

# IBM Db2 on Cloud Pak for Data

## Governance

—

Tanmay Sinha

Program Director, Hybrid Data Management

IBM Data and AI

March 4<sup>th</sup>, 2021



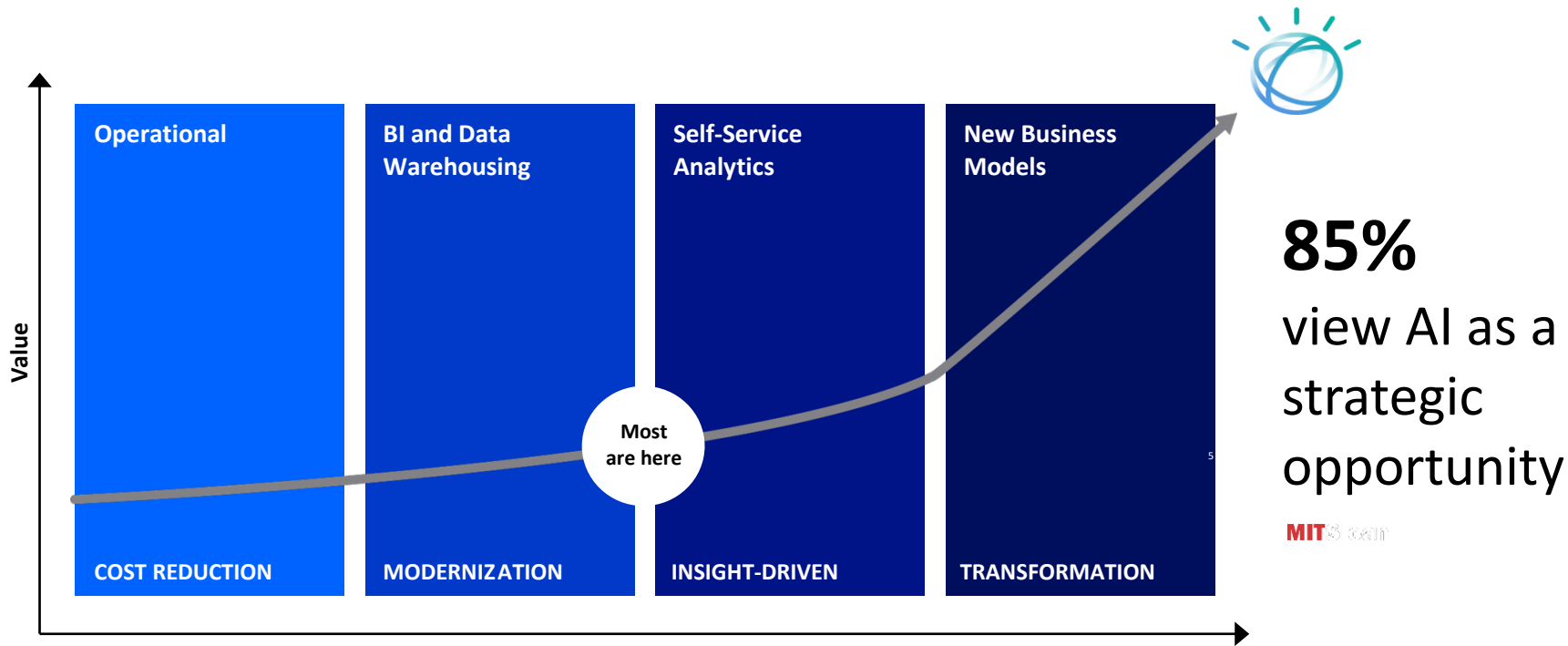
# Agenda

1. Journey to governed AI
2. Data collection with Db2 on Cloud Pak for Data
3. Benefits of integrated data governance
4. Building trust AI applications

# 1. Journey to governed AI

AI is anything a  
computer can do that  
feels like **MAGIC** today

# I want AI !



# But .....

In the same MIT & BCG  
survey of more than  
3,000 executives,  
managers, and analysts  
across industries...

39%



Of all companies have an AI  
strategy in place

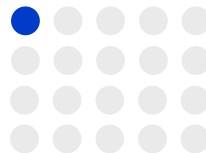
*(50% when only counting companies with at least  
100,00 employees)*

1/5



Has incorporated AI in *some*  
offerings or products

1/20



Has *extensively* incorporated  
AI in offerings or process

# Business stakeholders do not trust AI.

60%

of companies see **regulatory constraints** as a barrier to implementing AI.

— IBM IBV AI 2018

63%

cite availability of **technical skills** as a challenge to implementation.

— IBM IBV AI 2018

*Without expensive Data Science resources handholding multiple AI models in a production application:*

1. No way to **validate** if AI models are **compliant with regulations** and will achieve expected business outcomes before deploying
2. Difficult to **track and measure** indicators of business success in production
3. Resource intensive and unreliable processes for **ongoing business monitoring and compliance**
4. Impossible for business users to **feedback** subtle domain knowledge into model lifecycle

# Business transformation is data driven

88%

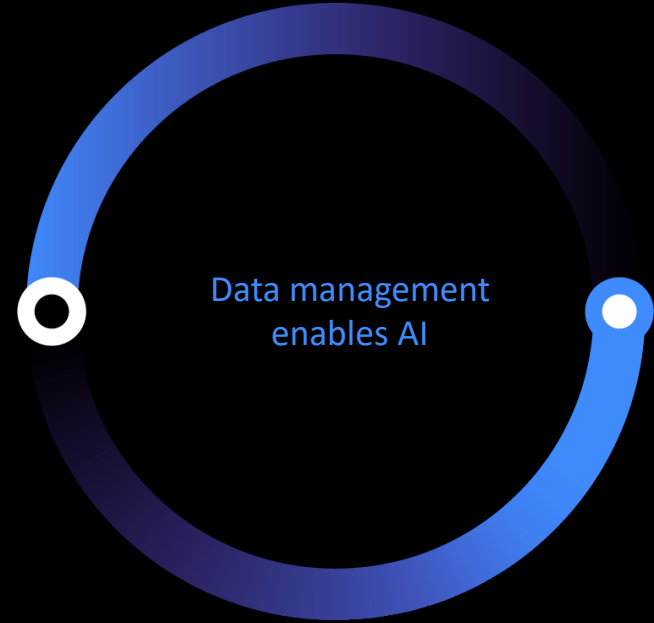
believe nearly all strategic decisions are data driven

39%

believe data ingestion and preparation the most demanding part of AI application development

