AI in the Enterprise World

—

Donna Dillenberger
IBM Fellow
CTO Hybrid Cloud Research

AI is impacting every single industry

According to a Forrester report, organizations that scale AI are 7x more likely to be the fastest-growing businesses in their industry
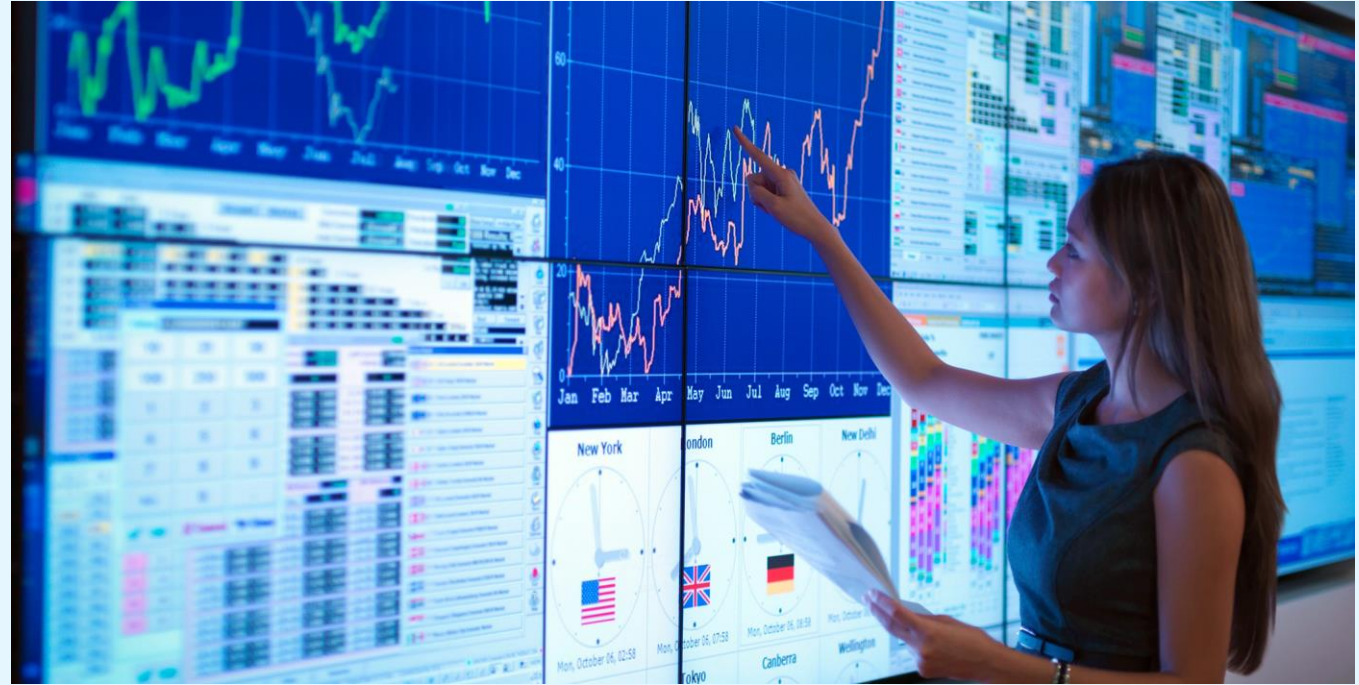
IBM z16 with

On-chip AI accelerator designed for high-speed, latency-optimized inferencing

### Speed to scale with transaction volume

Up to 3.5 Million AI inferences per second

1 millisecond response time,
19x higher throughput,
20x lower response time

Real-time insights when you need them

100% transaction AI scoring reduced fraud by $100M per year for a large bank

Enterprise AI -
Virtualization,
Energy reduction,
Cache coherency
Data Privacy, Auditing,
Provenance

Core

AI accelerator
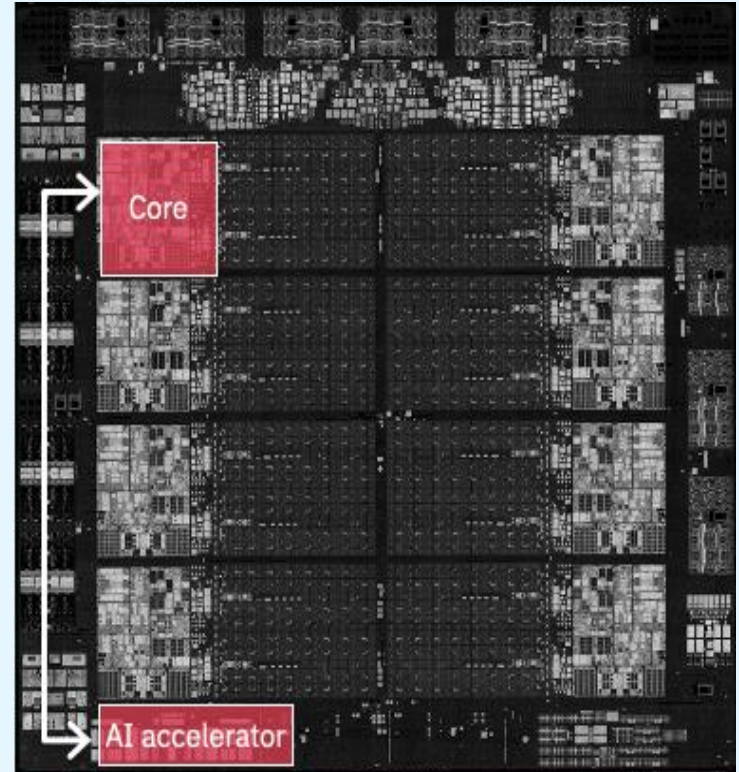
An optimized stack for AI
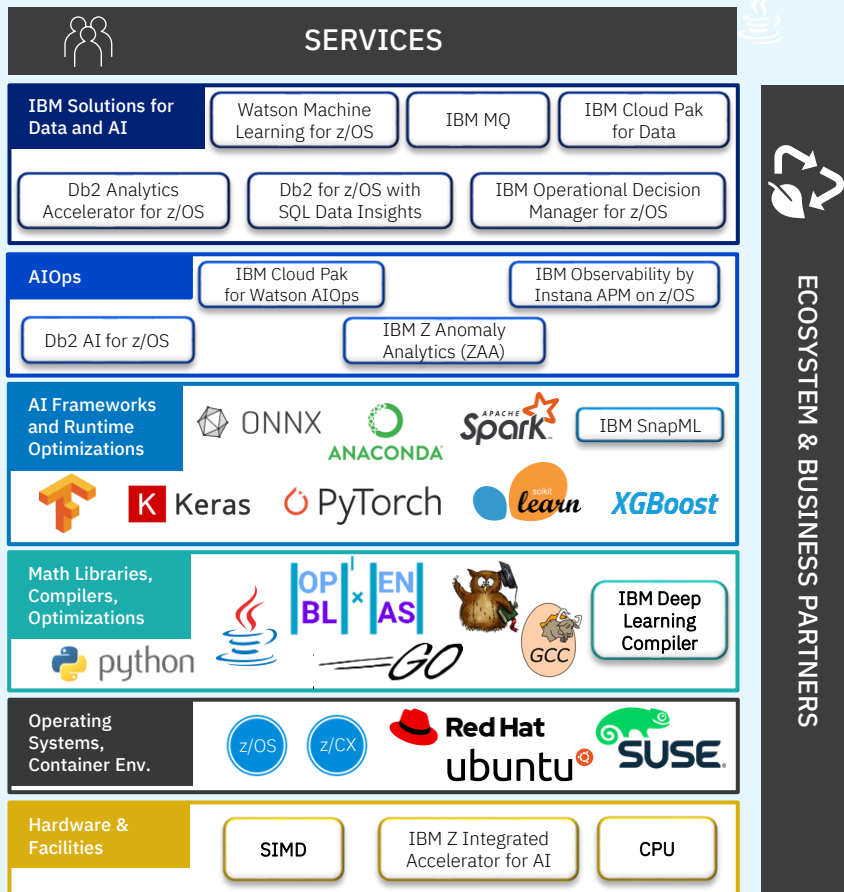
A holistic approach to AI:

- Enable popular open-source data science packages on IBM z16

- Optimize libraries, compiler, data transformation

- Use open source directly – or couple with IBM's AI offerings:

  - Watson Machine Learning
    - Collect, organize, analyze data
    - Infuse AI into your workloads.

