GOODBYE, DATA LAKE:

WHY CONTINUOUS ANALYTICS YIELD HIGHER ROI

STAC SUMMIT LONDON, 2018
ORI MODAI, VP R&D @IGUAZIO

iguazio

Business Challenges in The Digital World

Bringing new (online) customers

Reducing operational costs and risks
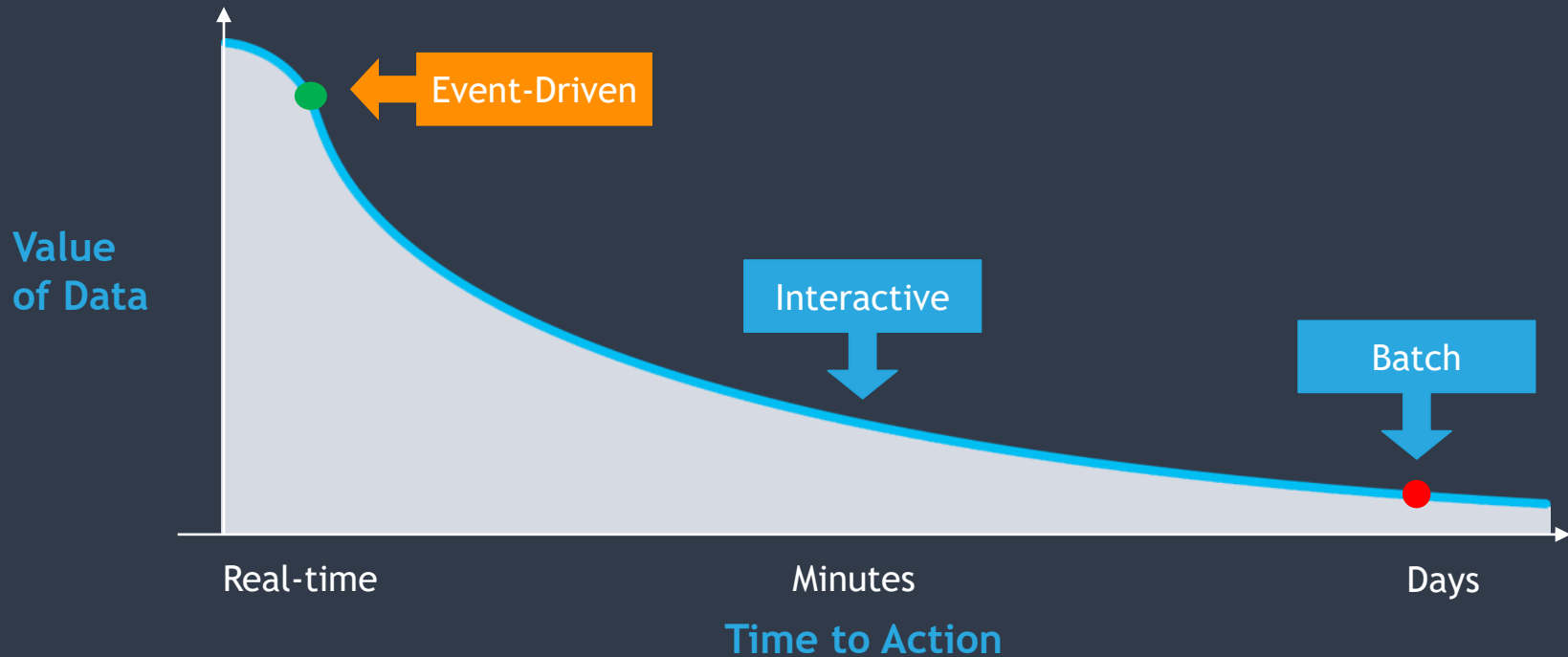
Increasing customers engagement

- Fast and intelligent data-driven actions

- Rapid and continuous innovation

- Building for scale and distribution

iguazio

The Data-Driven Business Challenge
From Reactive to Proactive

Value of Data (y-axis)

Time to Action (x-axis)

