



GIGASPACE
innovate with confidence

How NLP is Helping a European Financial Institution Enhance Customer Experience

Tal Doron
Director, Technology Innovation



Agenda





ABOUT ME



@taldoron



taldoron84



tald@gigaspace.com



Tal Doron

Director, Technology
Innovation



About GigaSpaces

We provide one of the leading in-memory computing platforms for real-time insight to action and extreme transactional processing. With GigaSpaces, enterprises can operationalize machine learning and transactional processing to gain real-time insights on their data and act upon them in the moment.



InsightEdge is an in-memory real-time analytics platform for instant insights to action; analyzing data as it's born, enriching it with historical context, for smarter, faster decisions



In-Memory Computing Platform for microsecond scale transactional processing, data scalability, and powerful event-driven workflows

300+

Direct customers

50+ / 500+

Fortune / Organizations

5,000+

Large installations in production (OEM)

25+

ISVs



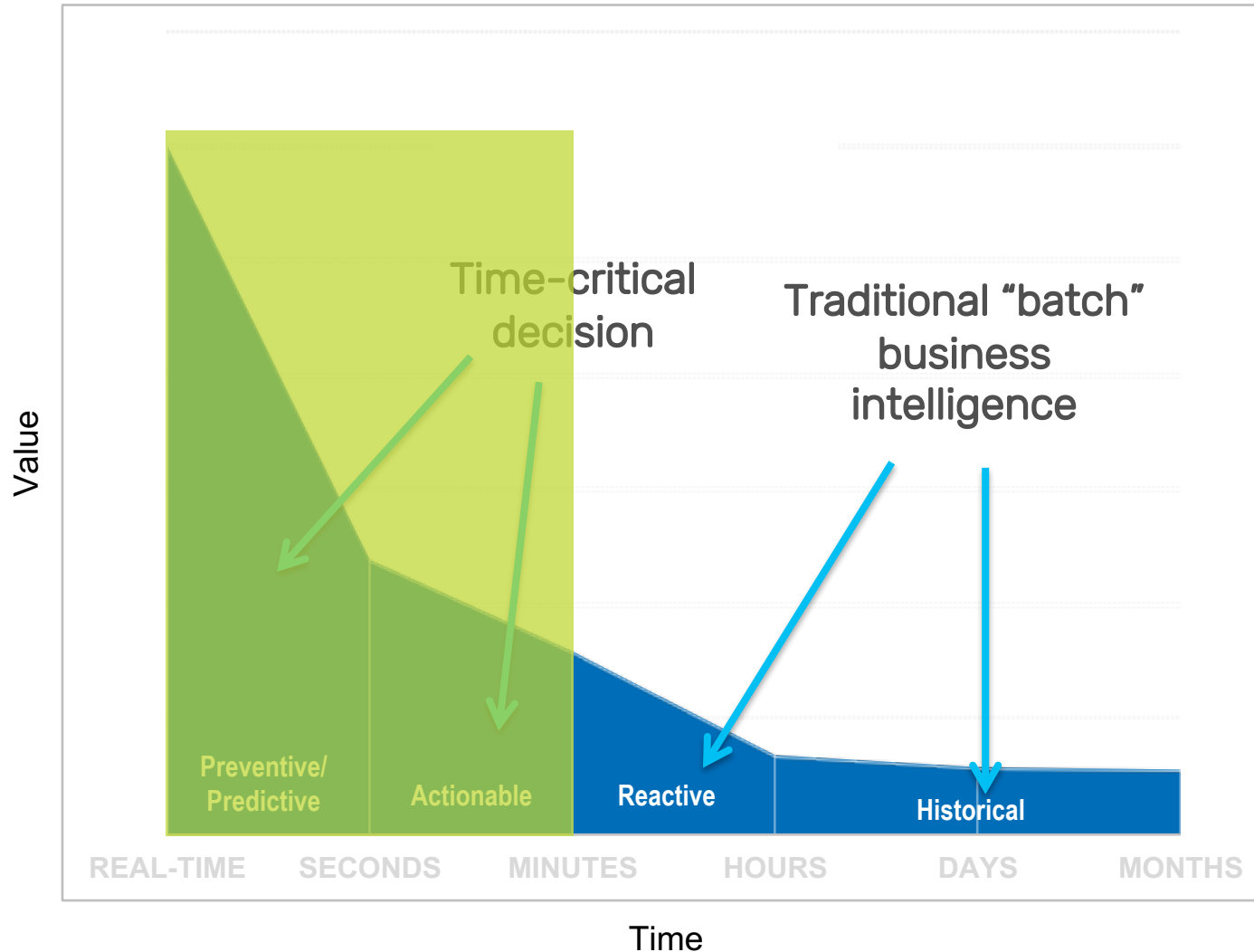
74%
want to be
data driven

only 23%
are successful,

FORRESTER®



How Can You Gain the Most Value from Your Data?



Near real-time data is highly valuable if you act on it on time



Historical + near real-time data is **more** valuable if you have the means to combine them









The Velocity of Business (once upon a time)

“To prevent fraud, anomaly detection needs to happen against 500,000 txn/sec in less than **200 milliseconds**”

“A typical e-commerce website will experience **40%** bounce if it loads in more than **3 seconds**, including personalization offers”

“A call center receives **450,000 calls/min**, across 200 phone numbers, each call needs to be routed in less than **60 milliseconds**”



**FINANCIAL
SERVICES**



ECOMMERCE



TELCO



ABOUT THE CUSTOMER

This Financial IT Service provider serves the leading banks in Germany with core solutions and services

Business Goals:



Enhance customer experience with quicker First Call Resolution



Reduce Average Handle Time for optimized efficiency



BUSINESS CHALLENGES

DISJOINTED CUSTOMER INTERACTIONS

Disparate data sources and systems, led to inefficient juggling between screen and systems and poor data quality & poor customer experience

KEEPING UP WITH EMPOWERED CUSTOMERS

Customers are smarter and have more insights into competitive products and services, raising expectations to a new standard

AN OMNICHANNEL EXPERIENCE

Customers want a consistent experience across all channels and agents, demanding faster resolution times



TECHNICAL CHALLENGES

HIGH PERFORMANCE

Ingestion of millions of CRM cases and data from other repositories into a unified analytics platform

MILLISECOND LATENCY

Customers demand an immediate response time, requiring high performance solutions that leverage ML models in real-time

CONTINUOUS ML TRAINING

Insights constantly need to adapt to changing conditions for smartest insights



PROPOSED SOLUTION

If a live agent is needed during a call, the NLP based solution automatically supplies the agent with articles and knowledge documents based on the conversation

