



# Kinesis Data Streams



**Brock Tubre**

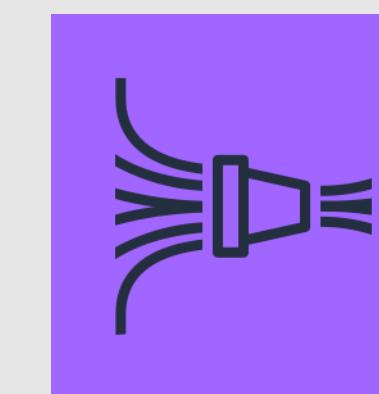
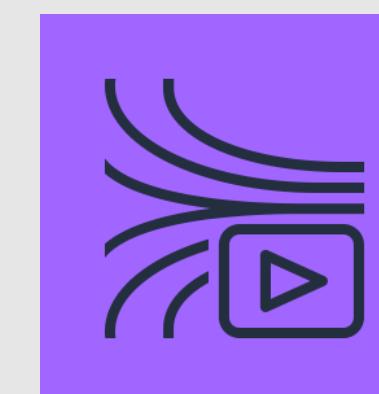
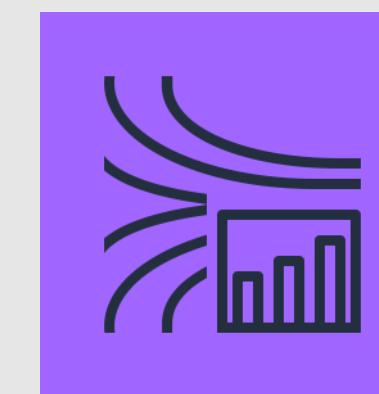
INSTRUCTOR

# The Kinesis Family

## The Kinesis Family



Kinesis

Kinesis  
Data StreamsKinesis  
Data FirehoseKinesis  
Video StreamsKinesis  
Data Analytics