Event-Driven

Interactive

Batch

Real-time — Minutes — Days

"Data is the new oil," Fortune 2016   "Data is coal," Amazon's Neil Lawrence
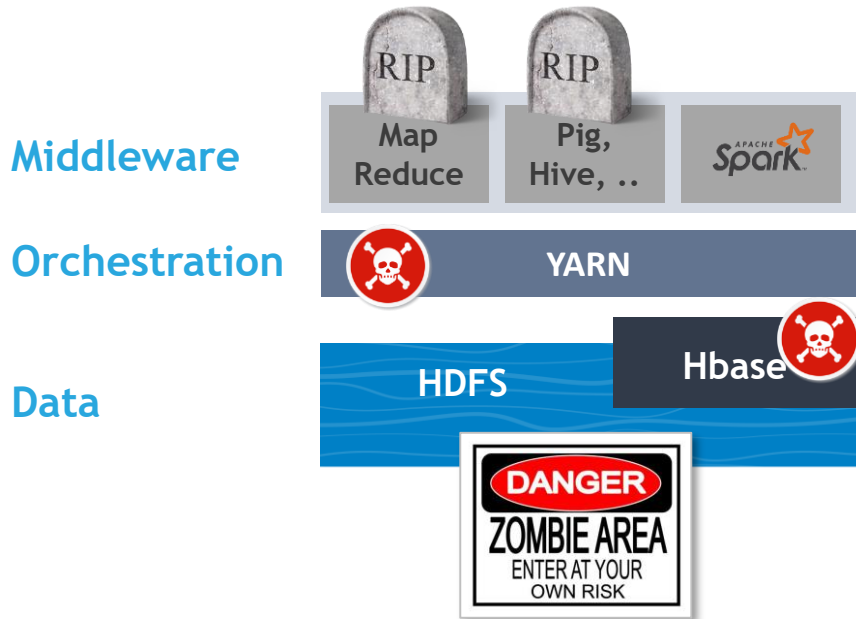
iguazio

Evolve Into a Future Proof Cloud-Native Architecture

Once upon a time there was a data lake
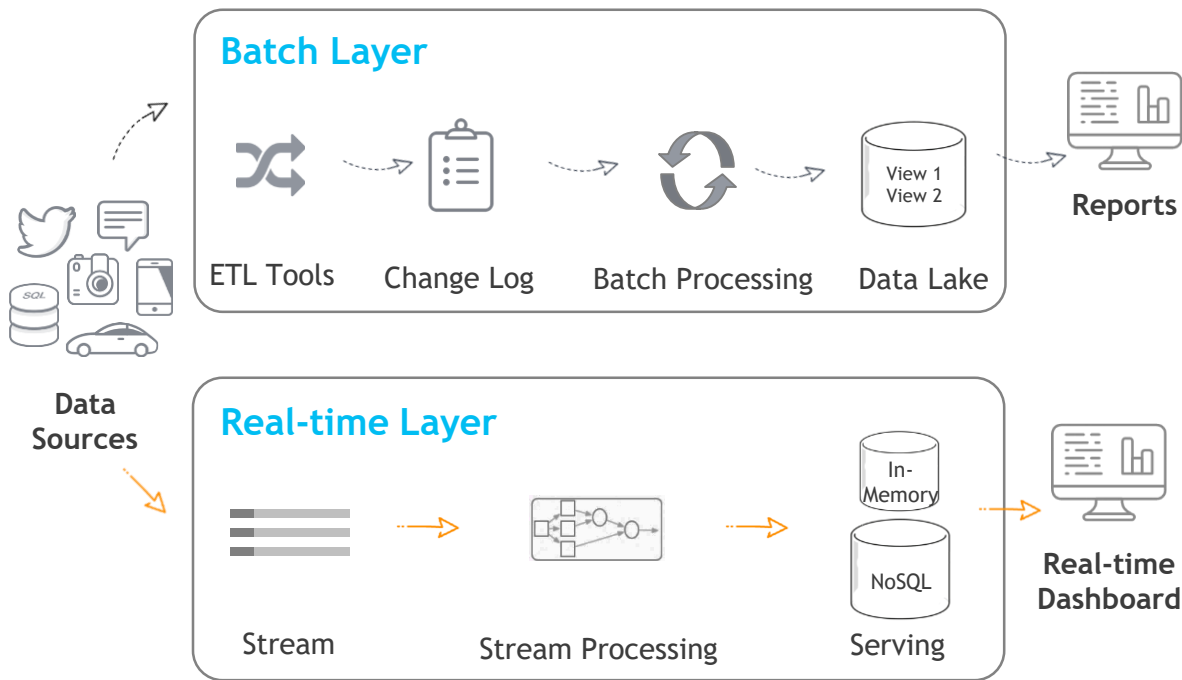(You spent years feeding it with DevOps)

Meanwhile in the cloud

Innovate

Consume

Middleware

RIP — Map Reduce

RIP — Pig, Hive, ..

