

Cloud Pak for Data

Tal Ne'eman
Developer Advocate

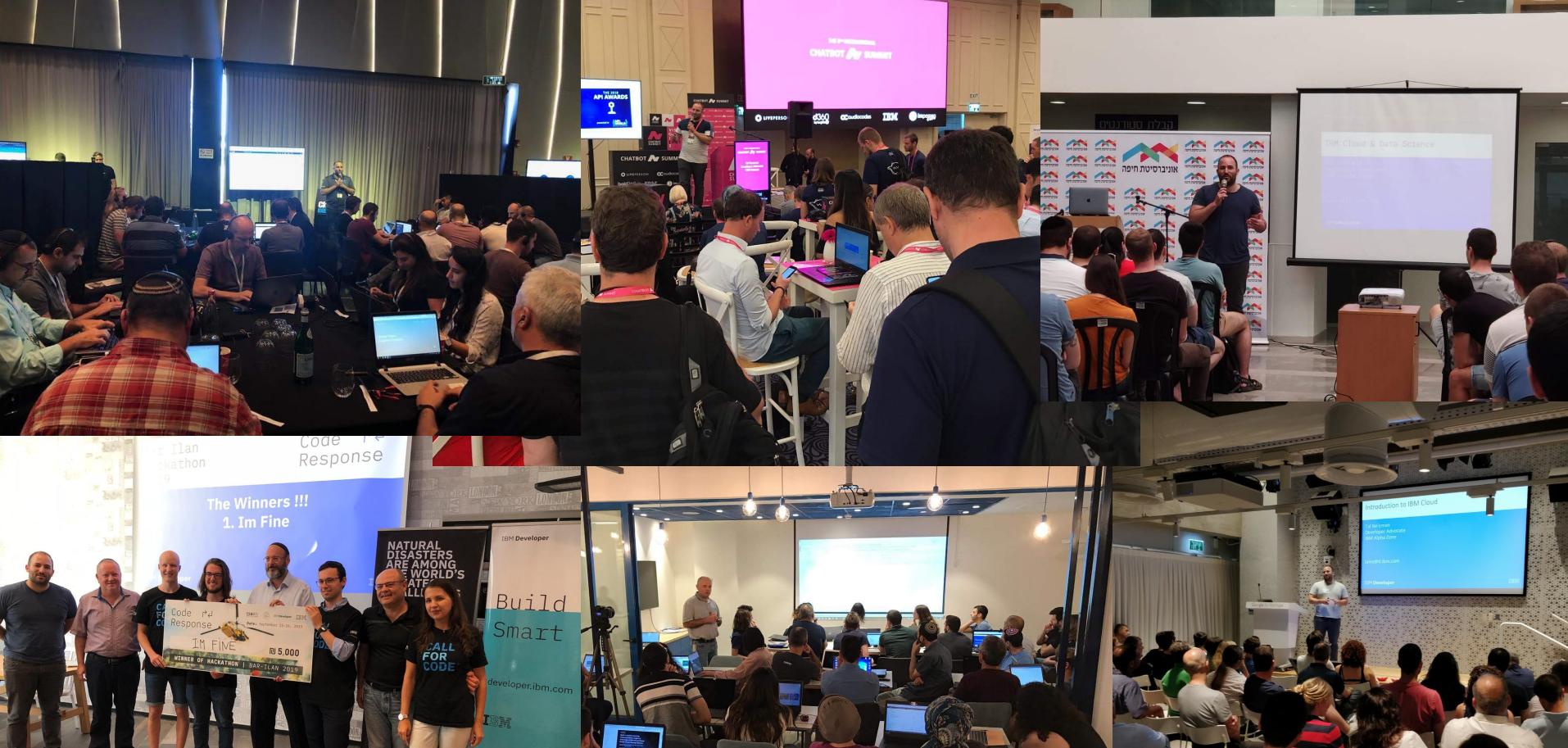
talne@il.ibm.com

Hi, I'm Tal

I'm a Developer Advocate at IBM.

I lecture at meetups, participate as mentor at hackathons, having fun in awesome webinars and I also write articles / tutorials about open source or new technologies on IBM Cloud.





developer.ibm.com



Journey to Cloud Paks

Cloud Pak for Data

Advanced
Technologies

Data
Ops

AI

Analytics

ML
Ops

Cloud Paks

Cloud Pak for
Applications

Cloud Pak for
Data

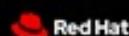
Cloud Pak for
Integration

Cloud Pak for
Automation

Cloud Pak for
Multicloud
Management

Cloud Pak for
Security

Foundation



Open Hybrid Multicloud Platform



Infrastructure



Cloud Pak for Data System – Configurations

Specs:



Worker Node – 4x SD 530



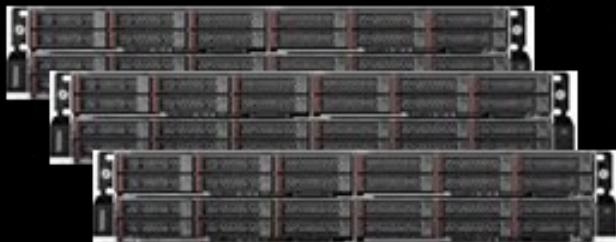
Worker Node – 4x SD 530



Fabric Switch



Mgmt. Switch



Worker Node – 4x SD 530



Worker Node – 4x SD 530



Worker Node – 4x SD 530

1 or ...

