

IBM Analytics

IBM Information Server

What is new -- what is next?

May 8th, 2018

Beate Porst – porst@us.ibm.com

Program Director Offering Management

IBM Unified Governance & Integration

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.

IBM INFORMATION INTEGRATION

Product Execution Update

IBM Analytics Portfolio

Collect

Hybrid Data
Management



**Write once,
access anywhere**
with a common access
layer to promote
application independence

Organize

Unified Governance &
Integration



**Prepare, publish,
integrate and protect**
your data to drive
insights while mitigating
compliance risks

Analyze

Data Science &
Visualization

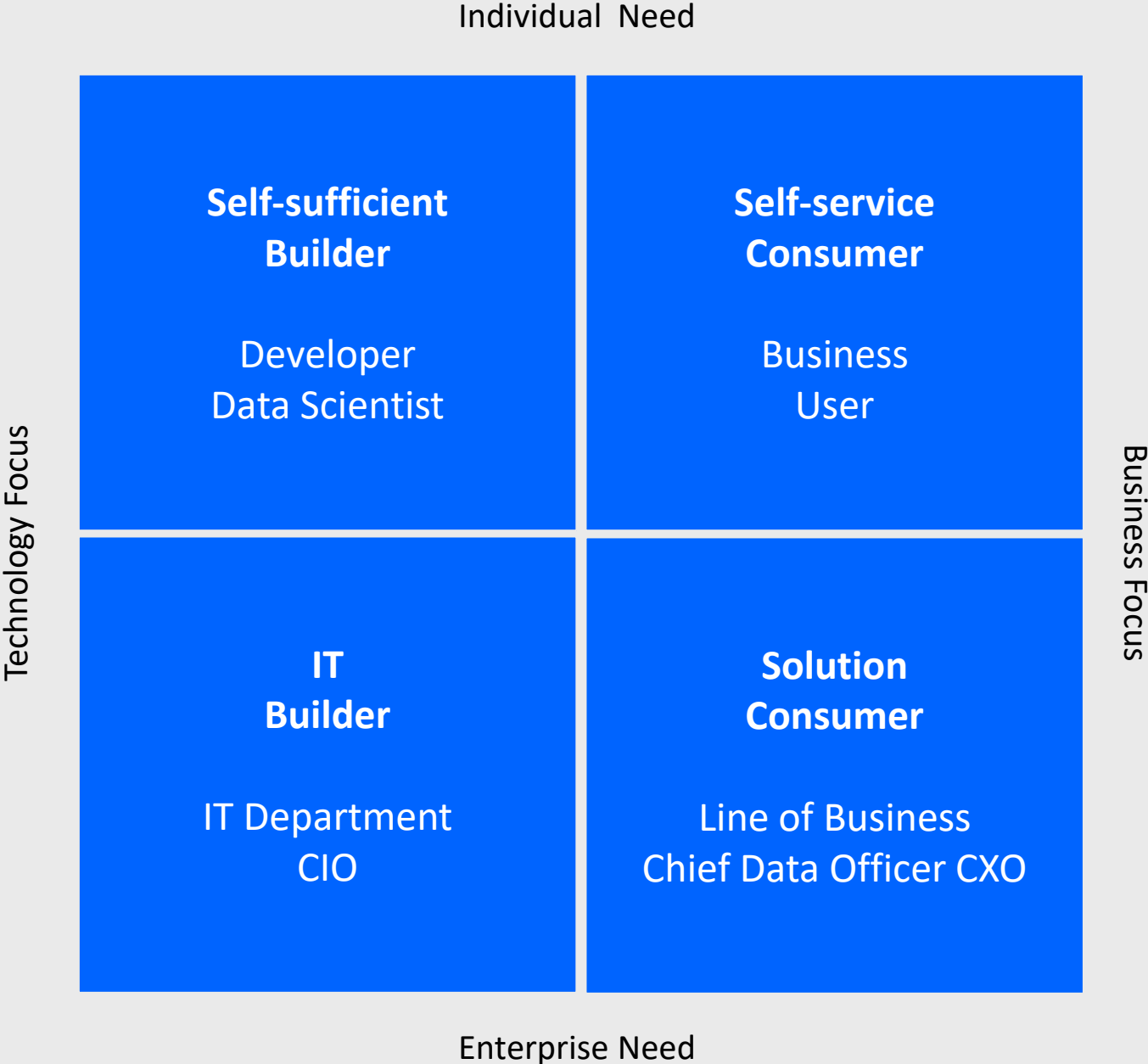


**Descriptive, predictive,
prescriptive** to
understand the current,
predict the future and
change the outcome

Automation through Machine Learning

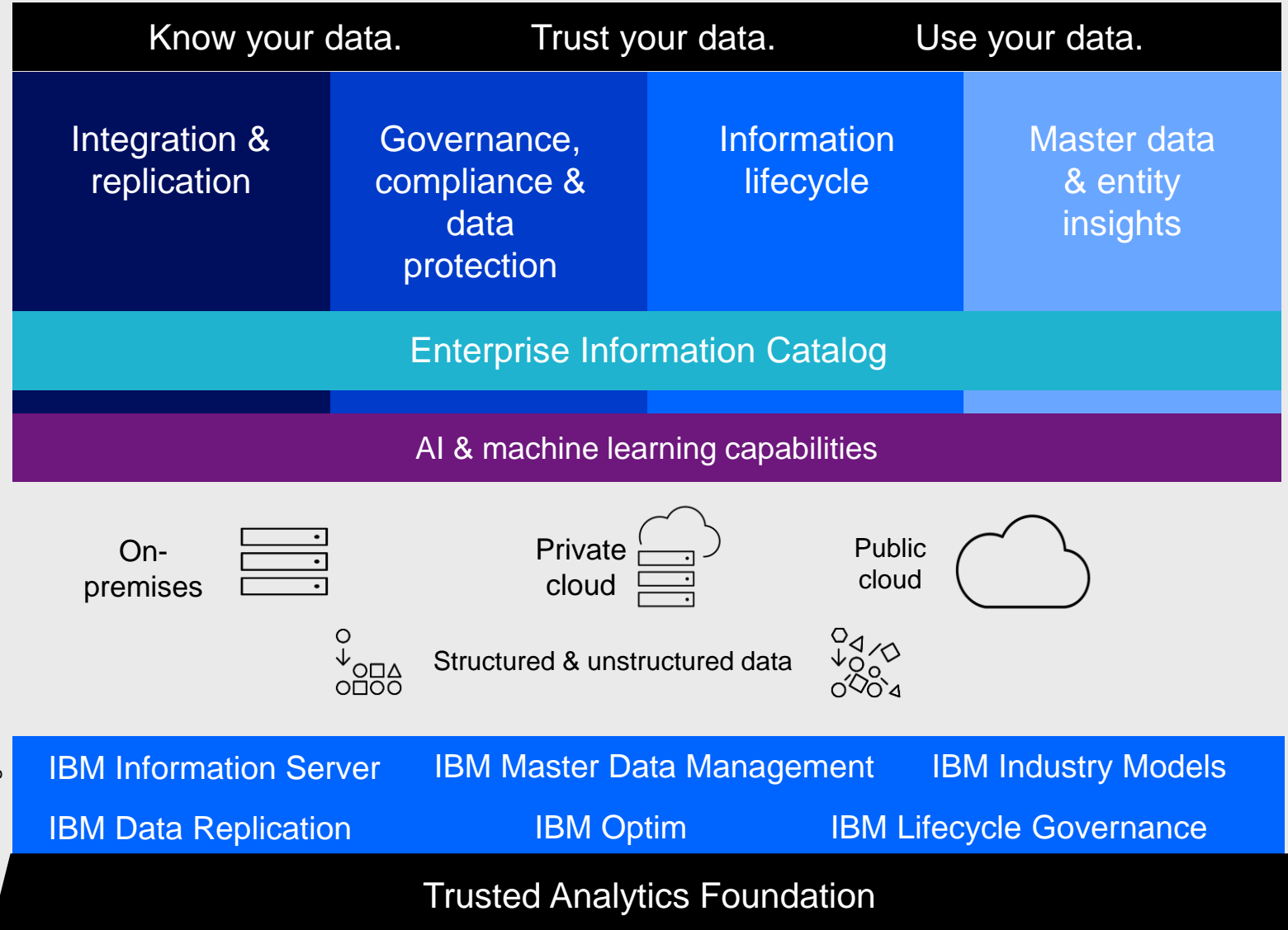
Hybrid Cloud Foundation

There is a growing need to provide trusted and business ready data to consumers across the enterprise

