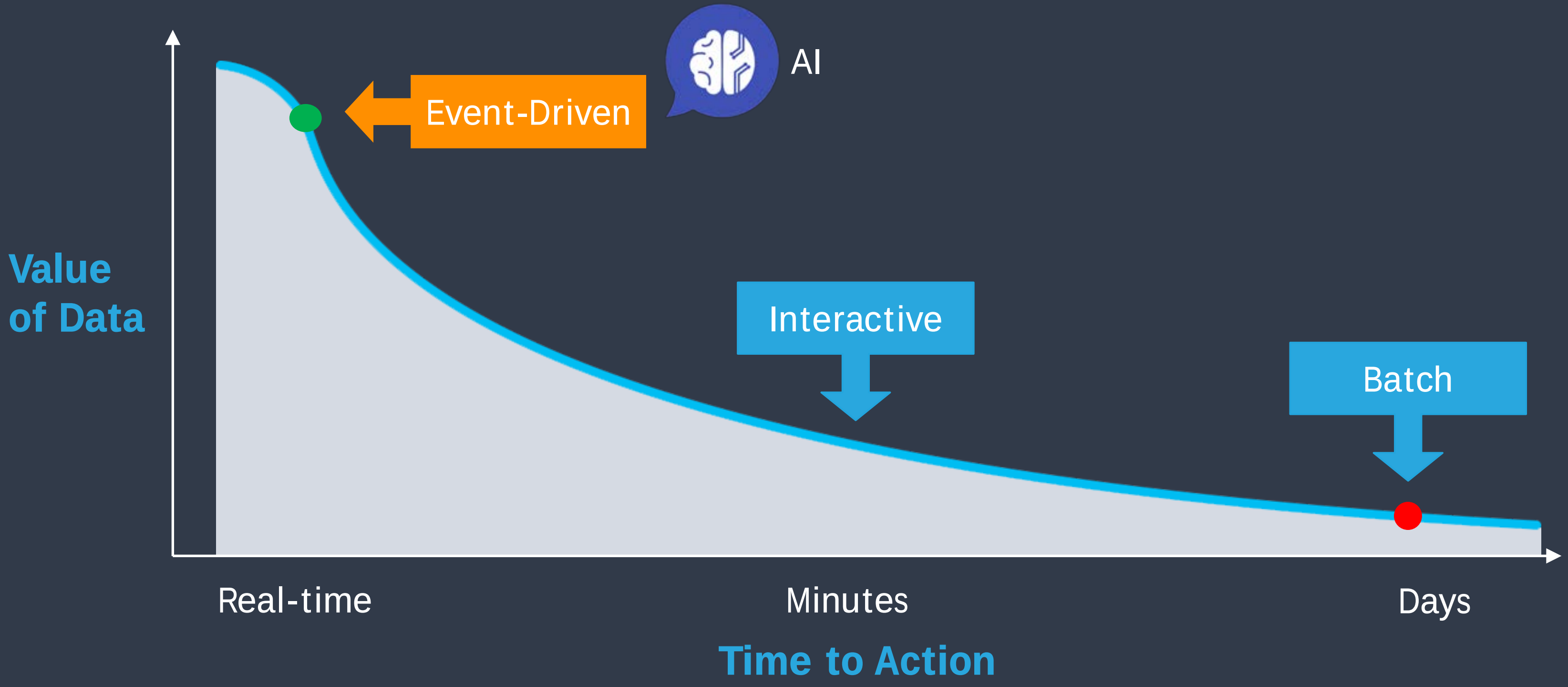
iguazio

In-mem DB Performance, Flash Cost
Enabling Real-time AI

June 2018

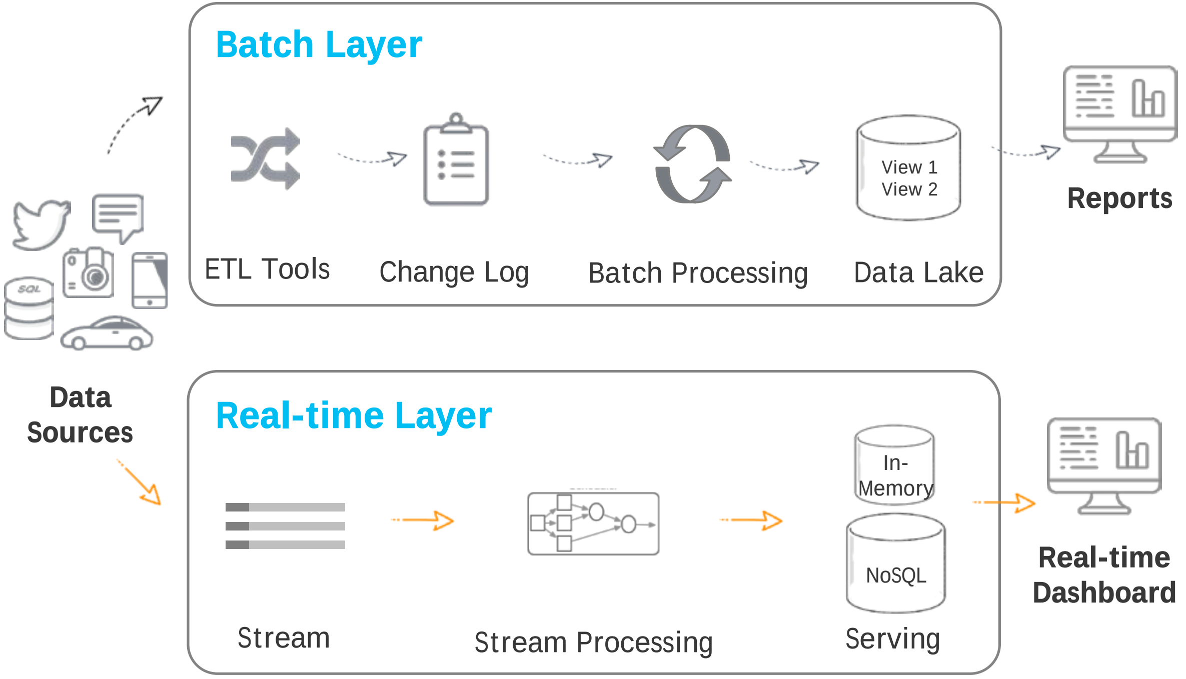
The Data-Driven Business Challenge

From Reactive to Proactive



iguazio

Big and Slow or Small and Fast



Too slow

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

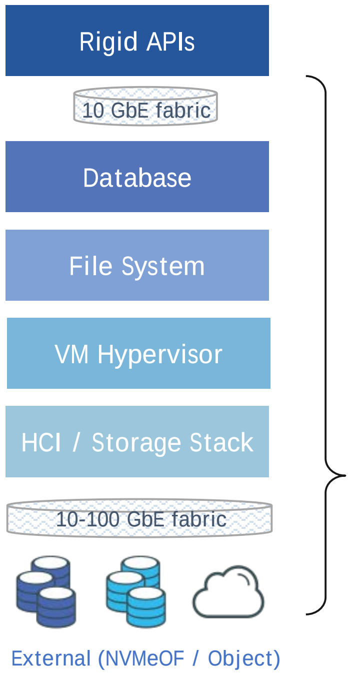
OR

Limited context

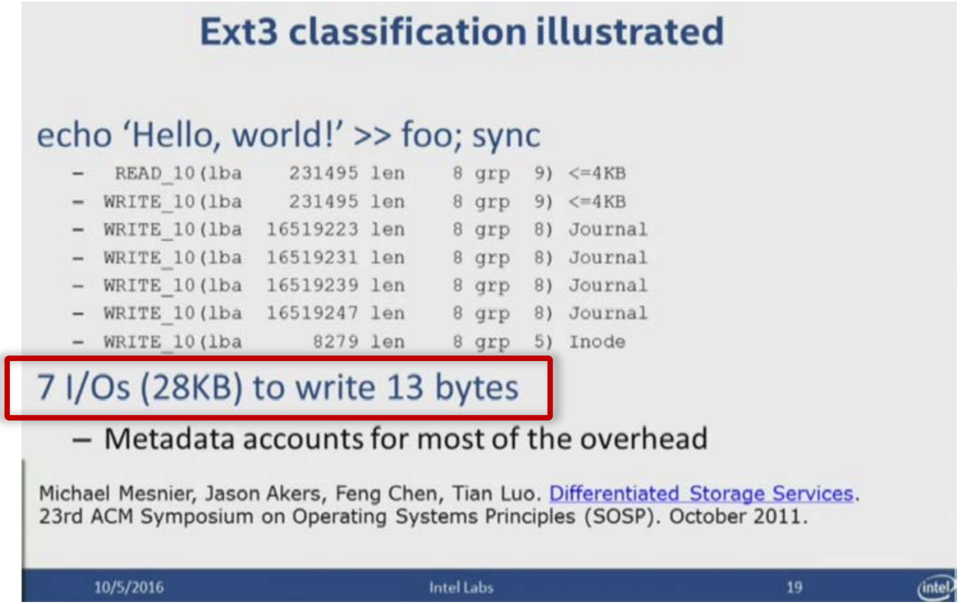
- Small amounts of data
- Expensive
- Lacks context

Traditional Approach, DB over File over Flash

Traditional Layered Approach



- Slow
- Complex
- Expensive



Ext3 classification illustrated

echo 'Hello, world!' >> foo; sync

```
–   READ_10(lba    231495 len   8 grp  9) <=4KB