## Data Producers

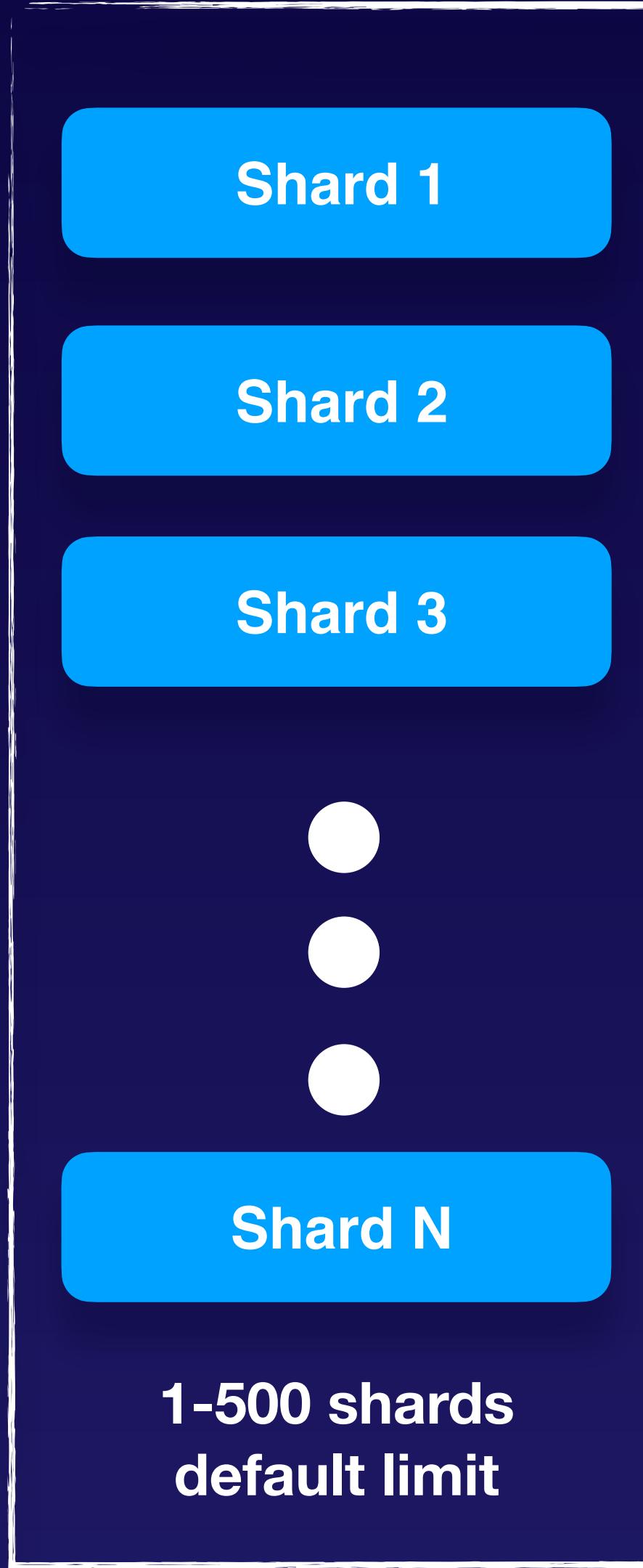


# Kinesis Data Streams

## Data Producers



## Kinesis Streams