Ticket ID #54367

Customer Name

#54367

Type

Enterprise

Support Level

Bronze

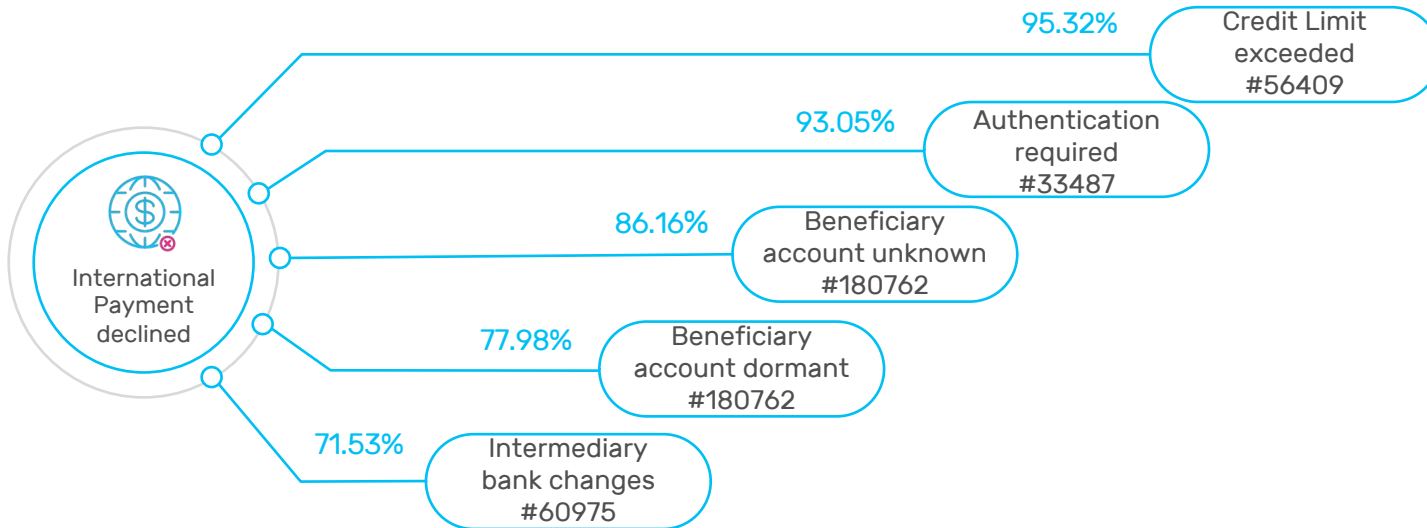
Last Contact Date

20.12.18

DATA SOURCES

CUSTOMER

CUSTOMER TICKET



Case Description

International payment to supplier declined

Read more

Case Resolution

Check that credit limit is not exceeded

Read more

Case Description

Check here to email instructions to customer

Email



Training the model
based on

2M CRM
records
in **27min**

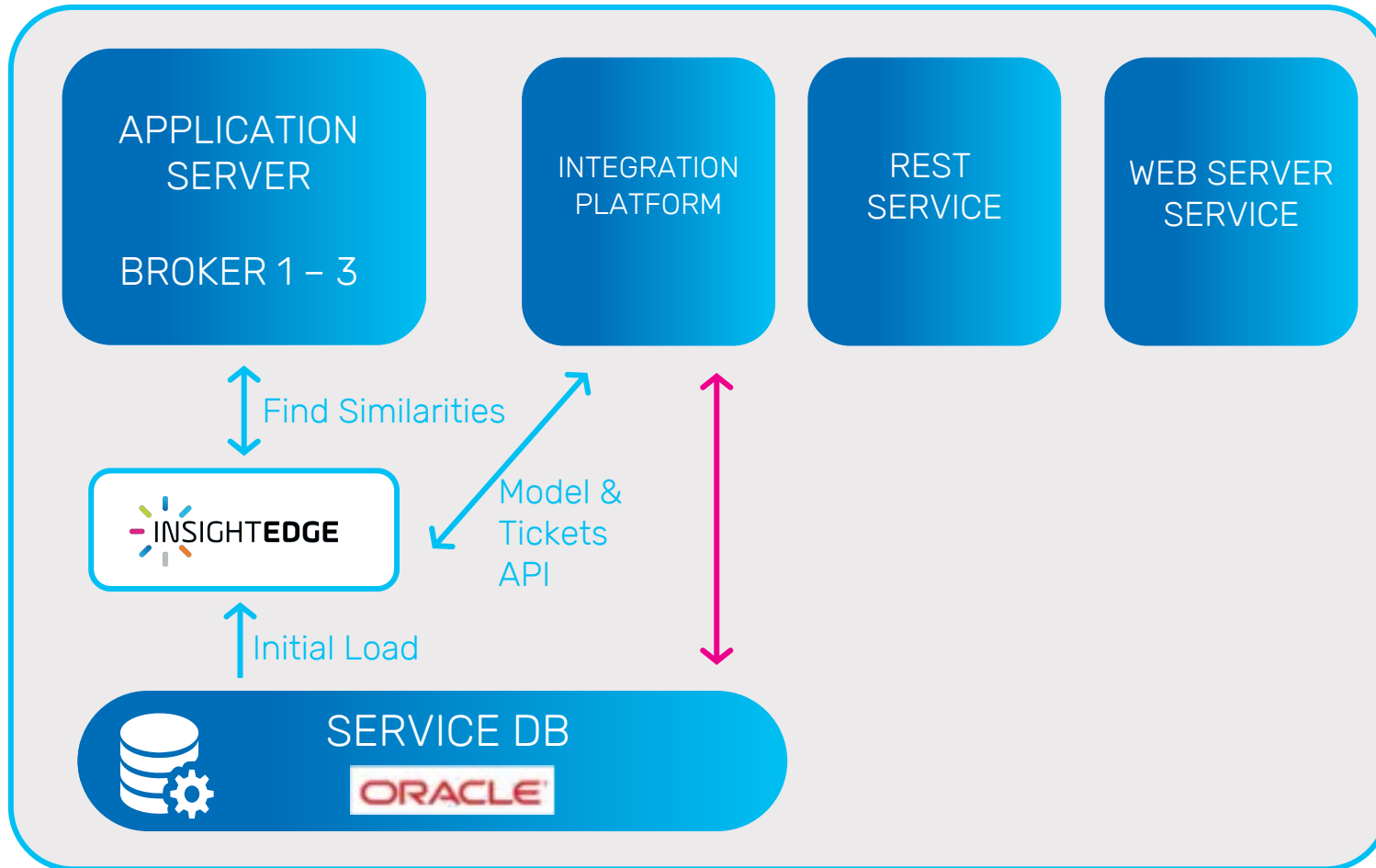
Time to results

~50ms



General Architecture & Data Flow

Server 1



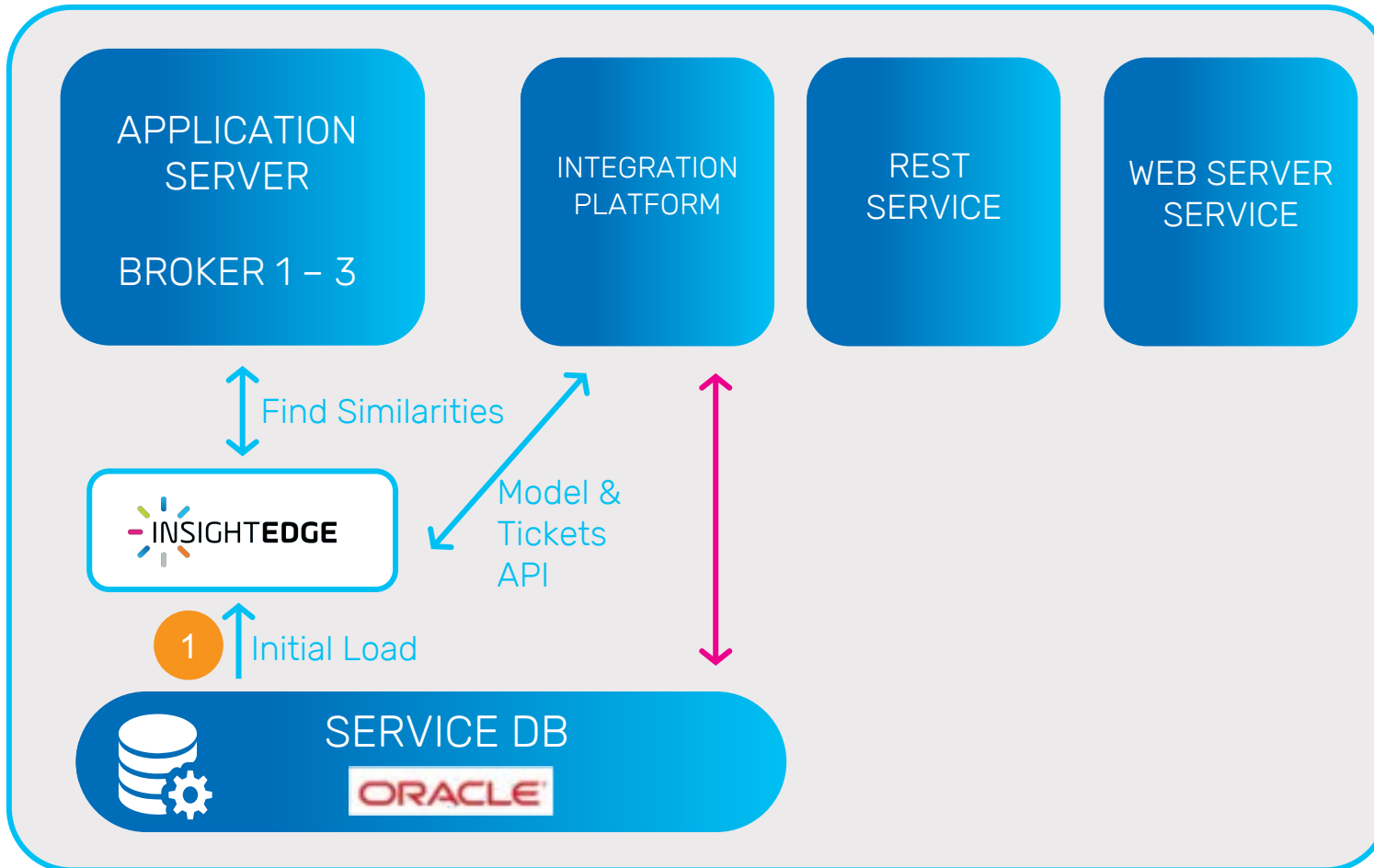
Server 2





General Architecture & Data Flow

Server 1



1 Initial Load

Hibernate on Object Store

```
@Id
private Integer ticketID;

private String searchText;

private Date openDate;

private Integer responsibleCompany;

private Integer customerDomain;

private Integer customerCompany;

private Integer ticketCreator;

private Integer addModelControl1;

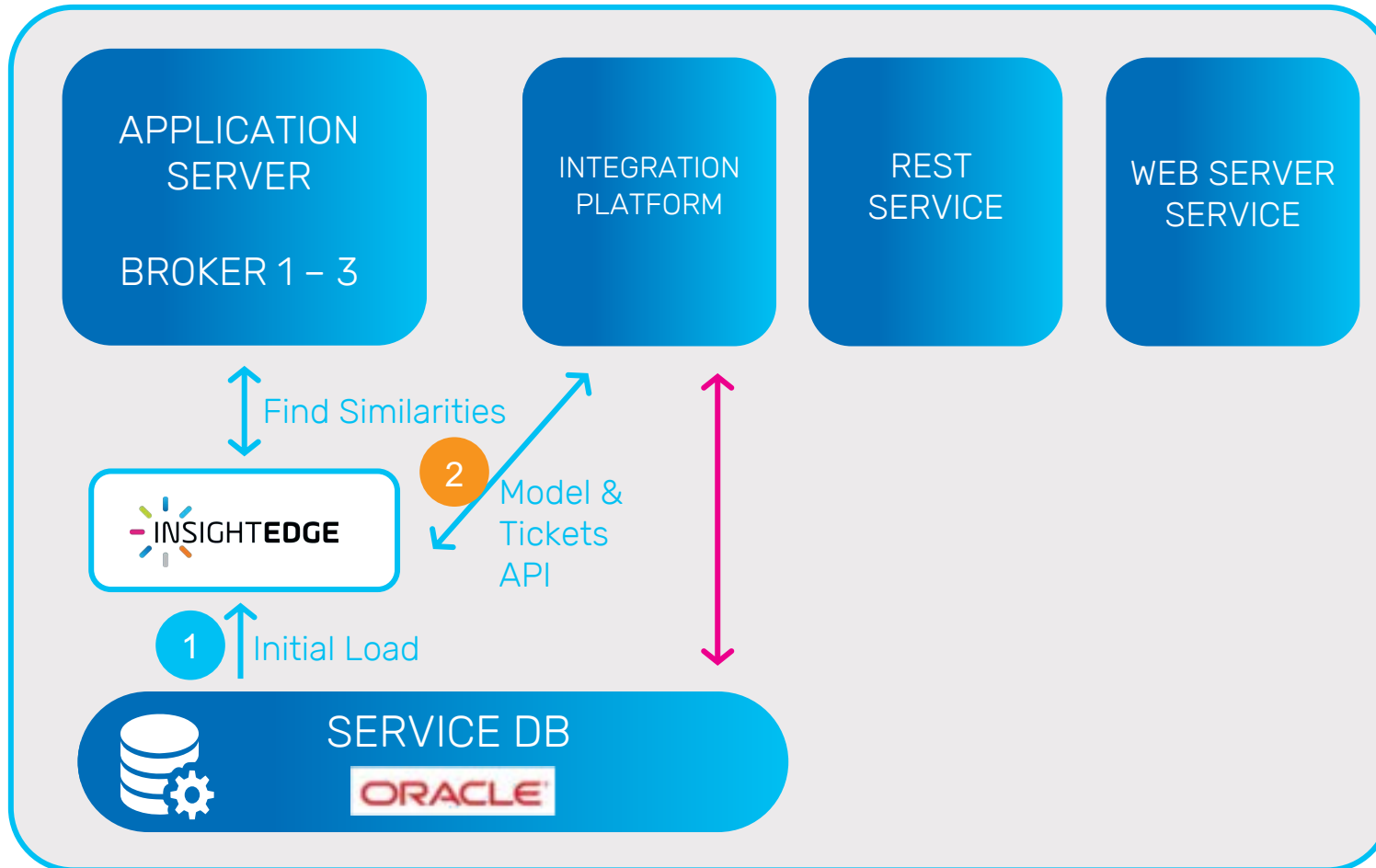
private Integer addModelControl2;

private Integer addModelControl3;
```



