

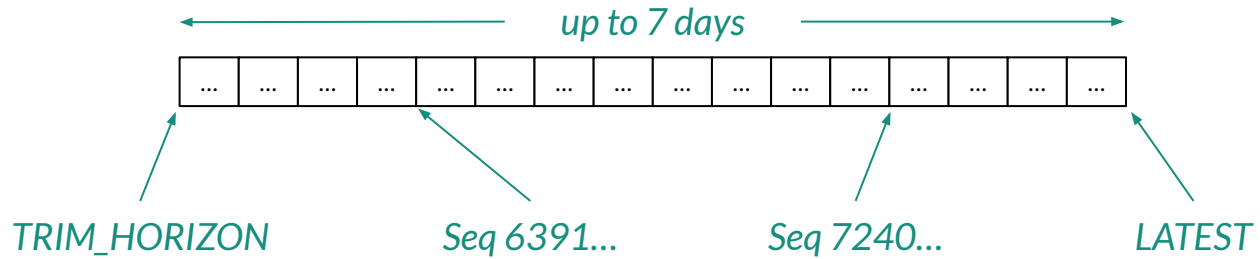
Building Data Pipelines with Kinesis

Keith Gregory
AWS Practice Lead
Chariot Solutions

IoT on AWS
A Philly Cloud Computing Event

CHARIOT
SOLUTIONS

Kinesis is a persistent log of messages



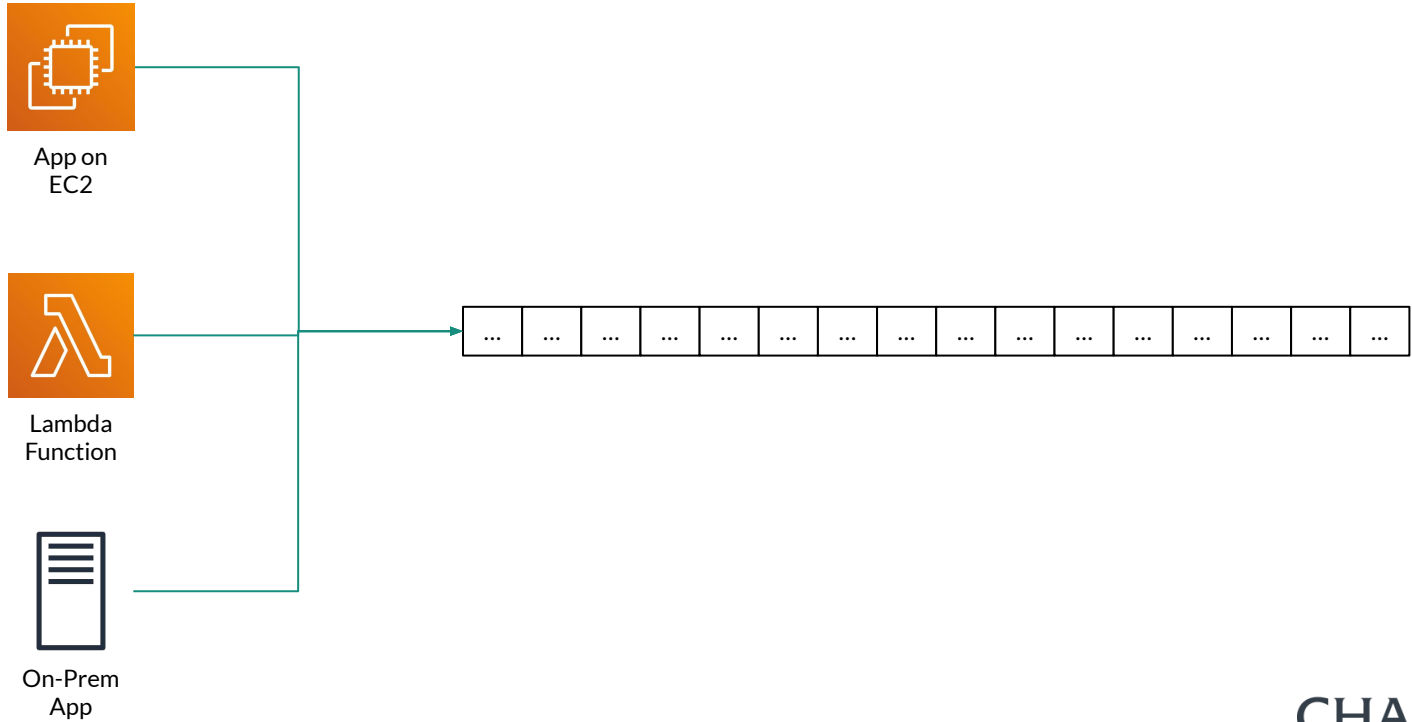
... that's scaled by adding shards

