



We help extract the knowledge hidden in  
your data.

# FEATURE STORE, FEATURES MANAGEMENT IN SCALE.

**Tomasz Sienkiewicz**

Machine Learning Engineer

[tsienkiewicz@tidk.pl](mailto:tsienkiewicz@tidk.pl)

**Maciej Rubczyński**

AI, ML & Cloud Development Leader

[mrubczynski@tidk.pl](mailto:mrubczynski@tidk.pl)

12

YEARS ON THE MARKET



TEAMS: ANALYTICAL,  
RESEARCH, ENGINEERING



ADVANCED SPECIALIZATION

Analytics on Microsoft Azure



Google Cloud  
Partner



## BUSINESS DATA PLATFORM

- DATA WAREHOUSE
- INTEGRATIONS & ESB
- BIG DATA, DATA LAKE, IoT
- BUSINESS INTELLIGENCE
- REPORTS & VISUALIZATIONS

## DATA SCIENCE

- DATA EXPLORATIONS
- ADVANCED ANALYTICS
- MULTI-OBJECTIVE OPTIMIZATION
- PROCESS SCHEDULING ALGORITHMS
- SIMULATIONS MODELS

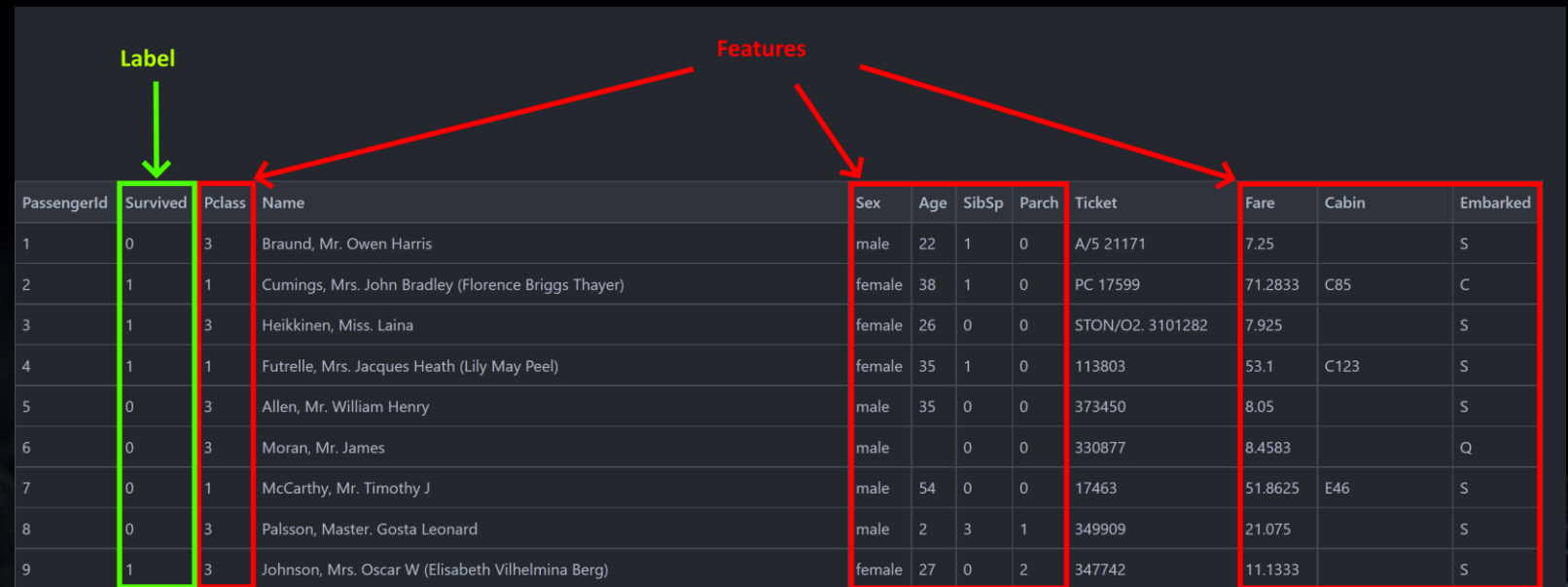
## ARTIFICIAL INTELLIGENCE

- MACHINE LEARNING
- DEEP LEARNING
- IMAGE PROCESSING
- NATURAL LANGUAGE PROCESSING
- INTELLIGENT CHATBOTS

# What is a feature?

**Feature** - in machine learning and pattern recognition, a feature is an individual measurable property or characteristic of a phenomenon. Choosing informative, discriminating and independent features is a crucial element of effective algorithms in pattern recognition, classification and regression.”