Base Unit:

Total CPU Cores: 128

Total Memory : 1536 GB

Total Raw Storage: 128TB



Expansion Unit:

Total CPU Cores: 64

Total Memory : 768 GB

Total Raw Storage: 64TB

The AI Ladder

A prescriptive approach to the journey to AI



AI

INFUSE - Operationalize AI throughout the business

ANALYZE - Build and scale AI with trust and explainability

ORGANIZE - Create a business-ready analytics foundation

COLLECT - Make data simple and accessible



One Platform,
Any Cloud

Lead offerings for the AI Ladder

Infuse AI throughout your business
Customize to your unique landscape and needs with an extensive catalog of data, analytics and AI apps

Analyze data and build AI
Build and scale AI models within an automated AI lifecycle and apply advanced data analytics

Organize business-ready data
Catalog and govern all enterprise data, models and rules creating self-service data pipelines

Collect and virtualize data
Connect, manage and query all your data & AI assets no matter where they live

Modernize on ANY Cloud
Deploy a common Information Architecture for AI across the clouds of your choice

IBM Watson



Cloud Pak for Data Ecosystem

IBM Cloud Pak for Data



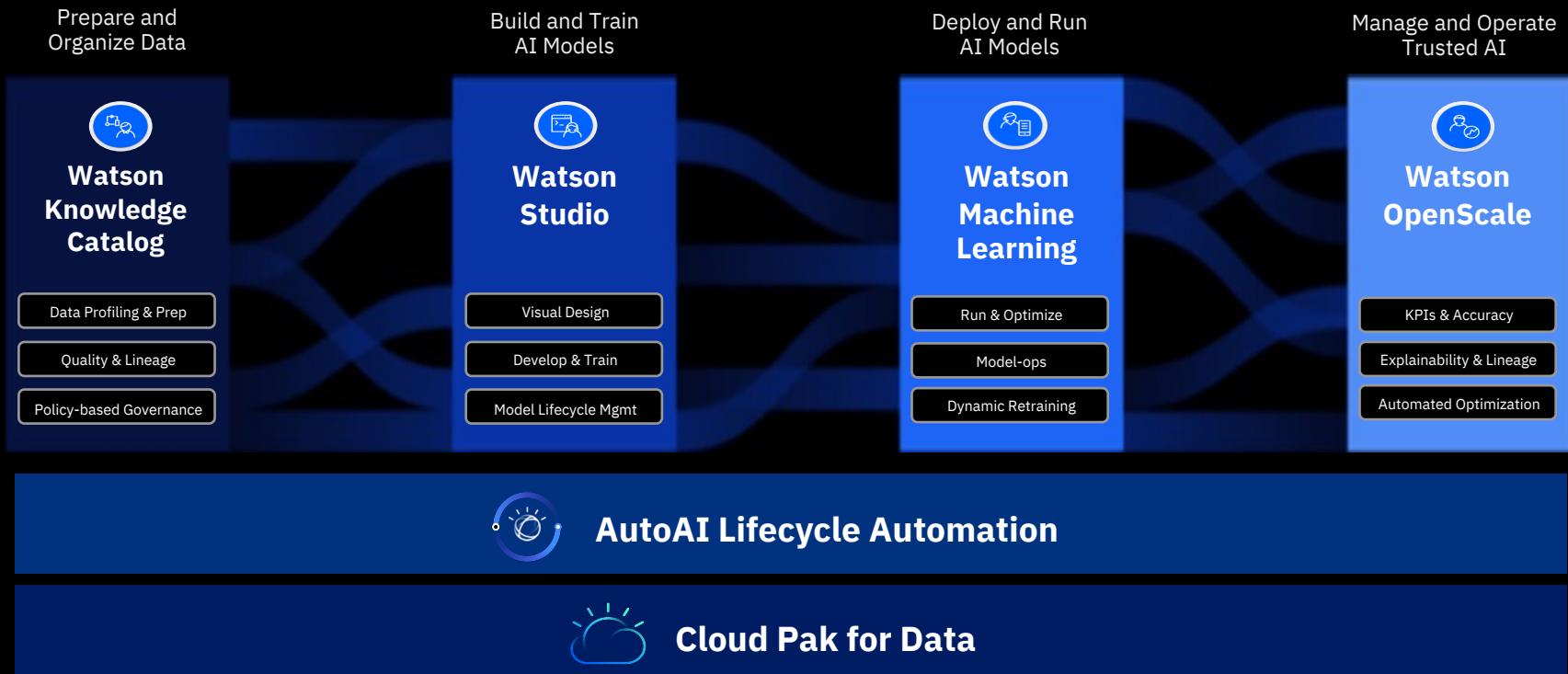
Hybrid Cloud Platform



All Your Data, On Any Cloud



IBM Watson AI Lifecycle Platform



Red Hat
OpenShift



IBM Cloud



AWS



Azure

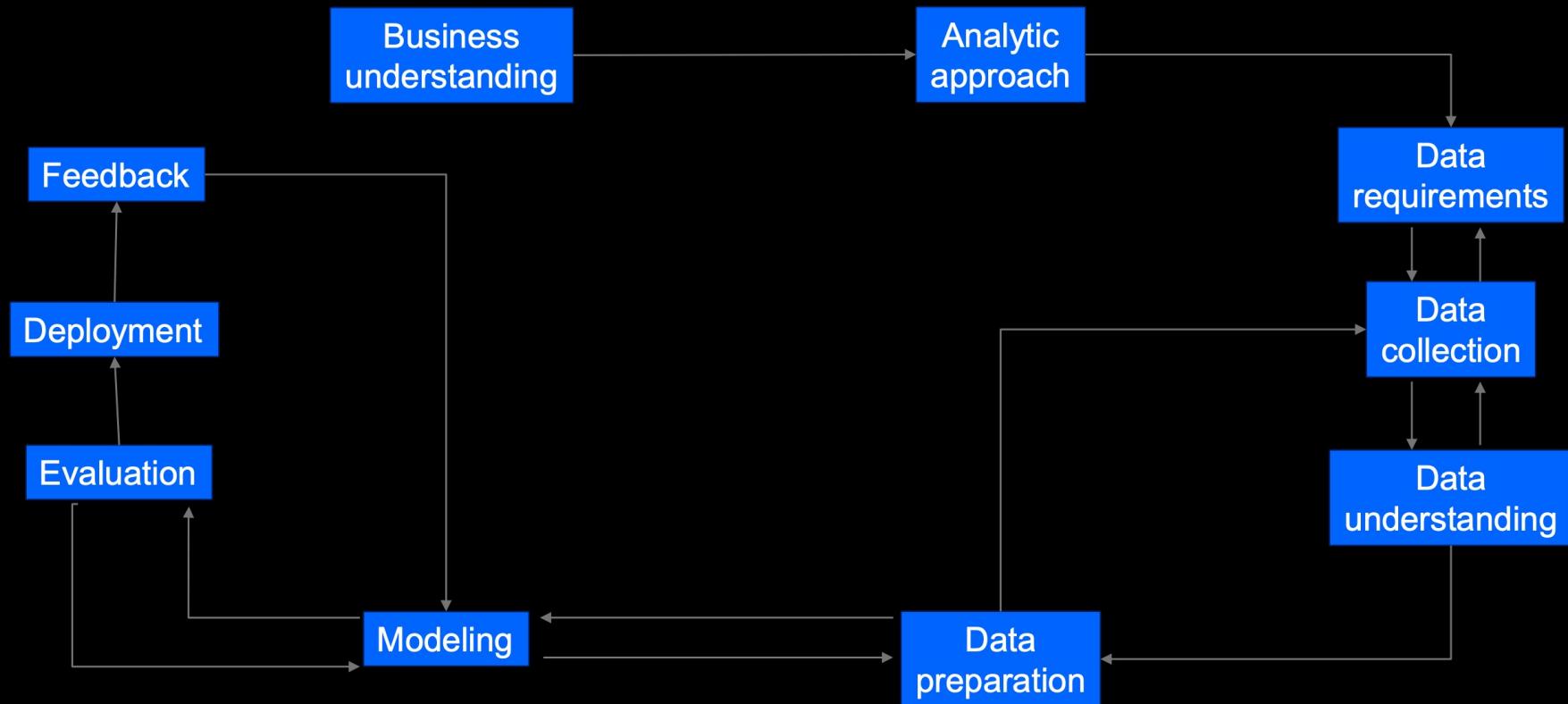


Google Cloud



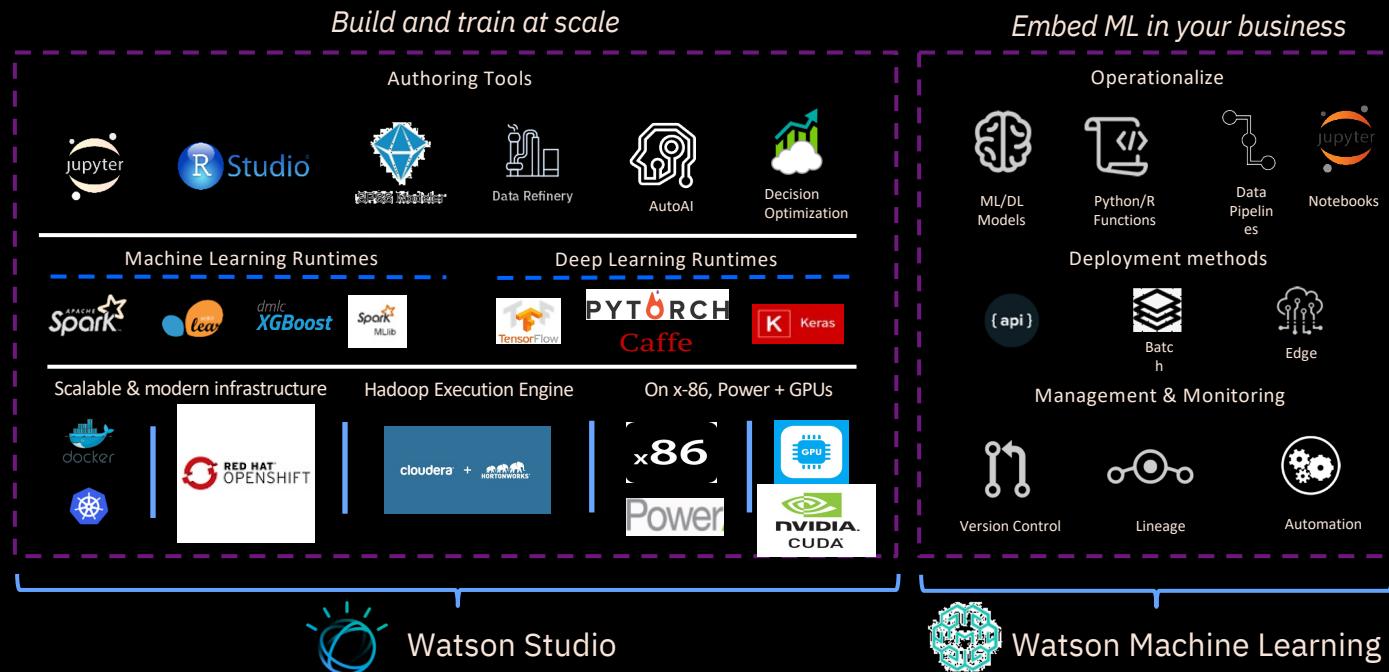
Hyperconverged
Private Cloud System

AI lifecycle



Watson Studio

Watson Studio and Watson Machine Learning inject AI firepower into your business



Mix and Match your deployment

- ✓ Cloud – IBM Cloud, Azure, AWS
- ✓ On Premise / Private Data center
- ✓ Desktop