  - Watson AIOps
    - Assess, diagnose and resolve incidents.

  - IBM Db2 13
    - A new approach to incorporating AI within Db2 for z/OS applications

Built in AI in the Database

- No data prep, copying, transformation required

- No data science skills required

- No humans needed to select features to train on

- Similarity

- Dissimilarity

- Explainability

- Bring AI to the data

- Easy AI

Trillions of transactions per day

Db2 for z/OS
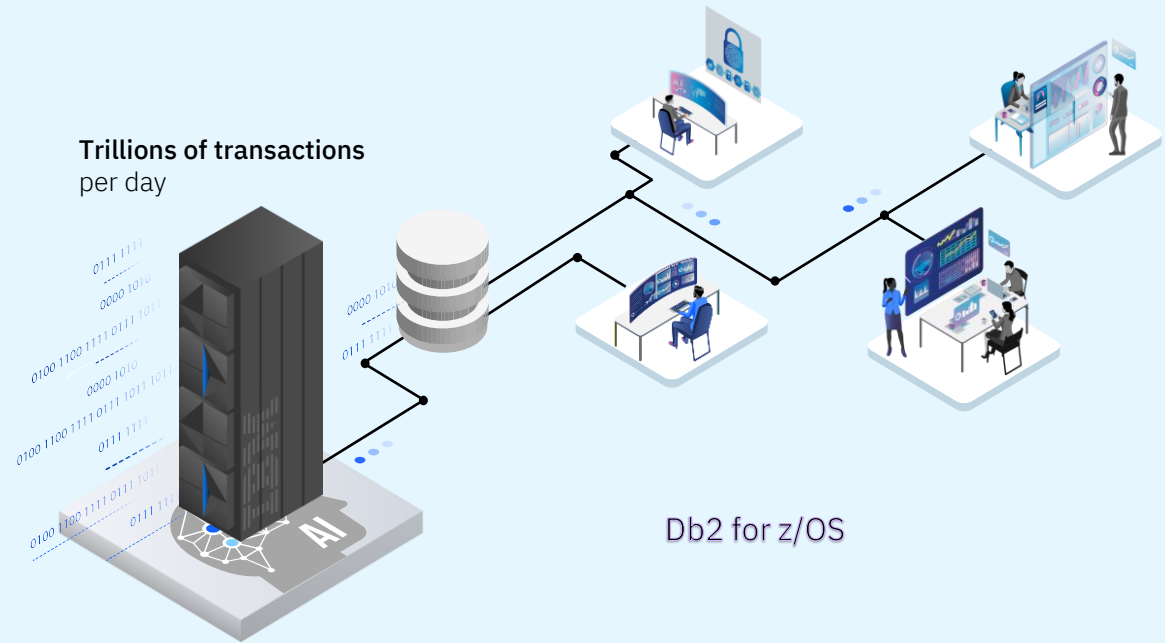
AI on IBM zSystems - Innovation delivered across the stack

AI at scale

INDUSTRY'S FIRST ENTERPRISE
AI ACCELERATOR
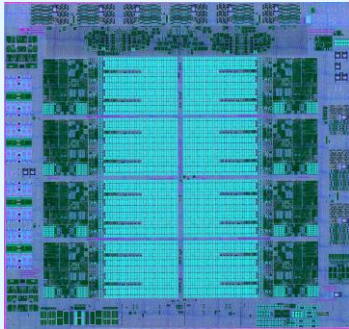
Game changing IBM z16 on chip inferencing
delivers up to 300 billion AI requests per day
with 1millesecon response time, 19x higher
throughput, 20x lower latency

AI Ecosystem

ROBUST ECOSYSTEM OPTIMIZED
FOR THE ACCELERATOR
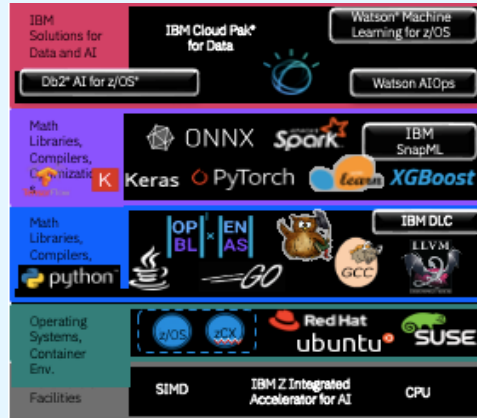
Data, transactions and accelerated AI in
IBM z16 effectively eliminates latency in
inferencing, with models built and trained with
familiar open-source tools

AI Agility

FRICTIONLESS AI

Train anywhere infer on IBM z16 and
embrace democratic AI

IBM zSystems AI in finance & banking

FRAUD DETECTION

Business Challenge
A large US bank was unable to score all transactions in real-time with an off-platform scoring engine. Causing 80% of transactions to go unscored

Business Impact
100% real-time scoring, saving > $20M in exposure risk per year

Capabilities Leveraged

CLEARING & SETTLEMENT

Business Challenge
A card processor needed to determine which trades and/or transactions have a high-risk exposure before settlement

Business Impact
Enhancement of rules-based approach with no impact to SLAs

Capabilities Leveraged
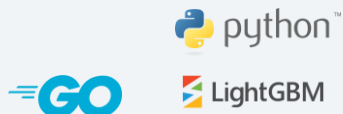
LOAN APPROVAL

Business Challenge
A large lender needed to speed up their consumer loan approval process, in order to maintain client satisfaction

Business Impact
Reduction in latency of over 1000X, with corresponding decrease in lender risk and exposure

Capabilities Leveraged

CLAIMS FRAUD

Business Challenge
A state government in the US realized that their process to determine fraudulent claims was manual & intensive and could not scale, taking up to 40 hours per case

Business Impact
Now able to process claims and detect fraud in under 1 minute per case and allocate resources to higher value tasks

Capabilities Leveraged

---

python™

GO

LightGBM

z/OS

zCX

TensorFlow

z/OS

IBM ODM

Watson Machine Learning for z/OS

z/OS

Watson Machine Learning for z/OS

IBM z16 Platform AI Stack

IBM Telum™   CPU

IBM Z Integrated Accelerator for AI

SIMD

z/OS

zCX

Red Hat

SUSE

ubuntu

OP|EN BL|×|AS

python™

GO

IBM Deep Learning Compiler

TensorFlow

ONNX

Keras

ANACONDA

Spark

IBM Cloud Pak for Data

Watson Machine Learning for z/OS
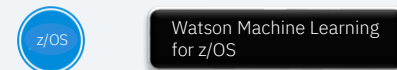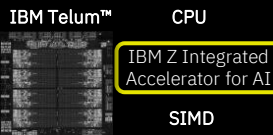
IBM zSystems AI in other industries

MEDICAL IMAGES

Business Challenge
A Health Insurer needs to analyze large volumes of medical records, including training computer vision models, while optimizing for energy efficiency

Business Impact
Replace x86 server environments with LinuxONE to reduce energy consumption 50-80%

Capabilities Leveraged

IBM Deep Learning Compiler
Integrated Accelerator for AI
ONNX
PyTorch
IBM LinuxONE

GEOSPATIAL ANALYSIS & Natural Language Processing

Business Challenge
A European land registry needs to determine which buildings have been added to, modified, or demolished for land surveys and tax purposes. Also embedding NLP into chat services.