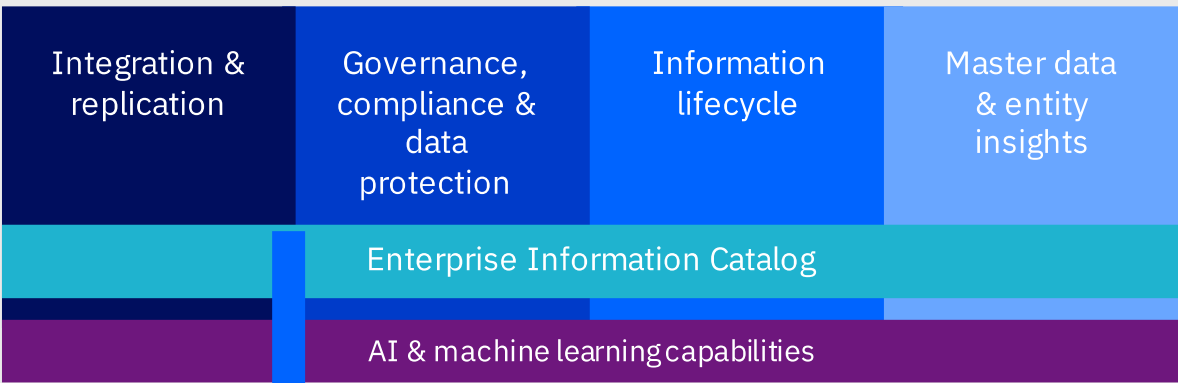
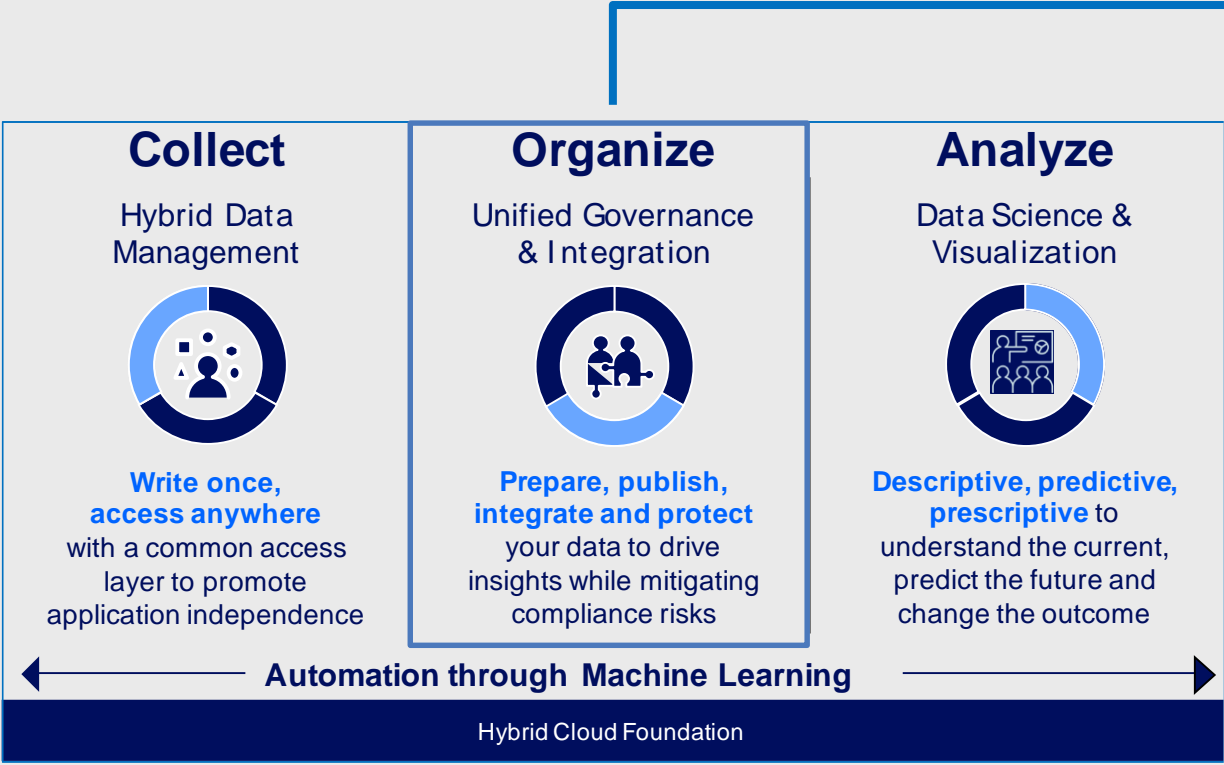


IBM Unified Governance & Integration Platform

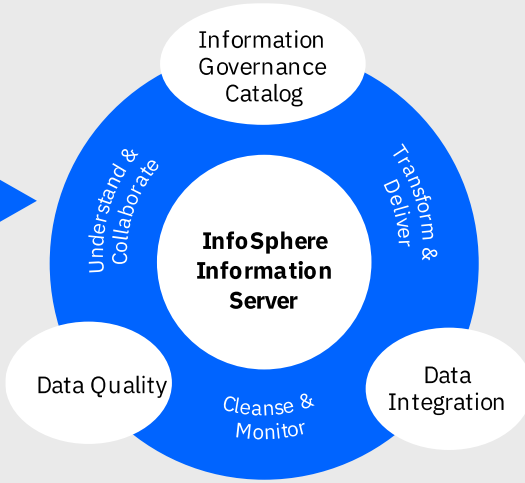
- Open and extensible platform
- Advanced machine learning
- Brings structured and unstructured together
- Scalability and parallel processing
- Smarter metadata drives embedded governance
- Pre-built industry data models
- Unified platform with adaptive deployment and licensing



Our Portfolio: Power behind and across the portfolio



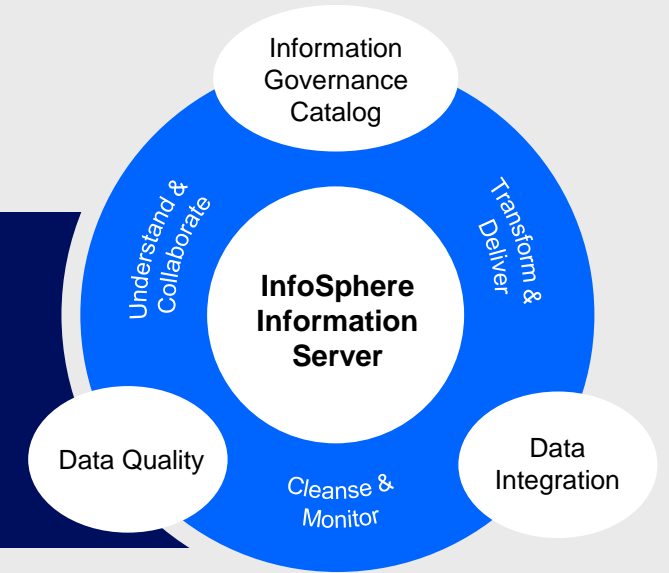
Enable better insight and compliance across all data through Unified Governance & Integration



IBM InfoSphere Information Server

Information Empowerment for Your Data Ecosystem

Integrating and transforming data and content to deliver accurate, consistent, timely and complete information through a unified platform with a common metadata foundation



Information Governance Catalog

Understand & Collaborate

- Catalog technical metadata & align w/ business language
- Manage (big) data lineage
- BCBS compliance reporting

Data Quality

Cleanse & Monitor

- Analyze, validate, classify
- Cleanse & standardize
- Define, manage & monitor data rules + exceptions