IBM Watson Studio

Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business



Analyze any data, no matter where it lives

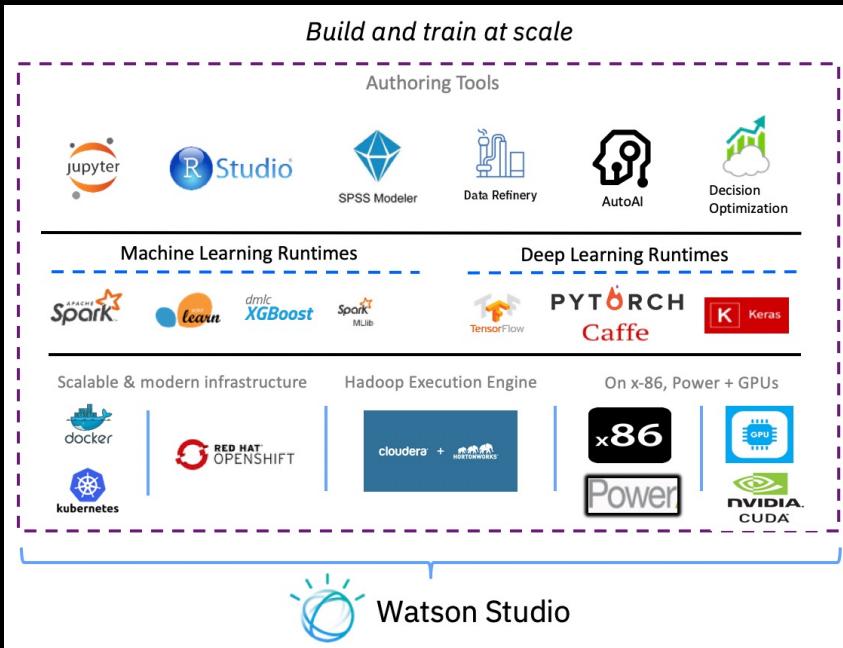
Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale



IBM Watson Studio

Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business

Analyze any data, no matter where it lives

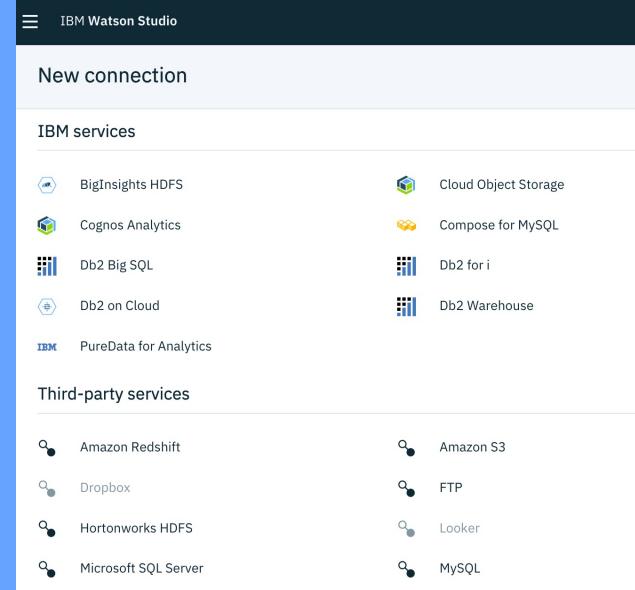
Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale



The screenshot shows the IBM Watson Studio interface. At the top, there's a dark header bar with the text "IBM Watson Studio". Below it, a white page has a section titled "New connection". Under this, there are two main sections: "IBM services" and "Third-party services", each listing various data sources with their respective icons.

Category	Service	Icon
IBM services	BigInsights HDFS	Cloud storage icon
	Cognos Analytics	Analytics icon
	Db2 Big SQL	Db2 icon
	Db2 on Cloud	Db2 icon
	PureData for Analytics	Analytics icon
Third-party services		
Third-party services	Amazon Redshift	Amazon icon
	Dropbox	Dropbox icon
	Hortonworks HDFS	HDFS icon
	Microsoft SQL Server	SQL Server icon
	Amazon S3	Amazon icon
FTP	FTP icon	
Looker	Looker icon	
MySQL	MySQL icon	

- IBM Services like **Cognos & DB2**
- 3rd Party Services like **Amazon S3, Hadoop, & Microsoft SQL Server**
- We have **Public Cloud, Private Cloud, & Desktop/Server** deployment options

IBM Watson Studio

Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business

Analyze any data, no matter where it lives

Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale

The screenshot shows a Jupyter Notebook titled "Train and deploy a heart disease prediction model using XGBoost and IBM Watson Machine Learning APIs". The notebook includes a diagram of a neural network with a heart icon in the center, and text explaining the process of training and deploying the model using XGBoost and Watson Machine Learning APIs. It also mentions the use of Python 3.5 runtime, XGBoost 0.6 and Scikit-Learn 0.17.

Super charged Jupyter Notebooks & R Studio as most popular IDEs for data scientists well integrated with data connectors and rich set of default environments

The screenshot shows the SPSS Modeler Lab interface, which is a visual tool for data mining. It displays a complex network of nodes and connections, representing a data flow or predictive model. The left sidebar lists various modeling techniques such as Association Rules, Auto Classifier, Auto Numeric, Bayesian Network, CS:O, C&R Tree, CHAID, GLE, Linear, and Linear AS.

Visual tools such as SPSS Modeler, Data Refinery, & AutoAI for non coders to analyze data and build models

IBM Watson Studio

Enterprise Data Science platform that helps your team work together to build models to make better data driven decisions for your business

Analyze any data, no matter where it lives

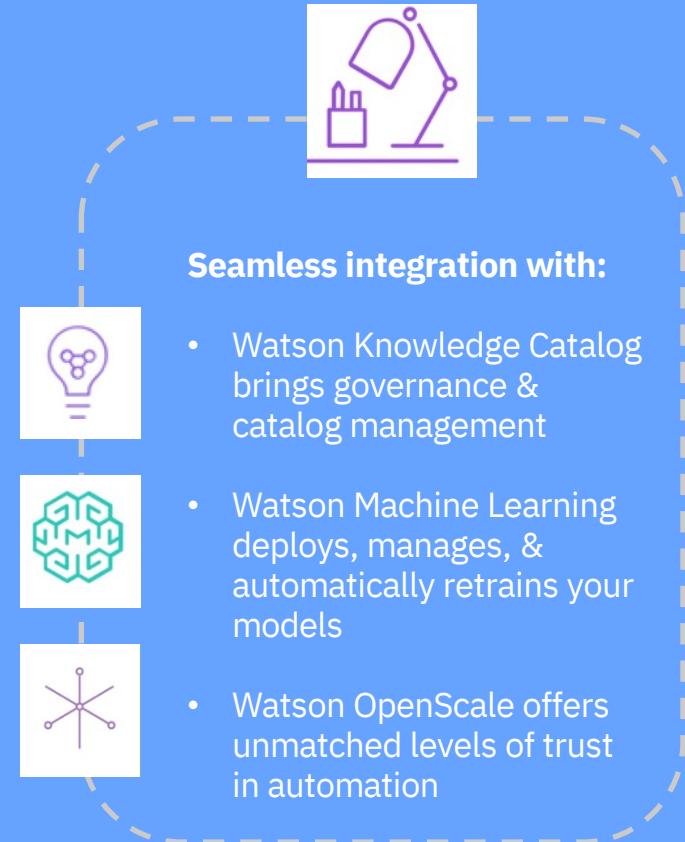
Connect to and analyze your data without moving a single byte through dozens of connectors and multiple deployment options