Business Impact
Deploy on IBM z16 to ensure security of data, and optimization of IT and operational costs

Capabilities Leveraged

z/OS
Watson Machine Learning for z/OS
HUGGING FACE
zCX
ONNX
RASA

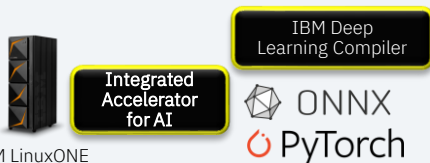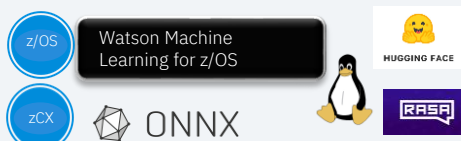## FEDERATED LEARNING

Business Challenge
Organized Retail Crime costs over $3.5B per year to US retailers, who lack the ability to fully model for the risk as they only have the data for their own business.

Business Impact
Federated learning enables sharing of models, without exposing sensitive data, to catch the bad actors

Capabilities Leveraged

IBM HPVS
Watson Federated Learning
IBM Cloud Pak for Data
Collect Data   Organized Data   Analyze Data   Infuse AI

DATA REDACTION

Business Challenge
Due to banking secrecy laws and slow post-processing performance, a major bank was unable to send diagnostic data to vendors to effectively maintain operational stability for their production environment.

Business Impact
Automatic redaction of PI & SPI, resulted in a 47x performance improvement in post-processing speeds

Capabilities Leveraged

z/OS
Data Privacy for Diagnostics

IBM z16 Platform AI Stack

IBM Telum
CPU
IBM Z Integrated Accelerator for AI
SIMD

z/OS
zCX
Red Hat
SUSE
ubuntu

OP|EN BL|AS
GO
python
Java

IBM Deep Learning Compiler
TensorFlow
ONNX
Keras
ANACONDA
Spark APACHE

IBM Cloud Pak for Data
Collect Data   Organized Data   Analyze Data   Infuse AI
Watson Machine Learning for z/OS
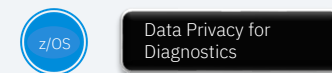
AI Flexibility with IBM zSystems
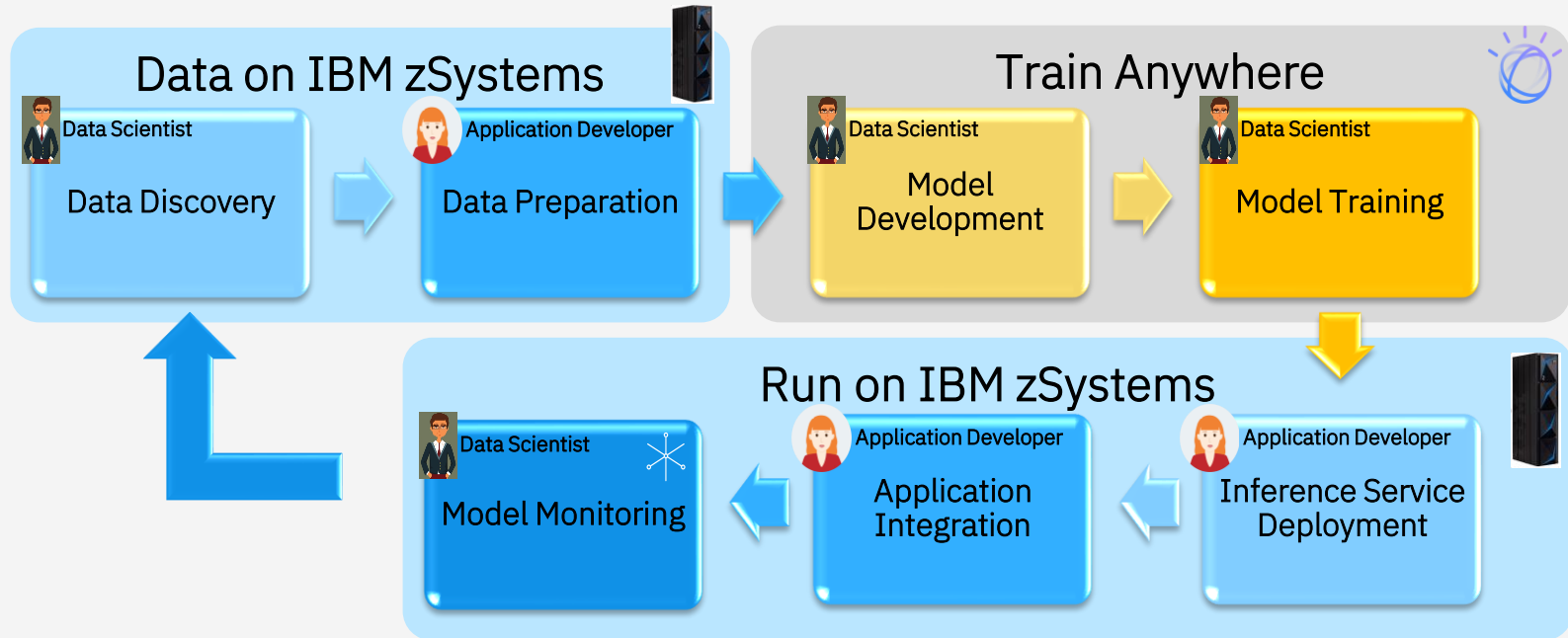
Demo – Secure AI on IBM zSystems for Fraud Detection

Business Challenge
A Large Retail company is unable to score all transactions in real-time with an off-platform scoring engine. Causing 80% of transactions to go unscored

Business Impact
100% real-time scoring, saving > $20M in exposure risk per year

Input Data:
Financial transactions

AI prediction required:
Which of these transactions are fraudulent ?

DB2/zOS

Core
Payments

Retail Payments

Fraudulent
Payment?

Production
AI Models

zOS Production

IBM zSystem –

Securely creating an AI model

Hyper Protect Data Controller
Data Policy
Encryption Keys

Watson Knowledge Catalog

Testing Data
Training Data
Testing Data

DB2/zOS

Model Training

Core Payments

Retail Payments

Fraudulent Payment?

Production AI Models

Watson Machine Learning /zOS

Create, Monitor, Manage, Update AI models

Test

Production

IBM zSystem – Securing Data with Hyper Protect Data Controller
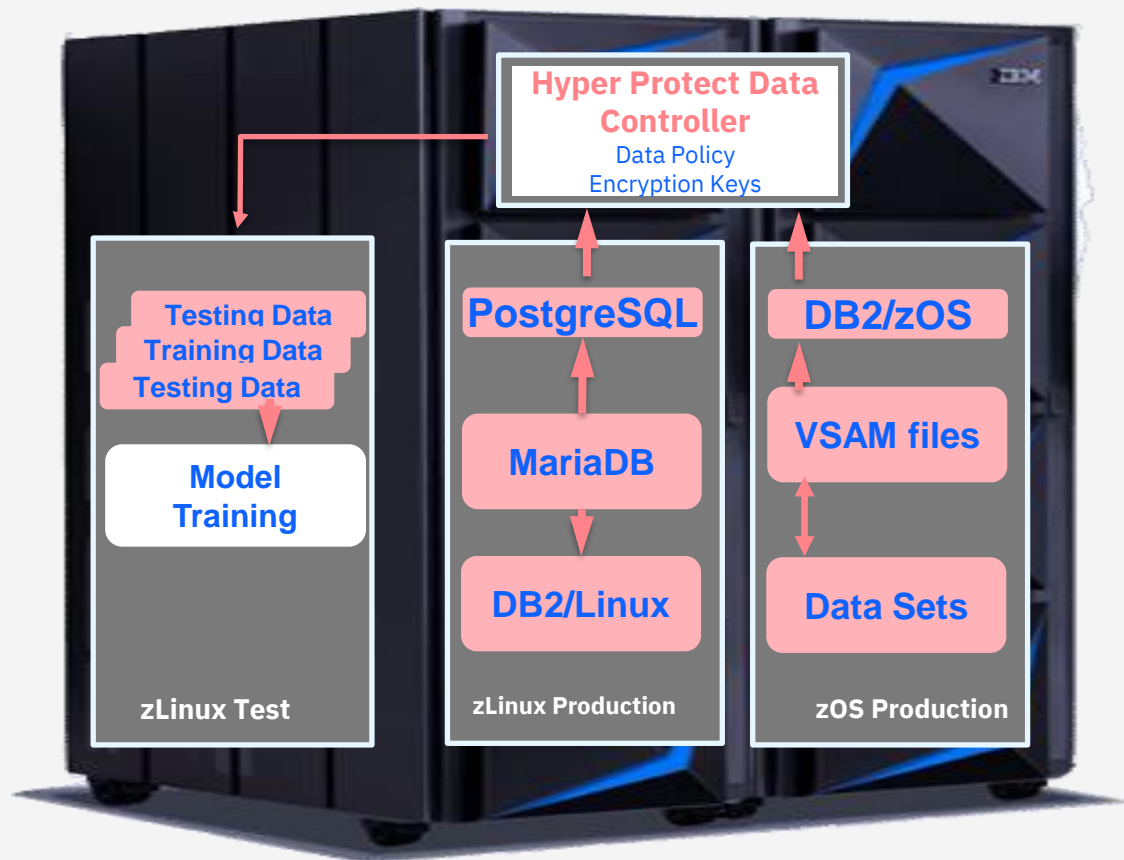
Data Privacy with Hyper Protect Data Controller