*„Applied machine learning is basically feature engineering.”*  
- Andrew Ng



PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	7.25		S
2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925		S
4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.4583		Q
7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	0	2	347742	11.1333		S

# Why we need a feature store?

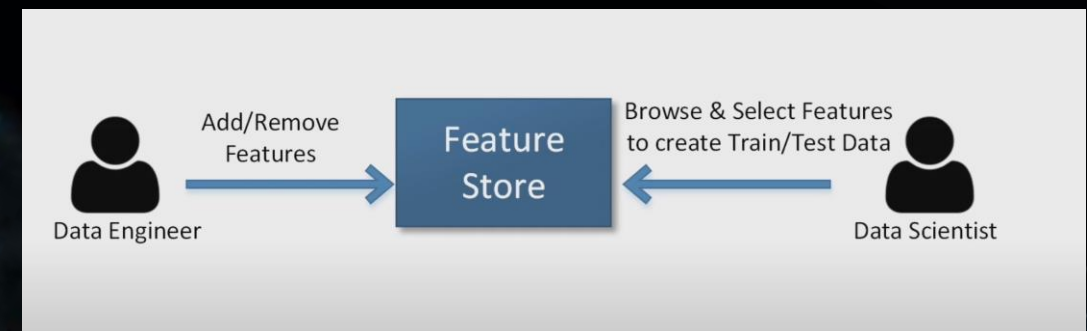
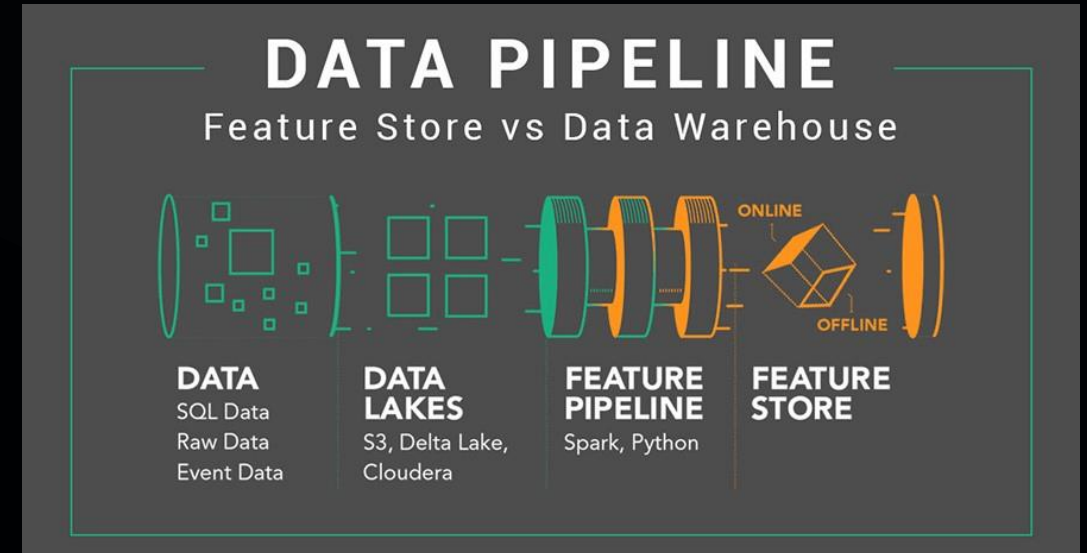
- prevents redundancy of features and feature pipelines
- unifies features and feature pipelines
- makes our features reusable across environments and teams





