
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ANT333-R

Build a serverless data streaming workload with Amazon Kinesis

Nihar Sheth (he/him)

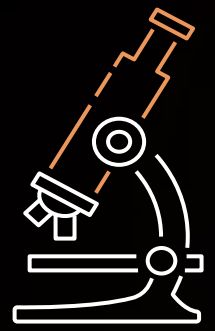
Senior Product Manager
Amazon Kinesis
AWS

Pratik Patel (he/him)

Senior Technical Account Manager
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer: Medical device company



Supporting healthcare and emergency services

1. Real-time dashboards for field sense
2. Reacting to real-time health events
3. Long-term storage for historical analysis

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

What is real-time analytics?

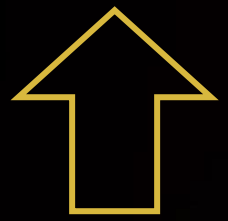
Amazon Kinesis

Serverless streaming architecture

Q&A

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What is streaming data?



High volume



Continuous



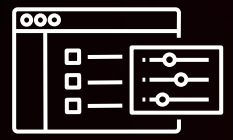
Ordered, incremental



Low-latency



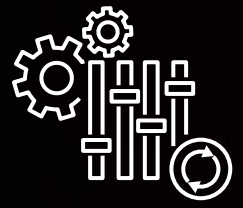
Mobile apps



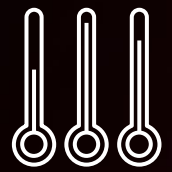
Web clickstream



Application logs



Metering records



IoT sensors

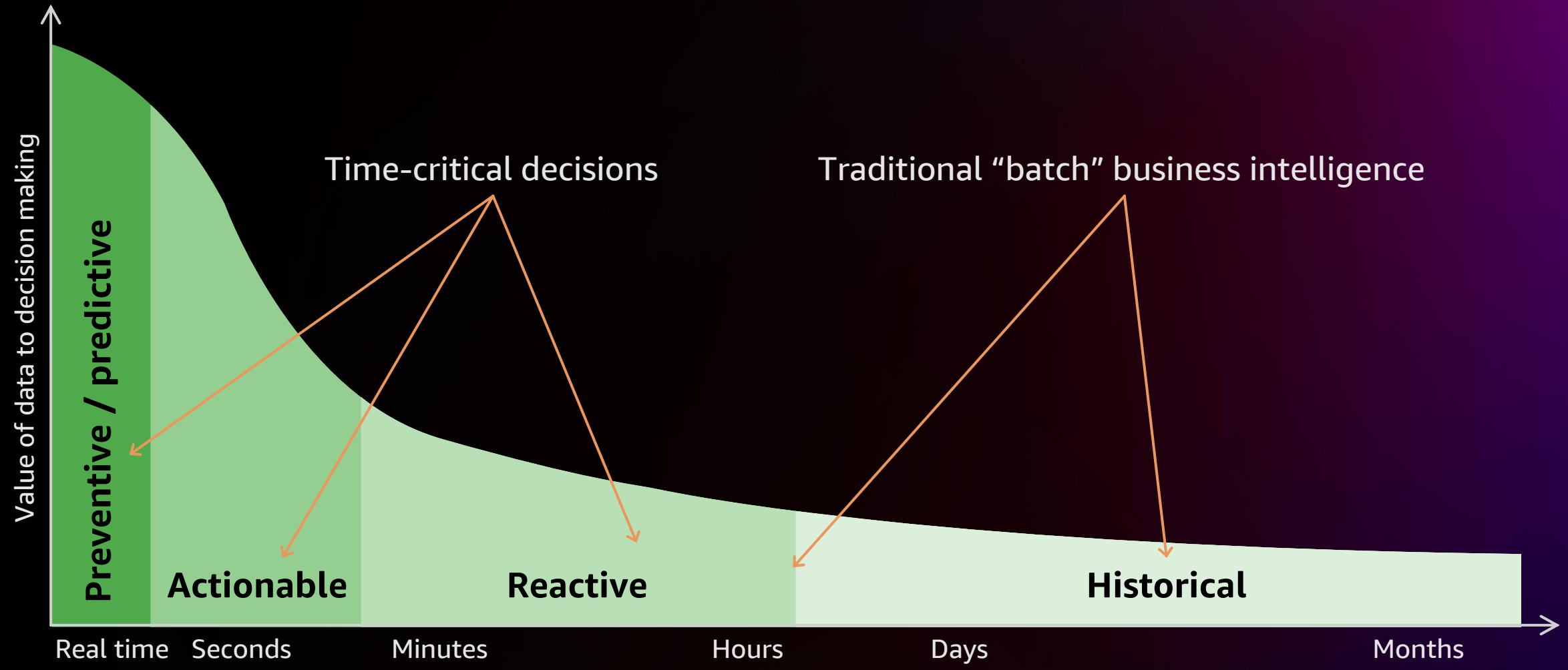


Smart buildings

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Why real-time analytics?