- Encrypt individual sensitive fields as **_Encrypted Data Objects (EDOs)_**

- Provide different masking or encryption for fields based on specific user

- Provides audits of who used the data

- Protects data when it leaves zSystem

- Can revoke access to Encrypted Data objects remotely

Hyper Protect Data Controller *

Encrypted Data

DBMS Table

Data Scientist

Data User
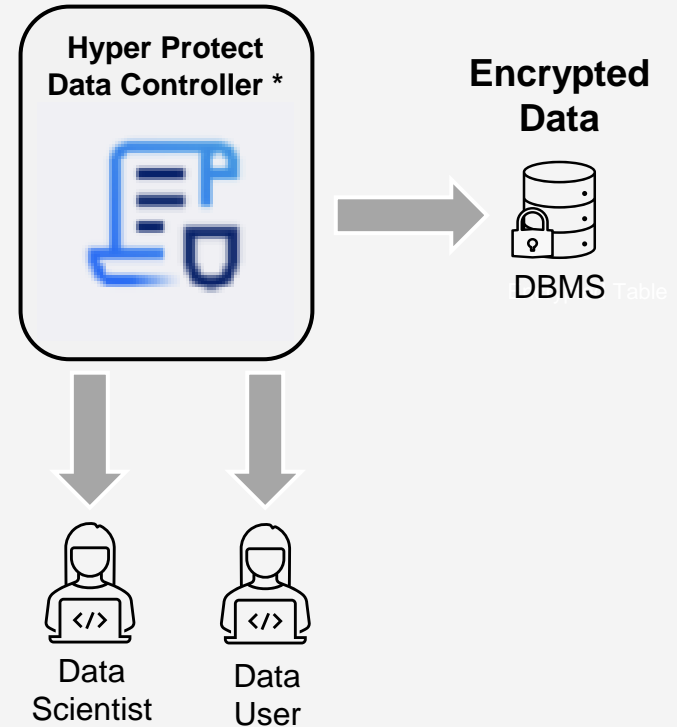
IBM Watson Knowledge Catalog
Cuts time to automate data discovery, quality and governance by up to 90%
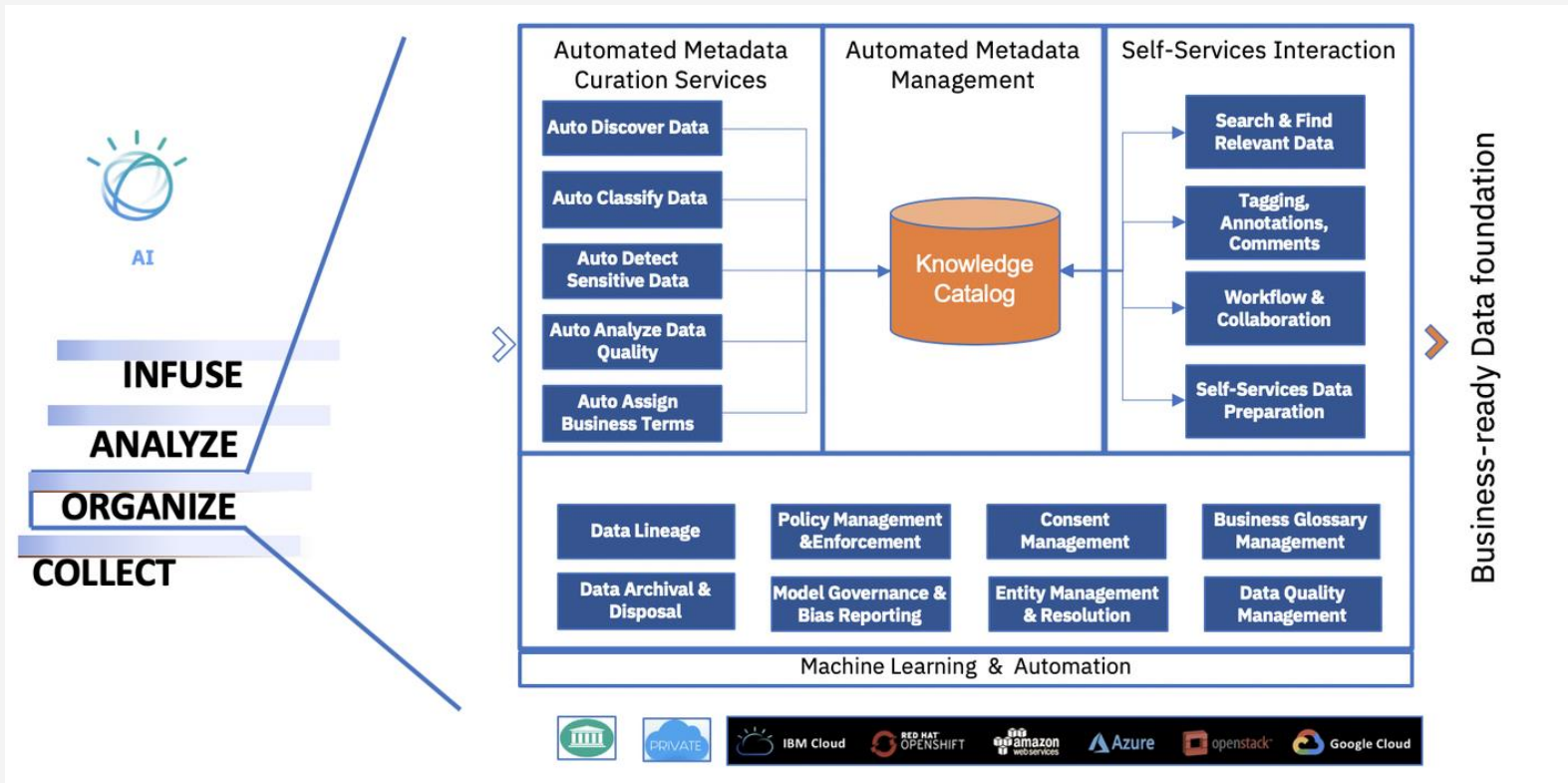
AI Ecosystem:
Seamlessly leverage AI on IBM zSystems

- Bring machine learning & deep learning models to IBM zSystems with the Deep Learning Compiler
- Exploit IBM Integrated Accelerator for AI for best inference performance.