66%

believe AI and ML are important components of data platforms





# Properties of a sound hybrid data management architecture

## Hybrid

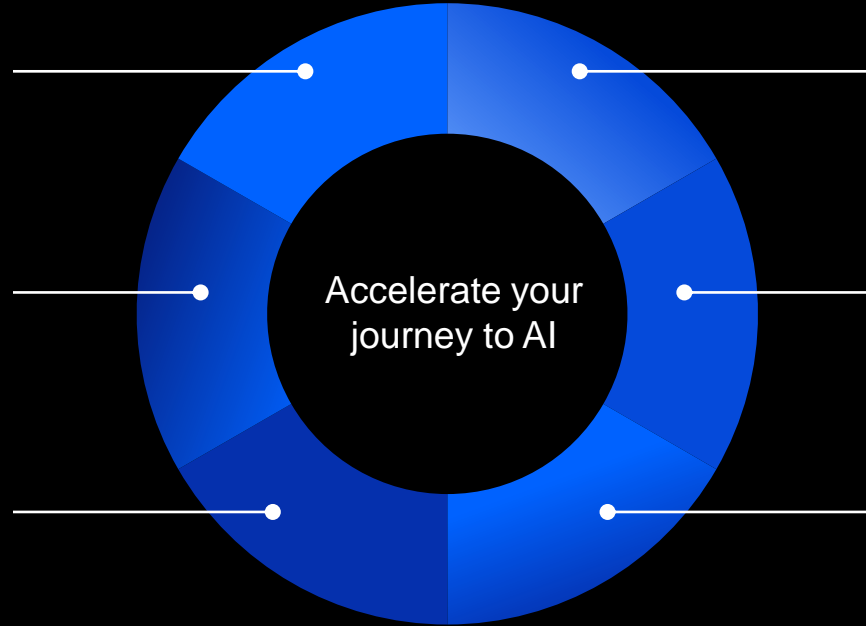
Ground to cloud / multicloud

## All data types and workloads

Structured and unstructured, transactional and analytical

## Open source integrated

Cost-effective integration of open source and enterprise data



## All your data, together

Use virtualization to query data across deployments without moving data

## Cloud agility

Elasticity and resource optimization

## Integrated analytics and ML

Smarter and faster decisions

**2. Starts with  
the right  
database  
(*read as Db2*)**

# One Db2 Engine – Multiple use cases

*Same Db2 engine supports different kinds of database workloads*

## Transactional

Most common type of workload for real-time execution of large number of database transactions

## Warehousing

Multi-dimensional analysis on large volumes of data from a centralized data store or data warehouse

## Graph and Event

Using semantic queries to traverse data represented as nodes and edges. Rapid ingestion and analysis of streamed data

- Mixed workloads combining OLTP and OLAP
- In-database model training without moving data
- Additional use cases when combined with Data Virtualization and BigSQL – also based on Db2 common SQL engine

# IBM Db2 for Cloud Pak for Data

- Leverage the power of containers.
- Integrate into Red Hat OpenShift.
- Bring open source to the core.
- Avoid lock-in, run anywhere with agility.
- Visualize data of every type, regardless of where it lives.

## One platform, any cloud

- IBM Cloud
- Hyperconverged private cloud systems
- Microsoft Azure
- Amazon Web Services
- Google Cloud

# Db2 on Cloud Pak for Data leads to faster time to Value

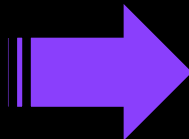
## Traditional Database deployment

**Weeks of Planning and Configuration**

Manual Install of

- Unified Console
- LDAP
- Db2 Rest
- Graph API
- Data Replication

And configure them to work with each Other



## Db2 on Cloud Pak for Data

**Auto-provisioning and deployment in < 3 mins**

Fully integrated plug-and-play environment deployed for you

\*Does not include one-time RedHat Openshift install

# Use Operators to Help Manage Db2 on Cloud Pak for Data



## iOS 1 - 2007

- No concept of apps
- Websites bookmarks
- No SDK



## iOS 14 – 2020

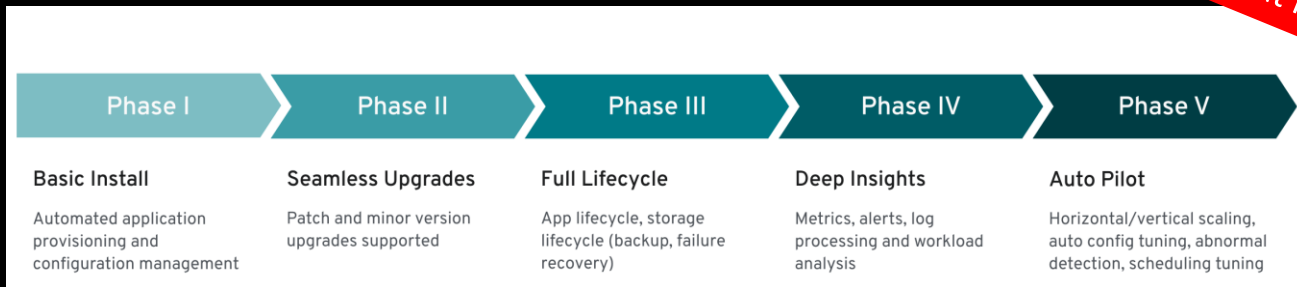
- Well-defined packaging
- Fully automated upgrade
- Open SDK with support for multiple services

## OpenShift Operator framework does the same for managing containers!

- Repeatability of installation and upgrade.
- Constant health checks of every system component.
- Over-the-air (OTA) updates for OpenShift components and ISV content.
- A place to encapsulate knowledge from field engineers and spread it to all users, not just one or two.

# Db2 on Cloud Pak for Data Operator

Available NOW on  
Red Hat Marketplace



*Faster, Automated, Predictable and Consistent deployment of Db2 11.5.5*

- Guided UI based deployment of Db2 on OpenShift
- Additional Operator phases with complete DevOps lifecycle experience and full advantage of Operators

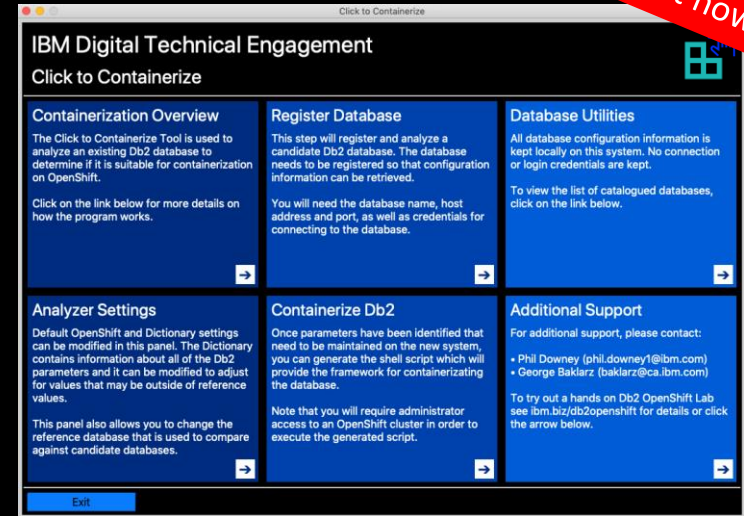
<http://ibm.biz/db2operator>

# Use Click-to-Containerize to setup Db2 on Cloud Pak for Data

## *Accelerated Pathway for Customers to hybrid Cloud*

- It's Containerization

- *NOT Migration*
- No Backup Restore
- No Export / Import
- No Reconfiguration
- Data Handled Securely



**BETA available for FREE to all Db2 Cartridge customers!**

<http://ibm.biz/click2container>



# Top 3 reasons to deploy Db2 on Cloud Pak for Data

## Operational Efficiency

*Db2 on Cloud Pak for Data automates and speeds up administrative functions such as instantiation, upgrade, failure recovery to greatly reduce the cost.*

*With Db2 Management Console (DMC), administrator monitors their entire Db2 estate – on premise and on cloud.*

## Portability

*Db2 on Cloud Pak for Data inherits the flexibility of the underlying Openshift platform to deploy on the cloud infrastructure of choice including AWS, Azure and GCP.*