Source: Mike Gualtieri, Forrester, *Perishable Insights*

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

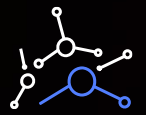
Common real-time analytics use cases



Anomaly and fraud detection



Log analytics



Empowering IoT analytics



Nourishing marketing campaigns



Real-time personalization



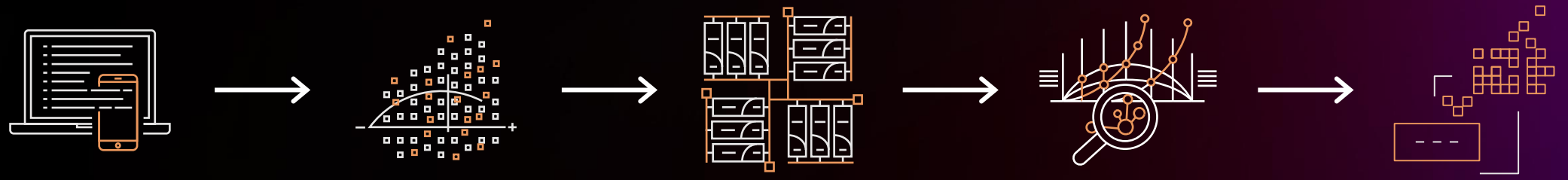
Supporting healthcare and emergency services

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data streaming workflow

INGEST, PROCESS, AND ANALYZE HIGH VOLUMES OF HIGH-VELOCITY DATA FROM A VARIETY OF SOURCES IN REAL TIME



Source

Devices and/or applications that produce real-time data at high velocity

Stream ingestion

Data from tens of thousands of data sources can be collected and ingested in real time

Stream storage

Data is stored in the order it was received for a set duration of time, and can be replayed indefinitely during that time

Stream processing

Records are read in the order they're produced, allowing for real-time analytics or streaming ETL

Destination

Data lake

Data warehouse (most common)

Database (least common)

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Challenges of self-managed data streaming