shard-001
shard-002
shard-003
shard-004

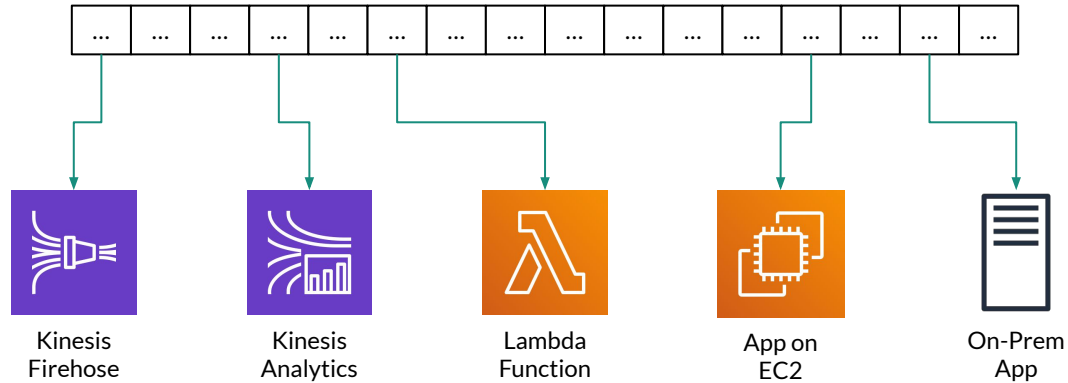
Each shard can accept 1 MB or 1,000 records per second

Incoming records distributed to shards by “partition key”

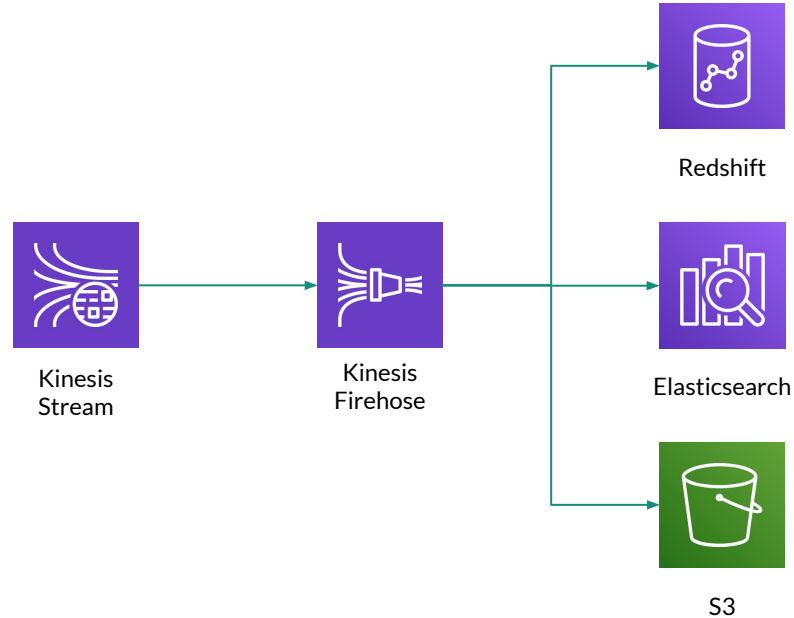
... can be written by multiple producers