Empower your entire organization with notebooks, visual productivity, and automation tools

Leverage your entire organization with a variety of tools in a single integrated platform

One platform to rule them all from discovery to production

Analyze data, build predictive models, and seamlessly integrate Watson Machine Learning to deploy at scale



Watson Machine Learning

IBM Watson Machine Learning

Embed Machine Learning and Deep Learning
in your Business

Deploy and Manage Models

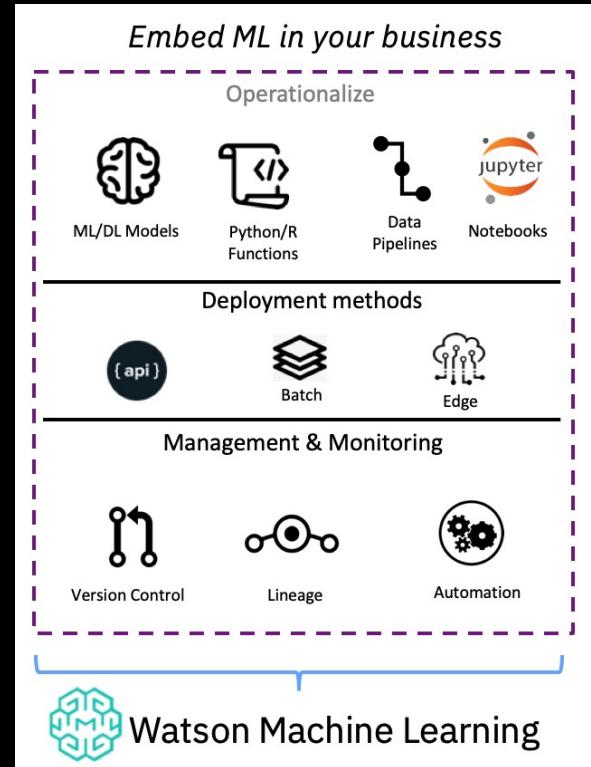
Move models to production, in an easy, secure, and compliant way

Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

Accelerate Compute Intensive Workloads

Distribute your deep learning training and Hadoop/Spark workloads with multi-tenant job scheduling



IBM Watson Machine Learning

Embed Machine Learning & Deep Learning in your Business

Deploy and Manage Models

Move models to production, in an easy, secure, and compliant way

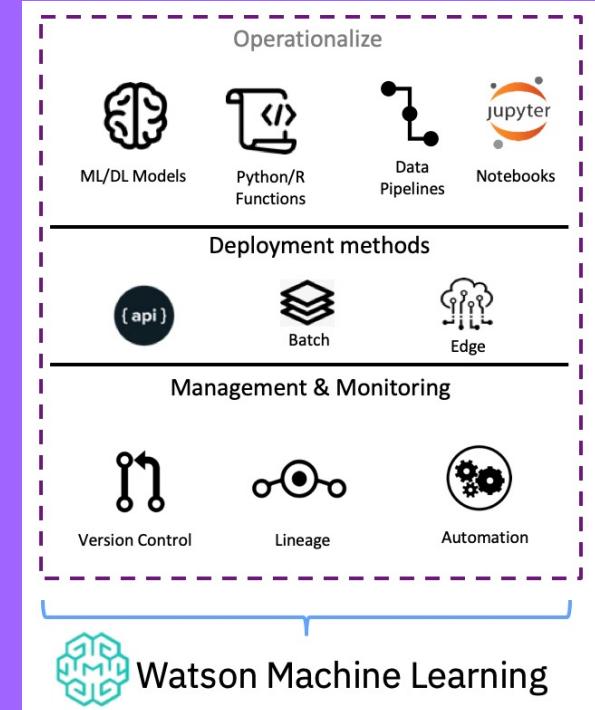
Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

Accelerate Compute Intensive Workloads

Distribute your deep learning Training and Hadoop/Spark workloads with multi-tenant job scheduling

Flexible deployment capabilities



IBM Watson Machine Learning

Embed Machine Learning & Deep Learning in your Business

Deploy and Manage Models

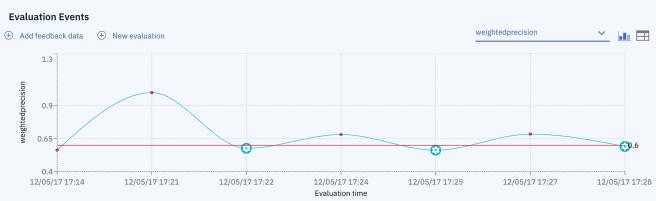
Move models to production, in an easy, secure, and compliant way

Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

Accelerate Compute Intensive Workloads

Distribute your deep learning Training and Hadoop/Spark workloads with multi-tenant job scheduling



- Configure learning systems
- Flexible management experience
 - User Interface
 - CLI
 - APIs
 - Python SDK