Apache Kafka

Apache Flink

Apache Spark



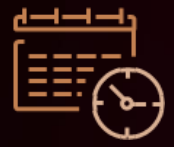
Difficult to set up



Tricky to scale



Hard to achieve high availability



Integration requires development



Error prone and complex to manage



Expensive to maintain

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Kinesis

EASILY COLLECT, PROCESS, AND ANALYZE DATA STREAMS IN REAL TIME

Amazon Kinesis Data Streams



Collect and store data streams for analytics

Amazon Kinesis Data Firehose



Load data streams into AWS data stores

Amazon Kinesis Data Analytics

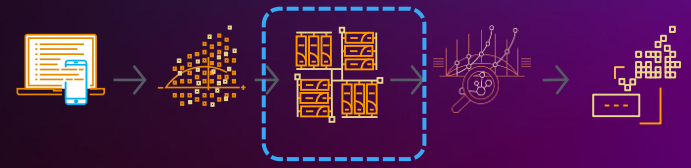


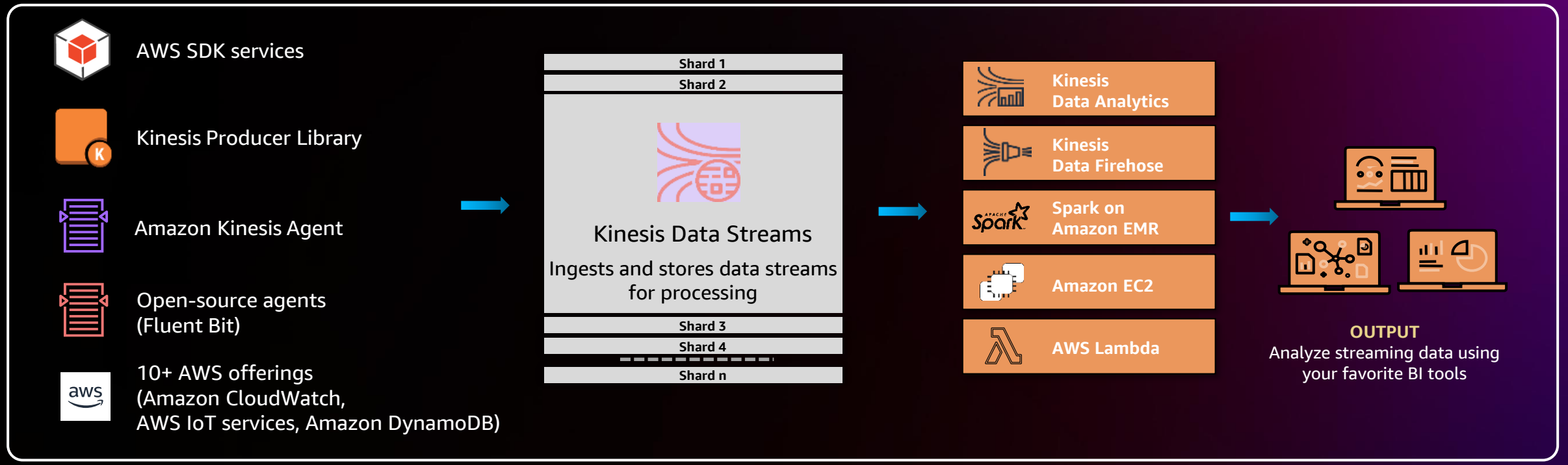
Analyze data streams with Kinesis Data Analytics Studio or Apache Flink

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Kinesis Data Streams

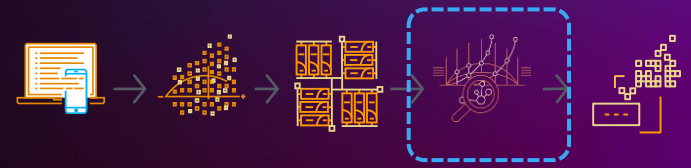


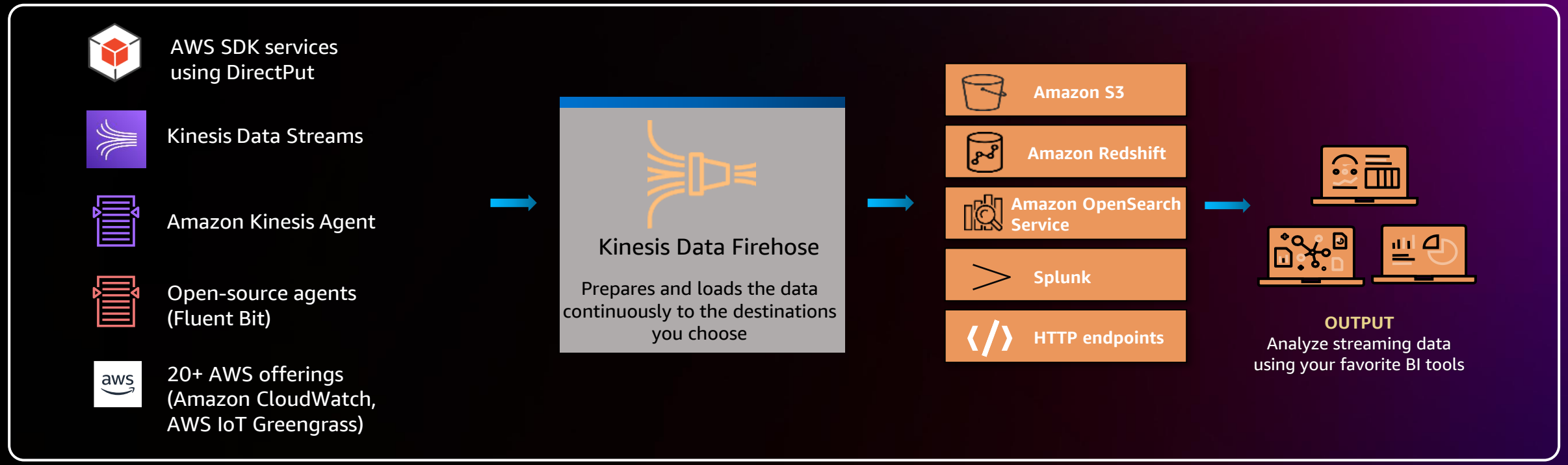


- Easy administration and low cost
- Serverless scaling
- Security, durability, and availability out of the box
- Performance at scale
- Concurrent consumers at low latency
- Data retention up to 1 year