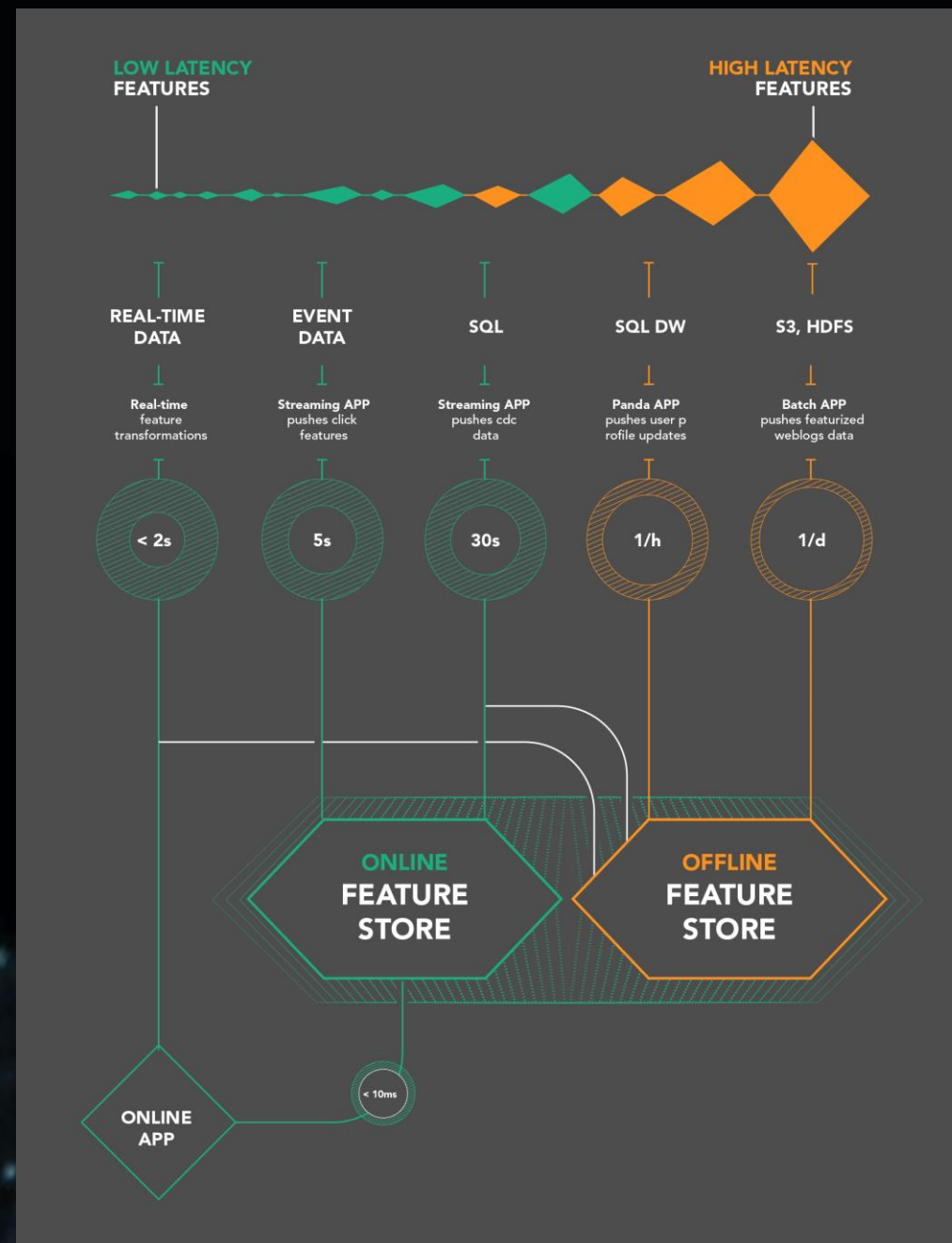
# What is a feature store?

- Data management layer in MLOps process
- Features repository
- Enables features to be registered and discovered across the organization
- Typically implemented as a dual-database system – online and offline
- Operational monitoring eg. monitoring feature drift
- Features versioning



# Offline vs online

Online	Offline
For features that require low-latency access (milliseconds) during inference	For training and testing purposes or for batch models
Typically a real-time database or a key-value store	Scale-out SQL database or data lakes
Real-time or near real-time models	Batch model
CosmosDB, Redis, BigTable, Cassandra	Azure Data Lake, Google Cloud Store, BigQuery, S3, Delta lake



# Feature store vs DWH

Feature store	DWH
Contains features for training or serving purposes in machine learning cycle.	Contains only data that is organised in a central repository that can be analyzed to make more informed decision.
Data used by Data Scientists	Data used by Analysts
Data ingestion using feature pipelines	Data ingestion using ETL processes
Data is prepared to use in ML models	Data needs additional processing before using it in ML modles

# Feature store solutions

- **Feast**
- Hopsworks
- Databricks Feature Store
- Amazon Sagemaker Feature Store
- Vertex Feature Store (Google)