Build and train model in any popular framework on any platform of your choice

Use ONNX, an opensource tool for framework interoperability Models are converted to the ONNX interchange format

The IBM DLC (Deep Learning Compiler), optimized for performance and new libraries, generates a program from the model for execution on z/OS or Linux® on IBM zSystems

Deploy on IBM zSystems and IBM LinuxONE and infuse model into workload application

Optimizing Machine Learning: SNAP Machine Learning

Snap Machine Learning: An open source optimized machine learning library

- Accelerates training/inference of popular ML algorithms
- Scales to very large datasets
- Maintains or improves accuracy
- Utilizes resource efficiently
- Offers popular scikit-learn API – seamless to the user

Linear Regression	Decision Trees
Logistic Regression	Random Forests
SVM	Boosting Machines

on IBM z16

on-Prem

On IBM Cloud

Demo – Secure AI on IBM zSystems for Retail Prediction

Business Challenge
Retail companies would like to offer customers a personalized list of additional services and products that customers would be interested in

Business Impact
General product recommendations have a very low hit ratio with users. Better predictions of what customers would purchase would lead to better inventory management and future product offerings

Input Data:
Shopping transactions, travel trips taken, travel services bought

AI prediction required:
What future trips, travel products and services would the user be interested in

DB2/zOS

Retail
Application

Purchases

Individualized
Offers

Production
AI Models

zOS Production

IBM zSystem –

Securely creating an AI model

Hyper Protect Data Controller
Data Policy
Encryption Keys

Watson Knowledge Catalog

Testing Data
Training Data
Testing Data

DB2/zOS

Model Training

Retail Application

Customer Purchases

Individualized Offers

Production AI Models

Watson Machine Learning /zOS

Create, Monitor, Manage, Update AI models

Test

Production

Demo – Inventory Management

Business Challenge
Businesses must determine a price for their products and services. This is difficult if the product or service is new or if the marketplace is changing and existing product prices change

Business Impact
Pricing too high may reduce volume purchases. Pricing too low would give up profit.

Input Data:
Brand, product name, and description of the product or service. Comparative product prices.

AI prediction required:
Suggested price

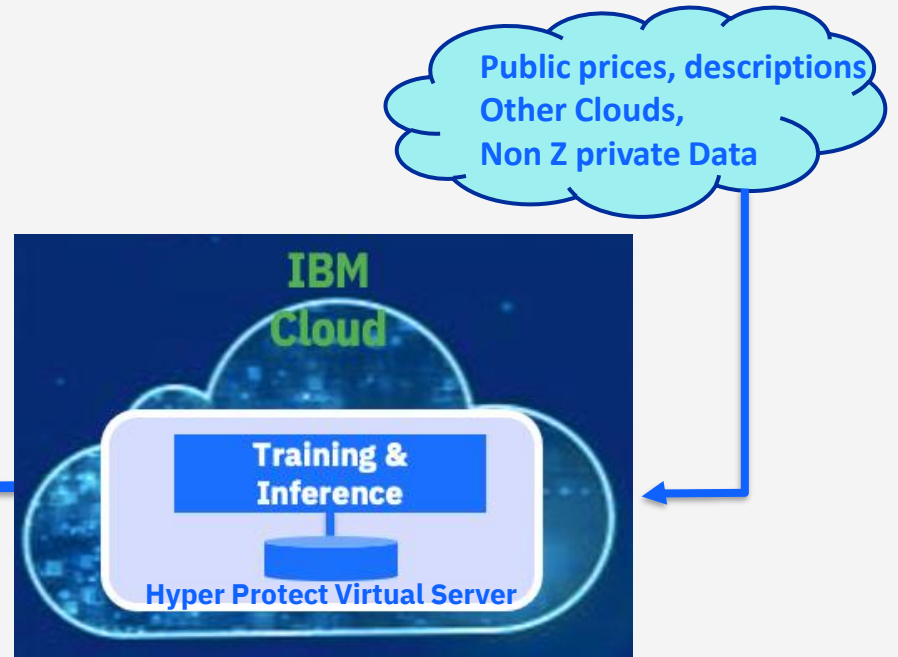# IBM zSystem with Hybrid Cloud

Public prices, descriptions
Other Clouds,
Non Z private Data

DB2/zOS

Inventory Economics

zOS Production

Suggested Price of product

IBM Cloud

Training & Inference

Hyper Protect Virtual Server

AI Protection with IBM Hyper Protect Virtual Server

Build with security

Data scientist can securely train their models a trusted environment with integrity.

Train model with trust

Software in virtual Machines and containers created with secure, signed, auditable Build process

Confidential Computing

Administrators are not able to access any applications or data, for example: AI models and AI data. All data is automatically encrypted, no application changes required.

Highest security in the Industry

Enforced with security hardware certified with Federal Information Processing Standard 140-2, level 4

Secure Service Container

Solution /Application

Application Interface

Management Backend

Base Operation System
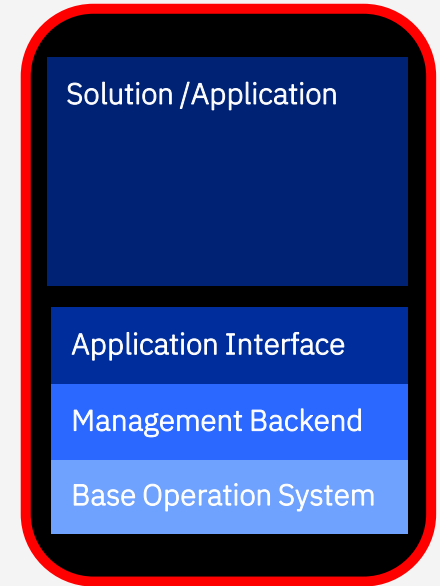
Federated Artificial Intelligence
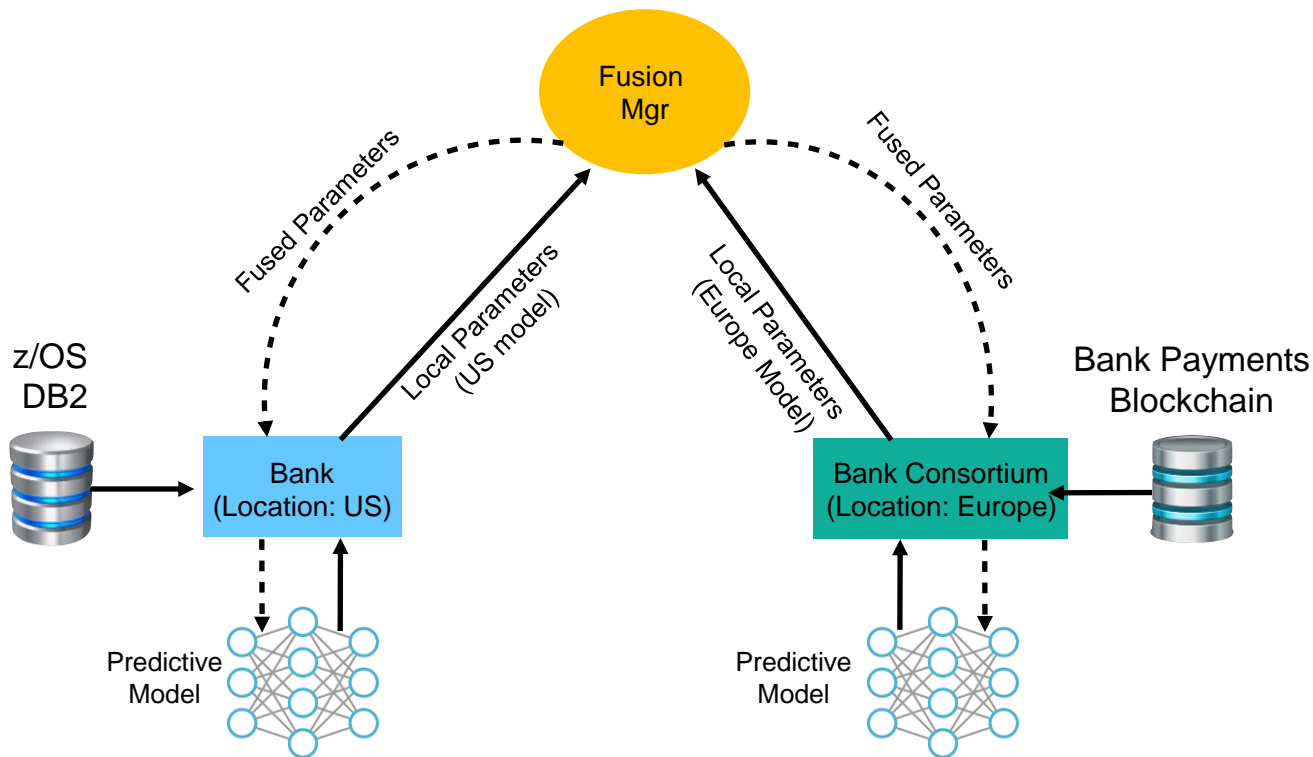
Benefits of Federated AI