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Kinesis Data Firehose

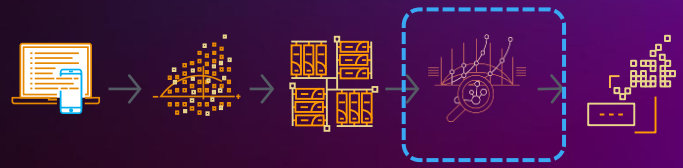


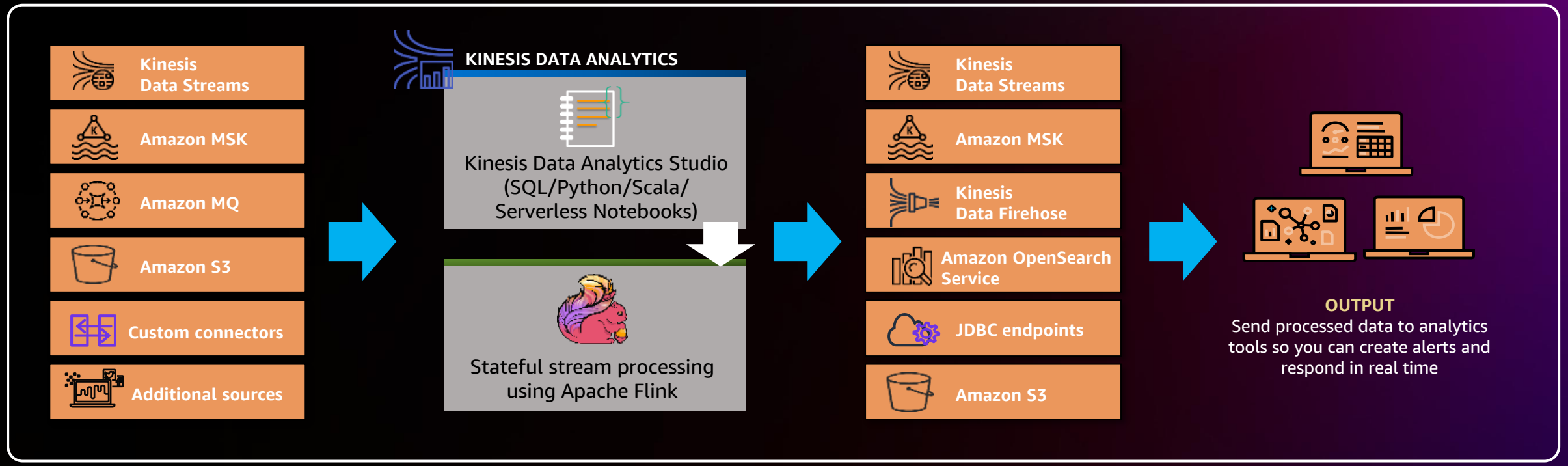


- Zero administration and seamless scaling
- Direct–to–data store integration
- Serverless data transformations
- Buffer and batching flexibility
- Data format conversion to Parquet/ORC
- Dynamic partitioning to Amazon S3
- Deliver data directly to 15+ destinations (Datadog, Sumo Logic, New Relic, MongoDB)

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Kinesis Data Analytics





- Interact with streaming data in real time using SQL, Python, Scala, and Java or integrated Apache Flink applications
- Deploy Kinesis Data Analytics Studio adhoc analysis as a durable state application within Kinesis Data Analytics for Apache Flink
- Build fully managed and scalable stream processing applications

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer references

VIZIO

Billions of events per day from TVs and connected devices

COMCAST

Migrated data bus to Amazon Kinesis

SONOS

1 billion events per day from connected devices

Zillow

Near real-time home valuation (Zestimates)



THOMSON REUTERS

Live clickstream dashboards refreshed under 10s



IoT predictive analytics

HEARST corporation

10 TB per day clickstreams from 250+ sites

AdRoll

50 billion daily ad impressions, sub-50 ms responses

NORDSTROM

Online stylist processing 100 million events per day

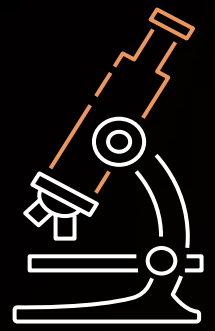
lyft

Facilitate communications between 100+ microservices

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customer: Medical device company