# Feast



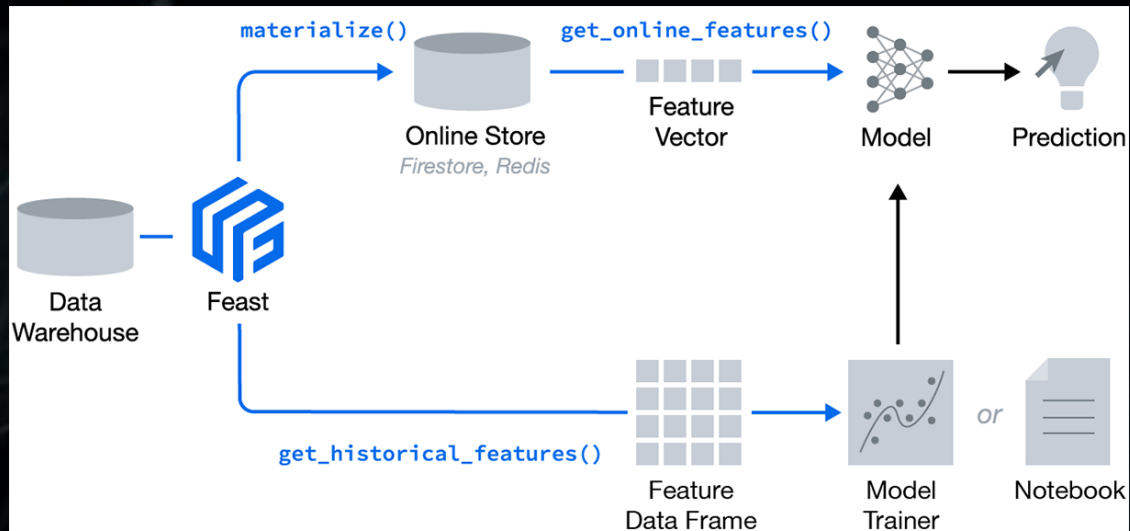
Star

2.5k



Fork

436



- Open-source feature store
- <https://github.com/feast-dev/feast>
- Current version (09.12.2021): 0.16.0

## Problems Feasts solves

- Models need consistent access to data
- Deploying new features into production is difficult
- Models need point-in-time correct data
- Features aren't reused across projects