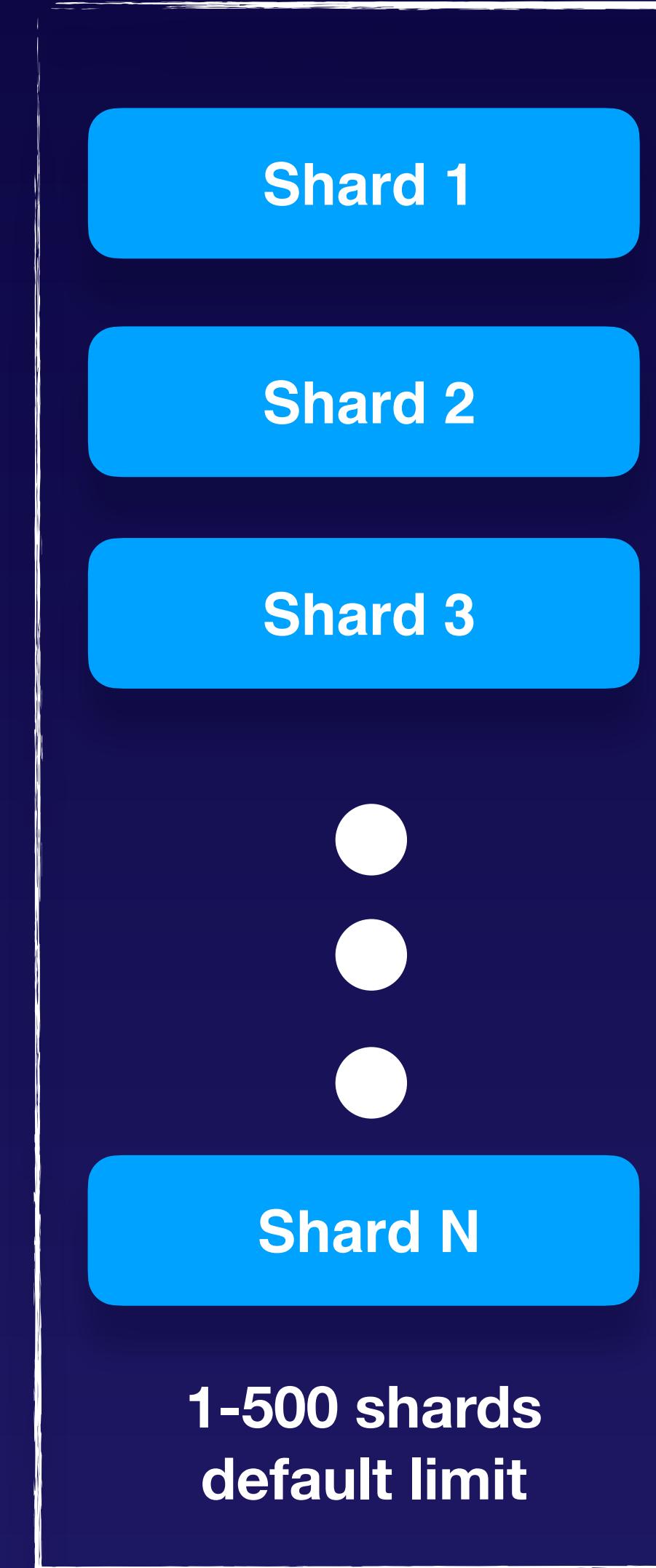


# Kinesis Data Streams

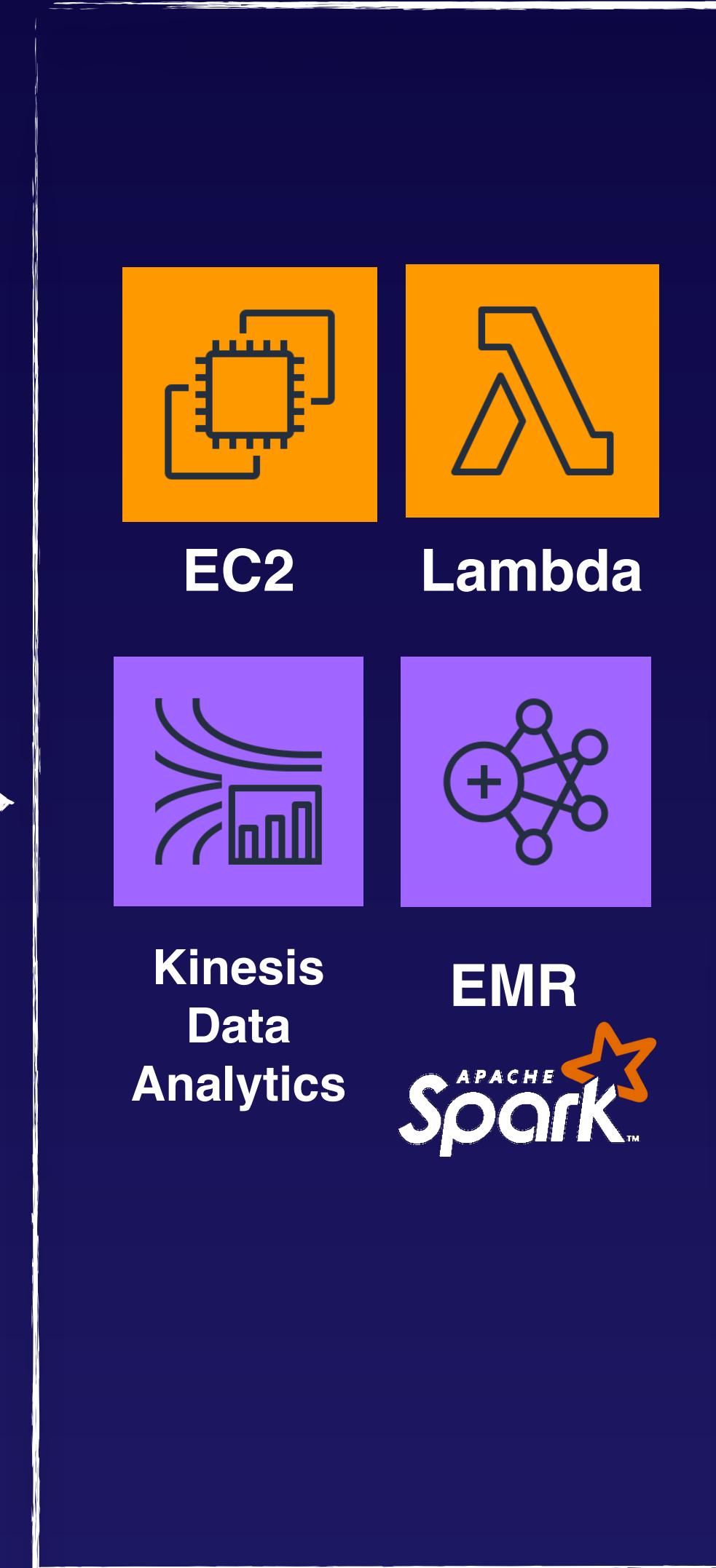
## Data Producers