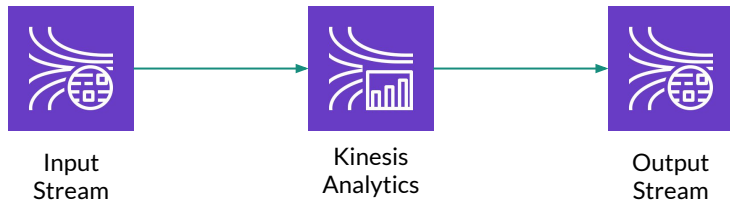
... and read by multiple consumers



... batched for delivery to other tools



... or transformed via SQL

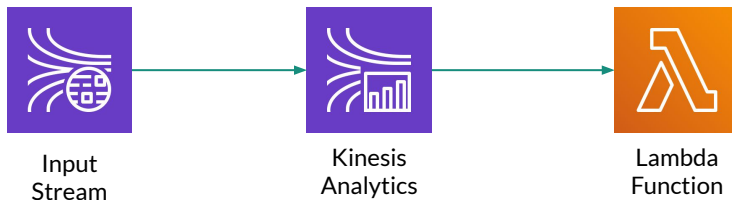


```
CREATE OR REPLACE PUMP "AVERAGE_TEMPERATURE_PUMP" AS
  insert into "OUTPUT_AVERAGE_TEMPERATURE"
    (device, start_timestamp, end_timestamp,
     num_readings, average_temperature,
     stdev_temperature)
  select STREAM "device",
    min(TO_TIMESTAMP("COL_timestamp")) OVER W1,
    max(TO_TIMESTAMP("COL_timestamp")) OVER W1,
    count(*) OVER W1,
    avg("temperature") OVER W1,
    stddev_samp("temperature") OVER W1
  from "SOURCE_SQL_STREAM_001"
  window W1 AS (
    PARTITION BY "device"
    RANGE INTERVAL '5' MINUTE PRECEDING);
```

Filter by column name

DEVICE	START_TIMESTAMP	END_TIMESTAMP	NUM_READINGS	AVERAGE_TEMPE
keith-arduino	2019-09-18 20:09:54.427	2019-09-18 20:14:53.991	29	79.05
keith-arduino	2019-09-18 20:10:14.538	2019-09-18 20:15:03.979	28	79.49
keith-arduino	2019-09-18 20:10:14.538	2019-09-18 20:15:14.171	29	79.61
keith-arduino	2019-09-18 20:10:34.495	2019-09-18 20:15:23.971	28	80.03
keith-arduino	2019-09-18 20:10:34.495	2019-09-18 20:15:23.971	28	80.03

... which can also trigger actions



```
CREATE OR REPLACE PUMP "OUTLIER_PUMP" AS
  insert into "OUTPUT_OUTLIERS"
    (device, average_temperature, actual_temperature)
  select STREAM src."device",
    avgtemp.average_temperature,
    src."temperature"
  from "SOURCE_SQL_STREAM_001" src
  join "OUTPUT_AVERAGE_TEMPERATURE" avgtemp
  on src."device" = avgtemp.device
  where src."temperature" < avgtemp.average_temperature - avgtemp.stdev_temperature
  or src."temperature" > avgtemp.average_temperature + avgtemp.stdev_temperature
```




References

Kinesis Streams User Guide

<https://docs.aws.amazon.com/streams/latest/dev/introduction.html>

Kinesis Firehose User Guide

<https://docs.aws.amazon.com/firehose/latest/dev/what-is-this-service.html>

Kinesis Data Analytics SQL Developer Guide

<https://docs.aws.amazon.com/kinesisanalytics/latest/dev/how-it-works.html>

Questions?