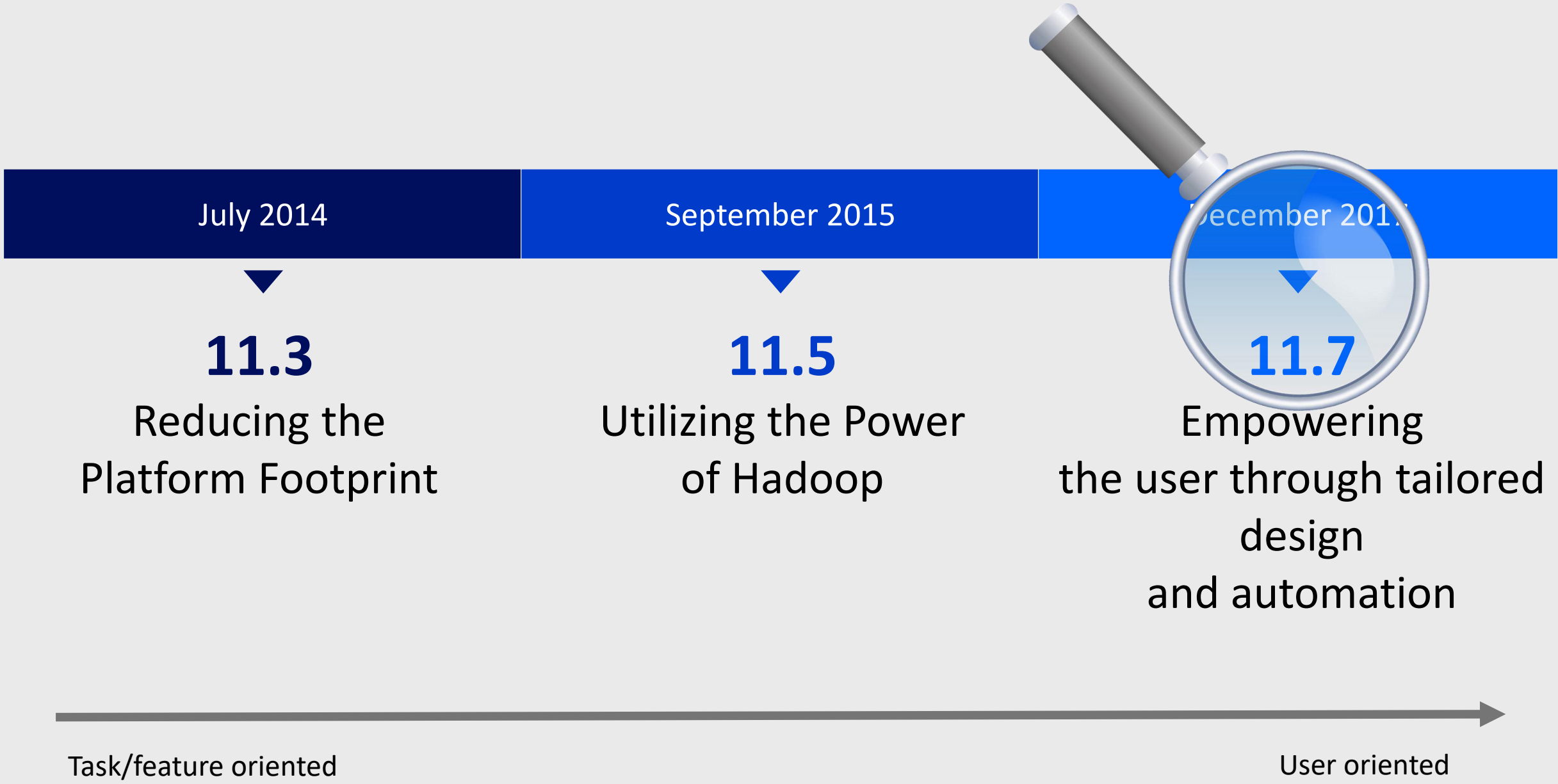
Data Integration

Transform & Deliver

- Massive scalability
- Power for any complexity
- Deliver in batch and/or real-time with change capture

Common Connectivity / Shared Metadata / Security /
Common Execution Engine With Flexible Deployments (Hadoop, Grid, Cloud)

IBM Information Server



Task/feature oriented

User oriented

IBM Information Server V11.7

... moving towards a user centric micro-service based architecture

Automation & M/L

Increased automation for the Governance & Data Quality process

Strengthen the Data Lake

Increasing speed and resilience on Hadoop

Empower the User

New Self-service / User centered experiences for Integration and Governance

Hybrid (Cloud) Deployment

More deployment options for Information Server components

Simplified Licensing

Single Flexible License for Unified Governance & Integration

Enable GDPR

Combining structured & unstructured data/content governance in ONE catalog

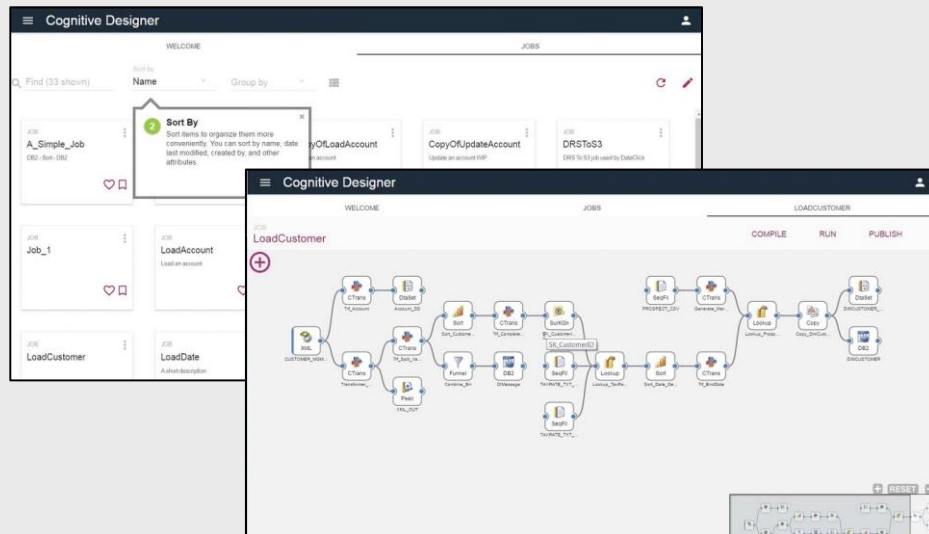
Expanding the Reach

More out of the box connectivity for Cloud, Hadoop & Enterprise

Empowerment of the User

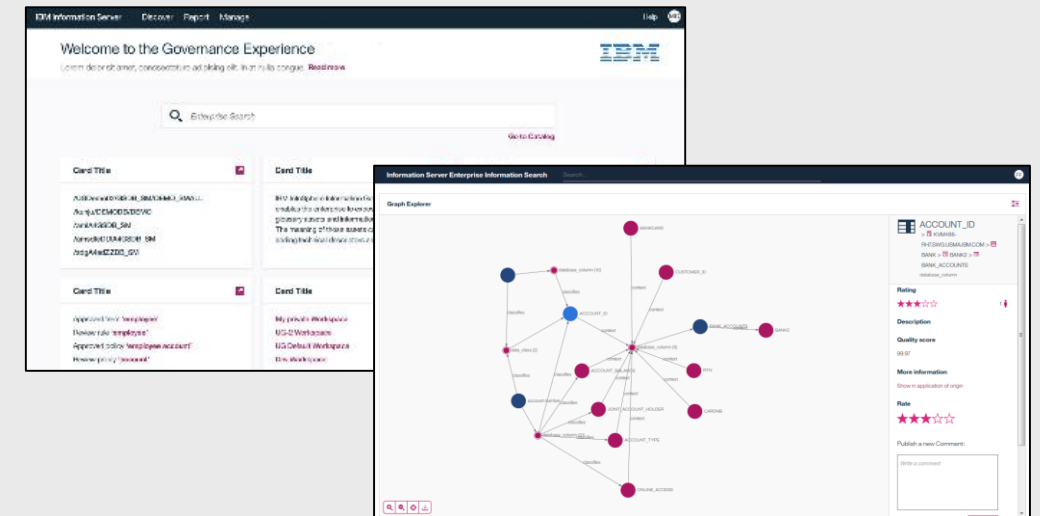
New Integration Experience for Data Engineers

- Intuitive, no-install experience
- Accelerating productivity through:
 - Automatic schema propagation
 - Highlighted design errors
 - Powerful type-ahead search
 - Full backwards compatibility



New User Experience for Business User

- Introducing simplicity and elegance when searching or exploring information
- Users may personalize/tailor the experience
- Easily visualize relationships to uncover insights