–  WRITE_10(lba    231495 len   8 grp  9) <=4KB
–  WRITE_10(lba  16519223 len   8 grp  8) Journal
–  WRITE_10(lba  16519231 len   8 grp  8) Journal
–  WRITE_10(lba  16519239 len   8 grp  8) Journal
–  WRITE_10(lba  16519247 len   8 grp  8) Journal
–  WRITE_10(lba      8279 len   8 grp  5) Inode
```

7 I/Os (28KB) to write 13 bytes

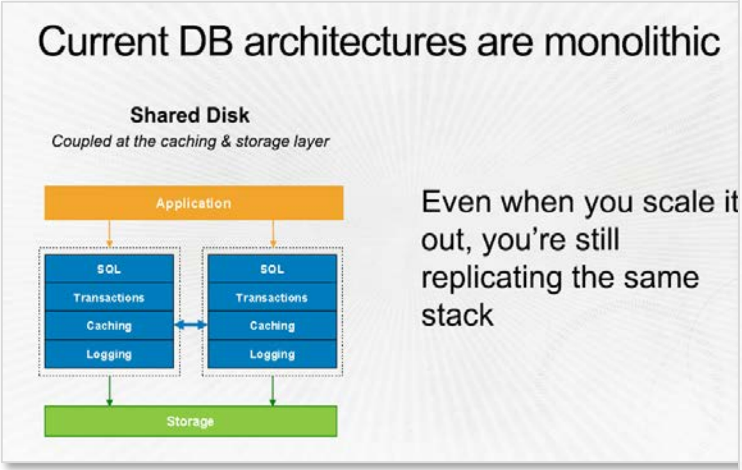
– Metadata accounts for most of the overhead

Michael Mesnier, Jason Akers, Feng Chen, Tian Luo. Differentiated Storage Services. 23rd ACM Symposium on Operating Systems Principles (SOSP). October 2011.

10/5/2016 Intel Labs 19

For every file IOs conducted by the DB
(Record, Redo/Undo, Metadata, ..)