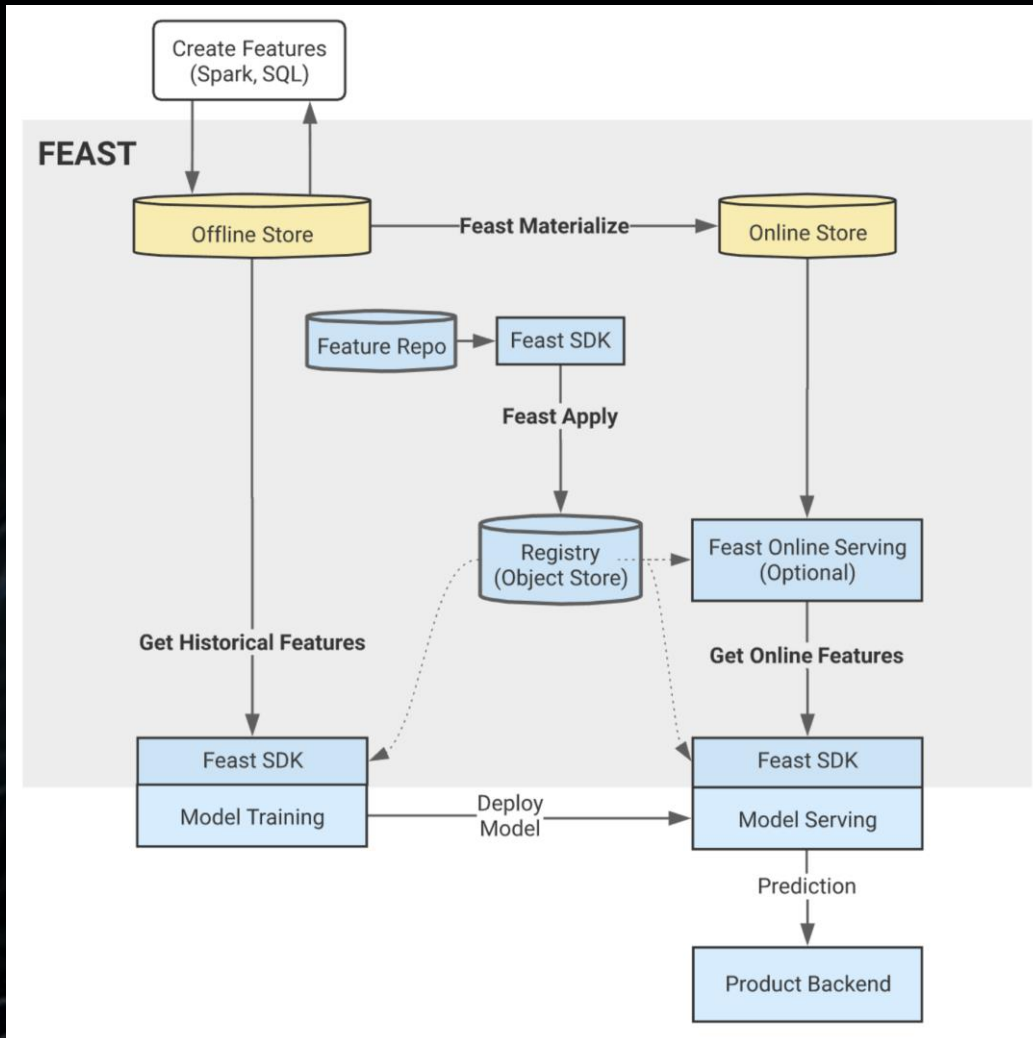
## Problems Feasts does not yet solve

- Feature engineering
- Feature discovery
- Feature validation

## What Feast is not

- ETL or ELT system
- Data warehouse
- Data catalog

# Architecture and components



## Registry:

- Local
- S3
- GCS

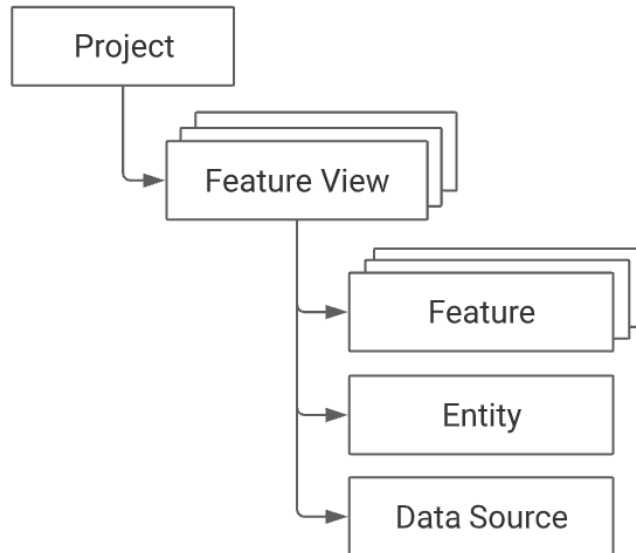
## Offline stores:

- File
- BigQuery
- RedShift

## Online stores:

- SQLite
- Redis
- Datastore
- DynamoDB

# Feast concepts



		entity			feature
row	timestamp	driver_id	trips_today	rating	
1	5/22/2021 0:00:00	driver_1001	5	4.8	
2	5/22/2021 0:00:00	driver_1002	0	4.2	
3	5/22/2021 0:00:00	driver_1003	8	4.5	
4	5/22/2021 0:00:00	driver_1004	3	4.2	

Diagram illustrating the mapping of table columns to Feast concepts:

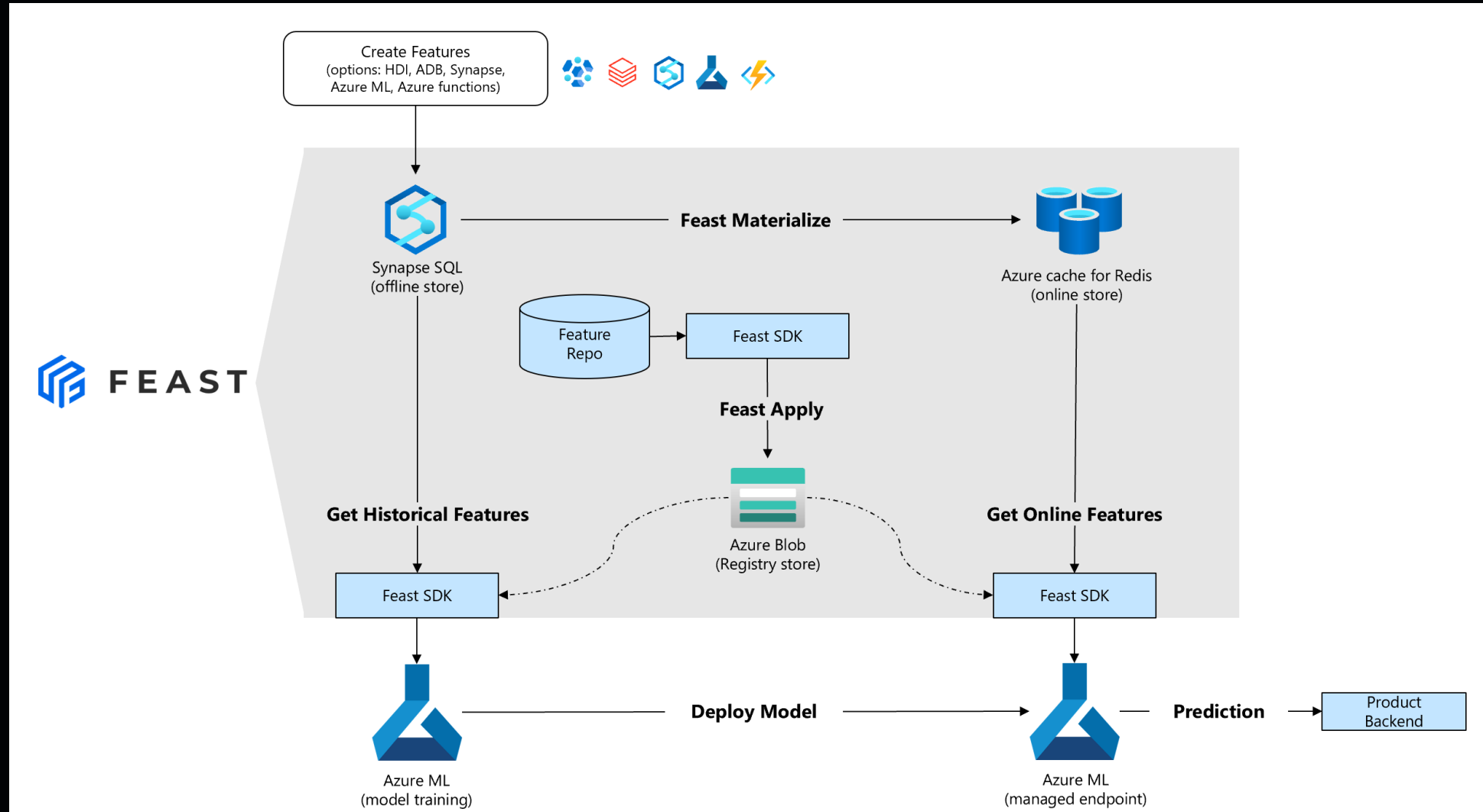
- timestamp maps to timestamp
- driver\_id maps to entity value
- trips\_today maps to feature value
- rating maps to feature value

		entity key			
row	timestamp	customer_id	driver_id	trips_today	rating
1	5/22/2021 0:00:00	customer_A	driver_1001	5	4.8
2	5/22/2021 0:00:00	customer_B	driver_1002	0	4.2
3	5/22/2021 0:00:00	customer_C	driver_1003	8	4.5
4	5/22/2021 0:00:00	customer_D	driver_1004	3	4.2

Diagram illustrating the mapping of table columns to Feast concepts, highlighting the entity key:

- customer\_id is the entity key
- driver\_id is the entity value
- trips\_today is the feature value
- rating is the feature value

# Feast on Azure





# LIVE DEMO

<https://github.com/TSienki/feast-demo-pytech>

<https://github.com/Azure/feast-azure/tree/main/provider/tutorial>





Q&A

TECHNOLOGY  
INNOVATION  
DATA  
KNOWLEDGE

**tidk**

We advise, design  
and implement  
advanced data-based  
solutions.



**Ciech**

**ABB**



**LOTOS**



**JEPPesen**  
A BOEING COMPANY



**arvato**  
BERTELSMANN



**InPost**  
out of the box

**NEUCA**



**SELENA**  
GLOBAL EXPERIENCE



grupa pracuj

**allegro**



Bank Polski



**LEWIATAN**



**MASPEX**

**ING**  
Liczą się ludzie



**zabka**

**eurocash**.pl  
TWOJE NOWE MOŻLIWOŚCI



**mediaexpert**

**GRUPA LUXMED**

**eBilet.pl**

**PROMEDICA24**



www: [TIDK.PL](http://TIDK.PL)

e-mail: [kontakt@tidk.pl](mailto:kontakt@tidk.pl)

Linkedin: [TIDK](https://www.linkedin.com/company/tidk)



Ministerstwo  
Finansów