Integrate the AI Insights from data found in different systems, formats and locations
- Secure data on premise IBM Z, Power, x86
- Public data in the cloud
- Data in different countries

No Need to move the data
- Move the AI insights instead
- 5x improvement in network transit and AI Modelling time
- Keep private data private

More Accurate and Powerful AI models

Regulatory Compliance

AI Explainability
AI Bias Detection
AI Transparency
AI Auditability

Summary

AI on IBM z16
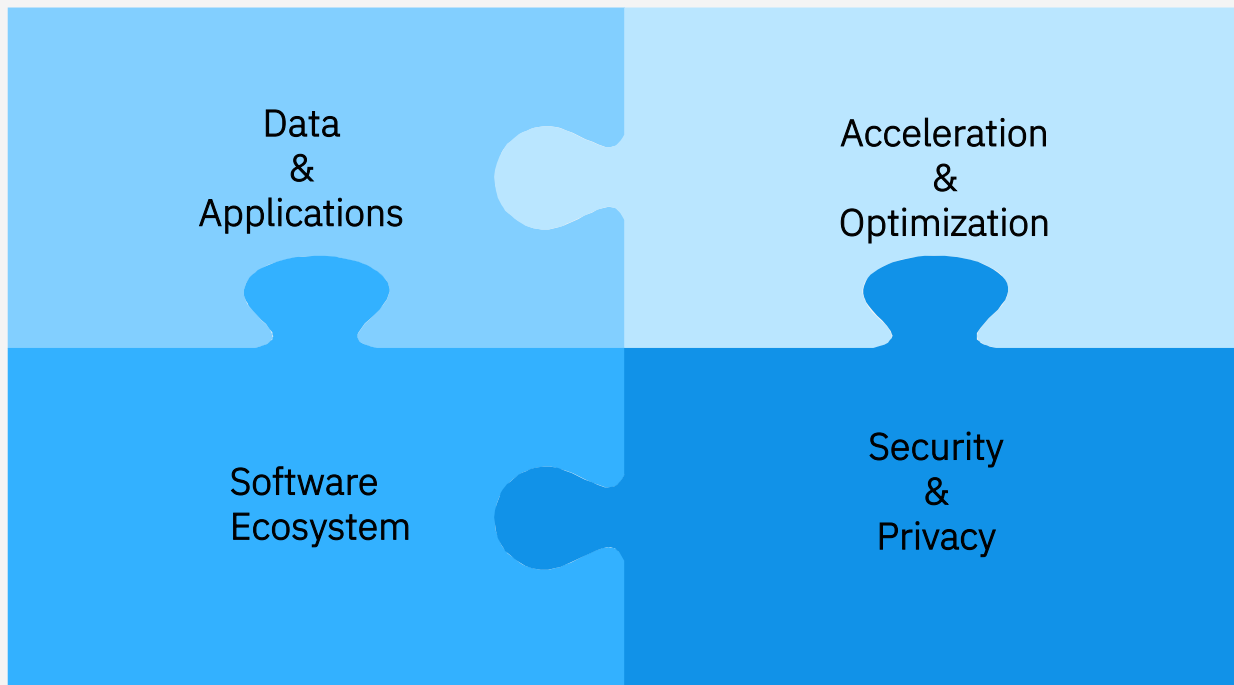
Data & Applications

Acceleration & Optimization

Software Ecosystem

Security & Privacy

Fraud Detection

◑ Before   ○ Installation   ○ Data Security Setup   ○ Data Protection   ○ Training   ○ After

Transactions in the last 60 Seconds
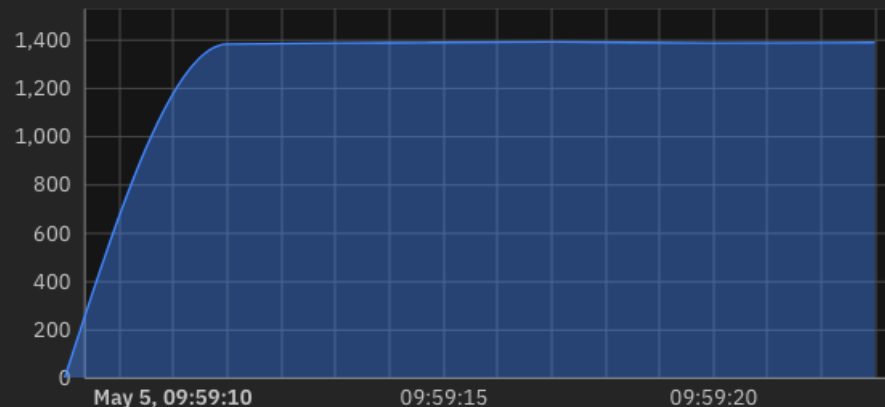
451,484	0
Total Transactions	Total Fraud Transactions

■ Fraud Transaction ■ Normal Transaction

Amount	Bank Account Number	Receiver ID	Sender ID	Timestamp	Transaction ID
$3588.29	DE4875000064899001	FOR62219753	61643	Thu, 05 May 2022 13:59:23 GMT	451472
$816.85	DE4875000048142457	FOR1031047	16707	Thu, 05 May 2022 13:59:23 GMT	451473
$301.95	DE4875000099027765	10099	6363	Thu, 05 May 2022 13:59:23 GMT	451474

Fraud Detection

Before ● Installation ○ Data Security Setup ○ Data Protection ○ Training ○ After

```
pip install snapml
```

Fraud Detection

	Watson Knowledge Catalog		Hyper Protect Data Controller	
Column Name	Auto Classification		Data Scientist	Data Owner
amount	Amount		Clear	Clear
iban	Bank Account Number		Clear	Clear
is_fraud	Is Fraud		Clear	Clear
receiverid	Receiver ID		**Hash**	Clear
senderid	Sender ID		**Hash**	Clear
time_	Time		Clear	Clear
transactionid	TransactionID		Clear	Clear

Fraud Detection

● Before ● Installation ● Data Security Setup ◐ Data Protection ○ Training ○ After

Amount	Bank Account Number	Is Fraud	Receiver ID	Sender ID	Time	TransactionID
768.63	DE4875000039296682	0	r3NrVkedMxr5r9uWENgXLvIbVF4GfV/QAhZoIvMm78k=	BWa3WqxKo085/54F5ol66vk6AtEKhZK6dsJ9YTOTyyg=	Sun, 25 Jul 2010 21:23:00 GMT	229210
749.62	DE4875000043645278	0	nvAXayUDc/0u1m3Ec4uJVNTlXws1IZ0nfqxGYogQC2s=	d4v7nbkvKhWWD6N0fDm0v2o8+B164VWKlh2ipgHF0IU=	Tue, 20 Sep 2011 14:14:00 GMT	263743
10700.891224773593	DE4875000040317075	0	2EBonXp+ra8wPjaS614pSnUv723a/ZsMSe7YJKs9Mxg=	6Fn/s6XTztmFYOKlS0CBHT6Nxhidyxbfhso0UwYZI6M=	Tue, 08 Nov 2011 21:39:00 GMT	465277
76.92	DE4875000022965185	0	d14YBo1uMKOmrEn5tYaGJse4/QU7DosxOPoLryyBlyM=	J6gn0b2Hnn4TF5G8m36N8iepCC+snytIuULp0HAqQyc=	Fri, 19 Aug 2011 19:02:00 GMT	501175
3841.8871324905026	DE4875000027390674	0	JopgF0byuCSoX57boE/4gQfTOrqRNifCKh9G/t+74+E=	+obGOw1K/KOJHQmGRwxf41cMd+QF8GJedXnjP1ixPds=	Sat, 06 Aug 2011 15:07:00 GMT	507178

| Data Scientist | Data Owner |

Fraud Detection

● Before   ● Installation   ● Data Security Setup   ◐ Data Protection   ○ Training   ○ After