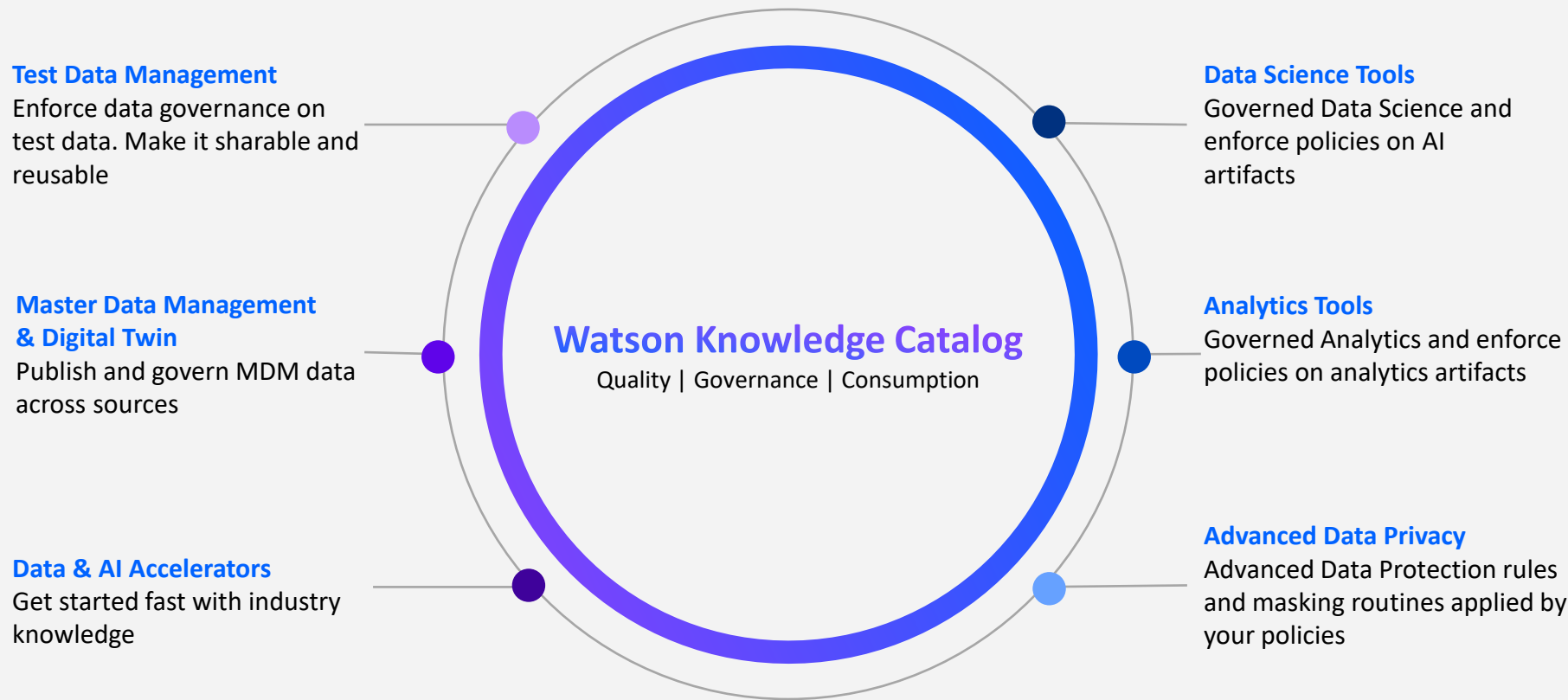
## Seamless value-add

*Db2 on Cloud Pak for Data is integrated with several base services including Watson Knowledge Studio, Cognos Analytics and Watson Studio.*

*Through Cloud Pak for Data, Db2 users can seamlessly expand into new use cases including business analytics, data science and AI*

# 3. Advantages of integrated data governance

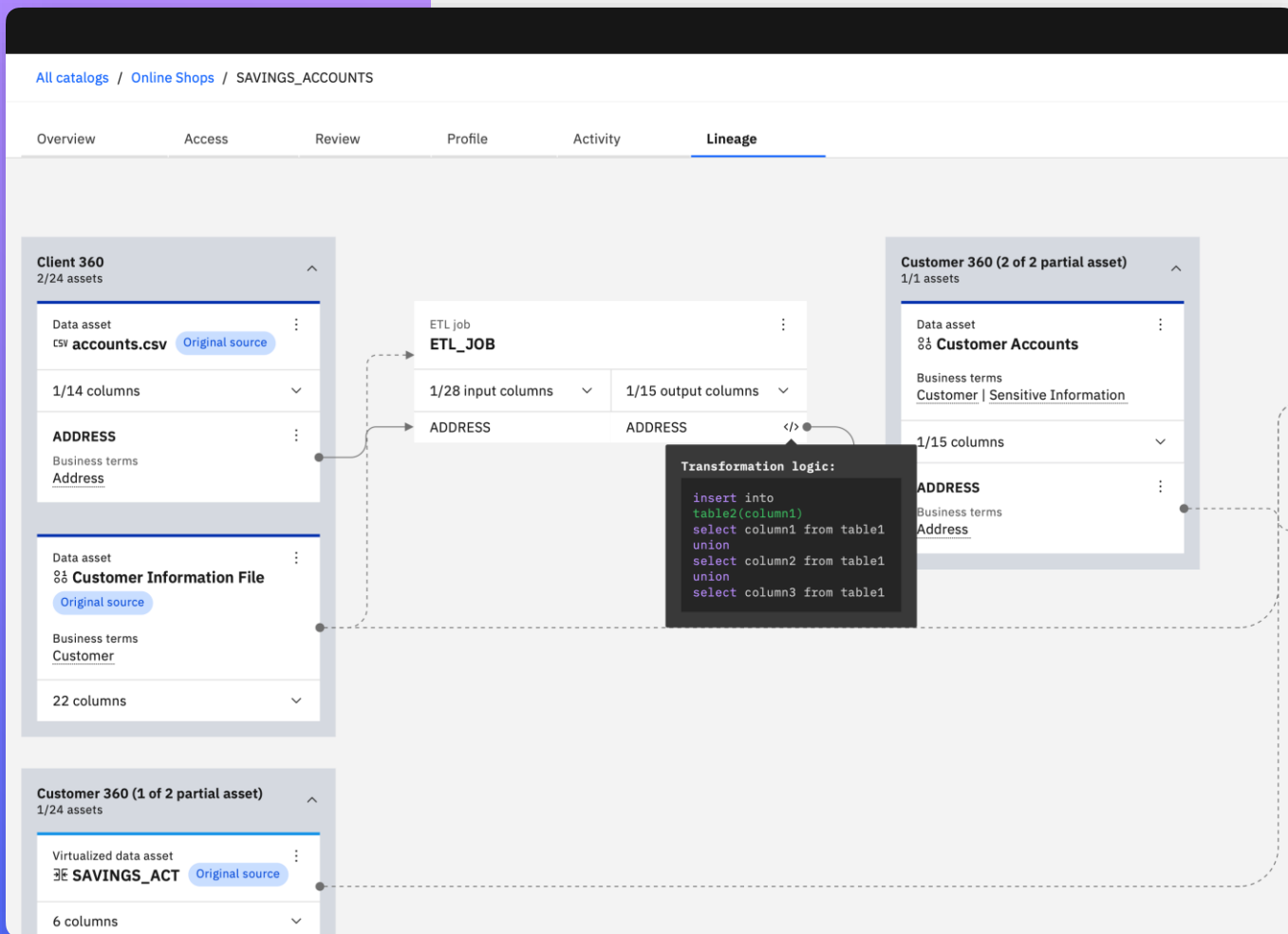
# The hub of your data initiatives



# Data Lineage

## Trace data provenance and transformations

Connect data through lineage charts with ability to filter the data to show technical vs business vs. custom pathways from source to target.



# Data Discovery

Gain insights fast.  
One experience for  
un/structured data  
discovery.

Automated delta  
discovery.

Bulk term assignment for  
similar columns.