Spark

Your Business Logic

Orchestration

YARN

kubernetes

Data

HDFS

Hbase

DANGER
ZOMBIE AREA
ENTER AT YOUR OWN RISK

S3 (object)

DBaaS

4

iguazio

Big and Slow or Small and Fast

Batch Layer

ETL Tools   Change Log   Batch Processing   Data Lake

View 1
View 2

Reports

Data Sources

Real-time Layer

Stream   Stream Processing   Serving

In-Memory

NoSQL

Real-time Dashboard

Too slow
- Big data but slow
- Not up to date
- Complex

OR

Limited context
- Small amounts of data
- Expensive
- Lacks context

iguazio

Addressing The Digital Transformation Challenges

Making AI Real-time

"Real-time data integration and access to data remain core challenges for data and analytics leaders looking to modernize data management ecosystems." - Gartner

One Solution Across Cloud and Edge

"Real-time decision-making and interaction requirements; growth in data being produced at the edge; and requirements for autonomy, security and privacy will increase the percentage of compute and storage capability that's closer to people, things and the edge." - Gartner
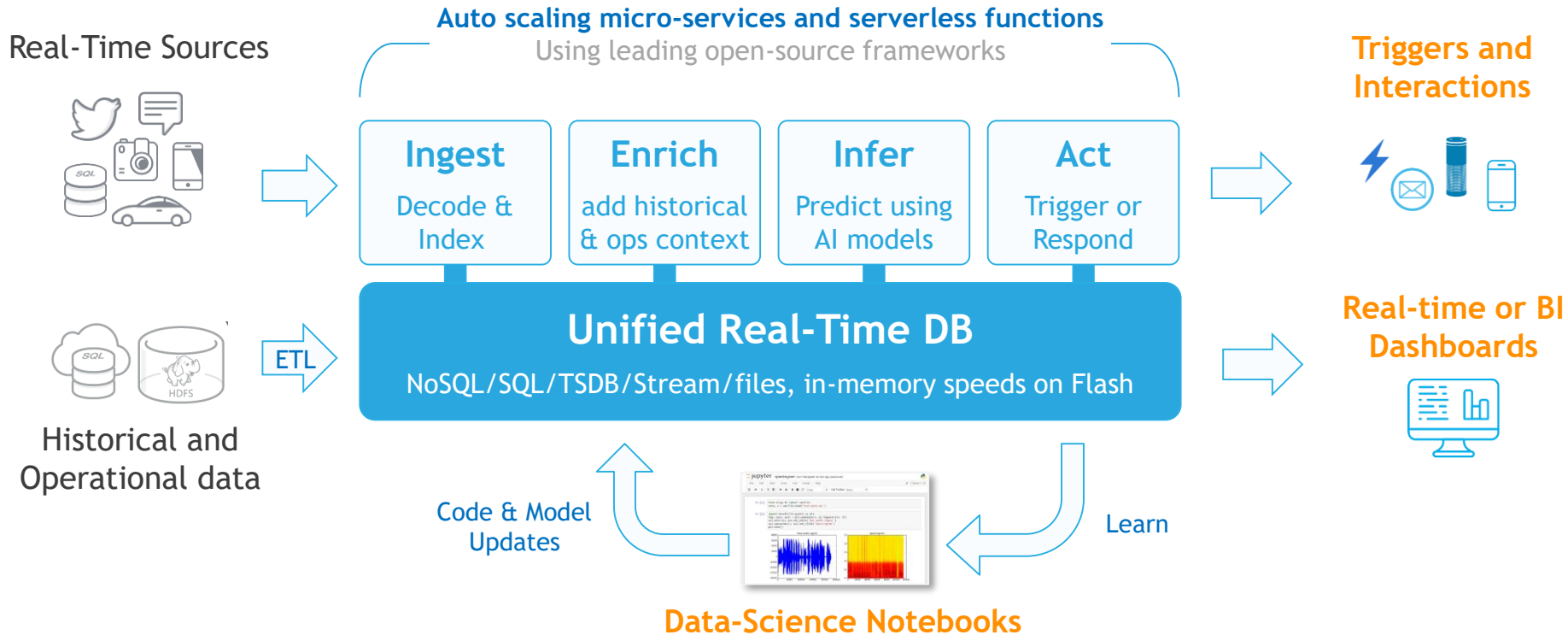
Simplified Dev&Ops with Serverless