Empower the User

DataStage Flow Designer– The New Integration Experience

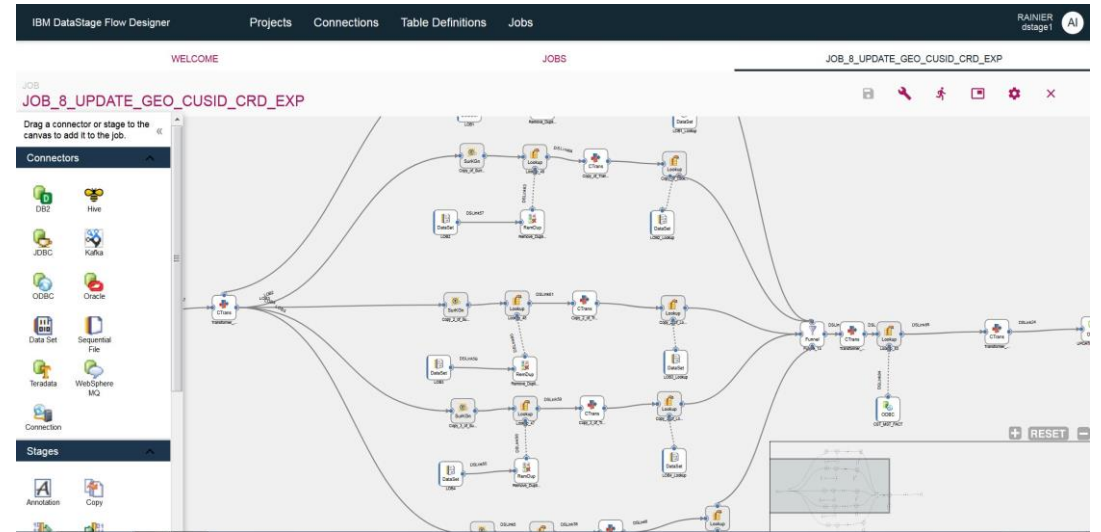
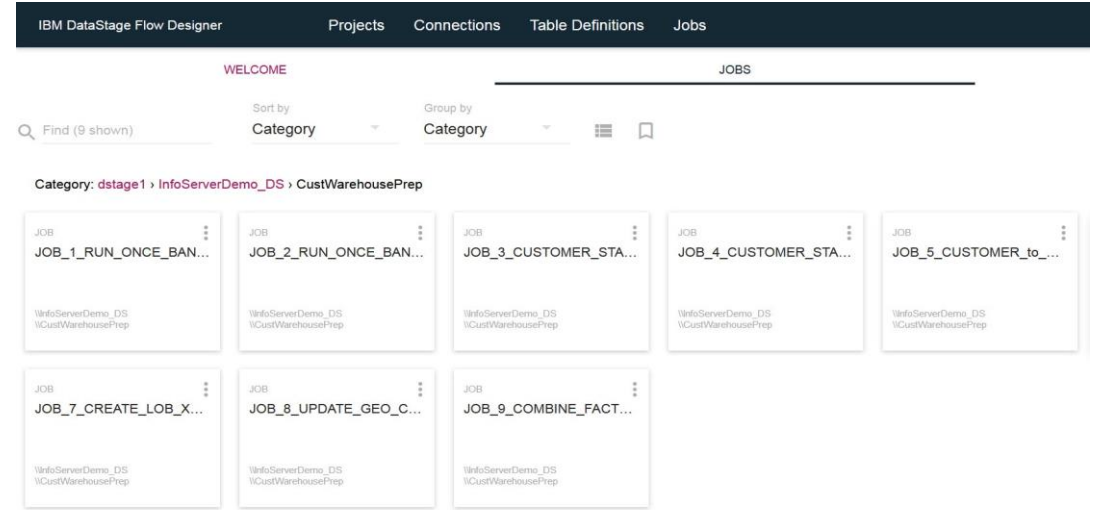
Intuitive, browser-based (no-install) experience

– Reducing total cost of ownership

Full backwards compatibility

Accelerated productivity through:

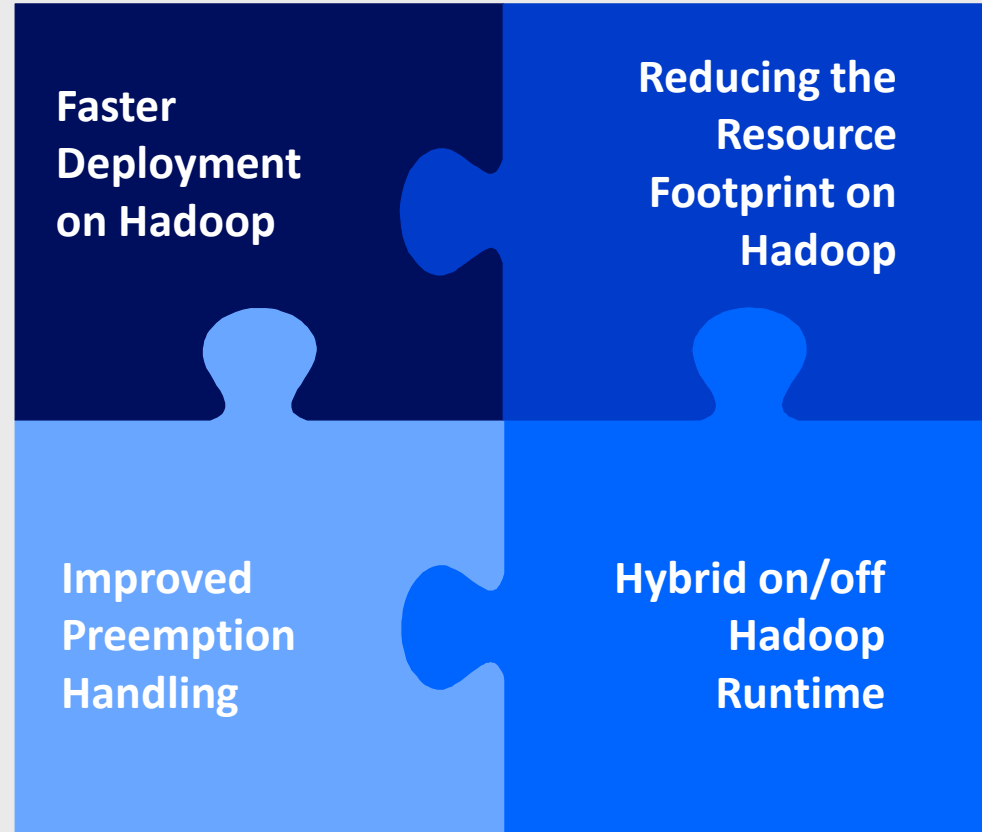
- Automatic schema propagation
- Highlighted design errors
- Powerful type-ahead search
- Server-side compilation



Strengthen the Data Lake

- Achieved through deeper integration into Ambari
- **10x accelerated deployment time**
- Automatically captures all parameters for node deployment

- **Accurate job/error handling** during container preemption
- Remembering preemption notification during container allocation
- Sending notification with diagnostic to conductor in preemption case



- Utilizing Hadoop Shuffle space as Information Server Scratch space
- Simple user choice through APT Configuration file setting

- Use a single Instance of Information Server to run Hadoop and non Hadoop workload
- **Optimized resource utilization** for dedicated workloads against non Hadoop sources/targets
- Simple APT Configuration option



Broader, Faster, Saver— Increasing Out of the Box Connectivity

Hadoop

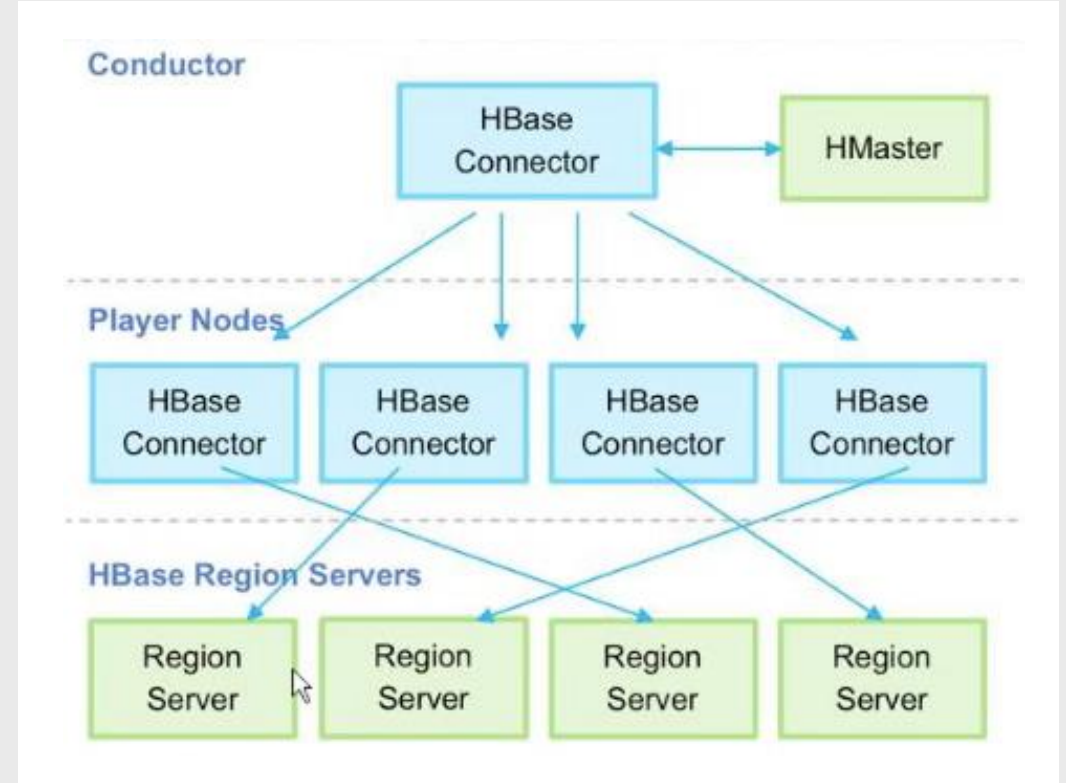
- New HBase connector
- Hadoop File Connector performance & security enhancements
- Kafka Connector security improvements
- Hive Connector
- MongoDB support