New Cloud Databases Are Built to Scale Ops & Capacity

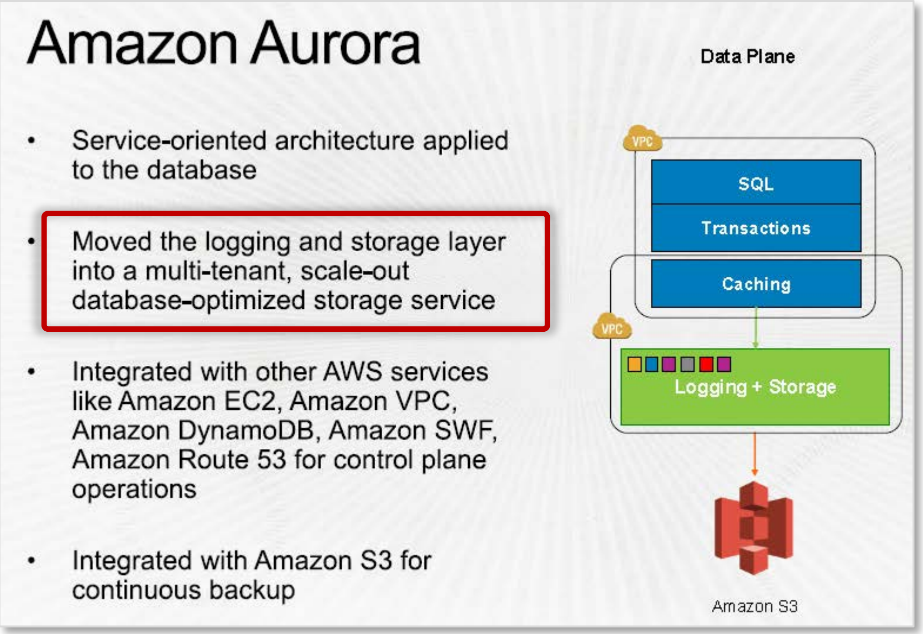


Current DB architectures are monolithic

Shared Disk

Coupled at the caching & storage layer

Even when you scale it out, you're still replicating the same stack



Amazon Aurora

- Service-oriented architecture applied to the database
- Moved the logging and storage layer into a multi-tenant, scale-out database-optimized storage service
- Integrated with other AWS services like Amazon EC2, Amazon VPC, Amazon DynamoDB, Amazon SWF, Amazon Route 53 for control plane operations
- Integrated with Amazon S3 for continuous backup

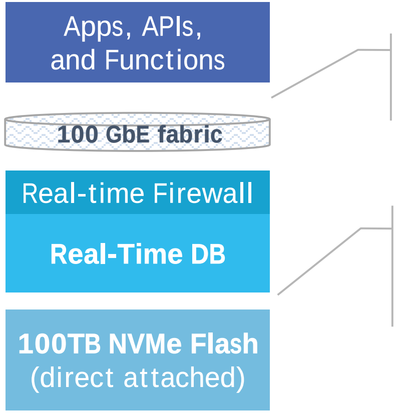
API & Transaction

Distributed Processing & Cache

Capacity (Object)

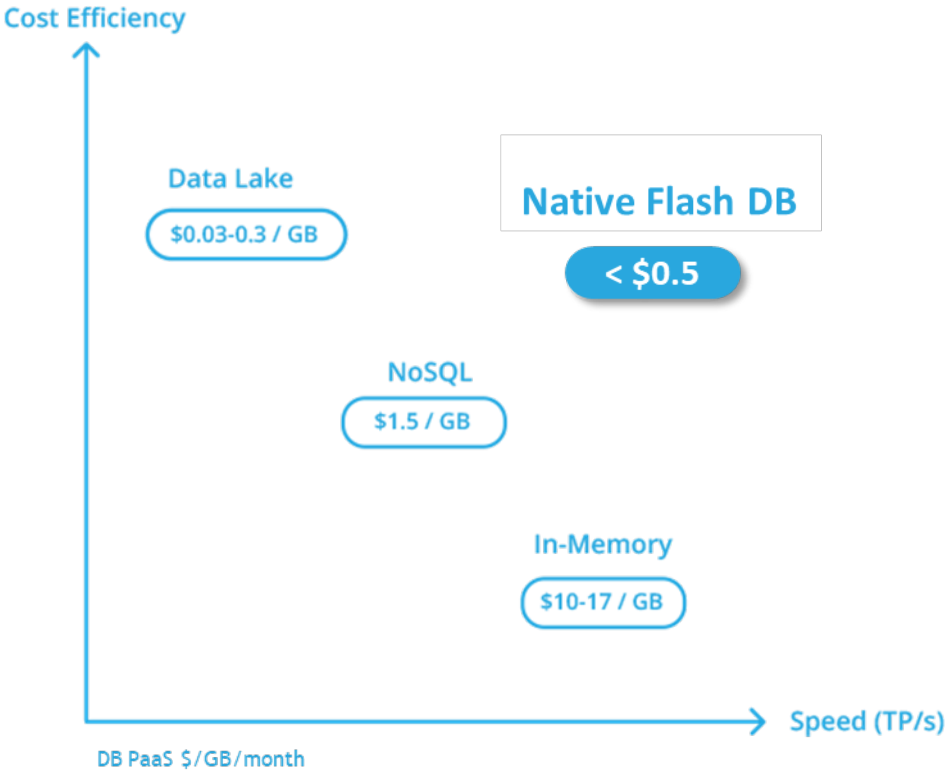
Decouple access, processing, and capacity and eliminate storage serialization

Breaking The Volume and Velocity Barrier



Support many standard APIs on a common DB Engine

Unique architecture which use NVMe Flash as an extension of OS Memory



Re-engineer the stack to deliver memory speed with Flash density

Breaking Performance Barriers – Design Principles

Never blocking, never locking, 100% parallelism

Latency optimized, QoS aware, data scheduler

Lockless, preempt less memory management

True scale out through parallelism

Zero processing wastes

CPU cache optimization and prediction

E2E zero buffer data flow (NIC to Disk, accelio)

Complete OS bypass

HW awareness

RDMA, NVMe (3DXP)