IBM Watson Machine Learning

Embed Machine Learning & Deep Learning in your Business

Deploy and Manage Models

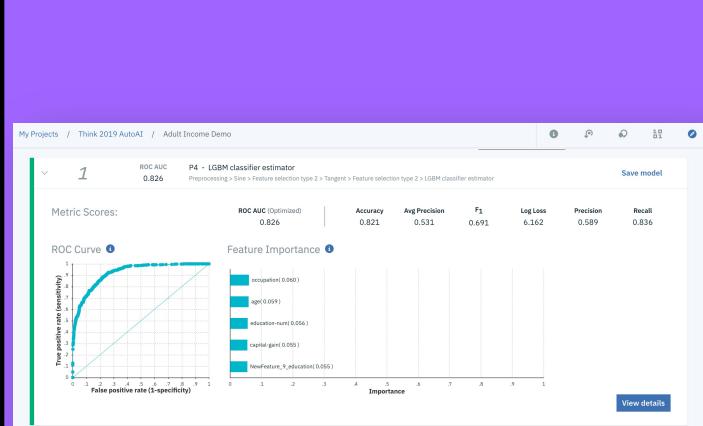
Move models to production, in an easy, secure, and compliant way

Intelligent Model Operations

Embed intelligent training services, with feedback loops that constantly learn from new data, regardless where it resides

Accelerate Compute Intensive Workloads

Distribute your deep learning Training and Hadoop/Spark workloads with multi-tenant job scheduling

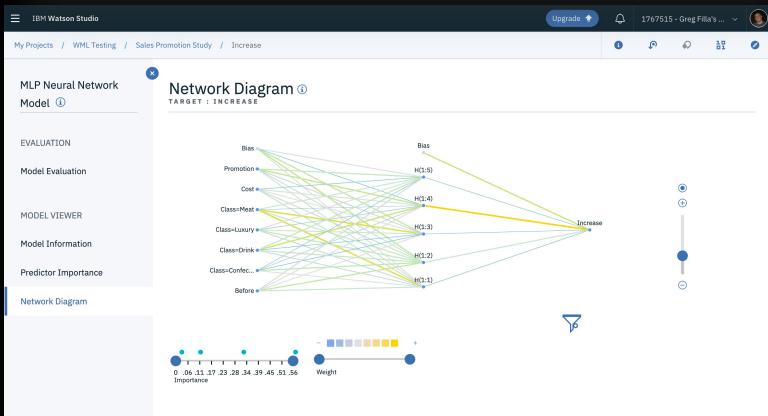
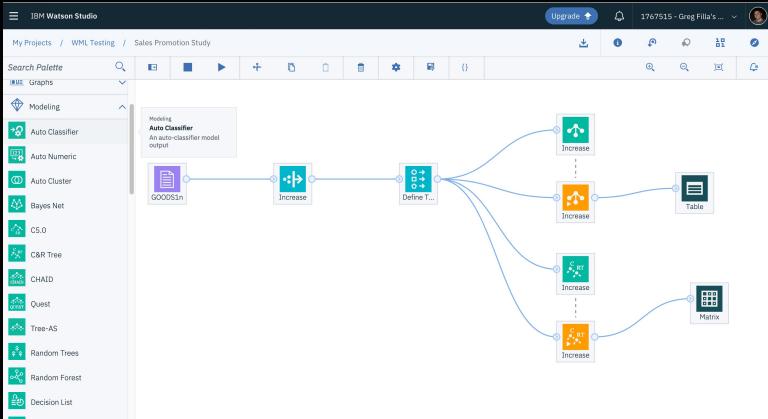


- Visually monitor training jobs
 - Experiments
 - AutoAI
- WML Accelerator for optimized Spark and GPU jobs
- Integration with Hadoop for in-place batch training and scoring

SPSS Modeler for Watson Studio

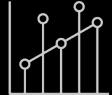
SPSS Modeler for Watson Studio

- SPSS Modeler is *embedded* in Watson Studio
- Watson Studio brings the best of both worlds by supporting legacy SPSS streams and also best in breed of open source
- Many new nodes in 2.0 – 28 for data preparation & 44 for Machine Learning. Includes Auto Data Prep and Auto Modeling, enhanced Data Visualization, Data Refinery integration and push to production options.



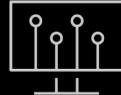
Decision Optimization for Watson Studio

Machine learning and optimization: better together



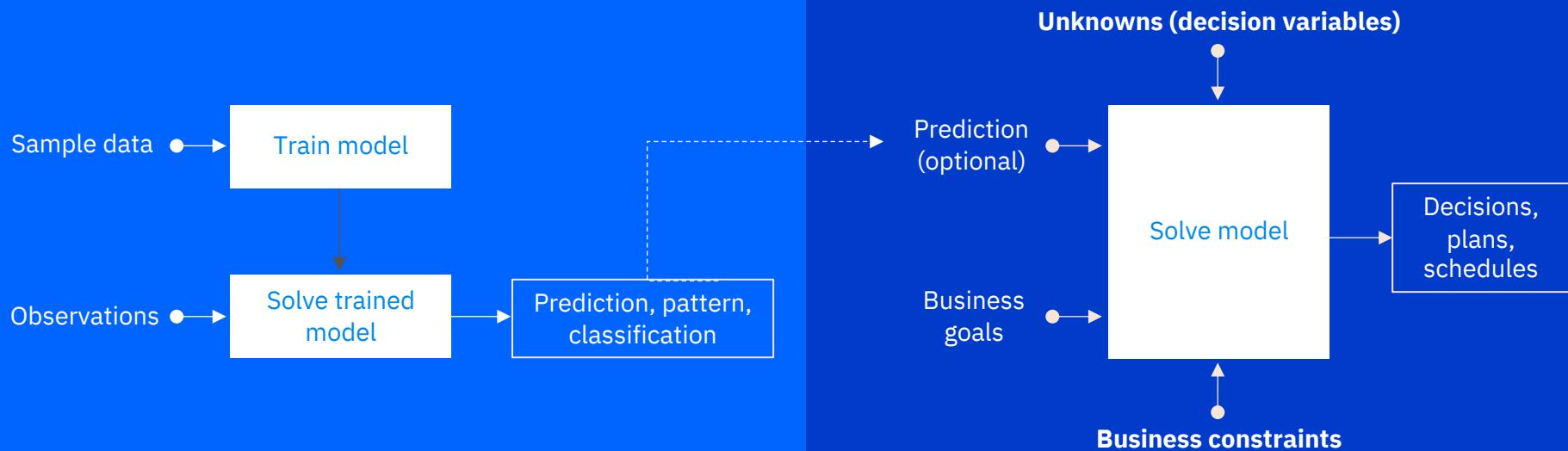
Predictive analytics

- Basic (supervised): You **know the answer**, and you **train the machine how to find it**.
- Advanced: Unsupervised, reinforcement, deep learning