General Architecture & Data Flow

Server 1



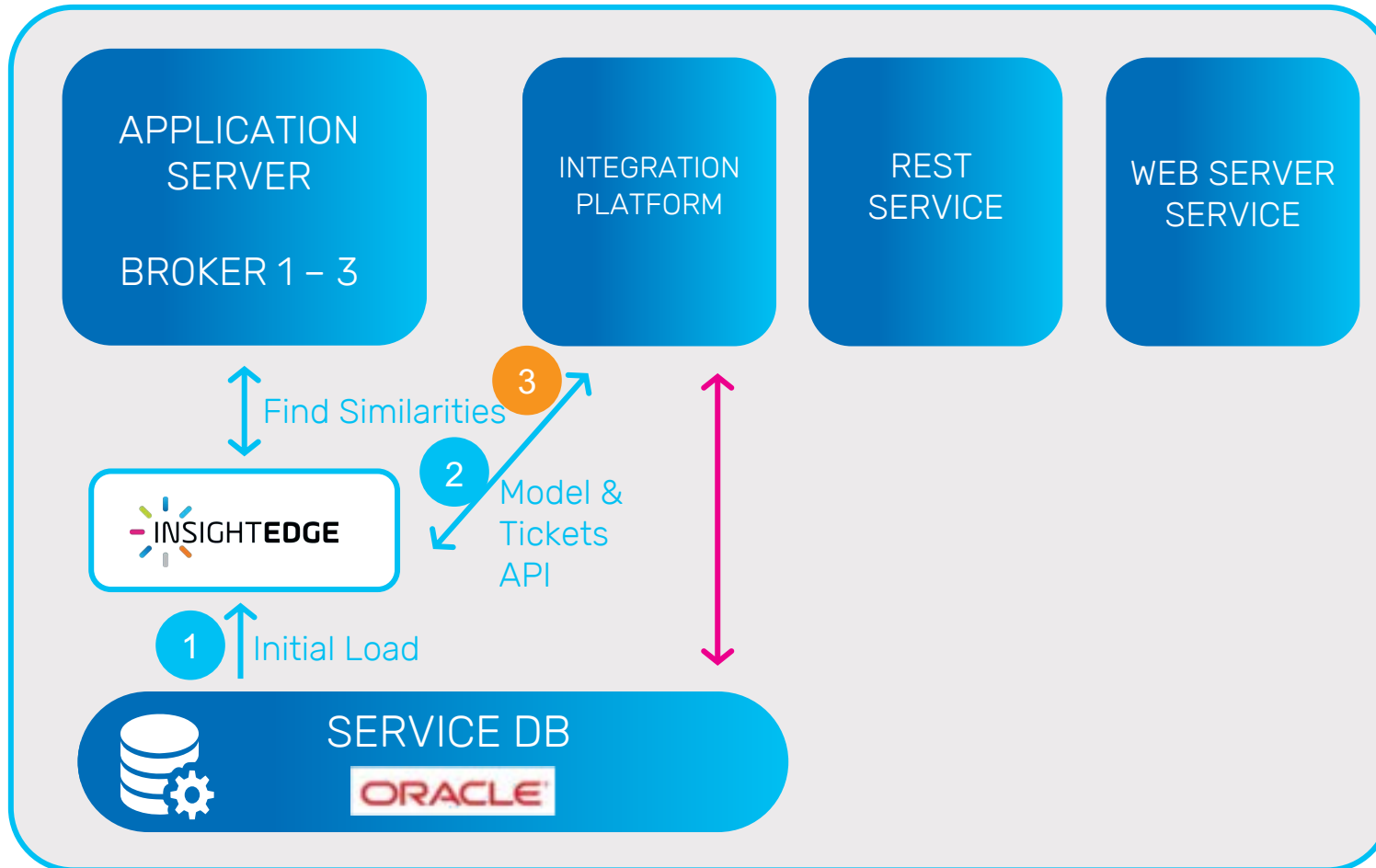
2 Training/building model

- Train
- stopTrainModel
- getTrainModelStatus
- checkModelInSpace
- destroyModel