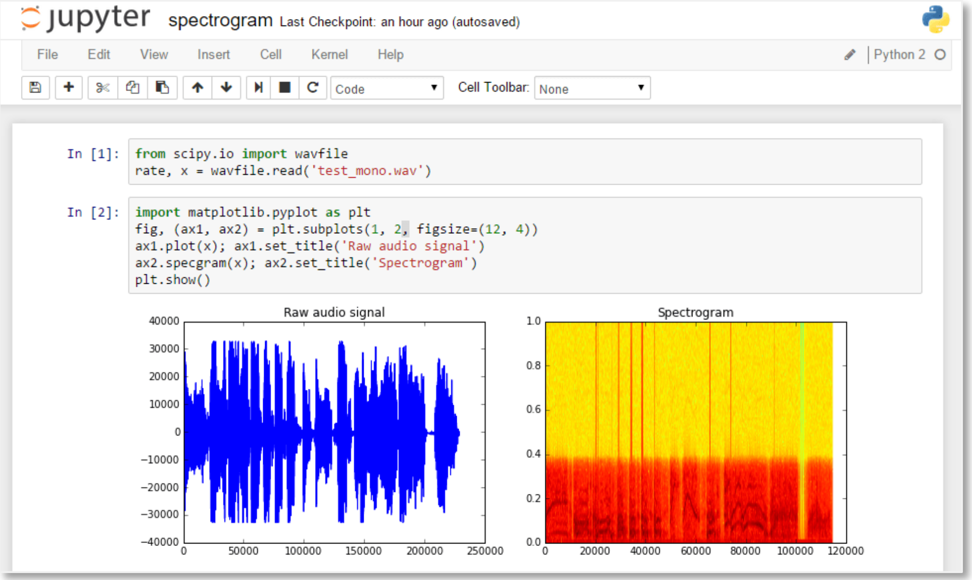
Vector processing operations

IRQ balancing and throttling

Ok, any other challenges on the way to real-time AI ?

90% of AI Today

Build feature vectors using batch and CSVs



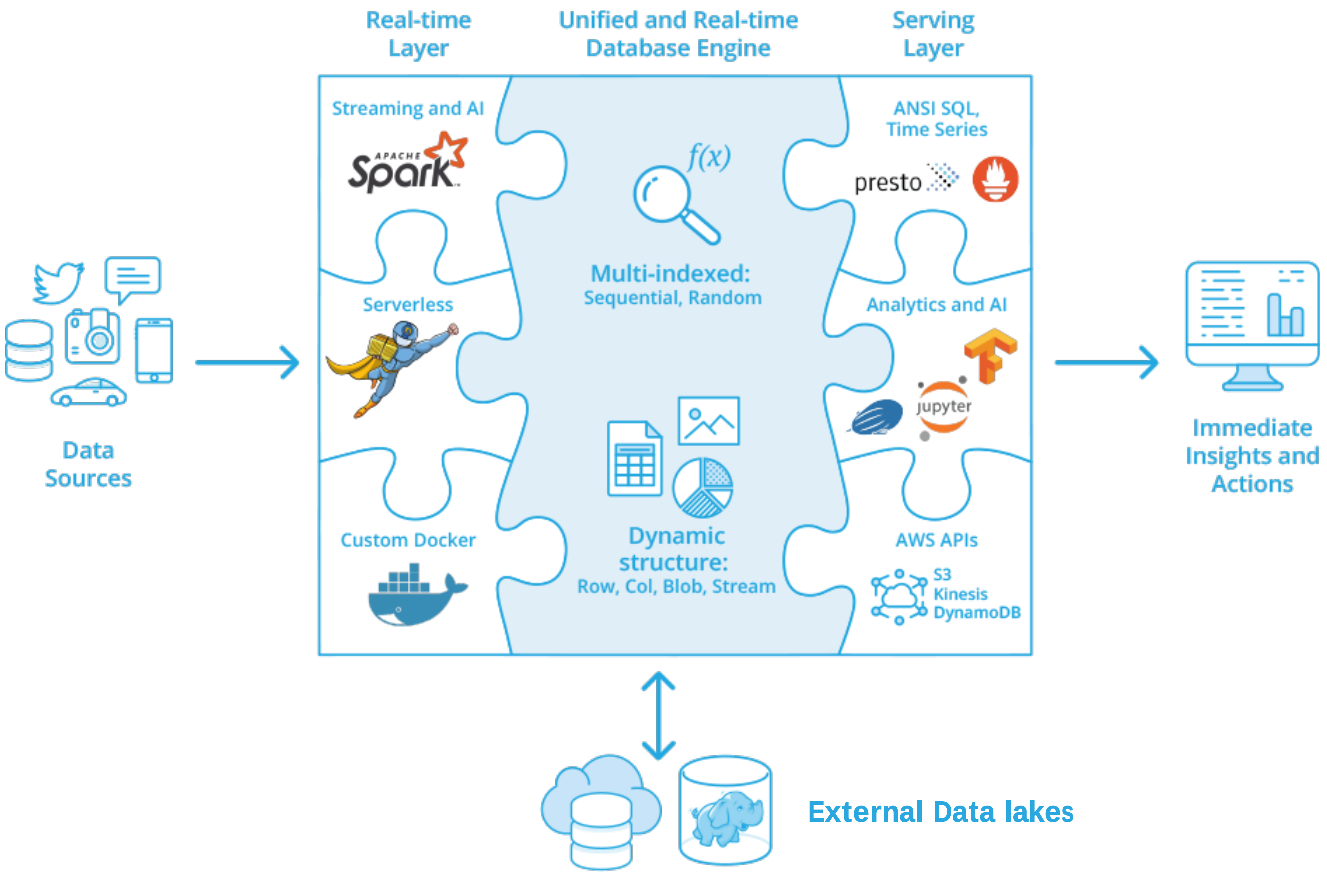
```
In [1]: from scipy.io import wavfile
        rate, x = wavfile.read('test_mono.wav')

In [2]: import matplotlib.pyplot as plt
        fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 4))
        ax1.plot(x); ax1.set_title('Raw audio signal')
        ax2.specgram(x); ax2.set_title('Spectrogram')
        plt.show()
```

Inspect, Improve

How do we form complex feature vectors in real-time?

How do we visualize or act on the results in real-time?