Prescriptive analytics

- You **don't know the answer**, and you **provide the machine the logic on what is a good and a bad solution**.
- Advanced: Robust, stochastic, etc



AutoAI

Watson Studio AutoAI

- Integrated with **Watson Studio** and **Watson Machine learning**
- Automatically ingest, clean, transform, and model with data prep, model selection, feature engineering, and hyper-parameter optimization
- Training feedback visualizations provide real-time results to see model performance
- Binary, Multiclass, and Regression support
- One-click deployment to Watson Machine Learning

IBM Watson Studio

My Projects / AutoAI Playpen / Demo with Elena

Upcoming

Completed 15 minutes

SOURCE TABLE: titanic_train.csv

PREDICTION COLUMN: Survived

PREDICTION TYPE: Classification

OPTIMIZED METRIC: ROC AUC

Classifying Titanic Survivors

Stop run

Gradient boosting classifier

Hyperparameter Optimization

Feature Engineering

Extra trees classifier

Hyperparameter Optimization

LGBM classifier

Hyperparameter Optimization

Feature Engineering

Model Selection

Read Dataset

Split holdout data

Read training data

Preprocessing

P1

P2

P3

P4

P5

P6

P7

P8

P9

P10

P11

P12

Pipeline leaderboard

RANK	ROC AUC	Pipeline Information	Compare models	Ranking based on:	ROC AUC
> 1	0.860	P11 - LGBM classifier estimator Transformers (6): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > LGBM classifier estimator	View details	Save model	
> 2	0.852	P9 - LGBM classifier estimator Transformers (2): Preprocessing > LGBM classifier estimator	View details	Save model	
> 3	0.852	P10 - Extra trees classifier estimator Transformers (2): Preprocessing > Extra trees classifier estimator	View details	Save model	
> 4	0.851	P7 - Extra trees classifier estimator Transformers (6): Preprocessing > Cosine > Feature selection type 3 > Standard scaler > Feature selection type 1 > Extra trees classifier estimator	View details	Save model	
> 5	0.839	P5 - Extra trees classifier estimator Transformers (2): Preprocessing > Extra trees classifier estimator	View details	Save model	
> 6	0.839	P6 - Extra trees classifier estimator Transformers (2): Preprocessing > Extra trees classifier estimator	View details	Save model	
> 7	0.838	P1 - Gradient boosting classifier estimator Transformers (2): Preprocessing > Gradient boosting classifier estimator	View details	Save model	
> 8	0.838	P2 - Gradient boosting classifier estimator Transformers (2): Preprocessing > Gradient boosting classifier estimator	View details	Save model	
> 9	0.831	P4 - Gradient boosting classifier estimator Transformers (6): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > Gradient boosting classifier estimator	View details	Save model	
> 10	0.826	P3 - Gradient boosting classifier estimator Transformers (6): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > Gradient boosting classifier estimator	View details	Save model	
> 11	0.744	P12 - LGBM classifier estimator Transformers (6): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > LGBM classifier estimator	View details	Save model	
> 12	0.500	P8 - Extra trees classifier estimator Transformers (6): Preprocessing > Cosine > Feature selection type 1 > Standard scaler > Feature selection type 1 > Extra trees classifier estimator	View details	Save model	

Benefits of Using AutoAI



Build models faster

Automate [data preparation](#) and model development



Find signal from noise

Auto-feature [engineering](#) makes it easy to extract more predictive power from your data



Jump the skills gap

[No coding?](#) No problem – get started with a couple clicks



Rank and explore models

Quickly compare [candidate pipelines](#) to find the best model for the job



Discover more use cases

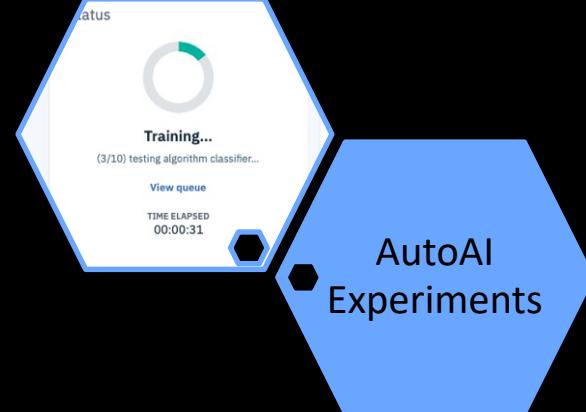
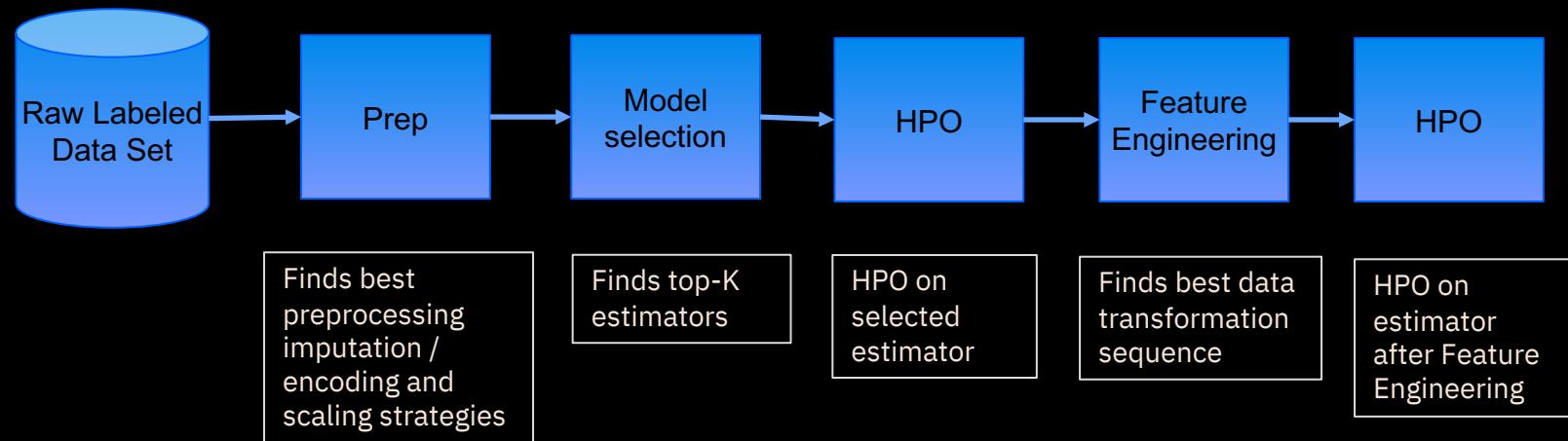
Supercharge [collaboration](#) with [AI everywhere](#) to disrupt & transform