"No need to spend time and resources on server provisioning, maintenance, updates, scaling, and capacity planning. Instead, all of these tasks and capabilities are handled by a serverless"
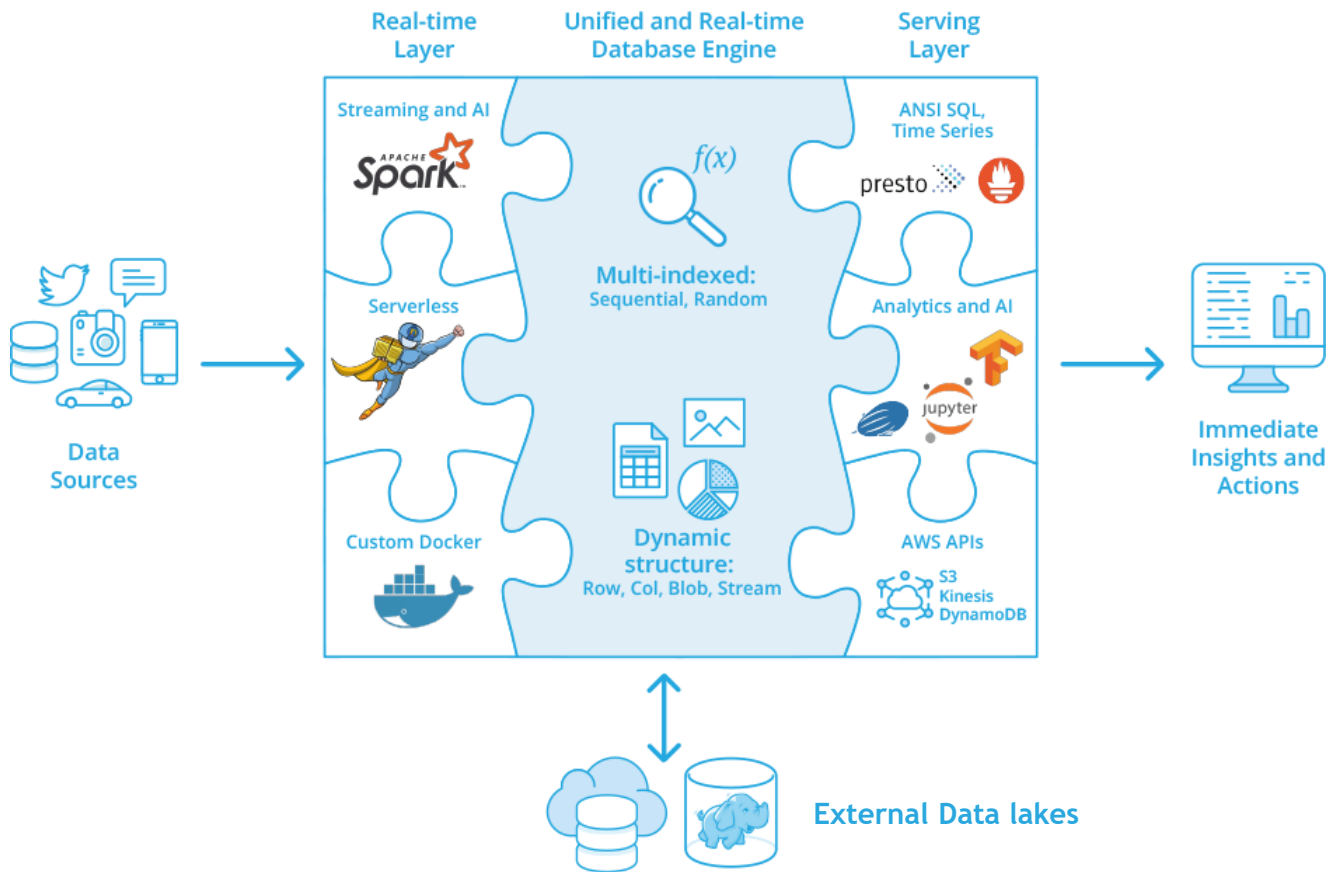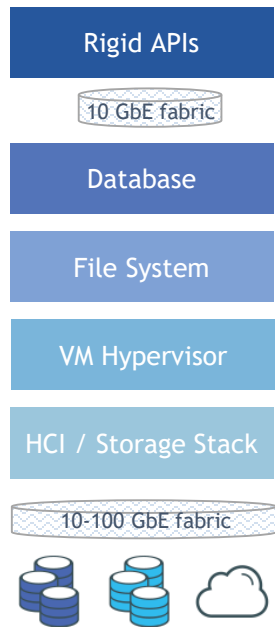*- CNCF Serverless Paper*

iguazio

The Real-Time AI Pipeline

Real-Time Sources

Auto scaling micro-services and serverless functions
Using leading open-source frameworks

Triggers and Interactions

Ingest Decode & Index	**Enrich** add historical & ops context	**Infer** Predict using AI models	**Act** Trigger or Respond

Unified Real-Time DB
NoSQL/SQL/TSDB/Stream/files, in-memory speeds on Flash

Historical and Operational data

ETL

Real-time or BI Dashboards

Code & Model Updates

Learn

Data-Science Notebooks

iguazio

Ingest, Enrich, AI, and Serve on One DB Engine

Real-time Layer

Unified and Real-time Database Engine

Serving Layer

Streaming and AI

f(x)

ANSI SQL, Time Series

Multi-indexed: Sequential, Random

Serverless

Analytics and AI

Dynamic structure: Row, Col, Blob, Stream

Custom Docker

AWS APIs
S3
Kinesis
DynamoDB

Data Sources

Immediate Insights and Actions

External Data lakes

8

iguazio

How To Deliver Volume, Velocity and Variety ?