Moving to Continuous Ingest + AI + Serve Flow



Data Sources

Real-time Layer

Unified and Real-time Database Engine

Serving Layer

Streaming and AI

Serverless

Custom Docker

Multi-indexed:
Sequential, Random

Dynamic structure:
Row, Col, Blob, Stream

ANSI SQL, Time Series

Analytics and AI

AWS APIs

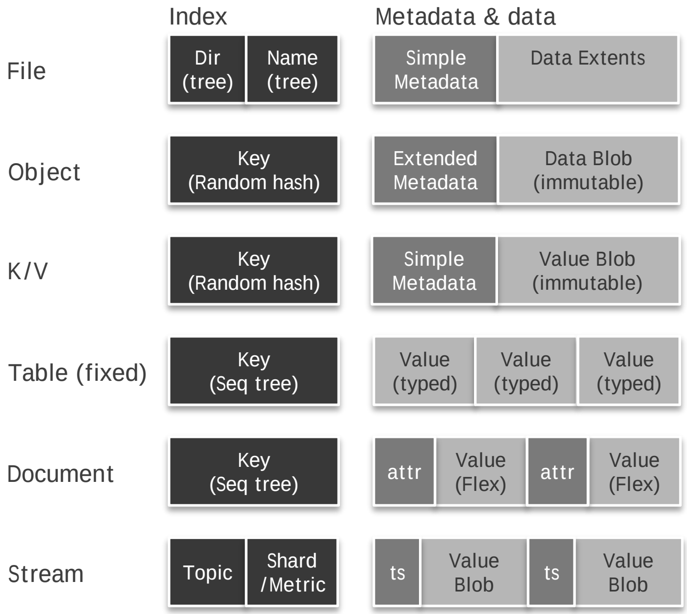
S3
Kinesis
DynamoDB

Immediate Insights and Actions

External Data lakes

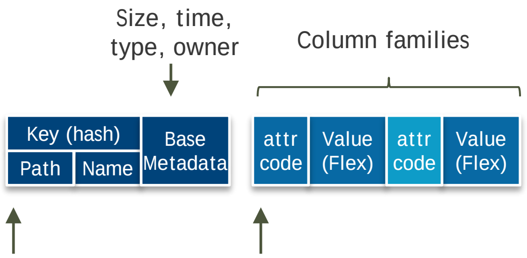
From Silos and ETLs to All-in-one DBs

Traditional: Unique Model Per Store



	Index	Metadata & data
File	Dir (tree) Name (tree)	Simple Metadata Data Extents
Object	Key (Random hash)	Extended Metadata Data Blob (immutable)
K/V	Key (Random hash)	Simple Metadata Value Blob (immutable)
Table (fixed)	Key (Seq tree)	Value (typed) Value (typed) Value (typed)
Document	Key (Seq tree)	attr Value (Flex) attr Value (Flex)
Stream	Topic Shard /Metric	ts Value Blob ts Value Blob

Multi-Model Store



Multiple Indexes

Random, sequential and hierarchical

Any Data Type

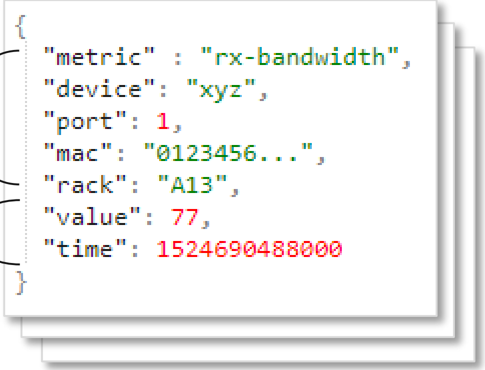
- Nested attributes (encoded)
- Flexible value types
- Can be organized and viewed as extents, rows, cols or logs

Independent tiering logic for indexes, metadata and data

Time Series Data Example

Raw time series sample data

Thousands of samples



```
{
  "metric" : "rx-bandwidth",
  "device": "xyz",
  "port": 1,
  "mac": "0123456...",
  "rack": "A13",
  "value": 77,
  "time": 1524690488000
}
```

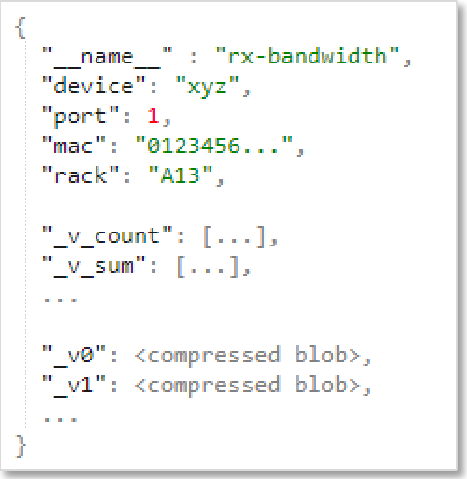
Labels

Data

Ingest / compress

In real-time

Optimized TSDB Layout (per unique metric)



```
{
  "__name__" : "rx-bandwidth",
  "device": "xyz",
  "port": 1,
  "mac": "0123456...",
  "rack": "A13",

  "_v_count": [...],
  "_v_sum": [...],
  ...

  "_v0": <compressed blob>,
  "_v1": <compressed blob>,
  ...
}
```

Filter based on labels

Pre-aggregation arrays:
(to accelerate queries)

T/V chunks with 10:1
Gorilla compression

Real-time
Consistency

50 : 1
Compression

10–100x
Faster Queries

Serverless, The New Stored Procedure

Traditional Dev and Ops Model

- **Write code + local testing**
- Build code and Docker image
- CI/CD pipeline
- Add logging and monitoring
- Harden security
- Provision servers + OS
- Handle data/event feed
- Handle failures/auto-scaling
- Handle rolling upgrades
- Configuration management

80%

“Serverless” Development Model

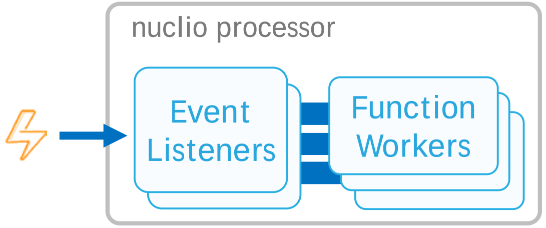
- **Write code + local testing**
- **Provide spec, push deploy**