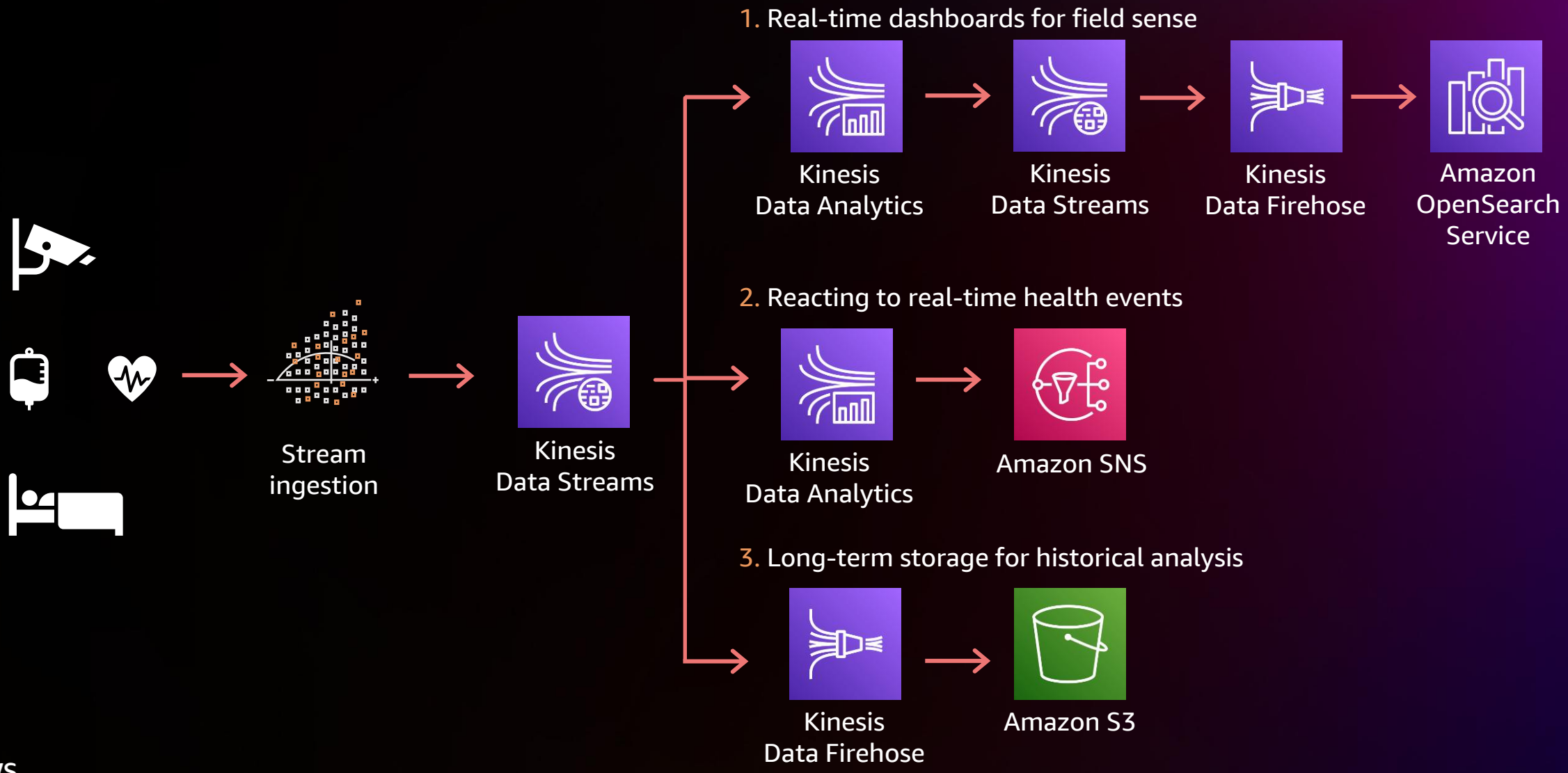
Supporting healthcare and emergency services

1. Real-time dashboards for field sense
2. Reacting to real-time health events
3. Long-term storage for historical analysis

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Serverless streaming architecture



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Q&A

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Nihar Sheth
niharsheth-aws (Linkedin)

Pratik Patel
pratikpatel-aws (Linkedin)



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.