Amount	Bank Account Number	Is Fraud	Receiver ID	Sender ID	Time	TransactionID
768.63	DE4875000039296682	0	FOR4669997	7821	Sun, 25 Jul 2010 21:23:00 GMT	229210
749.62	DE4875000043645278	0	15371	73473	Tue, 20 Sep 2011 14:14:00 GMT	263743
10700.891224773593	DE4875000040317075	0	9357	53589	Tue, 08 Nov 2011 21:39:00 GMT	465277
76.92	DE4875000022965185	0	7736	12541	Fri, 19 Aug 2011 19:02:00 GMT	501175
3841.8871324905026	DE4875000027390674	0	40601	7108	Sat, 06 Aug 2011 15:07:00 GMT	507178

| Data Scientist | Data Owner |

Fraud Detection

10,399,888

Total Transactions in
training dataset

0%

Percent
Fraudulent

Train Model

0

Fraudulent Transactions

0%

Training
Accuracy

Fraud Detection

Before   Installation   Data Security Setup   Data Protection   Training   After

10,399,888	0.93%
Total Transactions in training dataset	Percent Fraudulent

Train Model

96,407	17.9%
Fraudulent Transactions	Training Accuracy

Fraud Detection

Before • Installation • Data Security Setup • Data Protection • Training • After

10,399,888

Total Transactions in training dataset

2.24%

Percent Fraudulent

Train Model

232,669

Fraudulent Transactions

43.2%

Training Accuracy

Fraud Detection

10,399,888

Total Transactions in
training dataset

5.04%

Percent
Fraudulent

Train Model

524,045

Fraudulent Transactions

97.3%

Training
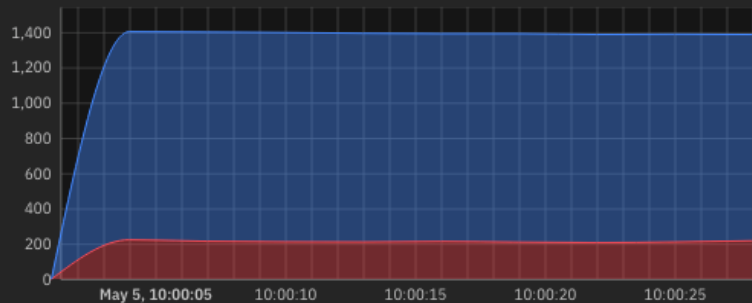Accuracy

Fraud Detection

● Before   ○ Installation   ○ Data Security Setup   ○ Data Protection   ○ Training   ◑ After

Transactions in the last 60 Seconds

452,985	**73,135**
Total Transactions	Total Fraud Transactions

■ Fraud Transaction ■ Normal Transaction

Amount	Bank Account Number	Receiver ID	Sender ID	Timestamp	Transaction ID	Is Fraud	Probability
$741.31	DE4875000027591130	FOR22532510	88261	Thu, 05 May 2022 14:00:28 GMT	452973	0	0.9938590589235534
$1474.13	DE4875000032450825	29491	90008	Thu, 05 May 2022 14:00:28 GMT	452974	0	0.9798126438808726
$1293.66	DE4875000082813794	79108	32972	Thu, 05 May 2022 14:00:28 GMT	452975	0	0.985230472933417
$9817.54	DE4875000029554035	69400	58717	Thu, 05 May 2022 14:00:28 GMT	452976	0	0.9749637201558711
$24.39	DE4875000093666235	43800	94881	Thu, 05 May 2022 14:00:28 GMT	452977	0	0.9718597669118737
$406.64	DE4875000043660768	FOR89920671	37343	Thu, 05 May 2022 14:00:28 GMT	452978	1	0.800275284553588
$1054.20	DE4875000076550523	FOR83531557	55306	Thu, 05 May 2022 14:00:28 GMT	452979	0	0.5508500562806968
$681.87	DE4875000054284443	10491	46617	Thu, 05 May 2022 14:00:28 GMT	452980	1	0.6464877521850038
$721.88	DE4875000044878406	92503	1974	Thu, 05 May 2022 14:00:28 GMT	452981	1	0.6763130688419042

Trip Classification

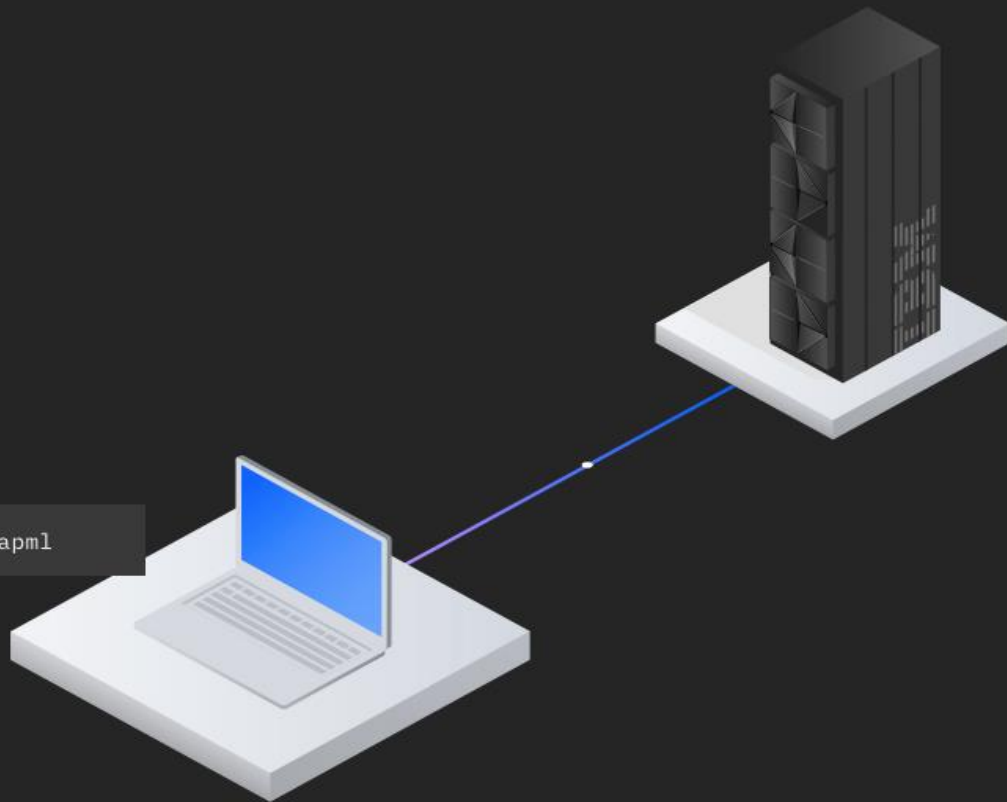Before | Installation | Data Security Setup | Data Protection | Training | After

Helen the hiker

```
pip install snapml
```

Trip Classification

● Before ● Installation ◑ Data Security Setup ○ Data Protection ○ Training ○ After

Watson Knowledge Catalog   Hyper Protect Data Controller

Column Name	Auto Classification	Data Scientist	Data Owner
date_	Date	Clear	Clear
o_customer_sk	Customer Surrogate Key	**Hash**	Clear
o_order_id	OrderID	Clear	Clear
trip_type	TripType	Clear	Clear
weekday	Weekday	Clear	Clear

