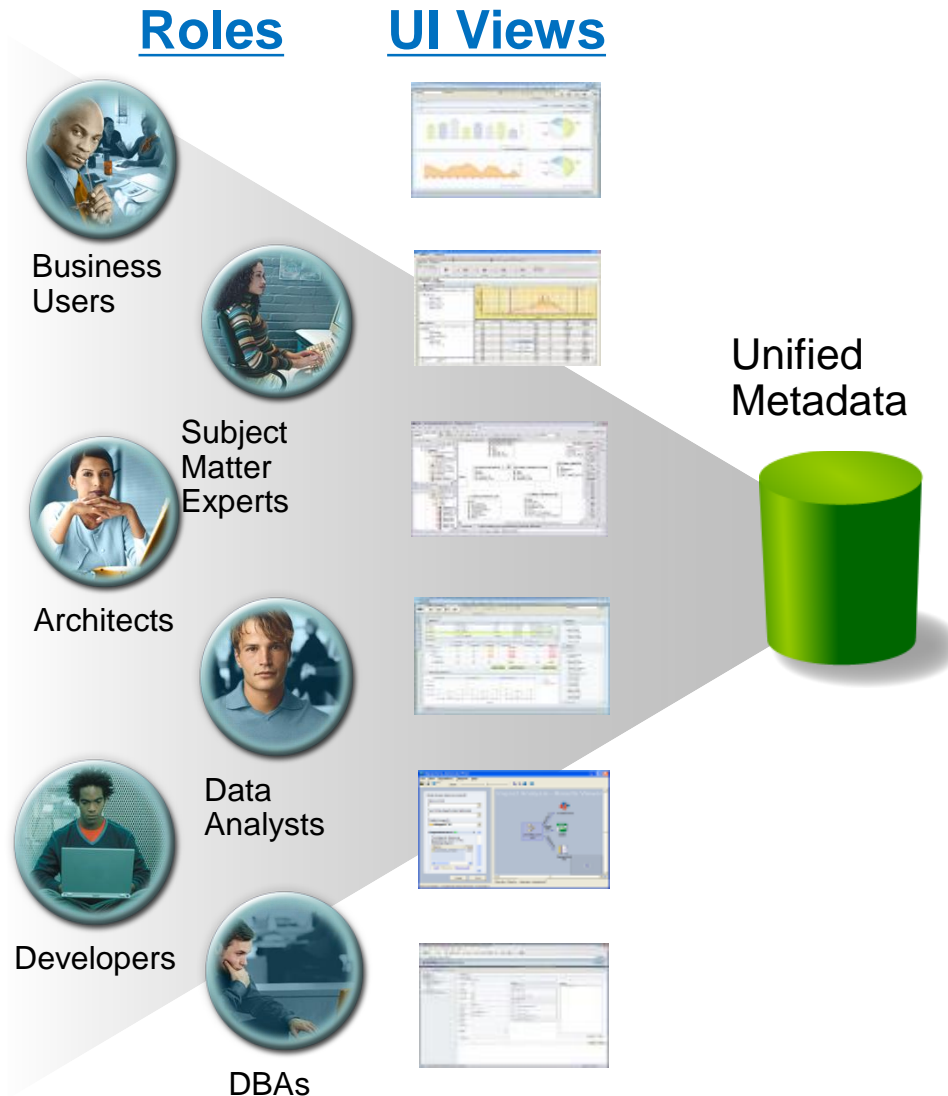
- **Dynamic**
Instantly get better performance as hardware resources are added to any topology
- **Extendable**
Add a new server to scale out through simple text file edit (or, in grid config, automatically via integration with grid management software).
- **Data Partitioned**
In true MPP fashion (like Hadoop) data persisted in the data integration platform is stored in parallel to scale out the I/O.
- **Hadoop Integrated**
Push all or parts of the process out to Hadoop to take advantage of it's scalability in ELT fashion.

InfoSphere DataStage Balanced Optimization

- Data processing operations can be pushed toward source or target data stores
- Optimizes job run-time by allowing the developer to control where the entire job or various parts of the job will execute
- Provides the same job design as traditional DataStage jobs so there is no recoding required



The Common Metadata Layer is the Key to Optimal Productivity and Collaboration



- Better and faster communication
- Powerful metadata-driven design tools
 - ▶ Advanced search
 - ▶ Impact analysis
 - What happens if I make a change?
 - ▶ Data lineage reports
 - Where does the data come from, where does it go?
- Increases compliance with standards
- Increases trust and confidence in information

Information Server Connects to All the Systems and Applications that You Use

Relational Databases

DB2 (on IBM System z®, System i®, System p® or System x®)

Oracle

IBM Informix® Dynamic Server and Informix Extended Parallel Server

Ingres

Netezza

Progress

RDB

RedBrick

SQL/DS

SQL Server

Sybase (ASE & IQ)

Teradata

Universe

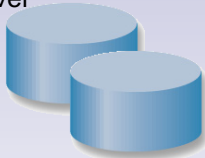
IBM UniData®

NonStop SQL

InfoSphere Federation Server

InfoSphere Classic Federation

And more.....



General Access

Sequential File

Complex Flat File

File Set

Data Set

Named Pipe

iWay

FTP

SFTP

Compressed / Encoded Data

External Command Call

Parallel/wrapped non-IBM apps

EMC InfoMover

Web logs

Email



Enterprise Applications

JDE/PeopleSoft OneWorld

Oracle Applications

PeopleSoft

SAS

SAP BW

SAP R/3

Siebel

Ariba

Manugistics

I2

Etc...



Standards & Real Time

WebSphere MQ

Java Message Service (JMS)

Java

XML and XSL-T

EBXML

Web Services (SOA)

Enterprise Java Beans (EJB)

EDI

FIX

SWIFT

HIPAA



Change Data Capture

DB2 (on System z, System i, System p or System x)

Oracle

SQL Server

Sybase

Informix

IMS

VSAM

ADABAS

IDMS

Datacom

Legacy

Allbase/SQL

C-ISAM

D-ISAM

Datacom/DB

DS Mumps

Enscribe

Essbase

FOCUS

IDMS/SQL

ImageSQL

Infoman

KSAM

M204

MS Analysis

Nomad

Nucleus

RMS S2000

Supra

TOTAL

TurboImage

Unify

And many more....

And It Connects to All the New Big Data Sources



Unlimited Data Scalability is Mandatory for Big Data Integration

- Unlimited Data Scalability means:
 - No limits on data volumes, processing throughput, numbers of processors and nodes
 - Simply add hardware to process larger data volumes
- Unlimited Data Scalability is a mandatory requirement for large enterprises and Big Data Integration
- Unlimited Data Scalability requires a specific software architecture
 - Anything else will not achieve Unlimited Data Scalability

Information Server is an Enterprise Class Integration Hub

- Most scalable integration runtime in the industry
- Connects to the broadest range of data sources
 - Traditional: Databases, applications, message queues
 - New: Hadoop, Streams, MongoDB, JSON, etc.
- Speeds productivity
- Reduces development costs
- Reduces development risks

Overall Leader in Gartner Magic Quadrant for Integration Tools

Gartner Report: <http://www.forrester.com/Forrester+Wave+Data+Governance+Management+Tooling+Q2+2014/-/E-WEB17623>