General Architecture & Data Flow

Server 1



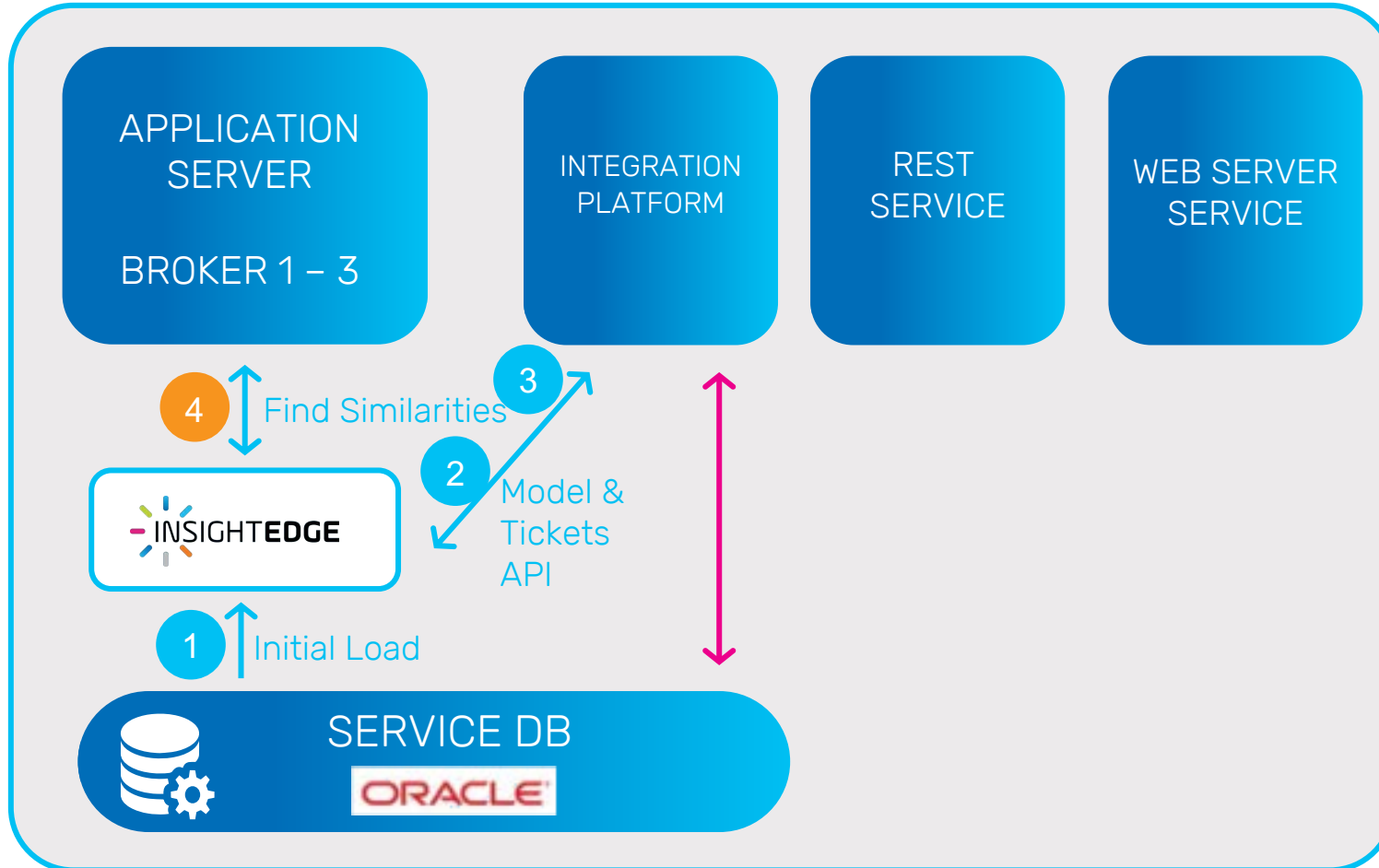
3 Long Running Spark Job API

- startModel
- stopModel
- checkModellsRunning
- getFindSimilaritiesStatus



General Architecture & Data Flow

Server 1



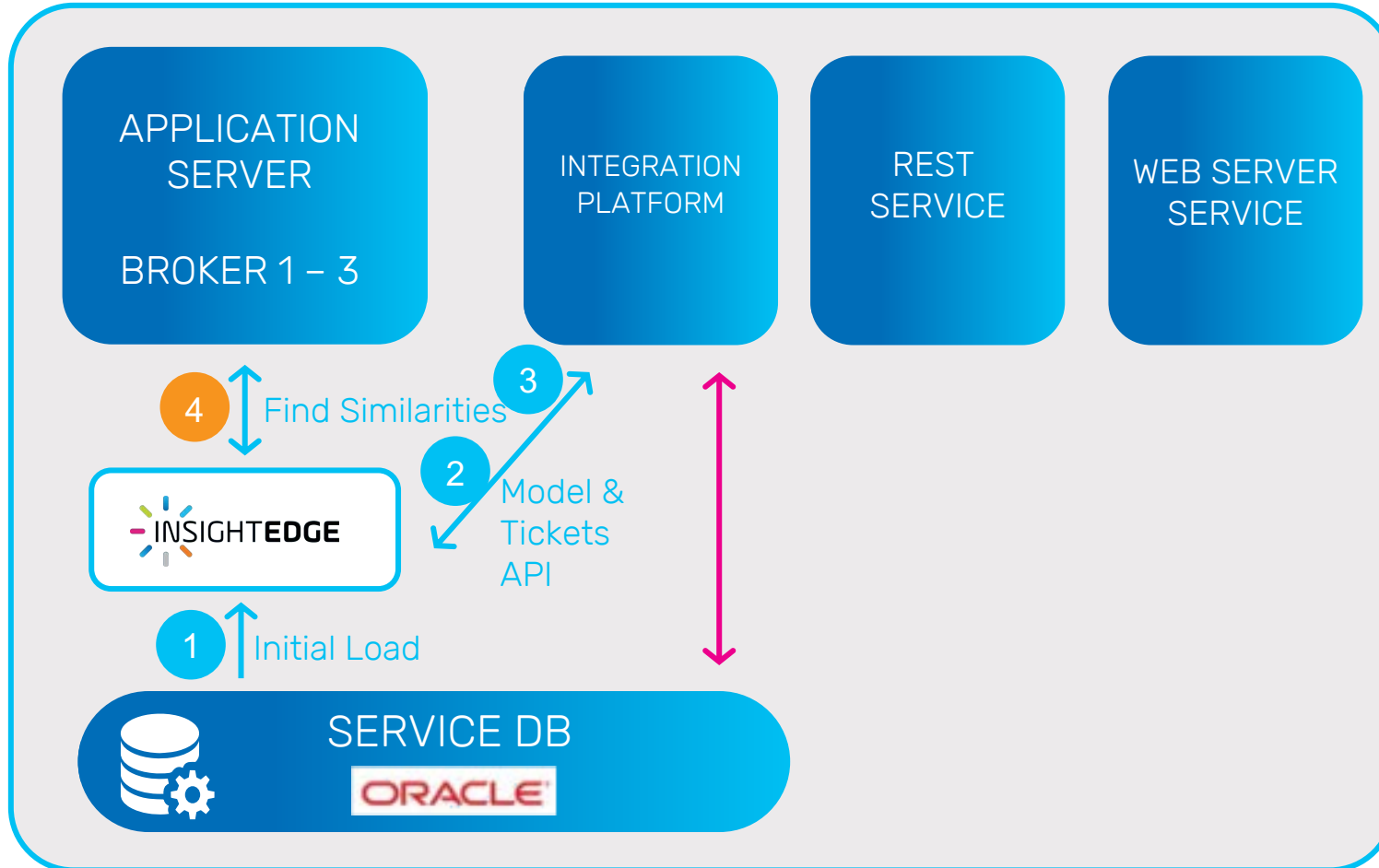
4 findSimilarities

- Write findSimilaritiesRequest object to the space using task
- Spark long time running job takes the object perform the find similarities action (set the object status to processed true)



General Architecture & Data Flow

Server 1



4 findSimilarities

```
ticketId>72018  
gs.exec(modelId, "my search")
```

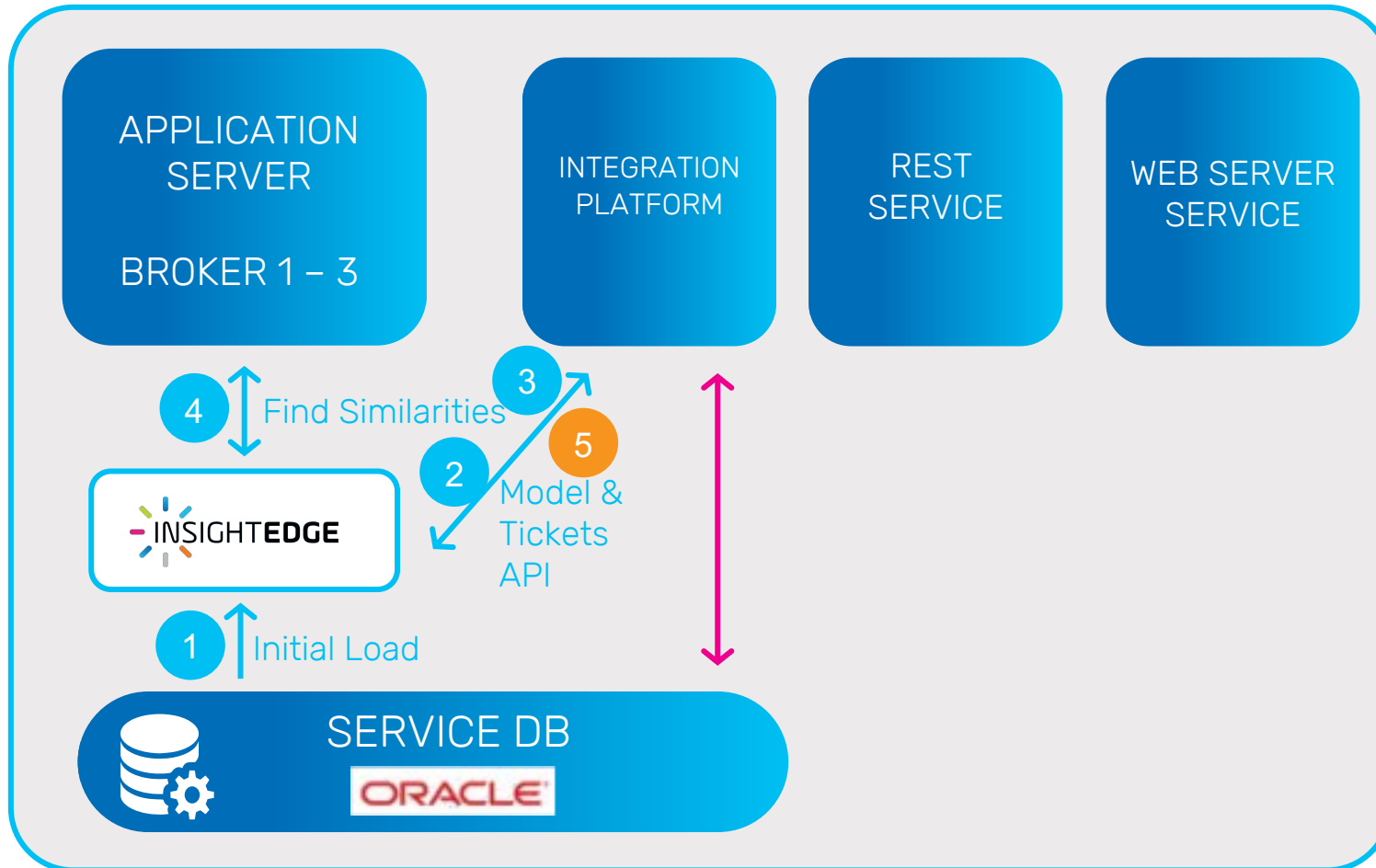
The result is the following similar cases:

```
70534 (0.823432215)  
70874 (0.726937532)  
70110 (0.719002341)  
70998 (0.528010191)
```



General Architecture & Data Flow

Server 1

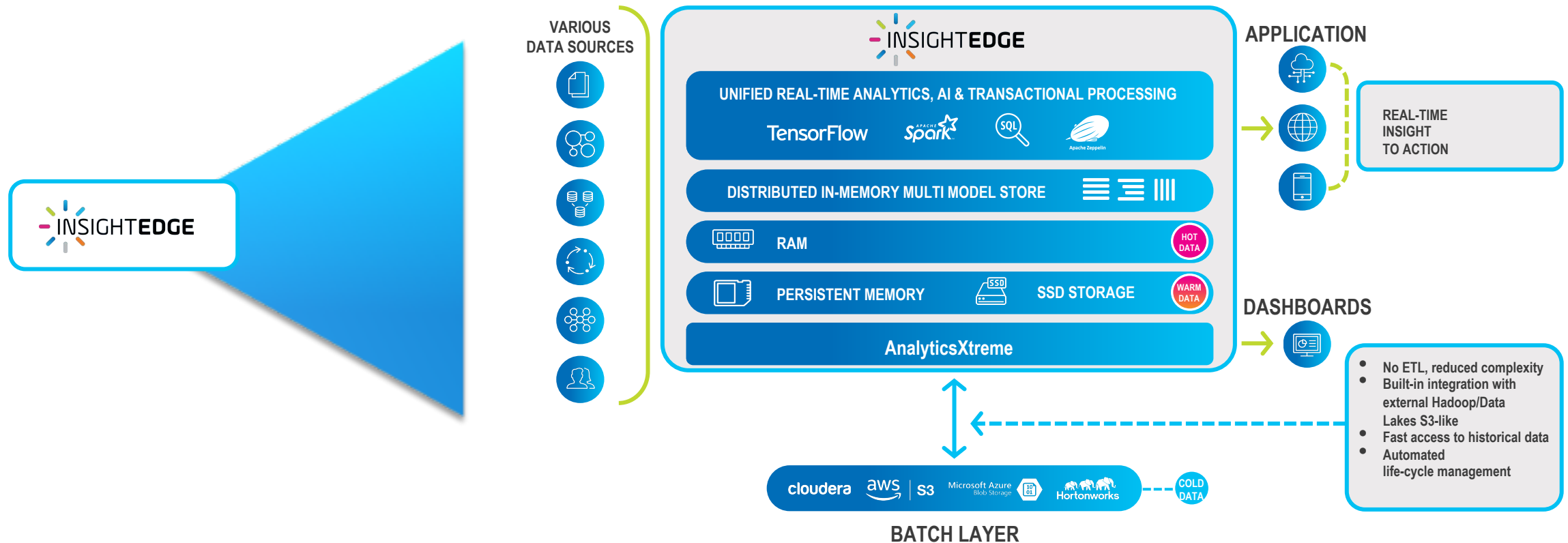


5 Support Tickets (the data)

- Incremental Feed
- Delete



Unified Transactional & Analytical Processing for Operationalizing ML





RESULTS

REAL-TIME

Average time of
50ms
to search and find
similar cases

EMPOWER THE AGENT

Allow the agents an
immediate response
time, **reducing mean
time to resolution**