Trip Classification

● Before   ● Installation   ● Data Security Setup   ◑ Data Protection   ○ Training   ○ After

Orders   Products   Line Items

Date	Customer Surrogate Key	OrderID	TripType	Weekday
Fri, 14 Jan 2011 00:00:00 GMT	gxdqDEkqscqlGx33IQ7Aga+pXo9BgtPF3 AUfXCXW3Gs=	1	Winter Sports	Friday
Fri, 09 Dec 2011 00:00:00 GMT	OYbpEQfynzc2EQJDyo+jA9SSlhrL7MJrIp 7wUy+ePr4=	2	City Tours	Friday
Wed, 23 Jun 2010 00:00:00 GMT	KXfexqCCfoW0pIvI7lzMkubfGLLBSnlXSG 4A+8C9rmc=	3	City Tours	Wednesday
Tue, 05 Oct 2010 00:00:00 GMT	t13mtbW5lCTMWAdzYN7Gs8hsYCqb0l+ SuXeKlSplnfM=	4	City Tours	Tuesday
Wed, 10 Mar 2010 00:00:00 GMT	wSRcZfLokH0pDCAHbZwlPl72SFOseAW rboTsdi/Kw8Y=	5	Winter Sports	Wednesday

| Data Scientist | Data Owner |

Trip Classification

- Before
- Installation
- Data Security Setup
- Data Protection
- Training
- After

Orders   Products   Line Items

Date	Customer Surrogate Key	OrderID	TripType	Weekday
Fri, 14 Jan 2011 00:00:00 GMT	47123	1	Winter Sports	Friday
Fri, 09 Dec 2011 00:00:00 GMT	24934	2	City Tours	Friday
Wed, 23 Jun 2010 00:00:00 GMT	30834	3	City Tours	Wednesday
Tue, 05 Oct 2010 00:00:00 GMT	32547	4	City Tours	Tuesday
Wed, 10 Mar 2010 00:00:00 GMT	66159	5	Winter Sports	Wednesday

| Data Scientist | Data Owner |

Trip Classification

Before   Installation   Data Security Setup   Data Protection   Training   After

London calling.
Will you answer?

Book now.

0%

Training Accuracy

Train Model

Before
Installation
Data Security Setup
Data Protection
Training
After

Surf's up.
Are you down?

Catch the waves with us.

38.64%

Training Accuracy

Train Model

Feel like flying?

Discover a new way to bike. Book your adventure now.

68.82%

Training Accuracy

Train Model

Trip Classification

- Before
- Installation
- Data Security Setup
- Data Protection
- Training
- After

Discover
new heights.

book your next
hiking trip with us.

95.89%

Training Accuracy

Train Model

Trip Classification

Before • Installation • Data Security Setup • Data Protection • Training • After

Real Time Inferencing	User Experience

Before

After
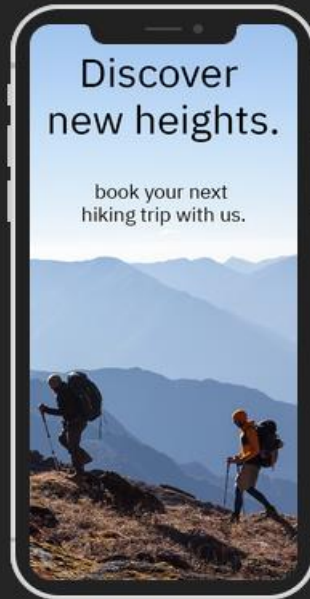
Helen the hiker
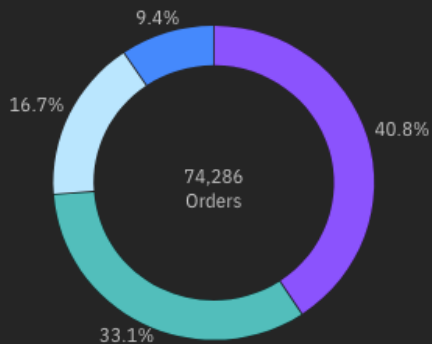
Trip Classification

Before ● Installation ● Data Security Setup ● Data Protection ● Training ● After ◐

| Real Time Inferencing | User Experience |

Real Time Orders by Trip Type

9.4%
16.7%
40.8%
33.1%

74,286
Orders

■ Winter Sports ■ Outdoor Adventures
■ Water Sports ■ City Tours

Customer Surrogate Key	OrderID	TripType	Weekday	Store	Probability
12923	75659	City Tours	Thursday	7	0.7288236894700603
91710	75658	City Tours	Tuesday	2	0.40213533884794167
12622	75657	City Tours	Wednesday	1	1
54740	75656	Winter Sports	Wednesday	11	0.1155403724399887
46809	75655	Winter Sports	Thursday	4	1
93441	75654	Outdoor Adventures	Wednesday	6	0.23690096676889727
24755	75653	Winter Sports	Thursday	2	1
13782	75652	Winter Sports	Wednesday	4	1
81811	75651	Winter Sports	Wednesday	11	1
70205	75650	Winter Sports	Saturday	7	1

Price Forecasting

$39.89

Retail Store A

$50.15

Retail Store B

$2.75

Our Store

- Before
- Installation
- Training
- After

Price Forecasting

- ● Before
- ● Installation
- ◑ Training
- ○ After

Description	Price
City Streets Super cute shirt Super cute Good fit for	26.72
Monarchy Men's monarchy size large tshirt. Graphic tee	35.00
Fabletics Fabletics black shirt top Fabletics new black crochet striped shirt top nwt size large	50.80
Oakley Oakley women's shirt size large	52.25
Deep Wavy Shirt Pepsi-esque type graphic. Slight fading in graphic but no cracking	41.51

$2.80

Our Store

0%

Training Accuracy

Train Model

Description	Price
City Streets Super cute shirt Super cute Good fit for	26.72
Monarchy Men's monarchy size large tshirt. Graphic tee	35.00
Fabletics Fablectics black shirt top Fabletics new black crochet striped shirt top nwt size large	50.80
Oakley Oakley women's shirt size large	52.25
Deep Wavy Shirt Pepsi-esque type graphic. Slight fading in graphic but no cracking	41.51

$16.47

Our Store

35.25%

Training Accuracy

Train Model

Price Forecasting

● Before   ● Installation   ◑ Training   ○ After

Description	Price
City Streets Super cute shirt Super cute Good fit for	26.72
Monarchy Men's monarchy size large tshirt. Graphic tee	35.00
Fabletics Fablectics black shirt top Fabletics new black crochet striped shirt top nwt size large	50.80
Oakley Oakley women's shirt size large	52.25
Deep Wavy Shirt Pepsi-esque type graphic. Slight fading in graphic but no cracking	41.51

$29.10

Our Store

69.25%

Training Accuracy
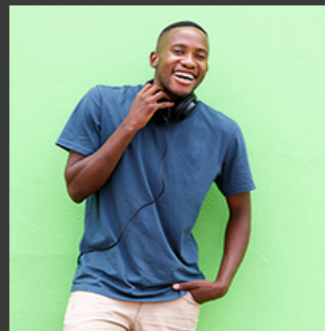
Train Model

Price Forecasting

- Before
- Installation
- Training
- After

Description	Price
City Streets Super cute shirt Super cute Good fit for	26.72
Monarchy Men's monarchy size large tshirt. Graphic tee	35.00
Fabletics Fablectics black shirt top Fabletics new black crochet striped shirt top nwt size large	50.80
Oakley Oakley women's shirt size large	52.25
Deep Wavy Shirt Pepsi-esque type graphic. Slight fading in graphic but no cracking	41.51

$42.37

Our Store

97.23%

Training Accuracy

Train Model

Price Forecasting

Before   Installation   Training   After

$41.41
Retail Store A

$50.80
Retail Store B

$42.69
Our Store