Ready, set, deploy

Pipelines generated with AutoAI can be deployed to REST APIs with [one click](#)

AutoAI: How does it work?



Watson OpenScale

IBM Watson OpenScale

Automate & Operate AI at Scale

Production monitoring for compliance

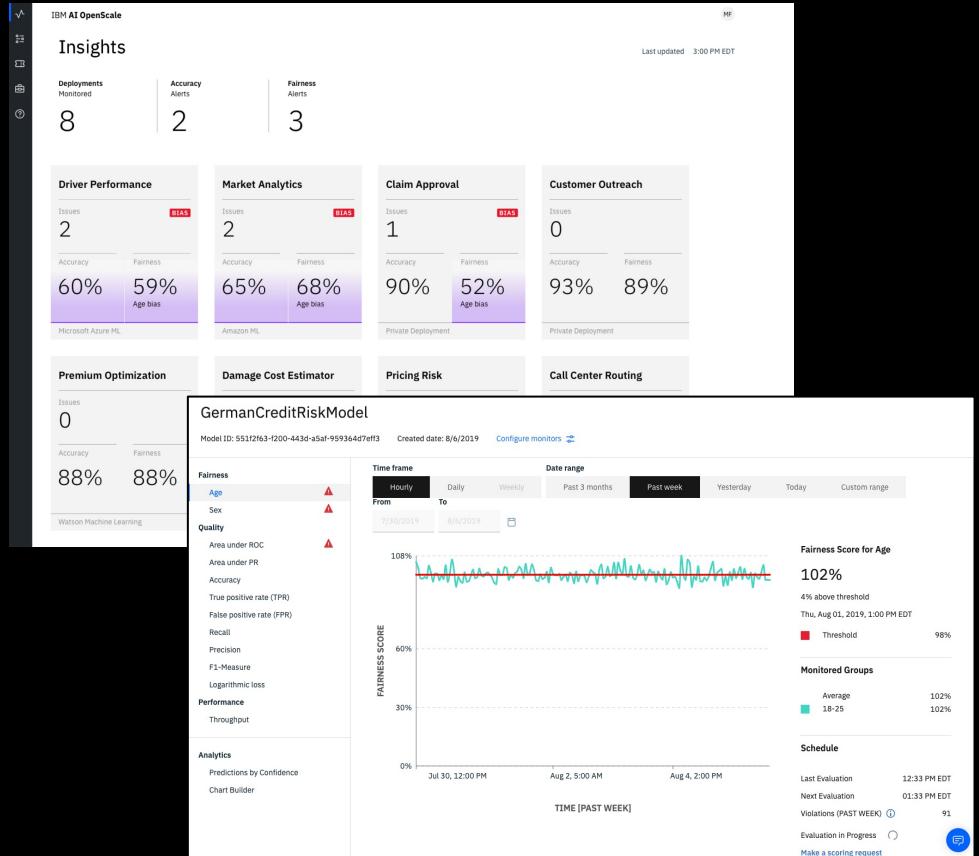
Detect and mitigate model bias; audit and explain model decisions

Ensure models resiliency in changing situations

Detect drift in data and anomaly in model behavior; specify inputs and triggers to model lifecycle

Align model performance with business outcomes

Correlate model metrics and business KPIs to measure business impact; actionable metrics and alerts



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

Why worry about AI and compliance?

A *leading bank* was **fined \$175M** for charging higher fees and interest rate to more than 30,000 minority borrowers.

A *leading credit card company* was **fined \$75M** for selling products to Spanish speaking customers at a much higher interest rate compared to others.

“... **50%** of our bonuses are linked to how we manage risk ...”
– VP, AI platforms, Large American bank

SR 11-7: Guidance on Model Risk Management



BOARD OF GOVERNORS
OF THE FEDERAL RESERVE SYSTEM
WASHINGTON, D.C. 20551



Ethics Guidelines for Trustworthy AI

Join AI Ethics Guidelines

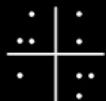


<http://www.fsb.org/wp-content/uploads/140407.pdf>



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

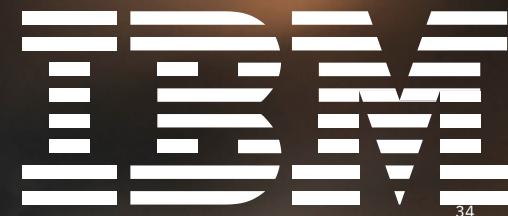
Detect drift during runtime

Generate specific model retraining inputs

Align model performance with business outcomes

Correlate model metrics and business KPIs

Actionable metrics and alerts



34