CONTINUOUS ML TRAINING

27 Minutes
background training
time for 2 million
records

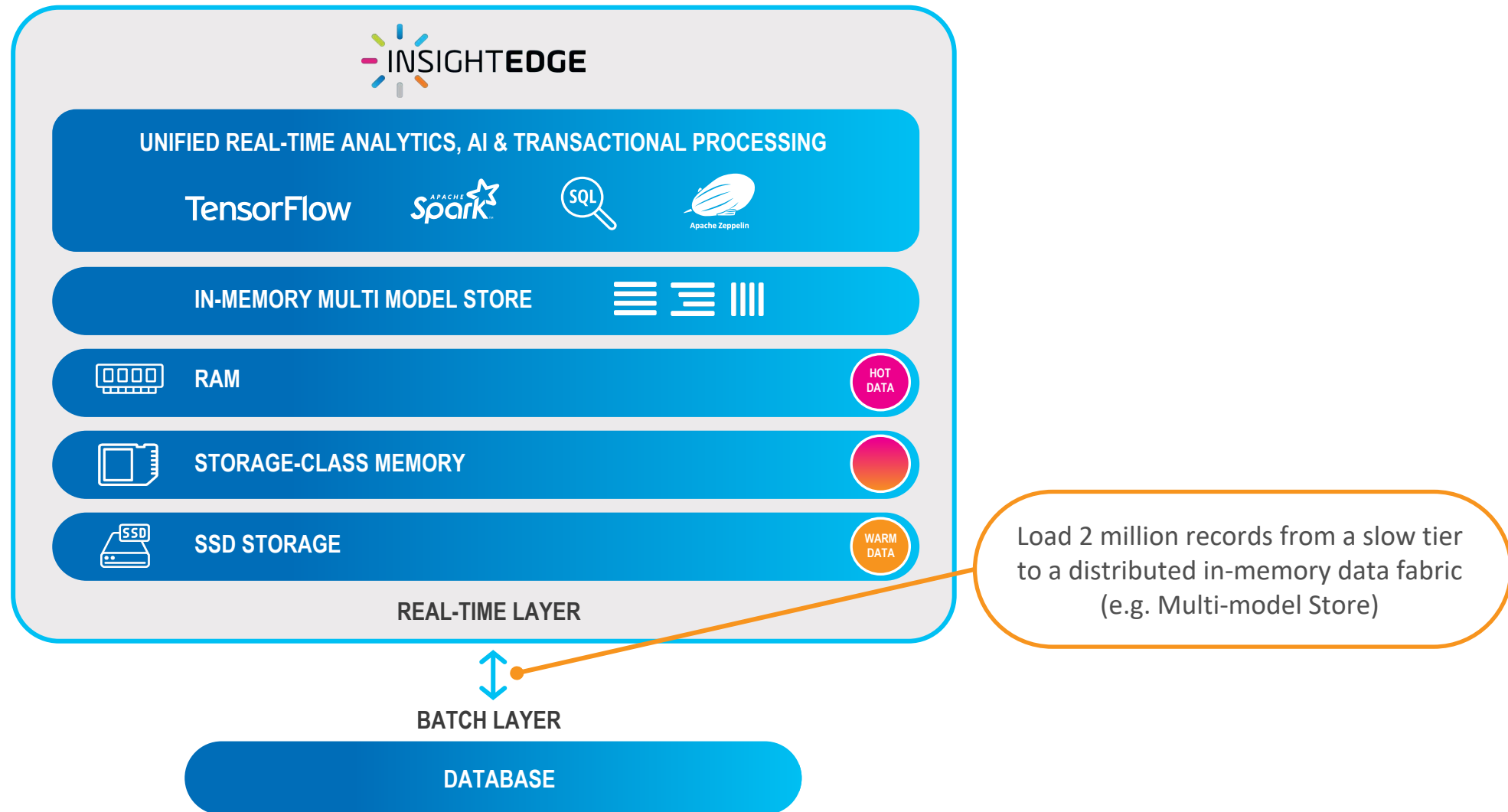


Overcoming Challenges



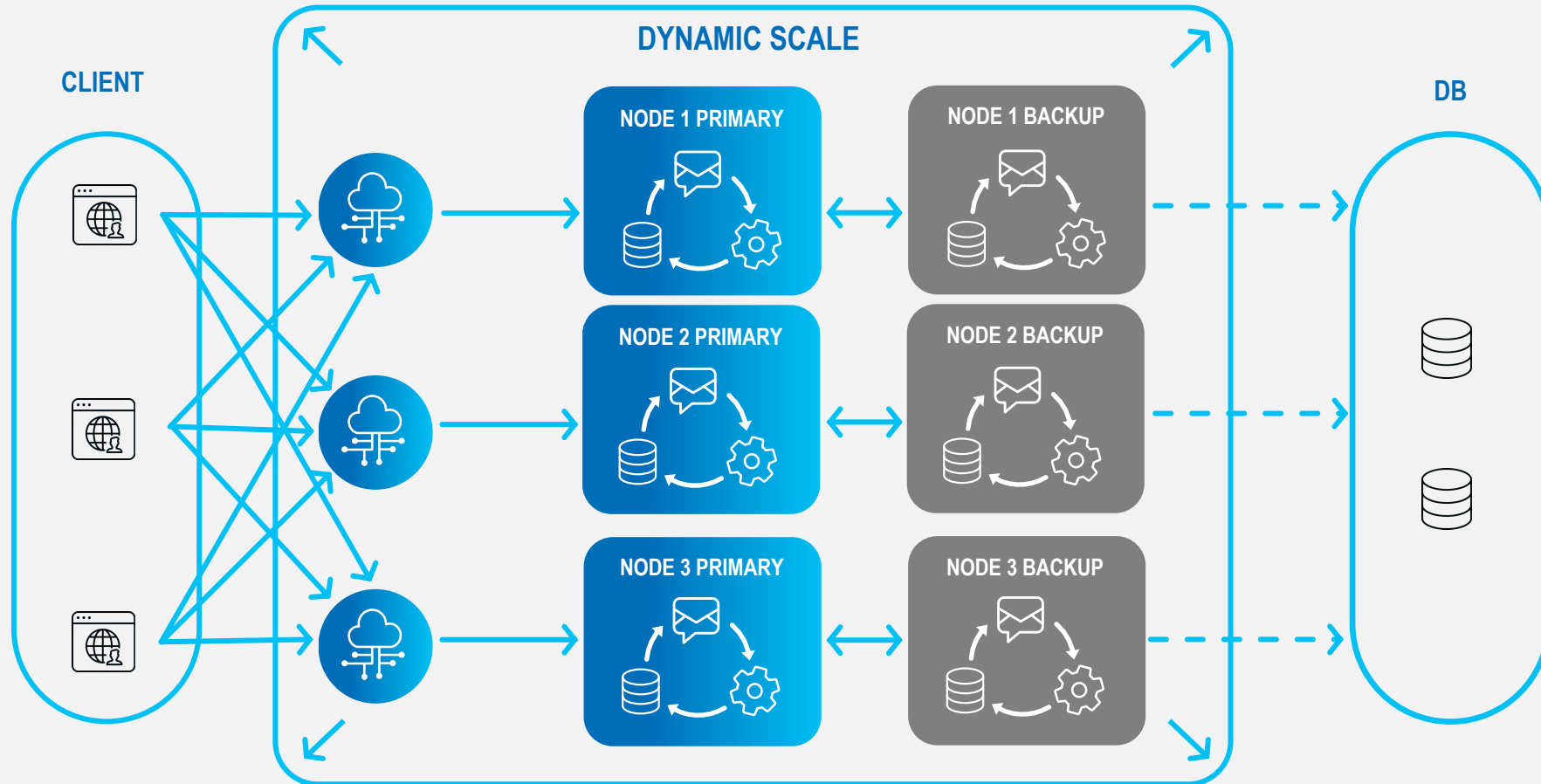


Step 1: Initial Load



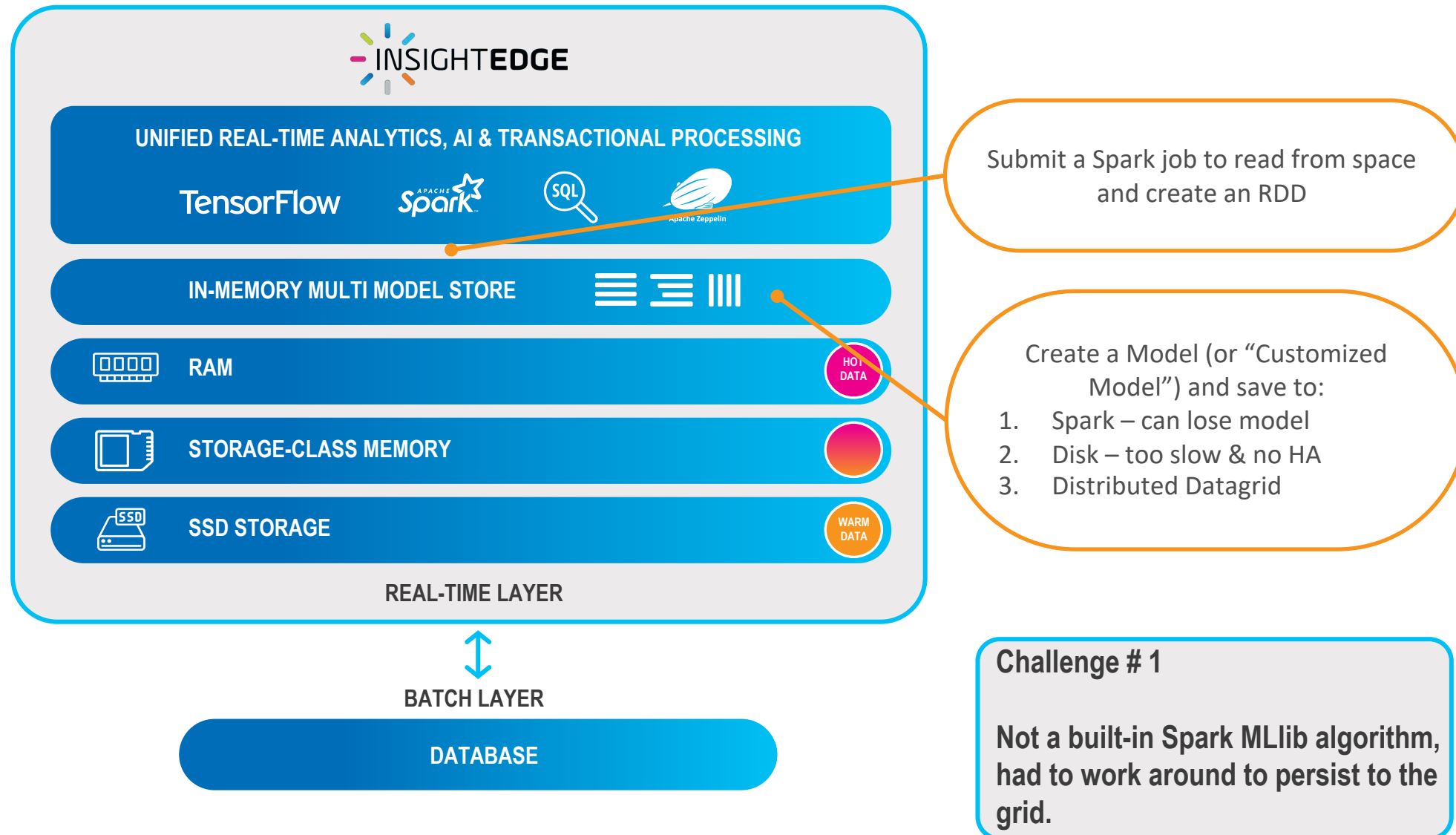


Distributed Multi-Model Object Store

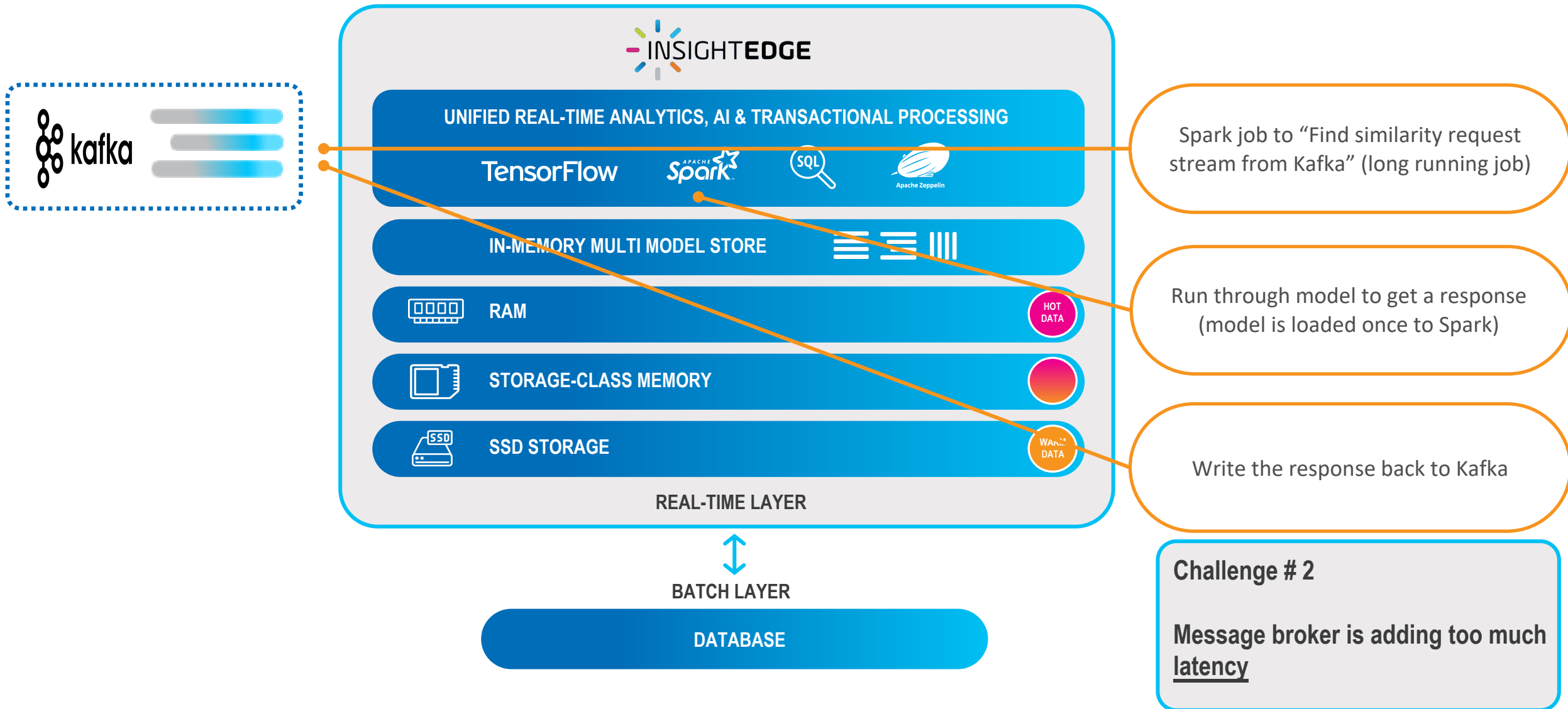




Step 2: Create Model and Save to...