Cloud

- Amazon S3 enhancements
- Amazon EMR/Hive
- Amazon Redshift

Enterprise

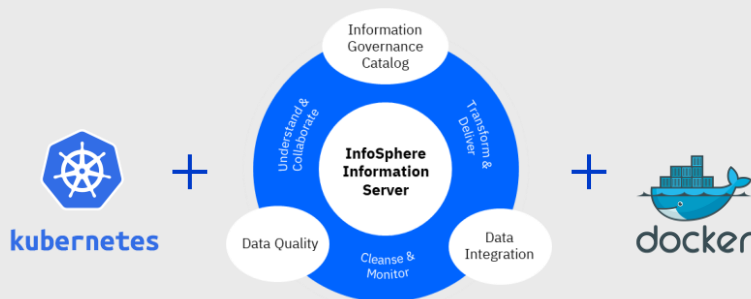
- Oracle PDB and CDB
- Siebel 8.2.2.4 certification
- Sybase datatype enhancement
- Security enhancement for metadata import
- New SAP BW feature pack
- Data Masking ODPP v11.3 support
- Distributed Transactions: MQ Client mode
- MQ Connector version update
- ILOG Connector Decision Engine



Hybrid Deployment and Execution

Container-based Deployment

- Easy 1-Click style deployment for Information Server “microservice” components
 - Starting with Enterprise Search



IBM Cloud Private

- IBM Cloud Private (ICP) delivers the speed of public with the control of private
- Information Server v11.7 is now available free of charge on ICP Community Edition

Hybrid Execution

- Industry unique hybrid execution feature to balance integration workloads across a Hadoop and non-Hadoop environment
- Aimed at minimizing data movements and optimizing integration resources



Capturing the **context** to aid in self-service data search & discovery

Source Systems

slack IBM

Social Networking Tools

IBM Watson Analytics X tableau

Business Analyst Tools

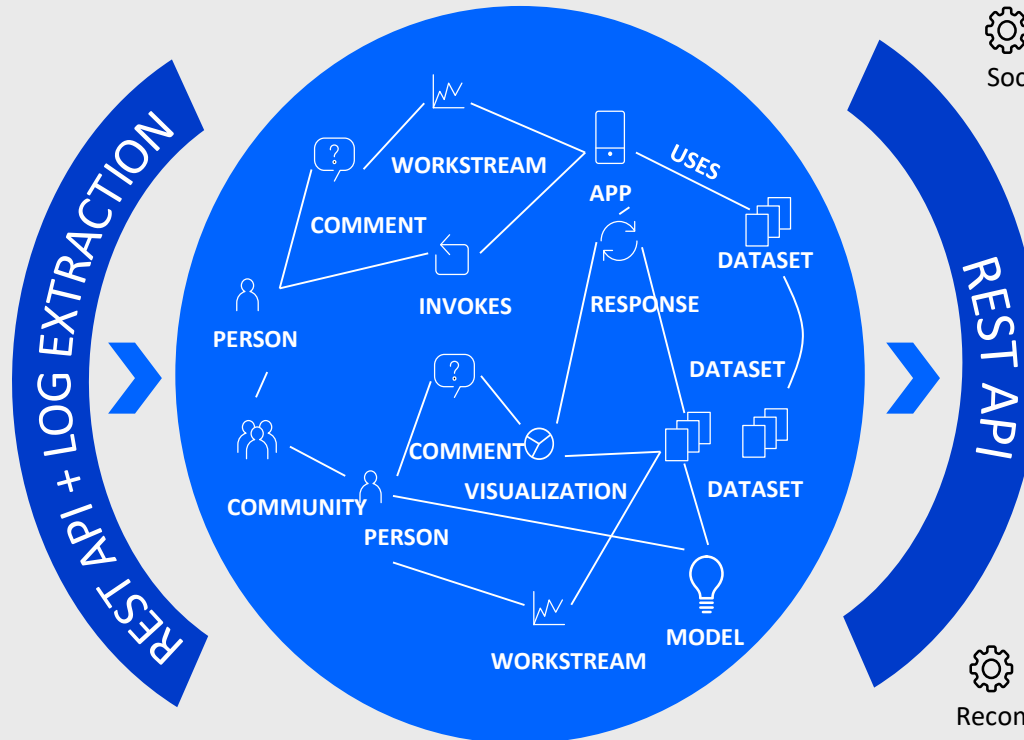
R Zeppelin

Data Science Tools

IBM InfoSphere Spark

Data Integration Tools

Contextual Usage Graph



Embeddable Widgets

Social Widgets

Contextual Search

Activity Streams

Contextual Graph Browser

Recommendations

User Interfaces

slack

Business Users

IBM Watson Analytics

Business Analysts

Zeppelin

Data Scientists

IBM InfoSphere DellWorks

IT Staff

Accelerating the Governance Process

Automating the Data Quality Process

- Automatically discover ALL data at once
- Actionable Automation Rules triggering any custom DQ action based on business events

The screenshot shows a 'Create new actionable rule' interface. It has two main sections: 'Rule name' and 'Rule description'. The rule name is 'ar_DOB_or_DOD' and the description is 'Date of Birth or Data of Death field must exist and be'. Below these is the 'Rule logic' section, which is a drag-and-drop interface. On the left, there are two categories: 'Conditions' and 'Actions'. The logic is built as follows: an 'if' block containing two conditions: 'Asset is assigned to Date of Birth' and 'Asset is assigned to Date of Death', connected by an 'OR' operator. This is followed by a 'then' block containing three actions: 'Set Data Quality Score Threshold to 98 %', 'Automatically bind Data Rule Definition FieldExistsAndDate_1', and 'Use All Data Quality Dimensions'.

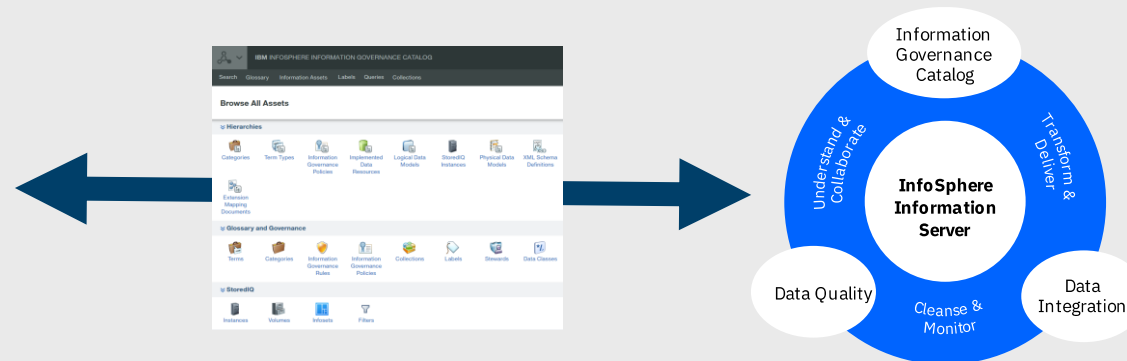
Automating the Governance Process

- Utilizing Machine Learning for an accelerated Metadata Classification Process (Auto Term assignment)
- Automatically classify data -- including understand PII risk

The screenshot shows the 'AA_Auto Discovery - BANK_CUSTOMERS' interface. It displays metadata for the 'BANK_CUSTOMERS' data set, including 10 columns and 2,000 rows. The data quality score is 99%, and the threshold is 80%. The 'CUSTOMER_ID' column is highlighted, showing a data quality score of 100%. The interface also shows a 'Column summary' section with 'NO NOTES WERE FOUND'. Below this, there is a '1 TERM(S)' section with 'Assigned terms' and 'Candidate terms'. The 'Assigned terms' section shows 'Customer Identifier' circled in green. The 'Candidate terms' section shows 'National Identifier 50%', 'Tax Identifier 50%', 'Customer 50%', and 'Customer Identifier 100%'. The 'Data quality' section at the bottom indicates the last data quality analysis was on 2018-01-25 01:21.

Support GDPR for ANY Data *Structured or Unstructured*

- Data class /taxonomy synchronization for common use across structured and unstructured.
- Easily detect & understand policy/regulation violations
- Data expert in StoredIQ gets additional insight into centrally managed data classes and associated attributes to apply the required actions