Traditional Layered Approach

Rigid APIs

10 GbE fabric

Database

File System

VM Hypervisor

HCI / Storage Stack

10-100 GbE fabric

External (NVMeOF / Object)

- Slow
- Complex
- Expensive

Optimized Approach

Apps, APIs, and Functions

100 GbE fabric

Real-time Firewall

Real-Time DB

100TB NVMe Flash (direct attached)

Decouple the APIs from the DB Engine

Use NVMe Flash as an extension of OS Memory

- In-memory speed
- Simple
- 1/3rd the TCO

iguazio

REAL-TIME DATA PIPELINES

CUSTOMER USE-CASES

iguazio

Latency Prevention for Electronic Trading Platforms

Global banks deployed Iguazio to predict and avoid latencies
in their electronic trading platforms

- Replaced traditional latency monitoring solutions – enforcing trading latencies SLO
- Iguazio has changed it to:
  - o Real-time predictions latencies trend
  - o Root cause analysis via multivariate analysis
  - o Actions triggered via serverless functions
  - o Trading stays continuous and uninterrupted
- Outcome – Improved SLOs, reduced trading down times

iguazio

Data Flow

Network Monitoring
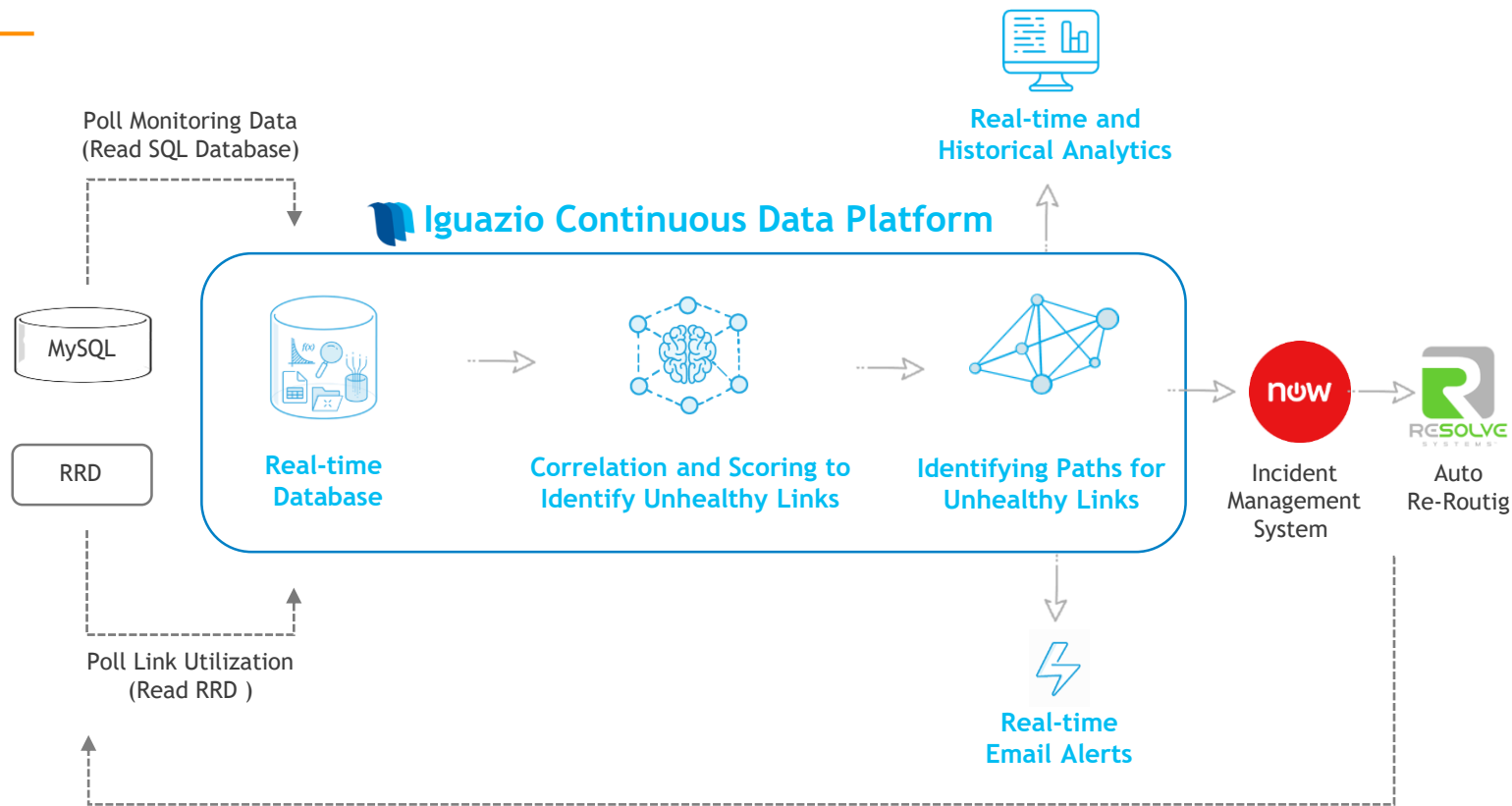Data, Latencies, etc...