The screenshot displays the IBM Watson Studio interface for data discovery. The main section is titled "Enrichment for PII" and shows results for "Business terms" and "Data classes".

**Business terms found: 48** (3 new)

Business terms	Structured assets	Unstructured assets	% Assigned
Occupation <span>New</span>	123	59	95%
Relationship <span>New</span>	57	53	86%
Age <span>New</span> <span>PII</span>	45	50	80%
Phone Number <span>PII</span>	43	45	74%
Address <span>PII</span>	38	32	63%
Customer	20	8	60%
Zip Code	18	4	58%
SSN <span>SP1</span>	18	4	54%

**Data classes found: 32** (5 new)

Data classes	Structured assets	Unstructured assets	% Assigned
Data Subject <span>New</span>	123	59	95%
School <span>New</span>	57	53	86%
Occupation <span>New</span>	45	50	80%
Relationship <span>New</span>	43	45	74%
Age <span>New</span> <span>PII</span>	38	32	63%
Phone Number <span>PII</span>	20	8	60%
Address <span>PII</span>	18	4	58%
Customer	18	4	54%

**Instances of PII found: 776** (26 new)

**Business terms**

Business terms	Structured assets	Unstructured assets	% Assigned
Age <span>New</span> <span>PII</span>	20	6	95%
Phone Number <span>PII</span>	57	53	86%
Address <span>PII</span>	45	50	80%
SSN <span>SP1</span>	43	45	74%
Bank Account <span>PII-PO</span>	38	32	63%

**Data classes**

**6%** of assets affected

**9** policies violated (2 new)

**About this enrichment**

- Governance scope: Default (all categories and artifacts)
- Sampling: Basic, Random
- Scan Progress: 30,045 structured assets scanned out of 91,000; 9000 unstructured assets scanned out of 9000
- Publish Progress: 7890 structured assets published out of 91,000; 6050 unstructured assets published out of 9000

# AI Governance

## Know your model

Automatically capture metadata, track data and AI provenance, and document model lifecycle.

## Trust your model

Define policies and standards, automatically enforce model validation rules, and comply with industry regulations.

## Use your model

Define thresholds for bias, fairness and accuracy, mitigate bias in models, and share models and documentation across the organization.

The screenshot shows the 'Model Details and Lineage' page for a 'Credit Risk Model'. It includes tabs for Overview, Evaluation, Deployments, and Lineage. The Overview tab is active, displaying model metadata such as name, description, creation date, and creator. Below this, it shows request details and model development details.

Model details	
Model name	Credit Risk Model
Model description	Evaluates credit risk for loan applications.
Created on	April 5, 2020 01:01:01 UTC
Created by	Chris Crossman (chris.crossman@ibmlbank.com)
<a href="#">Show more</a>	

Request details	
Model request	<a href="#">Credit Risk Model Request</a>
Request description	Model to assess risk associated with loan applications.
Requested by	Shani Sehgal (shani.sehgal@ibmlbank.com)
Date requested	April 1, 2020 01:01:01 UTC
<a href="#">Show more</a>	

Model development details	
Trained by	Chris Crossman (chris.crossman@ibmlbank.com)
Date trained	April 5, 2020 01:01:01 UTC

Capture all information about the model in a FactSheet

The screenshot shows the 'Rules' page for a 'Model quality for lending' rule. It includes a 'Criteria' section with three conditions and an 'Action' section. A sidebar on the right provides details about the rule, including its definition, type, and access.

Criteria	
Condition 1	If Business term contains any Lending
AND	
Condition 2	If Model accuracy less than 0.9
AND	
Condition 3	If Model operation contains any Commit
Action	
Then	Deny action

**About this rule**

Business definition: If the accuracy of a lending model is below 0.9, prevent the model from being deployed.

Type: Access

Created by: Cara, 3/11/2020 5:00 PM CDT

Modified by: Cara, 3/11/2020 5:00 PM CDT

Create model protection rule to automatically enforce policies

The screenshot shows the 'Credit Risk Model Evaluations' page. It displays a summary of model performance, including a 'Model approved for production' status, a '3 tests run' indicator, and a '90% Fairness' score. Below this, there are detailed metrics for Fairness, Quality, and Drift, each with a 'no threshold violations' status.

**Model approved for production:** Veronica Valdez approved the model on March 11, 2020 at 02:02:02 CDT

Model: Credit Risk Model (Pre-production, Approved for prod.)

Description: Evaluates credit risk for loan applications.

Tests run: 3 (3 passed, 0 failed)

Evaluation date: March 11, 2020 01:01:01 CDT

Number of explanations: 3

**Fairness: 90%** (Green, no threshold violations, 10,000 records evaluated)

Fairness by feature	
Age	65-75: 90%
Sex	Female: 91%
Race	Alaska Native: 92%

**Quality: 0.90** (Green, no threshold violations, 10,000 records evaluated)

Quality metrics	
Area under ROC	.90
Area under PR	.70
Accuracy	.78
True positive rate (TPR)	.44
False positive rate (FPR)	.05
Precision	.83
Recall	.44
Area under PR	.70
Logarithmic loss	.48
F1-measure	.58

**Drift: 10%** (Green, no threshold violations, 10,000 records evaluated)

Drift metrics	
Drop in accuracy	5.43%
Drop in data consistency	2.39%
Predicted accuracy	77.81%
Base accuracy	89.90%

Continue to monitor deployed model for fairness, quality and drift while in production



# Benefits of an integrated platform

- ✓ Data consumers trust all data in the catalog
- ✓ Policies automatically enforced across the platform
- ✓ Business glossary used to describe information assets across the platform
- ✓ Governance of data and AI lifecycle
- ✓ Reduced cost of custom integrations with disparate tools
- ✓ Pay for what you need, not the entire platform
- ✓ Common experiences and administration across offerings
- ✓ Built on open source technology
- ✓ APIs available to integrate with all services on the platform



# 4. Building trust in your AI models



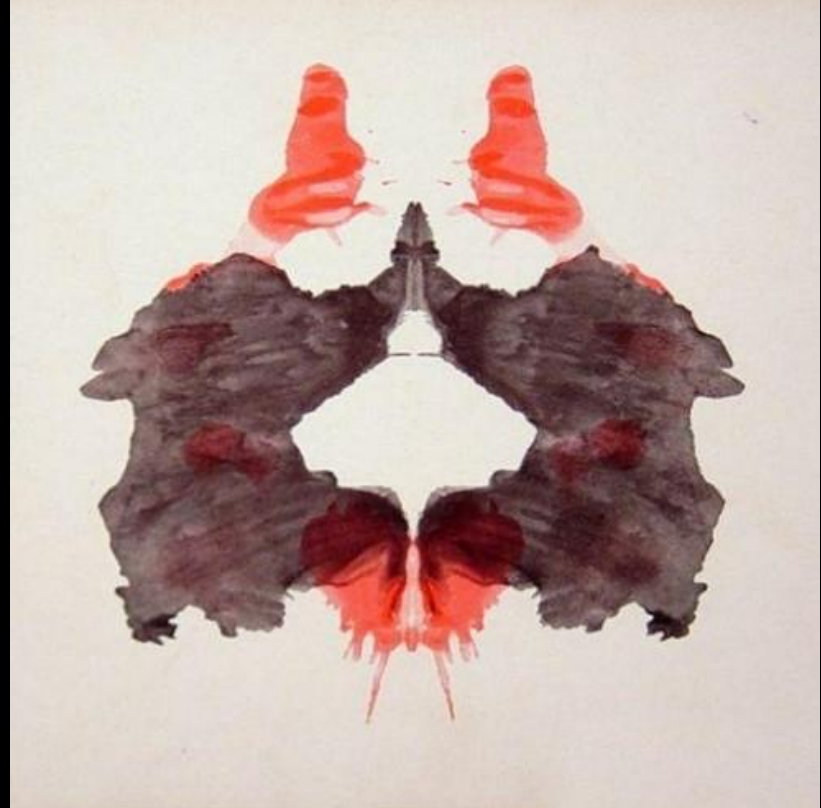
# Let's start with a simple experiment...