The Power of 3

Simple and Flexible consumption-based Licensing

Collect

Hybrid Data
Management

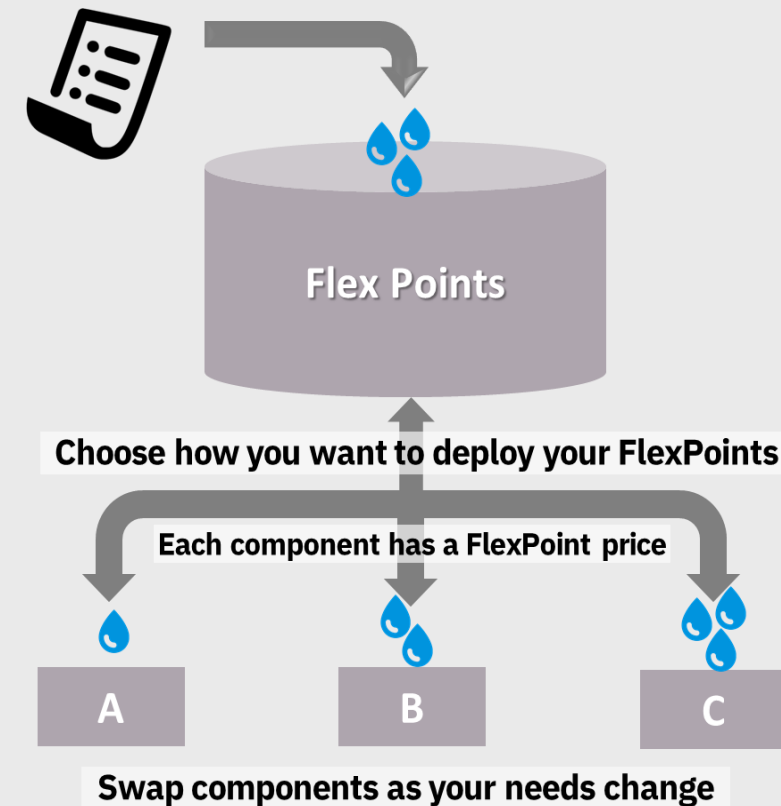
Organize

Unified Governance &
Integration

Analyze

Data Science &
Visualization

- Simple *FlexPoint* licensing for each of IBM's three analytics platforms
- Easily scale usage across teams and capabilities by simply buying additional FlexPoints
- Consistent planning & procurement across on premise and hybrid Cloud deployments



IBM INFORMATION INTEGRATION

Vision & Strategy Update

Most Complete Vision for the Marketplace

Development driven by Key Priorities

Simplify

Make products accessible
and easily consumable



Automate

Infuse data science and
machine learning into
everything we do



Hybrid Cloud

Flexible cloud deployment
and optimized workload



Compose

Enable the platform as
loosely coupled service for
fast & easy deployment

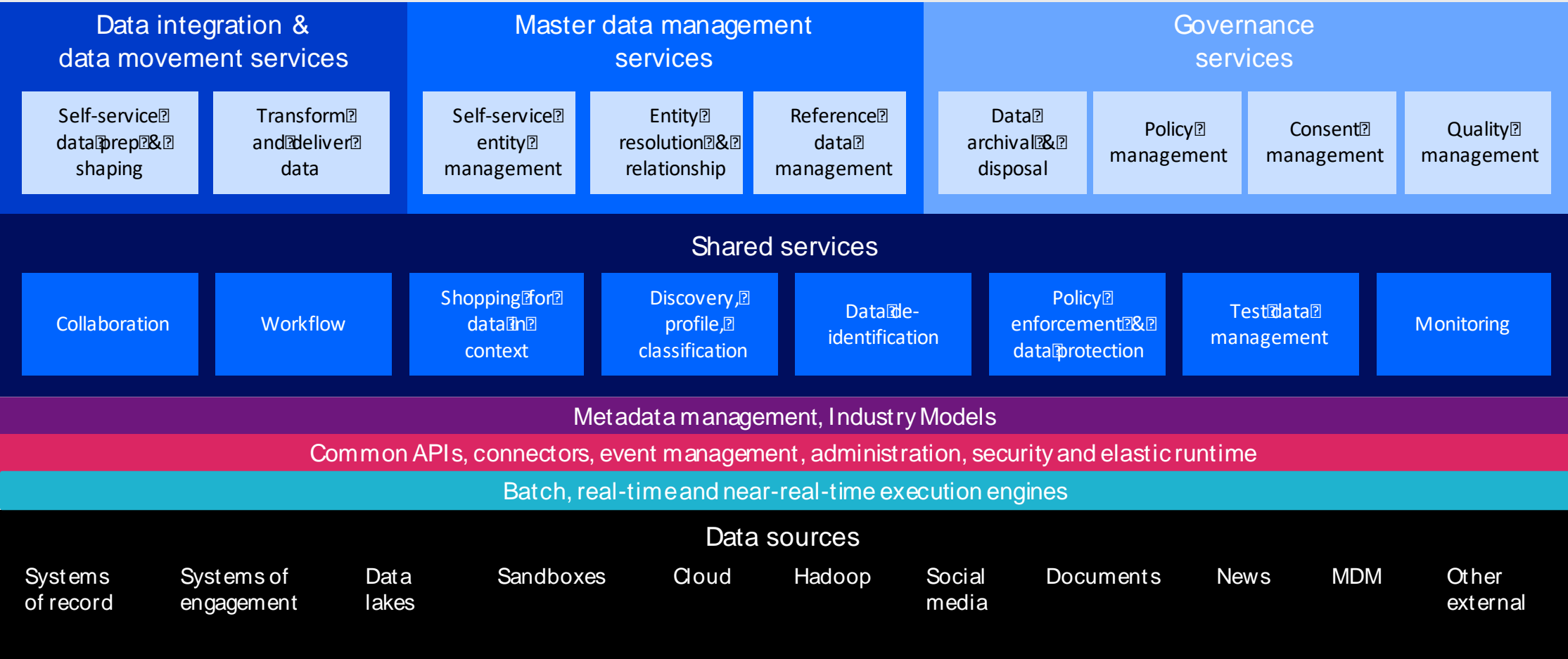


Unified Governance and Integration Platform

A service-based architecture underpinned by common Metadata & Governance foundation

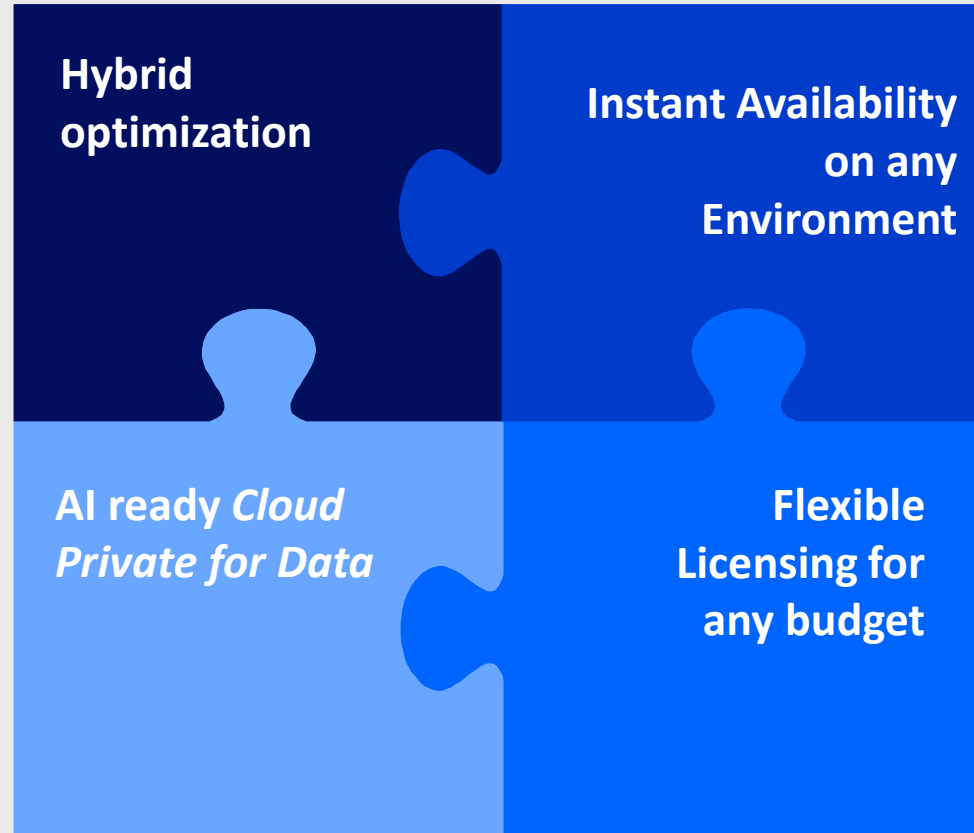


Data engineers Business users Data scientists Data stewards/curators Data quality analyst CDO