1. Automated by the serverless platform

2. Pay for what you use

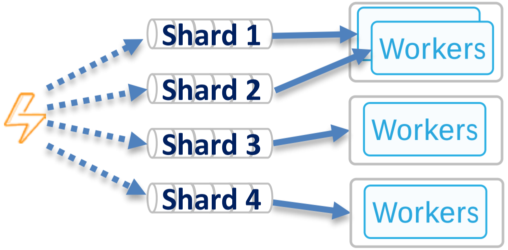
Addressing Serverless Limitations With Nuclio

Performance



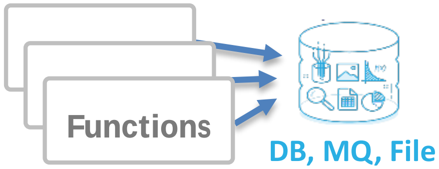
- Non-blocking, parallel
- Zero copy, buffer reuse
- Up to 400K events/sec/proc

Streaming and Batch



- Auto-rebalance, checkpoints
- Any source: Kafka, NATS, Kinesis, event-hub, iguazio, pub/sub, RabbitMQ, Cron

Statefulness

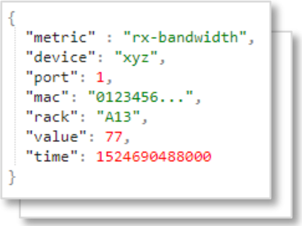


- Data bindings
- Shared volumes
- Context cache

Serverless for compute and data intensive tasks
100x faster than AWS Lambda !

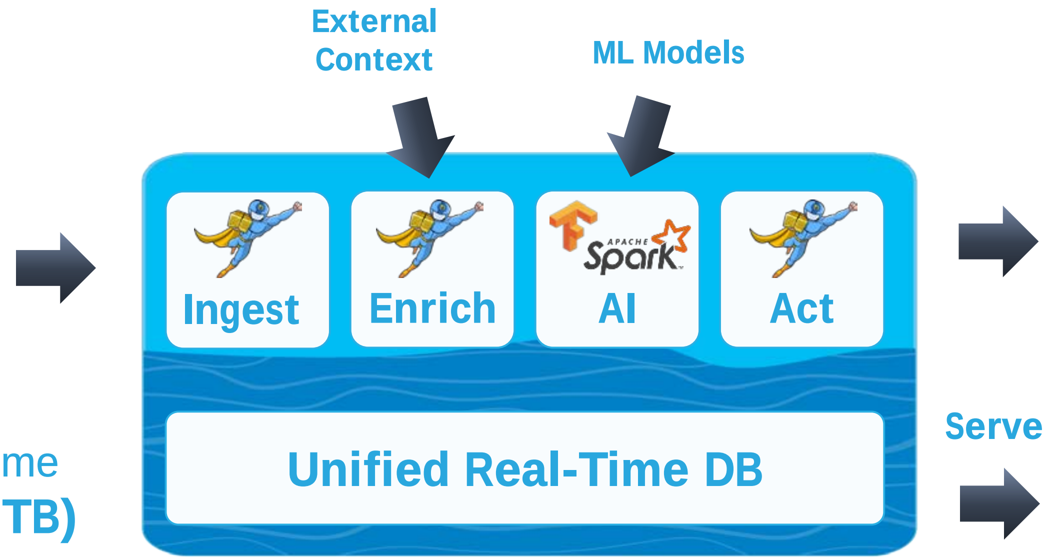
Delivering Intelligent Decisions in Real-Time

500TB of Raw Data



```
{
  "metric" : "rx-bandwidth",
  "device": "xyz",
  "port": 1,
  "mac": "0123456...",
  "rack": "A13",
  "value": 77,
  "time": 1524690488000
}
```

Ingested in real-time
(compressed to **10TB)**



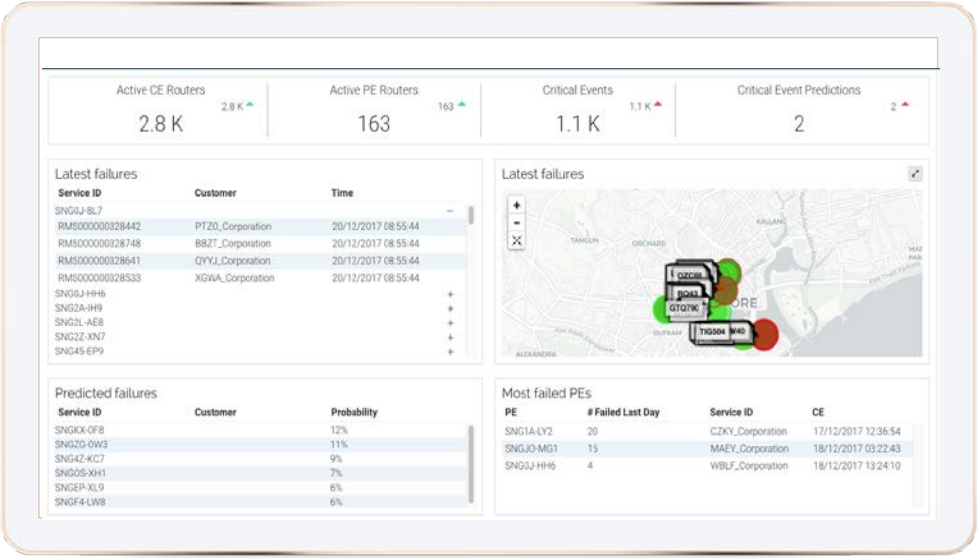
Real-time
triggers

Real-time and historical
dashboards

Cyber and Network Ops

A leading telco needs to predict network behavior in real-time:

- Processing high message throughput from multiple streams at the rate of > 50K events/sec
- Cross correlating with historical and external data in real-time
- AI predictions/inferencing conducted on live data
- Small footprint to fit network locations



Active CE Routers	Active PE Routers	Critical Events	Critical Event Predictions
2.8 K	163	1.1 K	2
2.8 K ▲	163 ▲	1.1 K ▲	2 ▲