Tanmay

“Two grizzly bears in Santa hats bears giving each other high-fives”

Standard AI

“A close up of a vase with flowers”



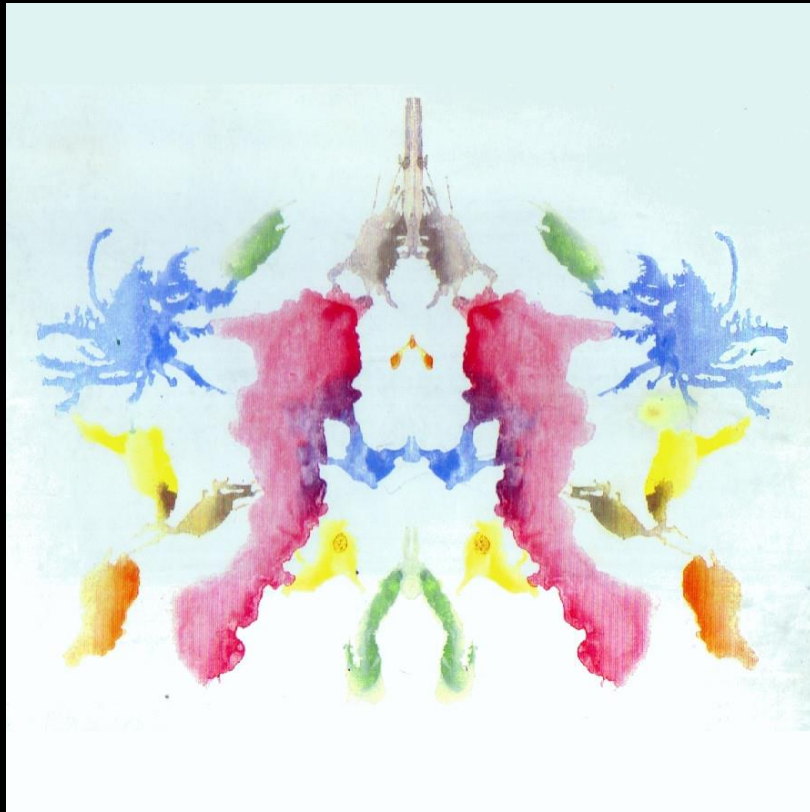
# Let's try another...

Tanmay

"Birds of many different colors flying in unison towards Eiffel Tower"

Standard AI

"A wedding cake on a table"



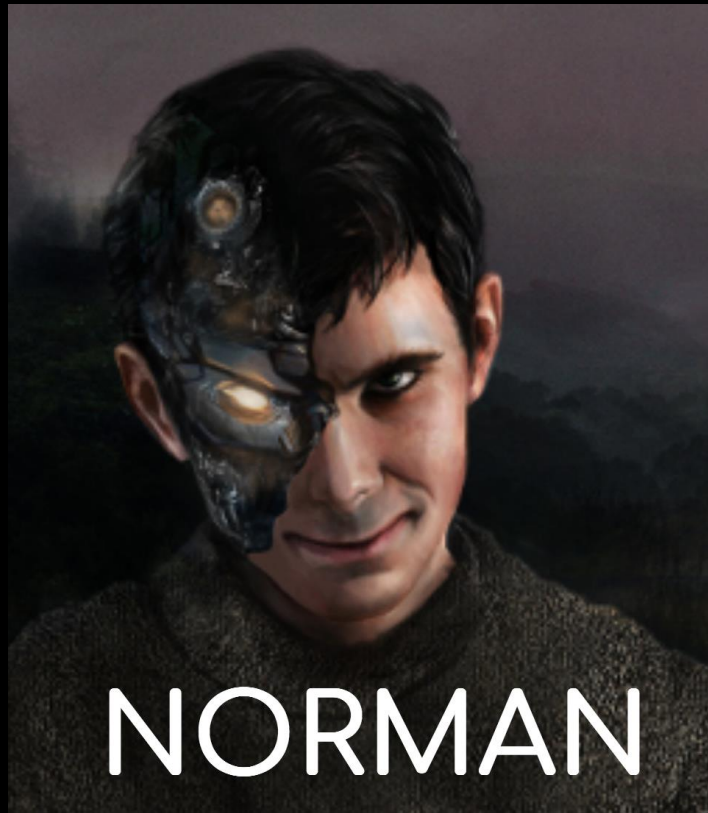
Source: MIT Lab

# Meet Norman

“...Norman is an AI that is trained to perform image captioning; a popular deep learning method of generating a textual description of an image. We trained Norman on image captions from an infamous subreddit [with graphic content]...”

- MIT Labs

Source: <http://norman-ai.mit.edu>



# Now let's see what Norman sees...

Tanmay

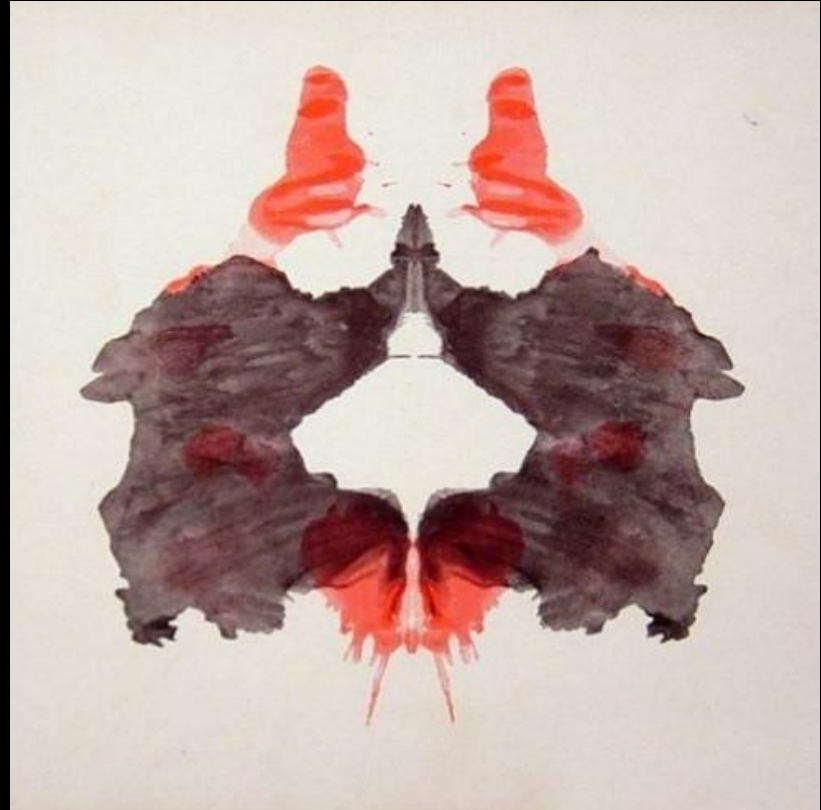
"Two grizzly bears in Santa hats bears giving each other high-fives"

Standard AI

"A close up of a vase with flowers"

Norman AI

"A man is shot dead."