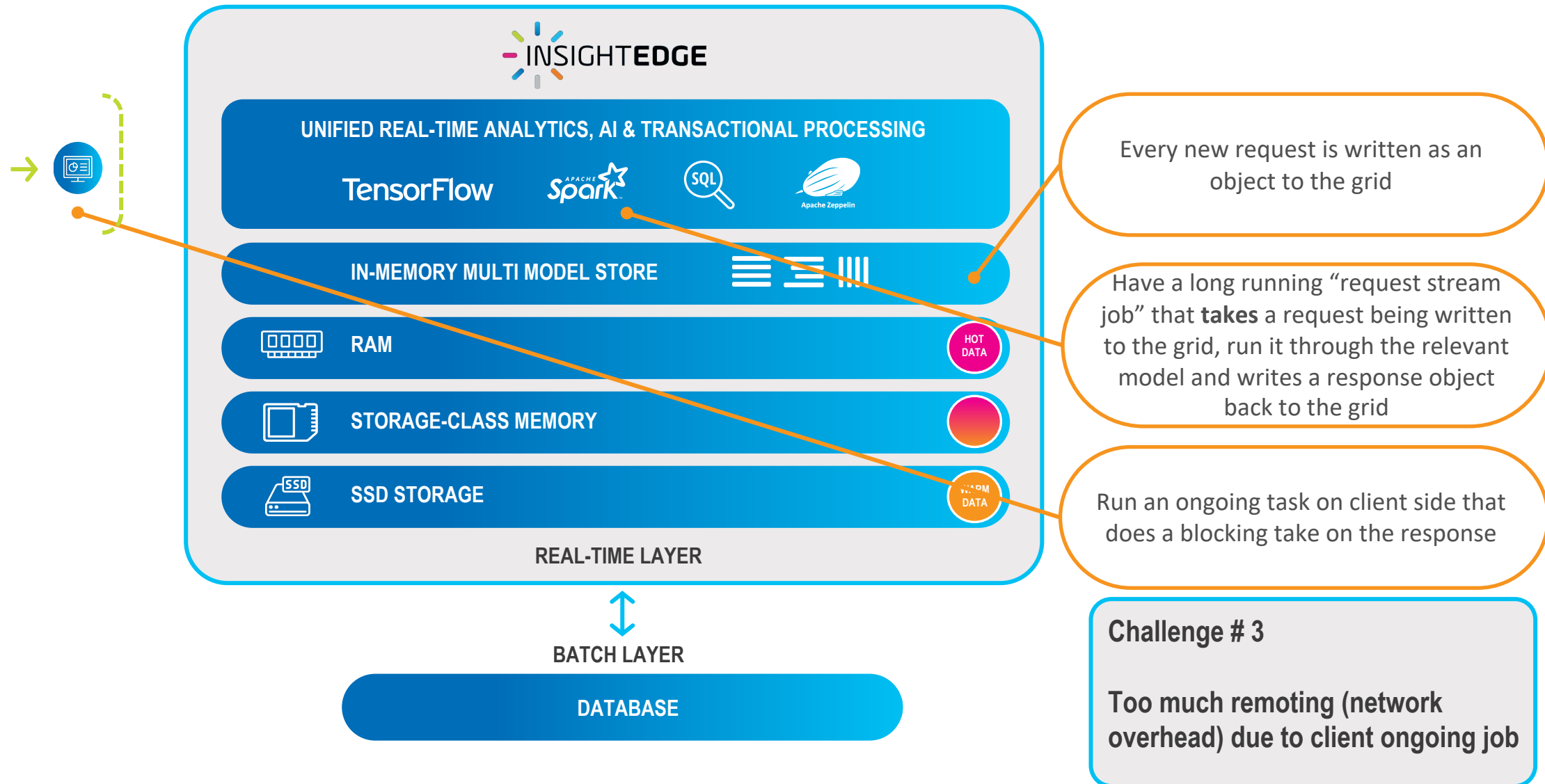


Step 3: Request/Response via Message Broker

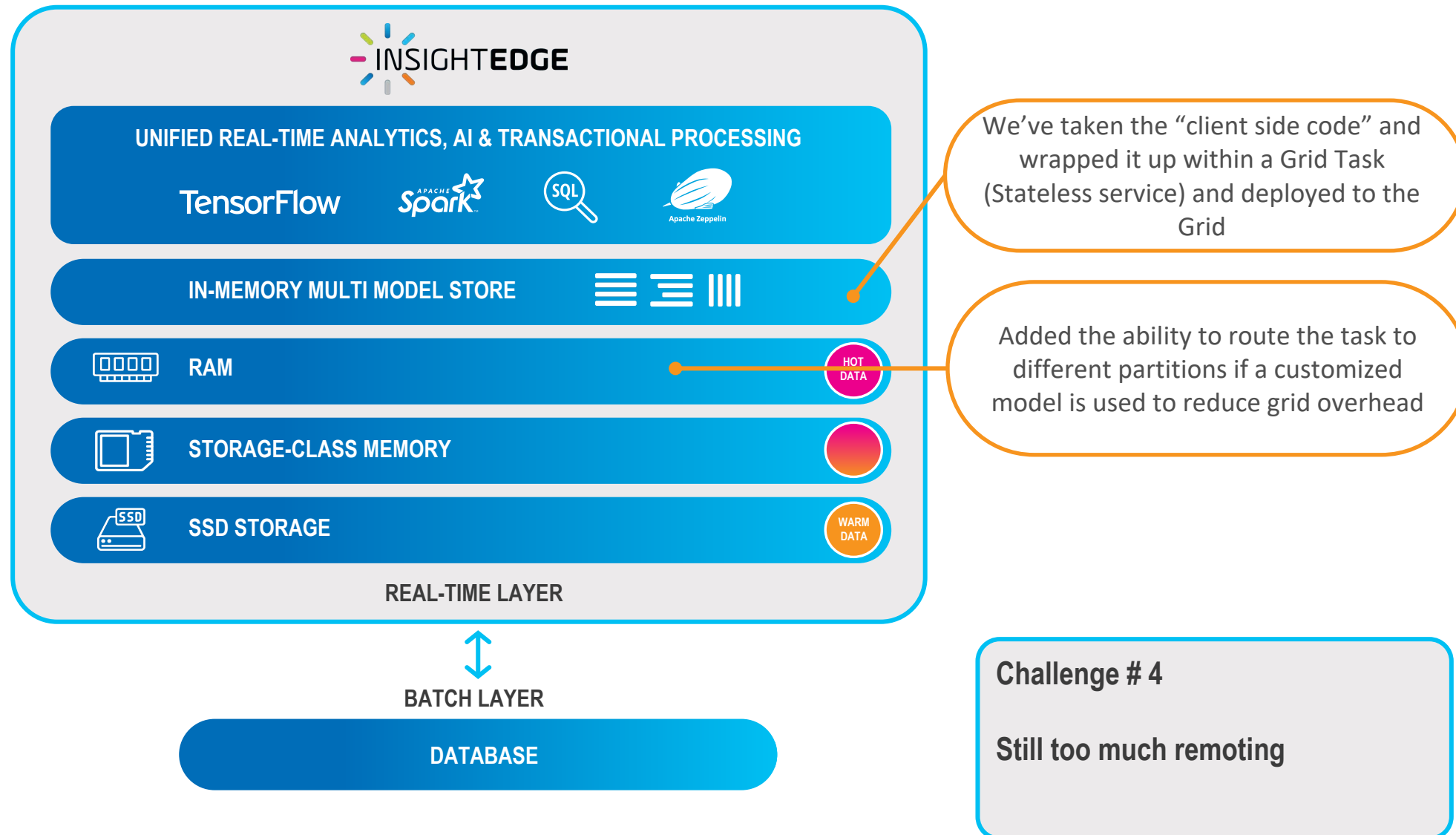




Step 4: Remove Message Broker

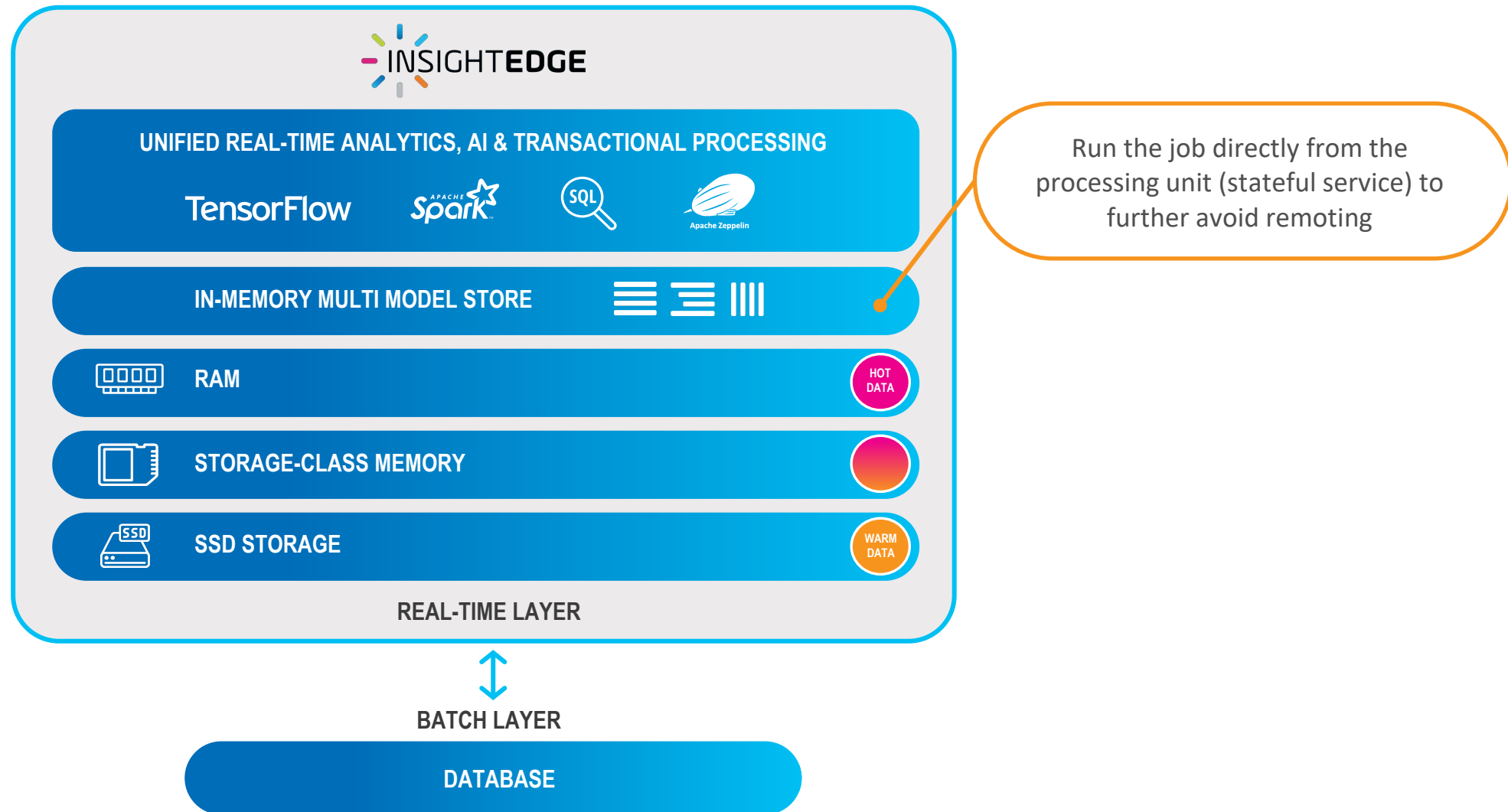


Step 5: Remove Remoting (as much as possible)

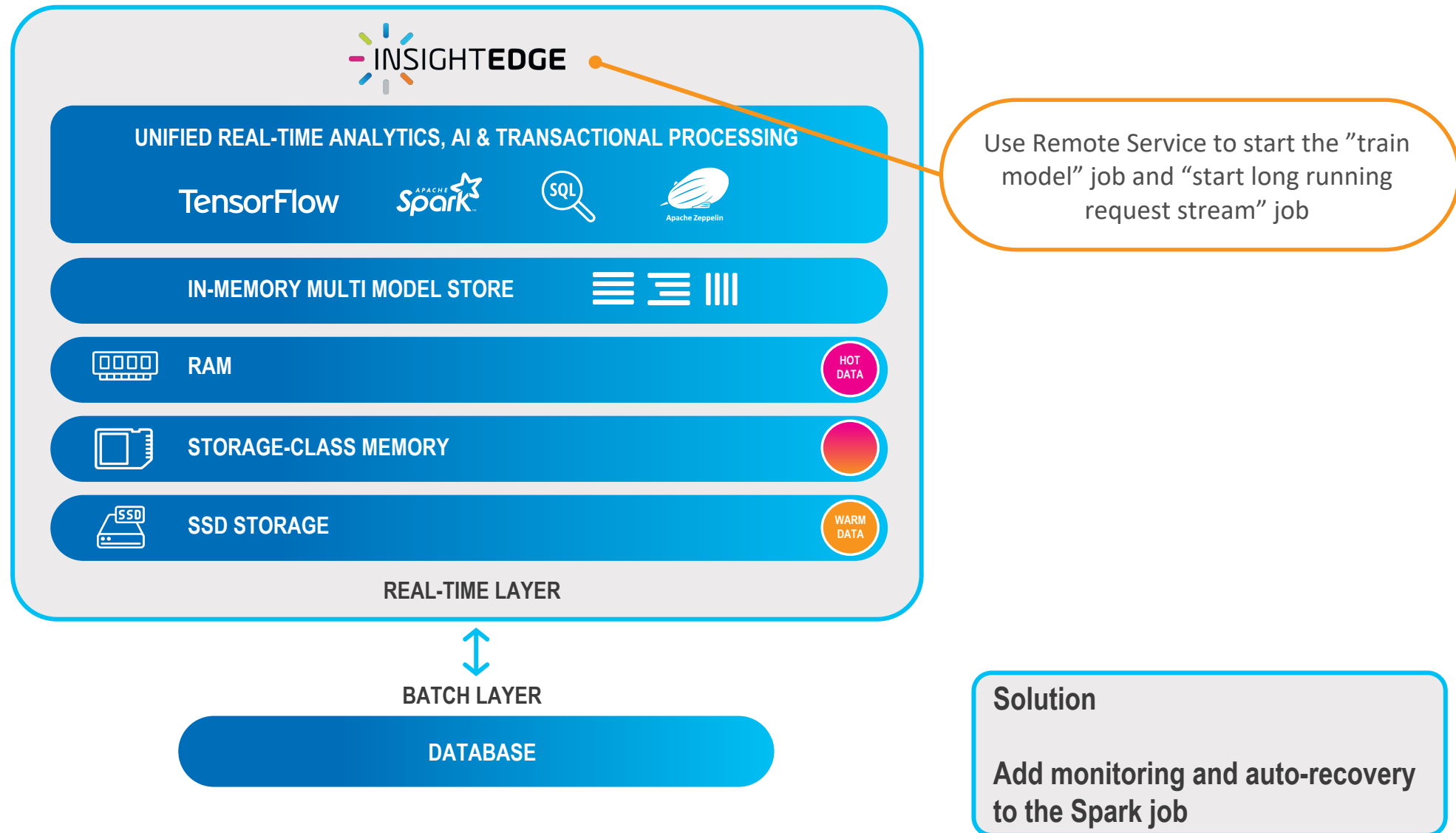




Step 6: Remove Remoting (cont.)



Step 7: Add Production Grade Capabilities to Spark





MOVING FORWARD FOCUS ON AUTOMATION

Automate Call Routing
(using Deep Learning Approach)

Automatic topic determination based on
text classification and sentiment analysis

Automatic CRM case creation





ABOUT THE USE CASE

This use case shows how to modernize existing software architecture for an efficient call center routing workflow