Iguazio Continuous Data Platform

Monitoring systems

Netflows

Collectors

Syslogs

Services Level
Events, Alerts, Data

**Real-time
Database**

**Cross Correlation to
Predict SLO Deviations**

**Identifying Root
Cause**

**Real-time and
Historical Analytics**

**Real-time
Email Alerts**

*REACT in
10s of Seconds*
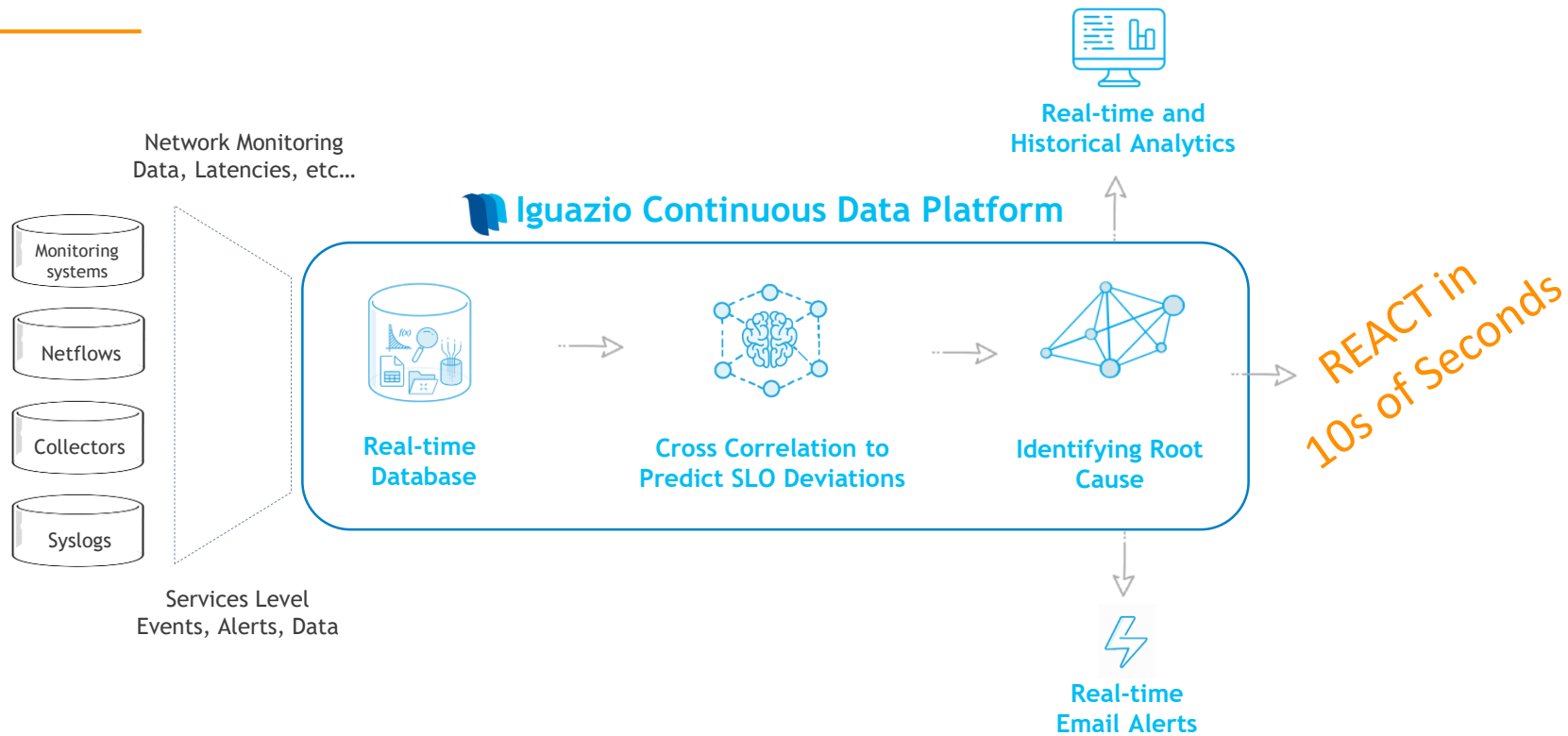
iguazio

Auto Healing Network Operations

Global network operator uses Iguazio to predict and avoid network outages in real-time

- Replaced a Hadoop data pipeline (that was never productized)