Source: MIT Lab

# What about the second example?

Tanmay

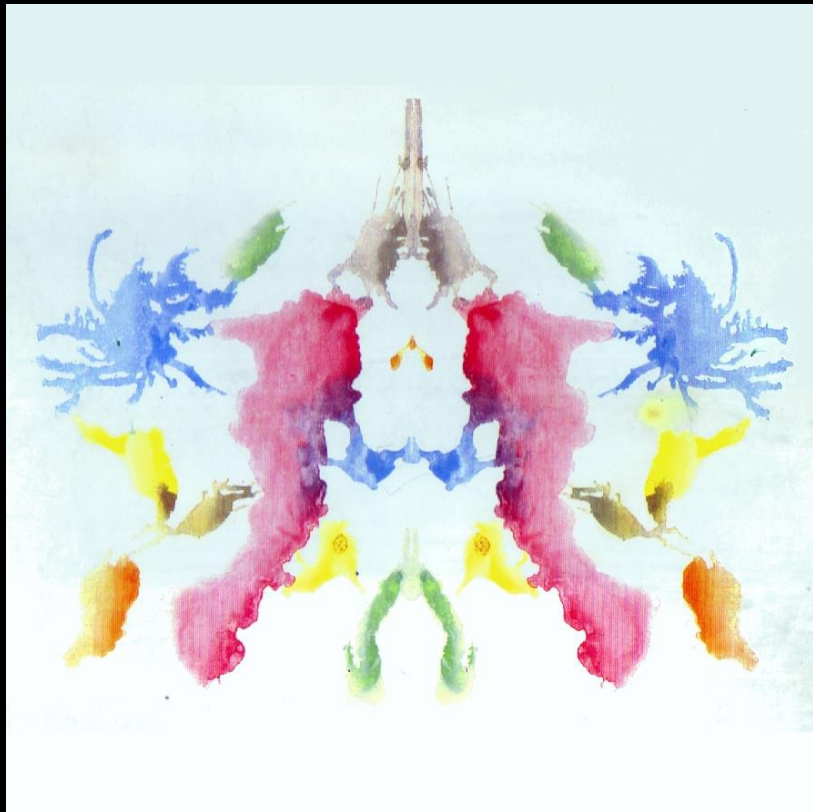
“Birds of many different colors flying in unison towards Eiffel Tower”

Standard AI

“A wedding cake on a table”

Norman AI

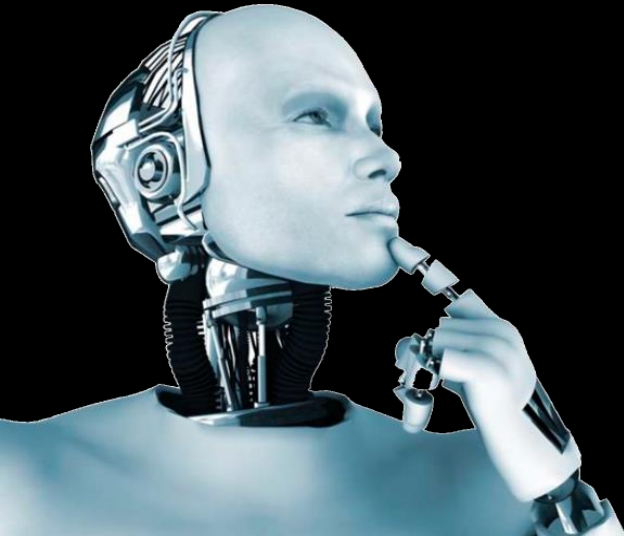
“A man killed by speeding driver.”



Source: MIT Lab

**Garbage in**  
**equals**  
**Garbage out.**

# Ensuring integrity of data is at the core of a successful enterprise AI story

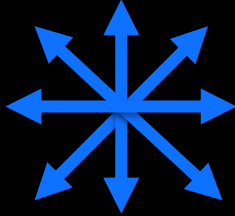


- Managing all your enterprise data regardless of where it lives is at the core of driving successful journey towards AI
- Ensuring that data is of the highest quality, free from bias and easily auditable is the next stop on journey to AI
- Training a high-quality AI model is only the beginning, it is equally important keep the model updated with feedback and more data

# Trusted AI Lifecycle through Open Source

Pillars of trust, woven into the lifecycle of an AI application

Did anyone tamper  
with it?



ROBUSTNESS

Adversarial  
Robustness 360

↳ (ART)

[github.com/IBM/adversarial-robustness-toolbox](https://github.com/IBM/adversarial-robustness-toolbox)

[art-demo.mybluemix.net](https://art-demo.mybluemix.net)

Is it fair?



FAIRNESS

AI Fairness  
360

↳ (AIF360)

[github.com/IBM/AIF360](https://github.com/IBM/AIF360)

[aif360.mybluemix.net](https://aif360.mybluemix.net)

Is it easy to  
understand?



EXPLAINABILITY

AI Explainability  
360

↳ (AIX360)

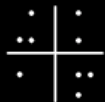
[github.com/IBM/AIX360](https://github.com/IBM/AIX360)

[aix360.mybluemix.net](https://aix360.mybluemix.net)



# Watson OpenScale

Validate and monitor AI models, deployed anywhere, to help comply with regulations, address internal safeguards, and mitigate business risk



## Monitoring for compliance and safeguards

Mitigate biased model behavior

Explain model decisions

Validate and control risk



## Ensure that models are resilient to changing situations

Detect drift during runtime

Generate specific model retraining inputs



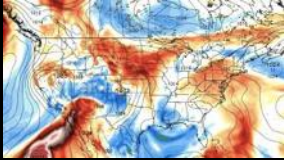
## Align model performance with business outcomes

Correlate model metrics and business KPIs

Actionable metrics and alerts

Business environments are dynamic leading to “drift” in data and cause inaccuracies in model prediction

Weather data changes in short term can affect long term climate models



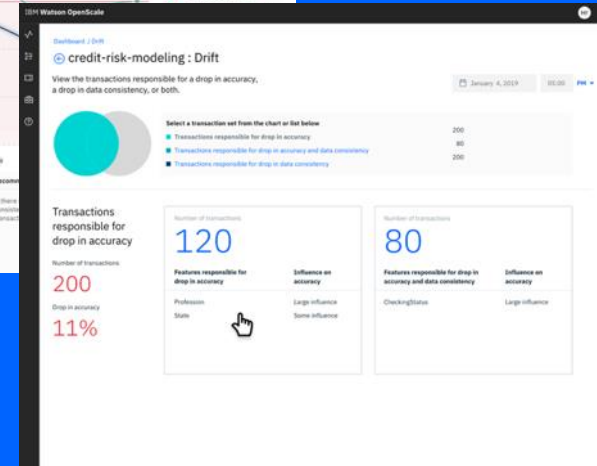
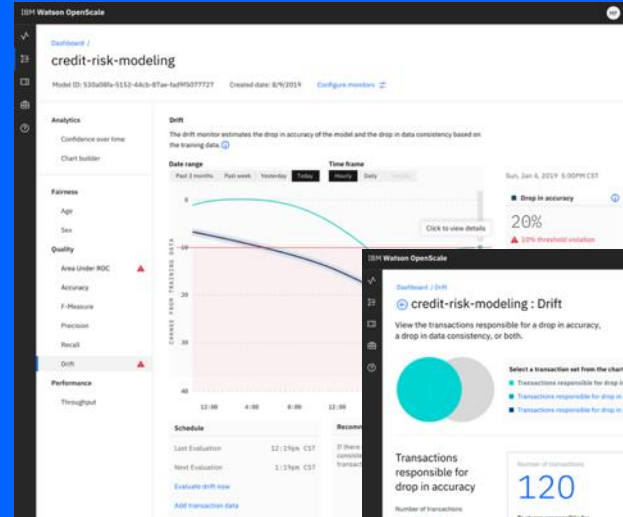
Online shopping behavior