Latest failures

Service ID	Customer	Time	
SNG0J-8L7			−
RMS000000328442	PTZ0_Corporation	20/12/2017 08:55:44	
RMS000000328748	BBZT_Corporation	20/12/2017 08:55:44	
RMS000000328641	QYYJ_Corporation	20/12/2017 08:55:44	
RMS000000328533	XGWA_Corporation	20/12/2017 08:55:44	
SNG0J-HH6			+
SNG2A-IH9			+
SNG2L-AE8			+
SNG2Z-XN7			+
SNG45-EP9			+

Latest failures

Predicted failures

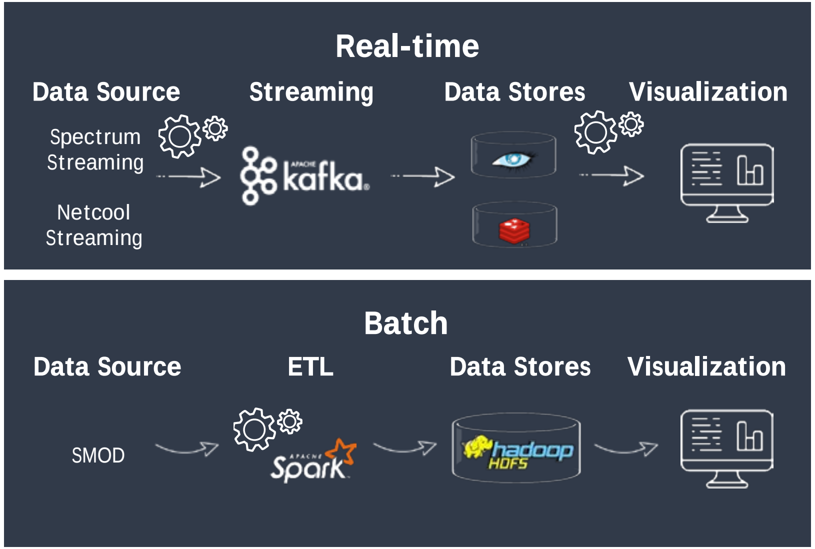
Service ID	Customer	Probability
SNGKX-0F8		12%
SNGZG-0W3		11%
SNG4Z-KC7		9%
SNG0S-XH1		7%
SNGEP-XL9		6%
SNGF4-LW8		6%

Most failed PEs

PE	# Failed Last Day	Service ID	CE
SNG1A-LY2	20	CZKY_Corporation	17/12/2017 12:36:54
SNGJO-MG1	15	MAEV_Corporation	18/12/2017 03:22:43
SNG0J-HH6	4	WBLF_Corporation	18/12/2017 13:24:10

Build and Operationalize Proactive Systems Faster

Traditional



Real-time

Data Source

Streaming

Data Stores

Visualization

Spectrum Streaming

Netcool Streaming

Batch

Data Source

ETL

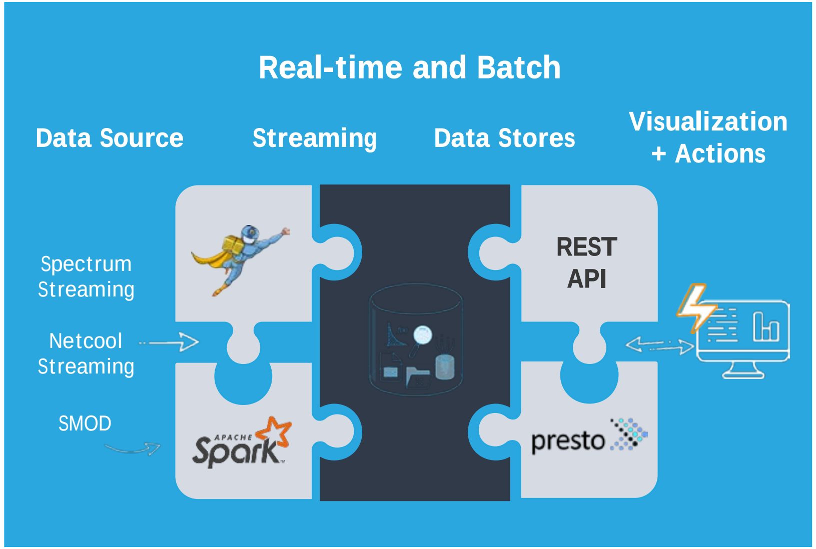
Data Stores

Visualization

SMOD

- Complex, skill gaps, slow to productize
- No single view of ops, real-time, history
- Reactive (no actions)