- Real-time cross correlation of data from multiple sources

- AI based predictions triggering pre-programmed network optimization algorithms
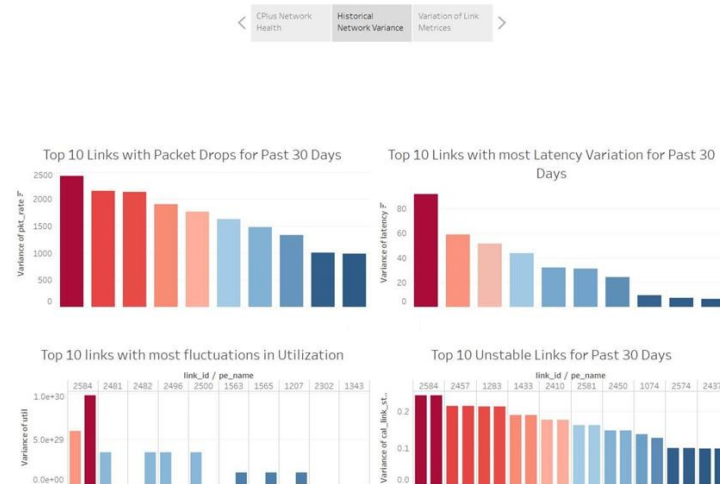
- Time to production < 4 weeks
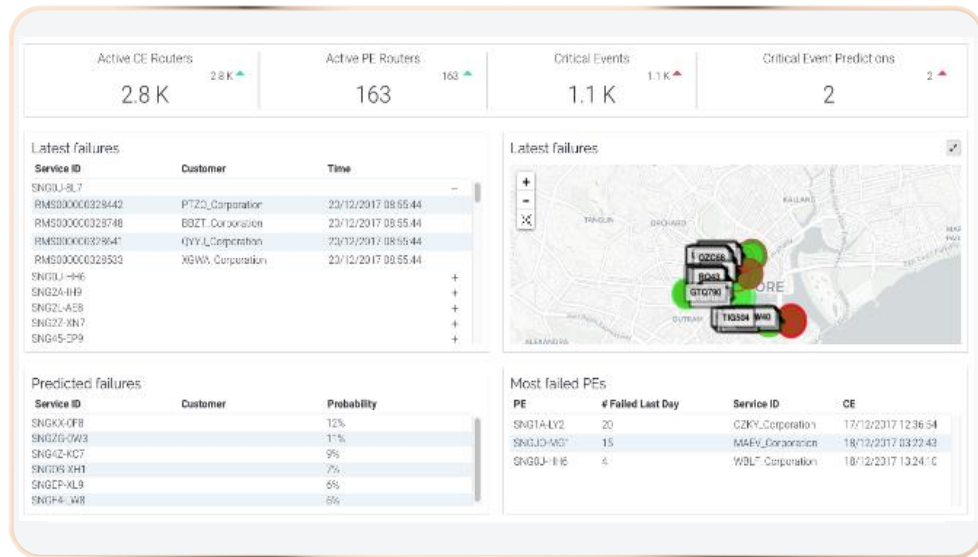
Data Flow

Poll Monitoring Data
(Read SQL Database)

Real-time and
Historical Analytics

Iguazio Continuous Data Platform

MySQL

RRD

**Real-time
Database**

**Correlation and Scoring to
Identify Unhealthy Links**

**Identifying Paths for
Unhealthy Links**

Incident
Management
System

Auto
Re-Routig

Poll Link Utilization
(Read RRD)