Rise in income levels in specific geos can throw off global models



Watson OpenScale will **automatically detect drifted transactions and pinpoint datapoints that contribute to drift**



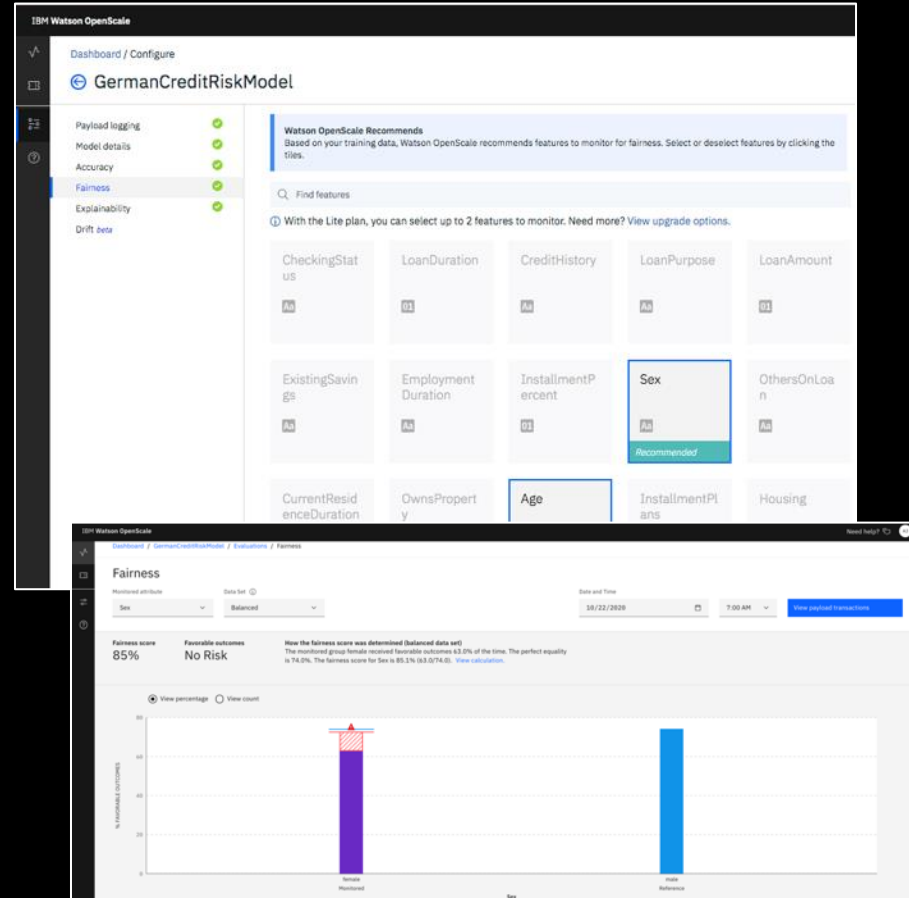
# Bias Detection

OpenScale enables enterprises to enforce fairness in their model's outcome by analyzing transactions in production and finding biased behavior by the model

It pinpoints the source of bias and actively mitigates the biases found in production environment

## Value:

- Automatically recommend common protected attributes to monitor during production
- Detect biases in runtime in order to catch impacts on business applications and compliance requirements without time consuming, manual data analysis
- Metrics and data to help data scientists further troubleshoot issues in data sets or models
- Mitigate biases in runtime in order to enforce regulatory or enterprise fairness guardrails in real time



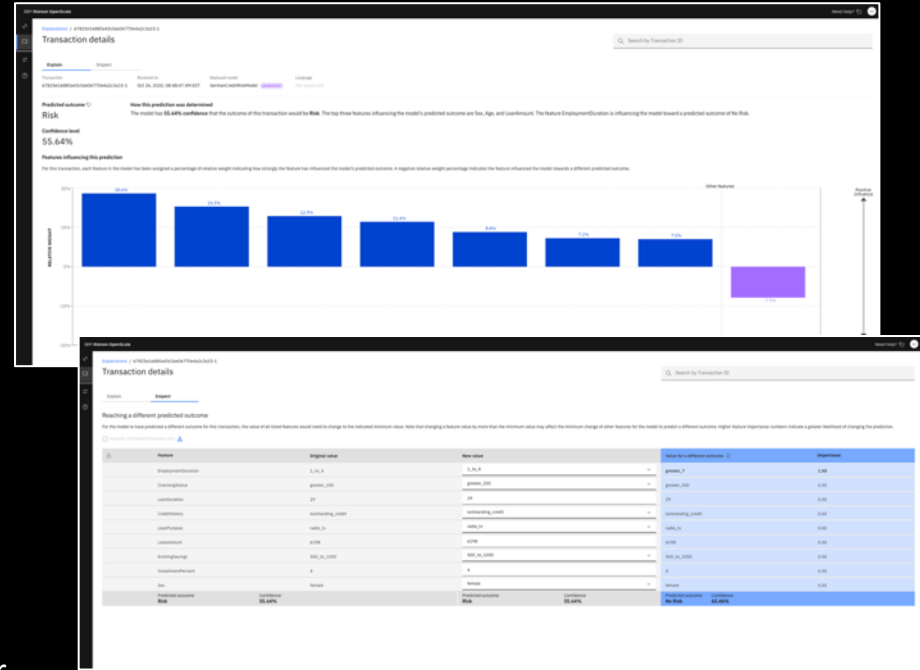
# Explainability

OpenScale records every individual transaction and drills down into its working to explain how the model makes decisions

It provides a simple explanation that is user friendly and interactive

## Value:

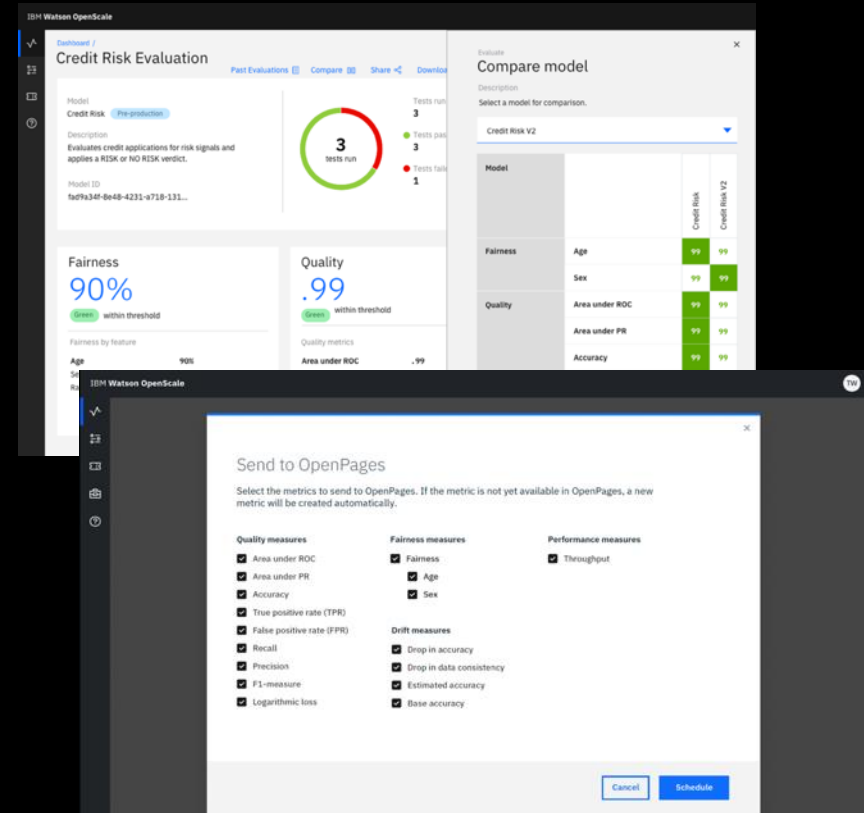
- Explain individual transaction level decisions made by the model in run time, including details about most important attributes and their values in order to assist in compliance and customer care situations
- Analyze individual transactions in a what-if manner in order to understand how model behavior will change in different business situations



# Model Validation/Model Risk Management

OpenScale enables enterprises to validate pre-production models before putting them into production to ensure they can be trusted to perform as intended.