Continuous Analytics



Real-time and Batch

Data Source

Streaming

Data Stores

Visualization + Actions

Spectrum Streaming

Netcool Streaming

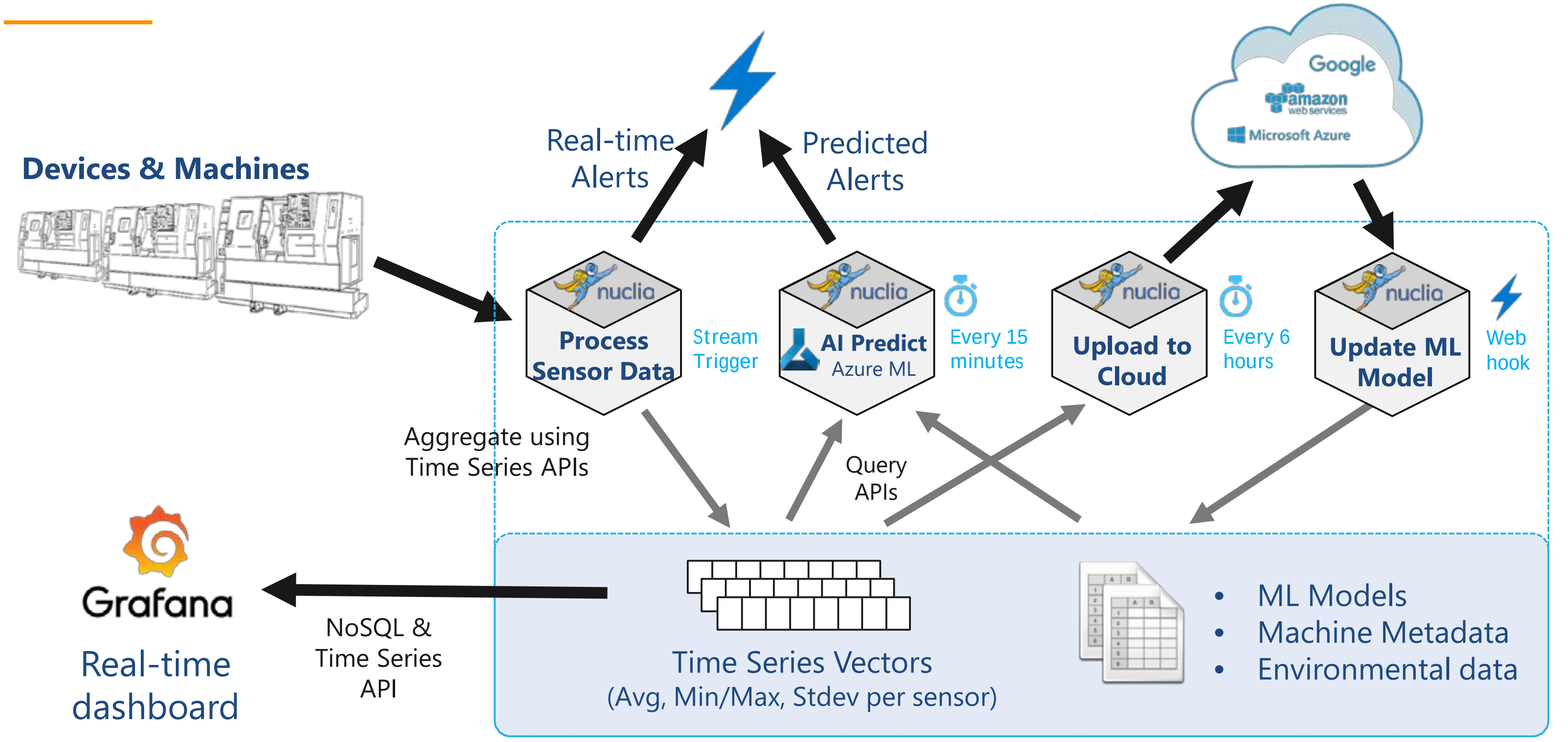
SMOD

REST API

- Simple, just a few weeks to a working app
- Unified view across ALL data
- AI driven, proactive

iguazio

Predictive Maintenance Based on Real-time + Historical + Ops Data



Devices & Machines

Real-time Alerts

Predicted Alerts

Google

amazon web services

Microsoft Azure

nuclio Process Sensor Data

Stream Trigger

nuclio AI Predict Azure ML

Every 15 minutes

nuclio Upload to Cloud

Every 6 hours

nuclio Update ML Model

Web hook

Aggregate using Time Series APIs

Query APIs

Grafana

Real-time dashboard

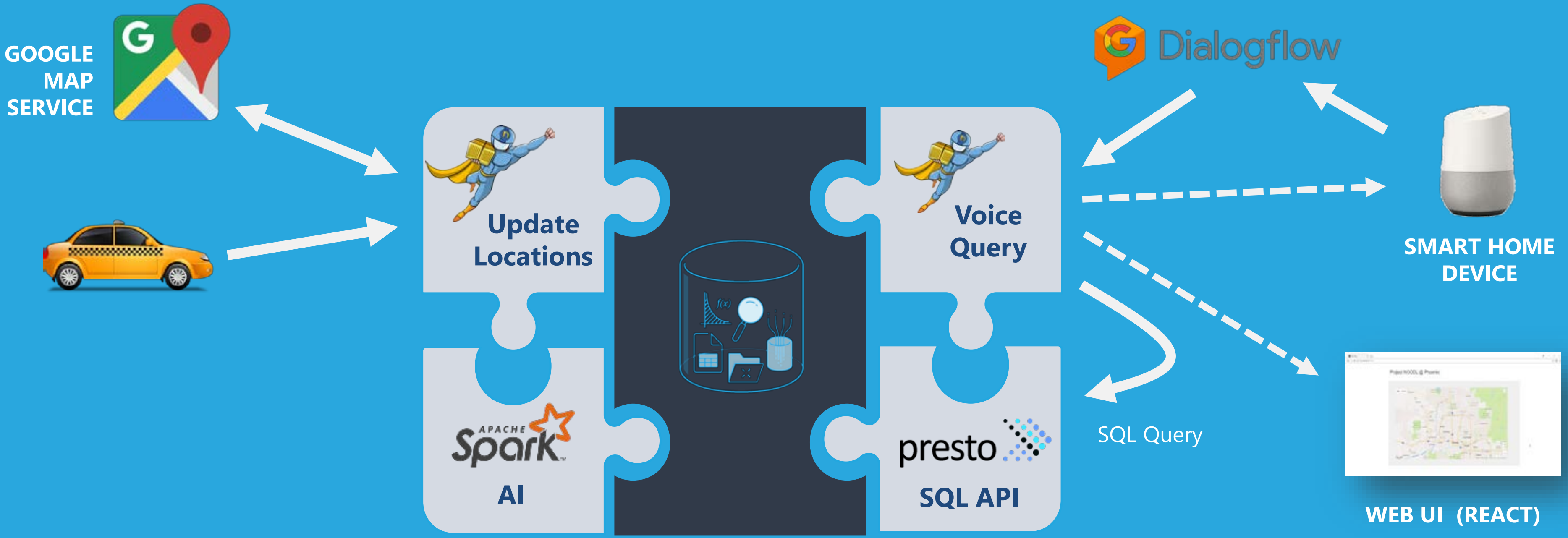
NoSQL & Time Series API

Time Series Vectors (Avg, Min/Max, Stdev per sensor)

- ML Models
- Machine Metadata
- Environmental data

intelligent edge

Demo: Voice Driven Real-Time Analytics



GOOGLE MAP SERVICE

Update Locations

AI

Voice Query

SQL API

Dialogflow

SMART HOME DEVICE

SQL Query

WEB UI (REACT)

Summary

Build continuous, data-driven and proactive apps

- Deliver real-time analytics on fresh, historical and operational data
- Optimize Flash usage 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

Thank You

info@iguazio.com | www.iguazio.com

iguazio