**Real-time
Email Alerts**

iguazio

Real time dashboards

15

Multivariate Real-time Analysis of Trade Data

Major financial institute instruments its trade data ML research with Iguazio

- Orchestrating multiple data feeds and data modalities

- Multi dimensional feature vectors aggregation

- Enabling ML research activities on operational recent data

- Advanced serving and dash-boarding from the same platform

- Time to POC ~ 2 weeks
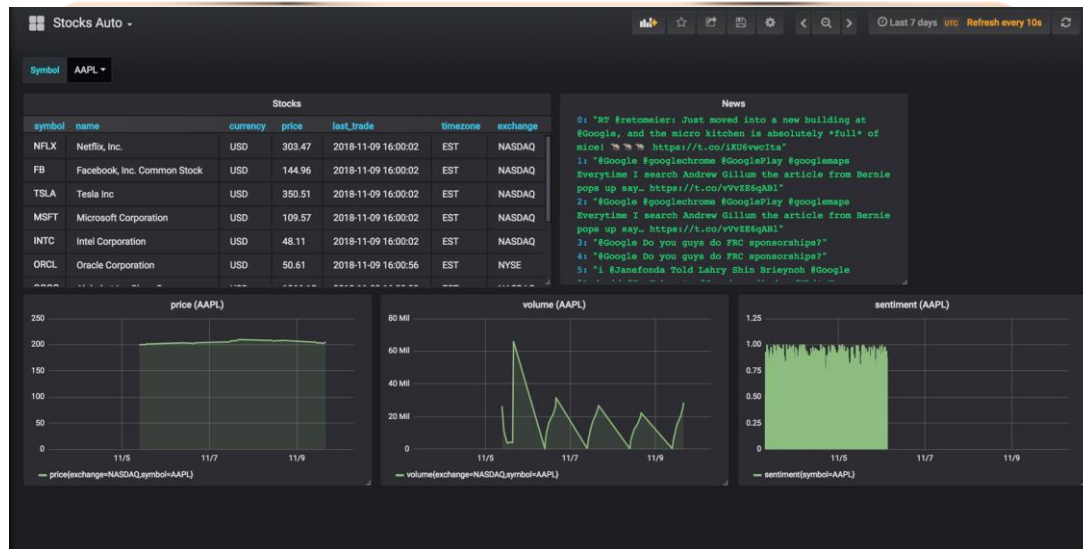
Real-Time Analysis of Trade Data
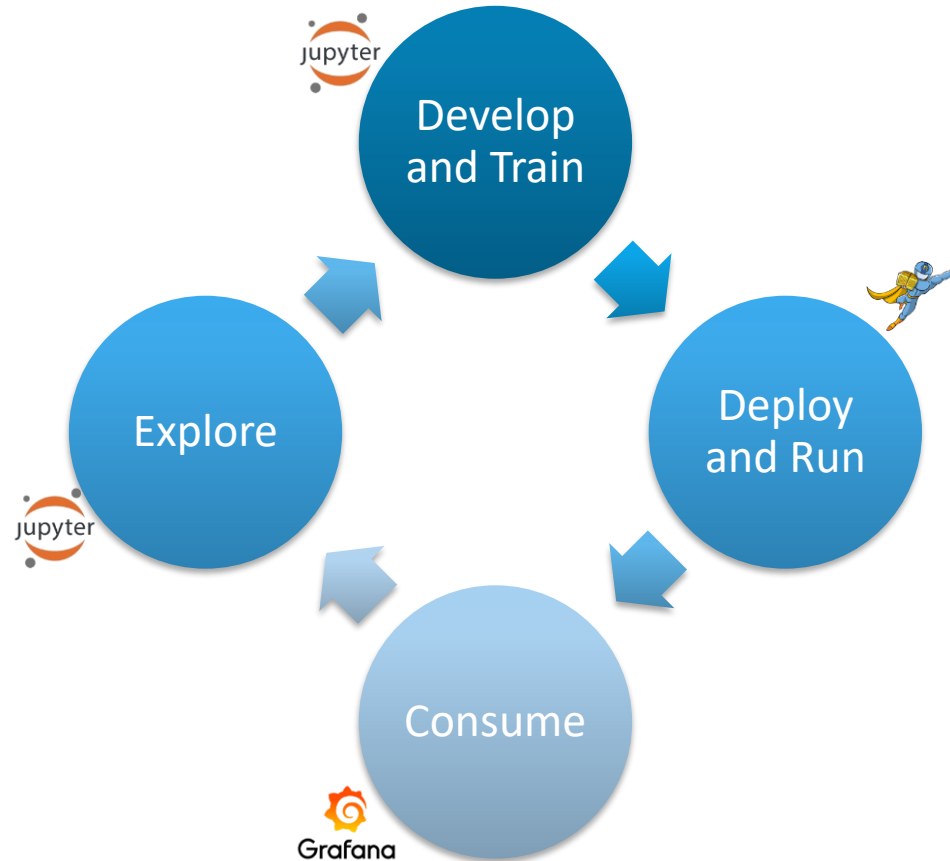
Real-Time Analysis of Trade Data

Real-time functions for
Ingestion, decoding and tagging

twitter → Tweet Sentiment Analysis

N
News API → News Analysis & Tagging

World Trading Data → Tick feed Analysis & Tagging

News stream viewer →

Grafana
Real-time Dashboard

jupyter
Real-time Analysis

Explore Real-Time Data on Iguazio via Integrated Notebooks

Analyze Real-Time Data Using Spark Streaming, SQL, and ML

Run Interactive SQL Queries on Real-Time Data

Read NoSQL Data as Real-Time Data Frame Stream

Read the Time-Series data (in Iguazio TSDB) as Pandas Data Frame

Iguazio Unified Real-Time DB Service
SQL | NoSQL | TSDB | Streams | Obj/files

Functions/Model Dev & Test

jupyter

deploy

iguazio

Summary

Build continuous, data-driven and proactive apps

- Deliver real-time analytics based on fresh and historical data

- Utilize Flash to deliver in-memory speed at much lower costs

- Create a unified data layer for stream processing, AI and serving

- Adopt cloud-native and serverless approaches to gain agility

- Enable your data scientists with access to fresh, large scale data

iguazio

Thank You

orim@iguazio.com | www.iguazio.com