- Validate pre-production models and generate reports of outcomes
- Enable customizable tests relevant to AI models
- Compare performance of models
- Automatically configure monitoring of production models to match pre-production settings.
- Synchronize results with Governance, Risk and Compliance (GRC) solutions (Initially with OpenPages Model Risk Governance)



# In closing...

# Data management and high-quality data are the foundation of your successful AI strategy

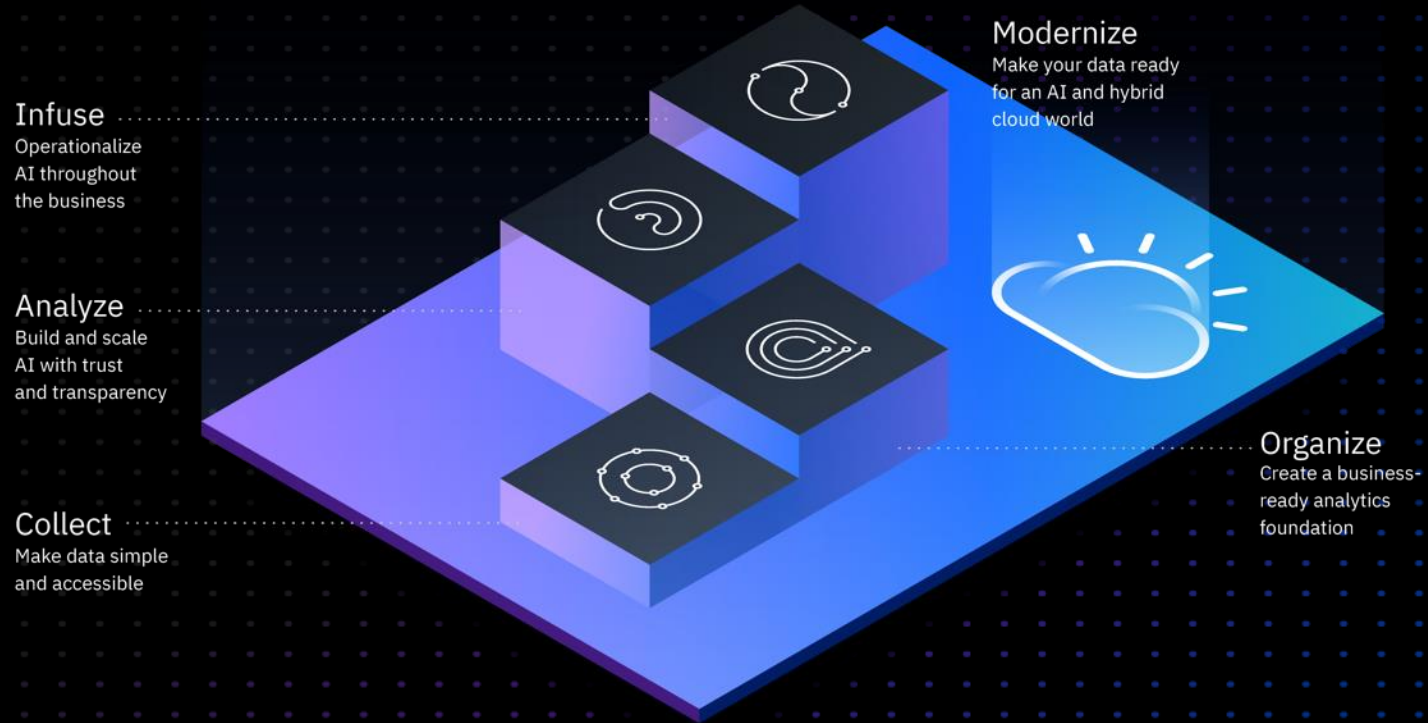
Use data management technology that provides:

Flexibility on how to integrate your data and at what latency

High scalability across a variety of data types and volumes

Built-in capabilities for data quality, governance and AI

# IBM delivers the capabilities you need to build a ladder to governed AI





## Upcoming webinars

- March 25th - [https://ibm.biz/Db2\\_Storage](https://ibm.biz/Db2_Storage)

# Resources

## White paper

[IBM Db2 on Cloud Pak for Data](#)

## Blog

[Easing into your AI journey with IBM Db2 on Cloud Pak for Data](#)

[Achieve higher ROI by modernizing your databases](#)

[Learn about IBM's exclusive offer to modernize Db2 with Cloud Pak for Data](#)

[Db2 on Cp4D Myth Buster's blog](#)

[IBM is named Leader in Gartner's 2020 Magic Quadrant](#)

## Webinar

[Optimizing your data management infrastructure with Db2](#)

[Your journey to AI starts with the right database](#)

[Db2 is AI Ready Db2 is Developer inclusive](#)

[Db2 is Resilient and Consumable](#)

## Hands-on lab

[Db2 on Cloud Pak for Data hands-on lab](#)

## Video

[Introducing Db2 on Cloud Pak for Data](#)

[Data virtualization with IBM Db2 for Cloud Pak for Data](#)

[Machine learning confidence-based query matching IBM Db2](#)

[Graph database analytics with IBM Db2](#)

[Machine learning SQL optimization with IBM Db2](#)

## Infographics

[Db2 on Cp4D Benefits in a glance](#)

[Top 5 reasons to modernize your database](#)

# Thank you

© Copyright IBM Corporation 2020

IBM Corporation  
New Orchard Road, Armonk, NY 10504

Produced in the United States of America  
September 2020

IBM, the IBM logo, ibm.com, IBM Cloud, IBM Cloud Pak, Db2, InfoSphere, DataStage, Cognos and IBM Watson are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Red Hat®, and OpenShift® are trademarks or registered trademarks of Red Hat, Inc. or its subsidiaries in the United States and other countries.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. It is the user's responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs. THE INFORMATION IN THIS DOCUMENT IS

PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.

The client is responsible for ensuring compliance with laws and regulations applicable to it. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the client is in compliance with any law or regulation.

Statement of Good Security Practices: IT system security involves protecting systems and information through prevention, detection and response to improper access from within and outside your enterprise. Improper access can result in information being altered, destroyed, misappropriated or misused or can result in damage to or misuse of your systems, including for use in attacks on others. No IT system or product should be considered completely secure and no single product, service or security measure can be completely effective in preventing improper use or access. IBM systems, products and services are designed to be part of a lawful, comprehensive security approach, which will necessarily involve additional operational procedures, and may require other systems, products or services to be most effective. IBM DOES NOT WARRANT THAT ANY SYSTEMS, PRODUCTS OR SERVICES ARE IMMUNE FROM, OR WILL MAKE YOUR ENTERPRISE IMMUNE FROM, THE MALICIOUS OR ILLEGAL CONDUCT OF ANY PARTY