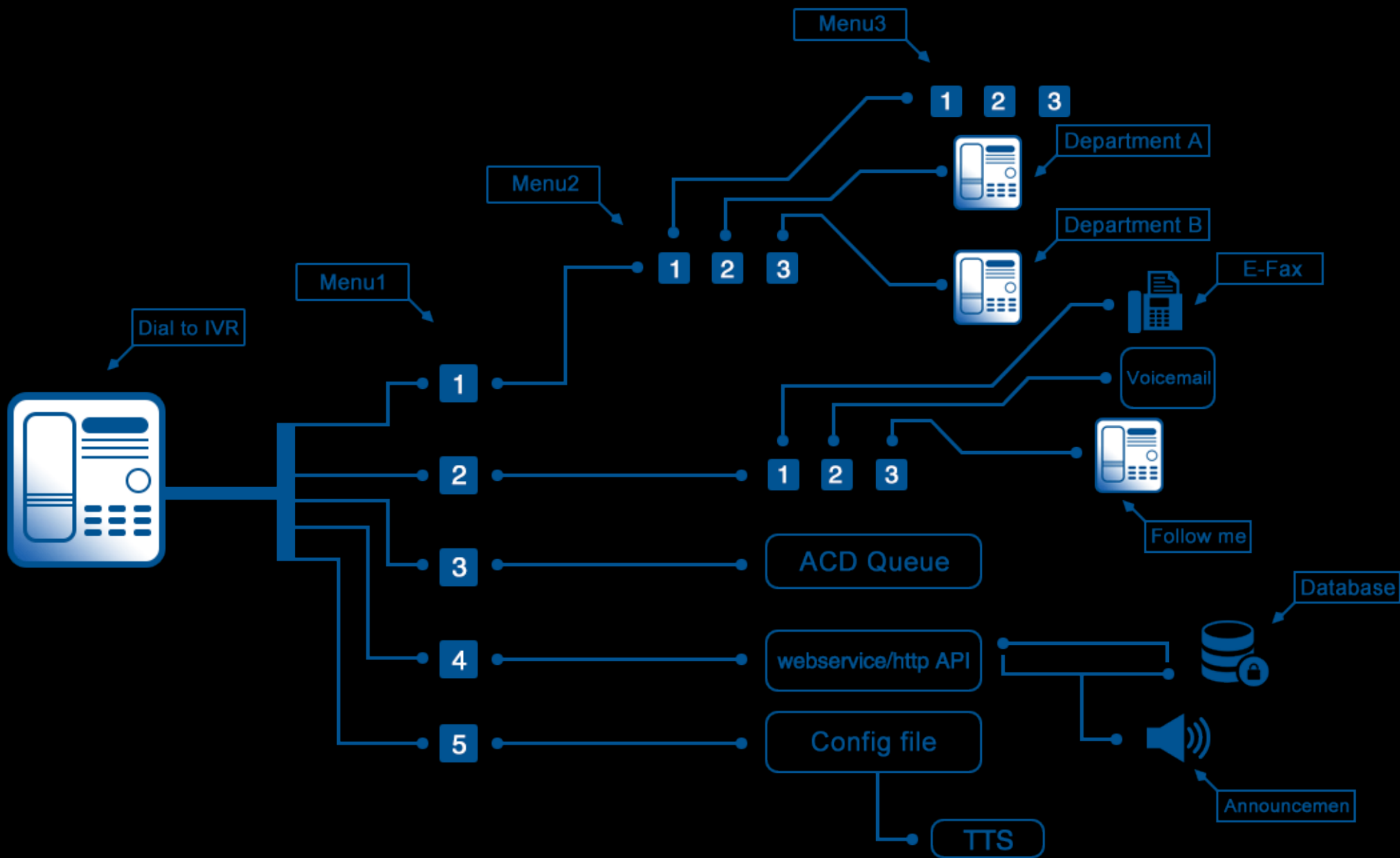
USE CASE BENEFITS:



Enhance Customer Experience with automatic routing that prevents customers from being buried in a hierarchical menu



Reduce Average Handle Time for optimized efficiency





60%

of callers bypass IVR for voice
(costs are 12x higher because of this)



BUSINESS CHALLENGES

Improve Customer Experience

Faster call routing to the correct agent means **more satisfied customers**

Reduce Costs: lower AHT

Faster call resolution:
Faster routing
+
Routing to correct agent

Enhanced System Agility

Higher agility when adding new categories or departments



TECHNICAL CHALLENGES

Performance

Event Driven Architecture based on prediction criteria is required for optimal performance supporting peak events

Simplification

Leveraging existing opensource frameworks such as BigDL in a unified platform simplifies architectural complexity

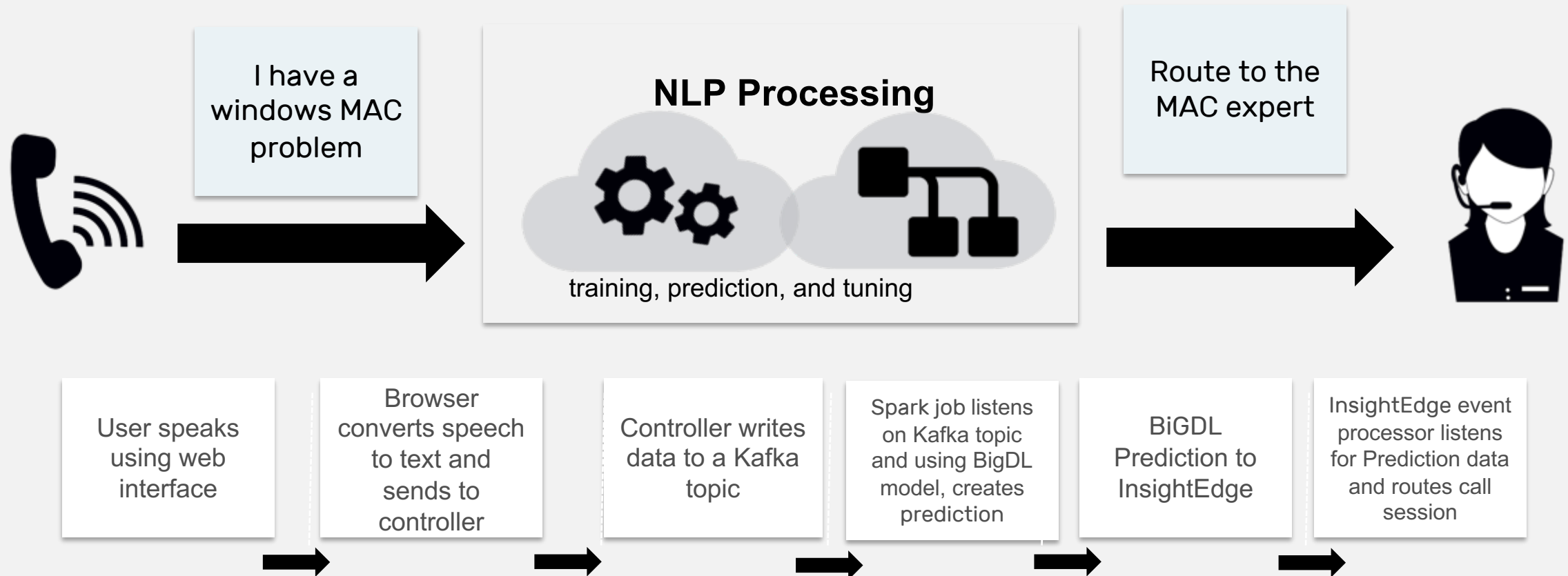
Continuous ML Training

Continuous model training based on previous transcribed calls + automatic training of alternative models ensure models with higher scoring



USE CASE: CALL CENTER ROUTING


Automatic routing to the right agent for the perfect personalized experience





Automatic Call Routing – Live Demo

Call session assistant



See the brain behind it!

Insightedge Web-UI

Spark Jobs

Model: TextClassification

Training time: 9 min

Accuracy: 0.7778865


BigDL v0.2.0

Search...

Help

Call Center BigDL/InsightEdge module

Hello it's almost evening and my Windows computer stopped working



In-process calls (Powered by Intel BigDL) 0

Id	Text
----	------

Call sessions 3

Id	Category	Agent Id	Time (ms)	Text
3	comp.os.ms-windows.misc	4	36	Hello it's almost evening and my Windows computer stopped working
2	comp.sys.mac.hardware	4	59	Hello it's a sunny day but my MacBook screen now just been broken
1	comp.sys.mac.hardware	1	86	Hello I have a problem with my Mac



WEB APPLICATION

CALL VOICE
RECOGNITION UI

CALL SESSION
ROUTING UI

SPEECH



- IN-PROCESS CALLS
- CLASSIFIED CALLS



STREAMING JOB



CLASSIFICATION
ALGORITHM

CLASSIFIED SPEECH

IN-MEMORY
MULTI-MODEL STORE



RESULTS

Performance

50ms

to route the call to the correct agent

Accuracy

75%-85%

model accuracy

Continuous Training

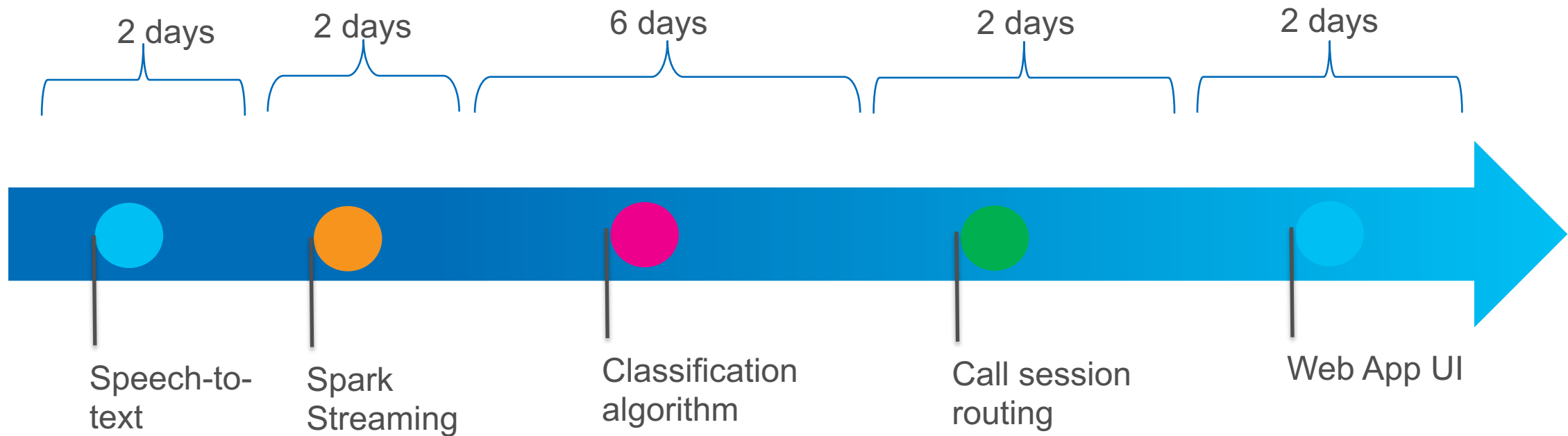
10 mins

Background processing and training to create a new model



Timeline to implement

end-to-end in 14 days





Real-time Insights and Actions Require High Performance

- Unifies analytics, AI and real-time transactions
- Triggers transactional workflows based on prediction criteria and scoring
- Efficient scale-out computing
- Distributed model training
- Lowers TCO/Decreases Deployment Costs – train and run large-scale deep learning workloads
- High performance



THANK YOU



@taldoron



taldoron84



tald@gigaspace.com



Tal Doron

Director, Technology
Innovation