Hybrid Cloud Focus

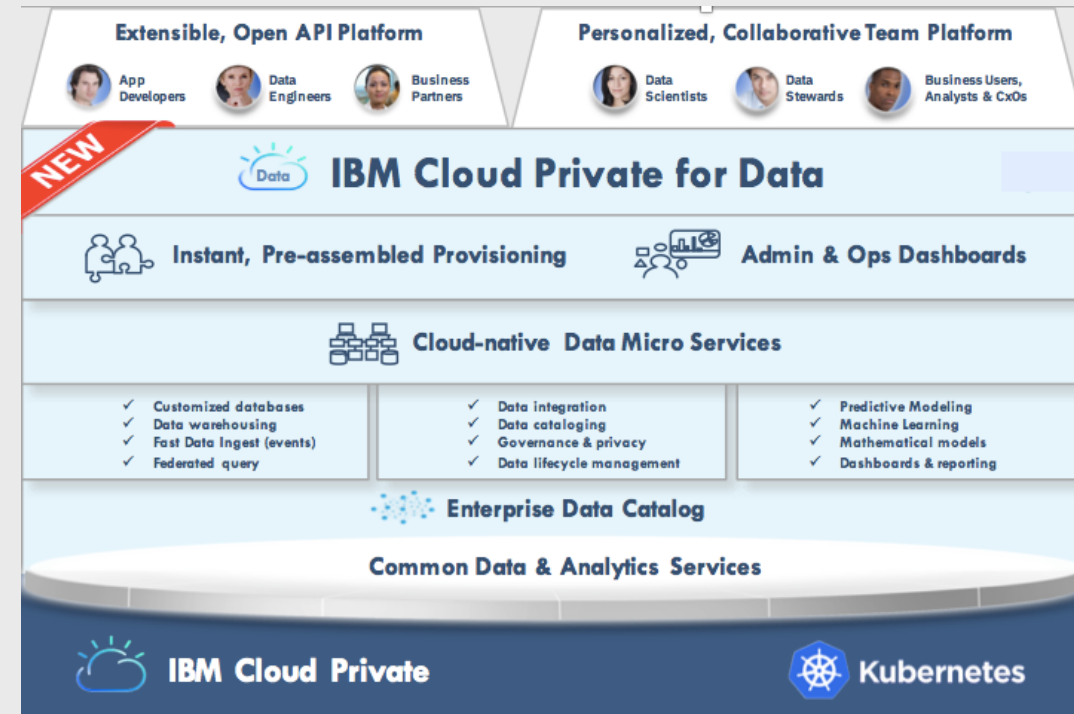
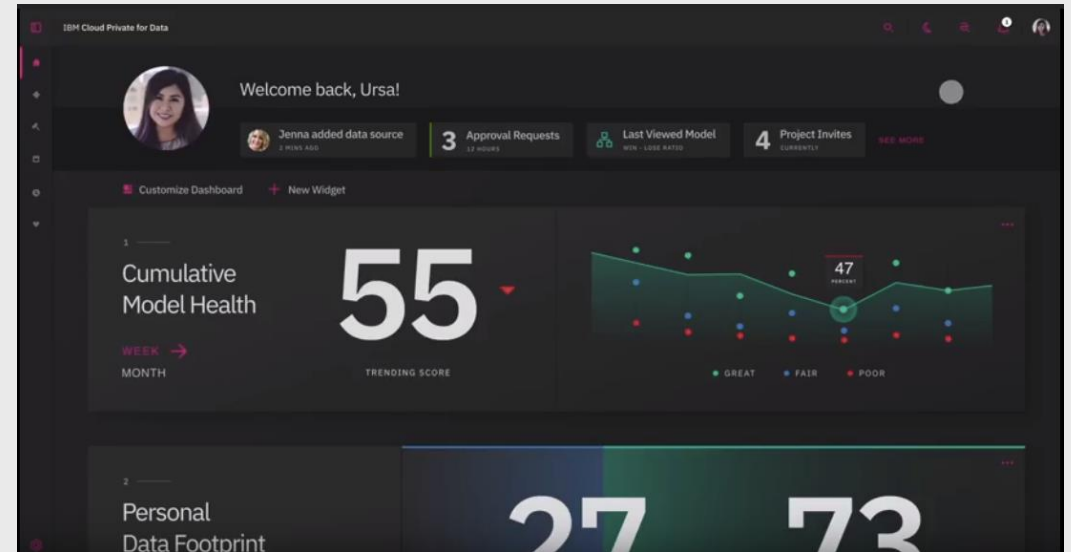
- Supporting ML based dynamic execution optimization across a hybrid landscape
- Ready to use person-based solution for private and public cloud
- Making data simple and accessible



- Micro-service based architecture supports instant availability on any IT environment (Public, Private or on-premise)
- Switch between classical and new license models including metered / usage based licensing

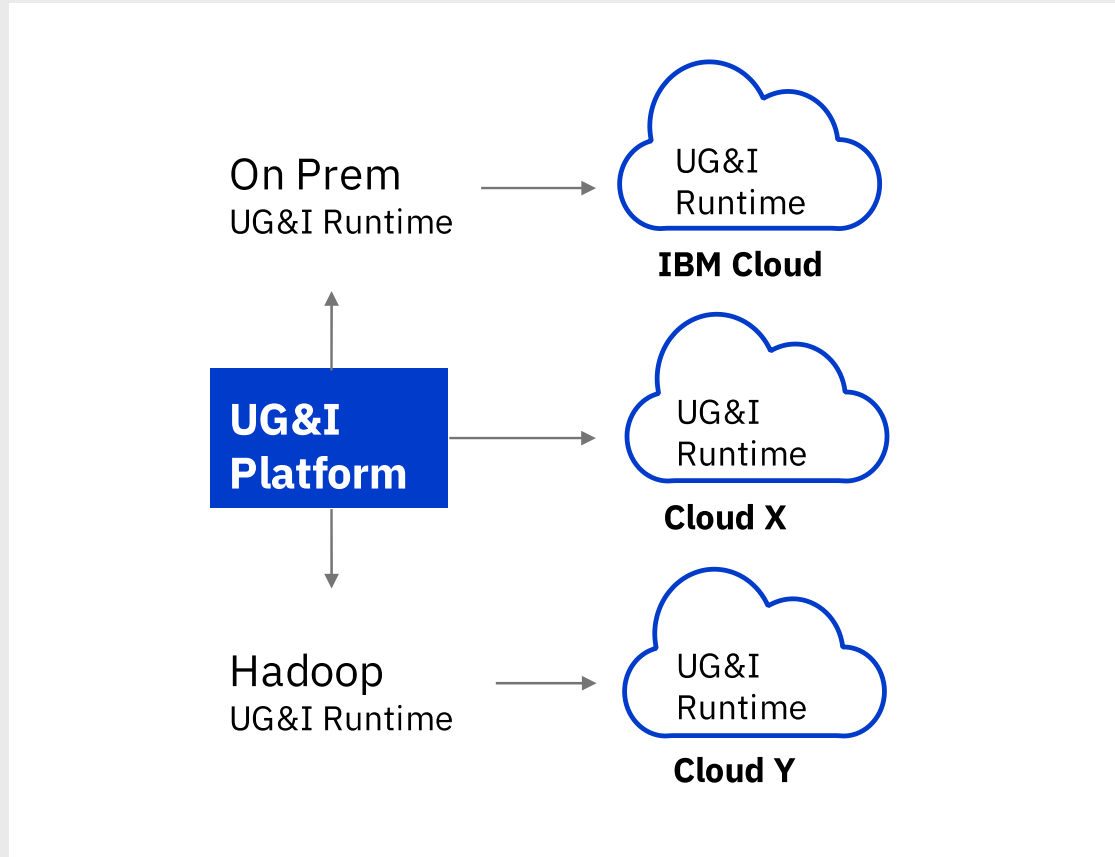
IBM Cloud Private for Data

- Ready to use solution that accelerates clients journey to AI with cloud agility and lightning speed runtime
- Collects every type of data, no matter where it lives, achieving freedom from ever changing data sources
- Organizes data into a trusted source of truth, delivering the agility to put data to work in new ways
- Empowers teams to analyze data in smarter ways to gain previously unobtainable insights
- Seamlessly integrates with Watson Studio



Hybrid Cloud Optimization

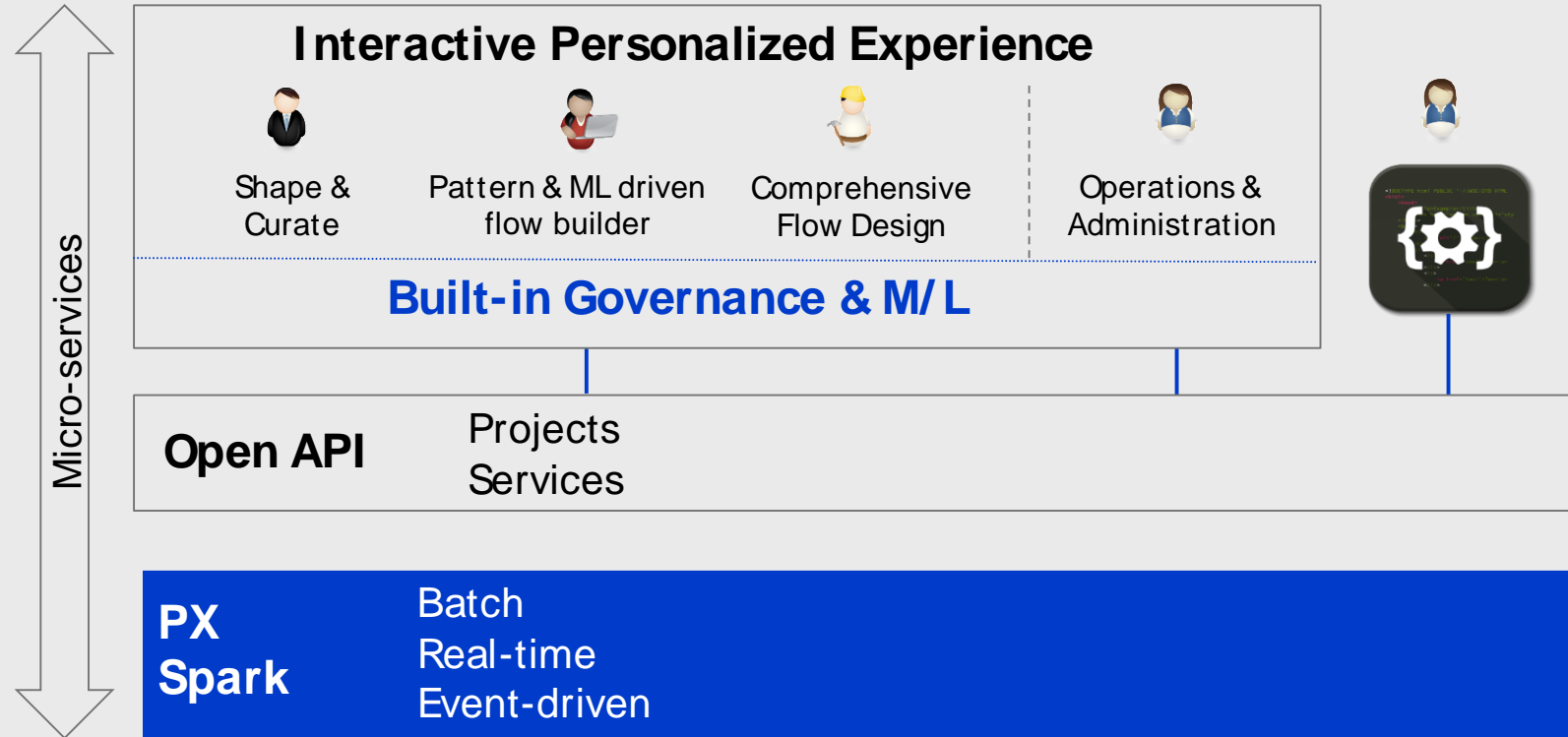
Customers are operating across environments in multiple clouds,



- Anywhere ad-hoc service provisioning anywhere
- Runtime/Deployment elasticity
 - Dynamically expand/shrink capacity based on workload requirements and data location
- Seamless interoperability between IBMs private & public cloud integration services
- Flexible licensing (metered or fixed)

Adaptable Integration Experiences

User experience adapting to users needs across the enterprise
--> NOT the user adapting to the experience



Any user leverages the same enterprise-ready foundation



Designing a modern Governance & Integration Platform ready for AI

Modernized architecture:

- Micro-service based architecture Enabling agile 1-Click container-based deployment
- Open API concept for easy integration into larger application stack

Advancing automation for every platform service utilizing M/L

- Auto-generating ingestion & integration
- Intelligent optimization & selection of integration style
- Advancing automation for data curation

Embracing Open Source as the Fabrics foundation

- IBM a top open source contributor (e.g. Atlas, Spark)

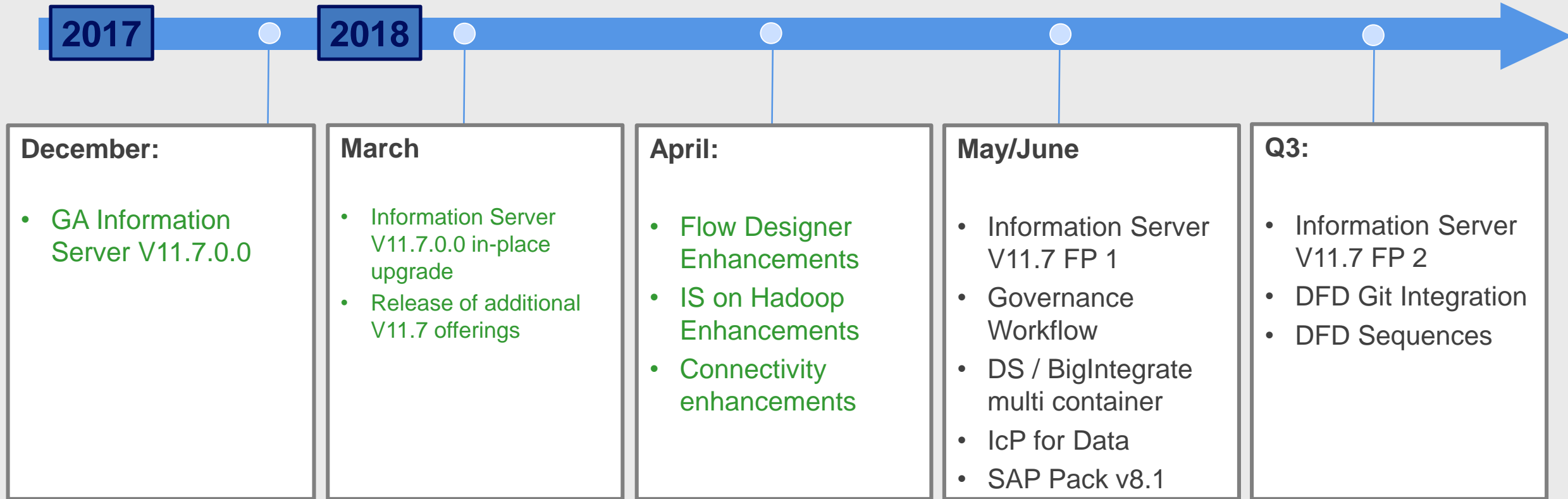
Applying Design Thinking principles

- Seamlessly integrated & user focused experiences

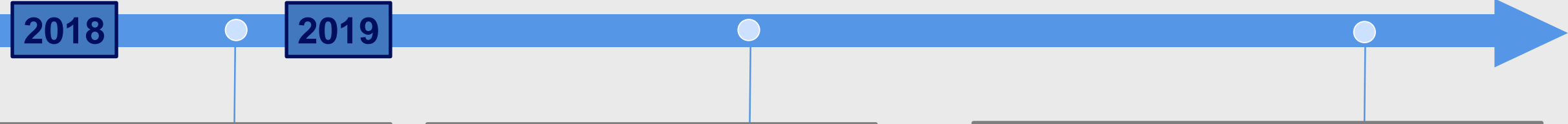
Embracing other innovations & technologies:

- MDM Blockchain

Information Server Release Plan 2018



Information Server Release Plan 2H/2018 - 2019



Q4:

- Information Server V11.7 FP 3
- MVP Business User driven data preparation & curation in DFD
- DS Administration in DFD

1H:

- Completion of DataStage Flow Designer migration
 - Including DQ

2H:

- Hybrid runtime optimization
- M/L based flow generation
- Automatic runtime selection

Enhancement released in the first 4 month:

DataStage Flow Designer:

- Delivered 2 Feature patches for DFD adding:
- *Smart Palette*: Uses M/L to automatically arrange stages in the palette based on usage. Ability to search for stages in the palette.
- *Mapper*: Ability to map input columns to output columns on links.
- *Additional Stages* in the DFD: Amazon S3, Lookup, Peek and Head.
- *Automatic Column Propagation*: Changes to column metadata, such as, rename, delete, or, change datatype are automatically propagated to down streams
- *Load Columns*: Ability to load columns from table definitions as part of a stage. Append or replace existing columns and have them automatically propagated
- *Rename Assets*: Support rename for connections, table definitions, jobs, links and stages.
- *Parameters*: Ability to create, edit and delete Job parameters, such as, encrypted, date, integer, float, pathname, date and time and configuration files.
- *View Data*: Preview a sample of data from relational connectors using a live connection.
- *Connections*: Create, edit and delete connections
- *Read-only Job*: A locked job can now be opened in a read-only mode

Connectivity:

- New Snowflake connector
- New Azure Cloud connector
- New Cassandra connector

Hadoop:

- Enable BDFS with Kerberos in a non-Hadoop environment
- Handle YARN Container Expiry when there aren't enough containers available
- Reduce Binary footprint from 1.7GB to 1.2GB
- Simplify Kerberos deployment by better integrating it in Ambari Console
- Support combination of RHEL 6 and 7 environments in a cluster as clients upgrade from 6 to 7
- Support CFF stage

IBM

Other noteworthy Enhancements in V11.7

IA Frequency Distribution

Now generates only one table per data set instead of per column

New Governance Monitor

New Governance Dashboard to quickly understand what is going on across projects

Open Discovery Framework

Lets you control the discovery process programmatically

Overlap Relationships

Can now be removed and be exported

Creating custom DQ Dimension

can now use metadata, values or row level data

New Data Class Types

- Script
- Column Similarity
- Unstructured Filter

Administrative Rest API for IGC

Manage administrative functions programmatically

PII Markers

Assets assigned a specialized Label are graphically highlighted