## Kinesis Streams



## Data Consumers (Processing Tools)

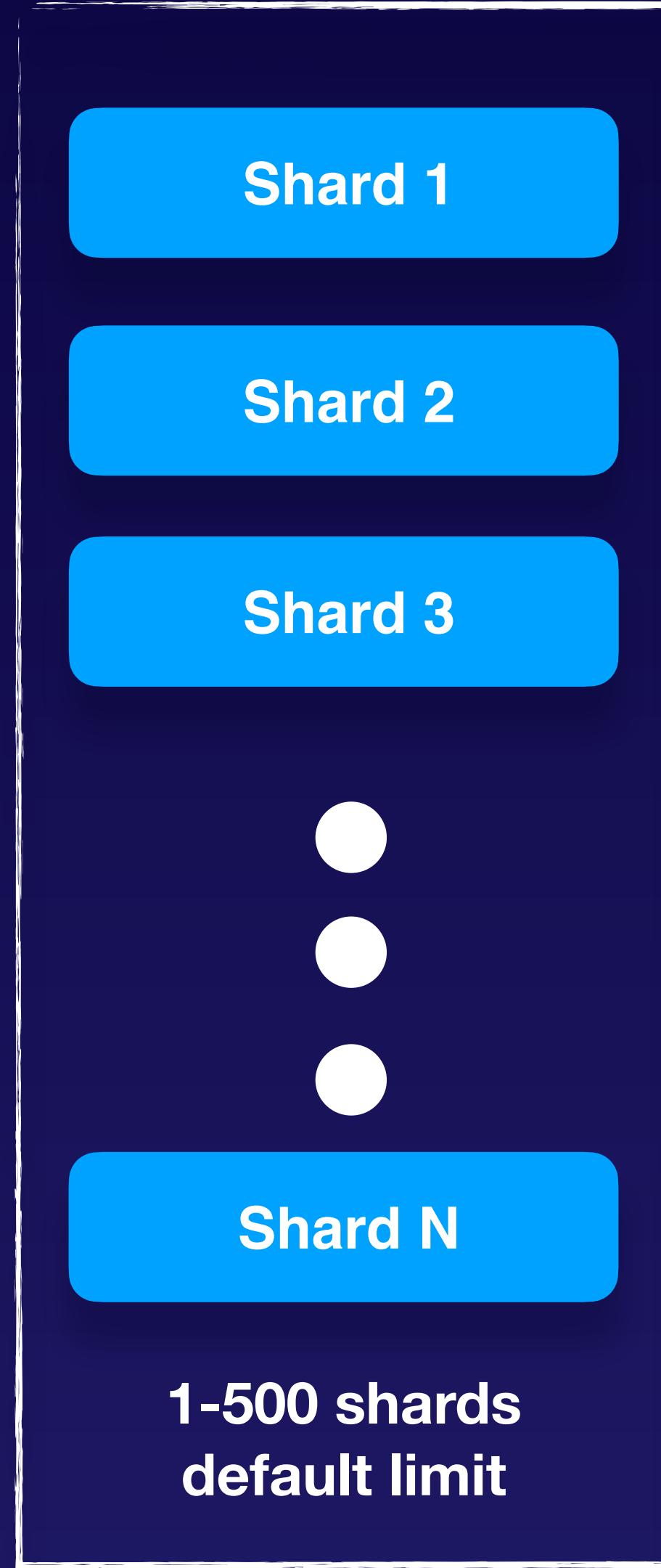


# Kinesis Data Streams

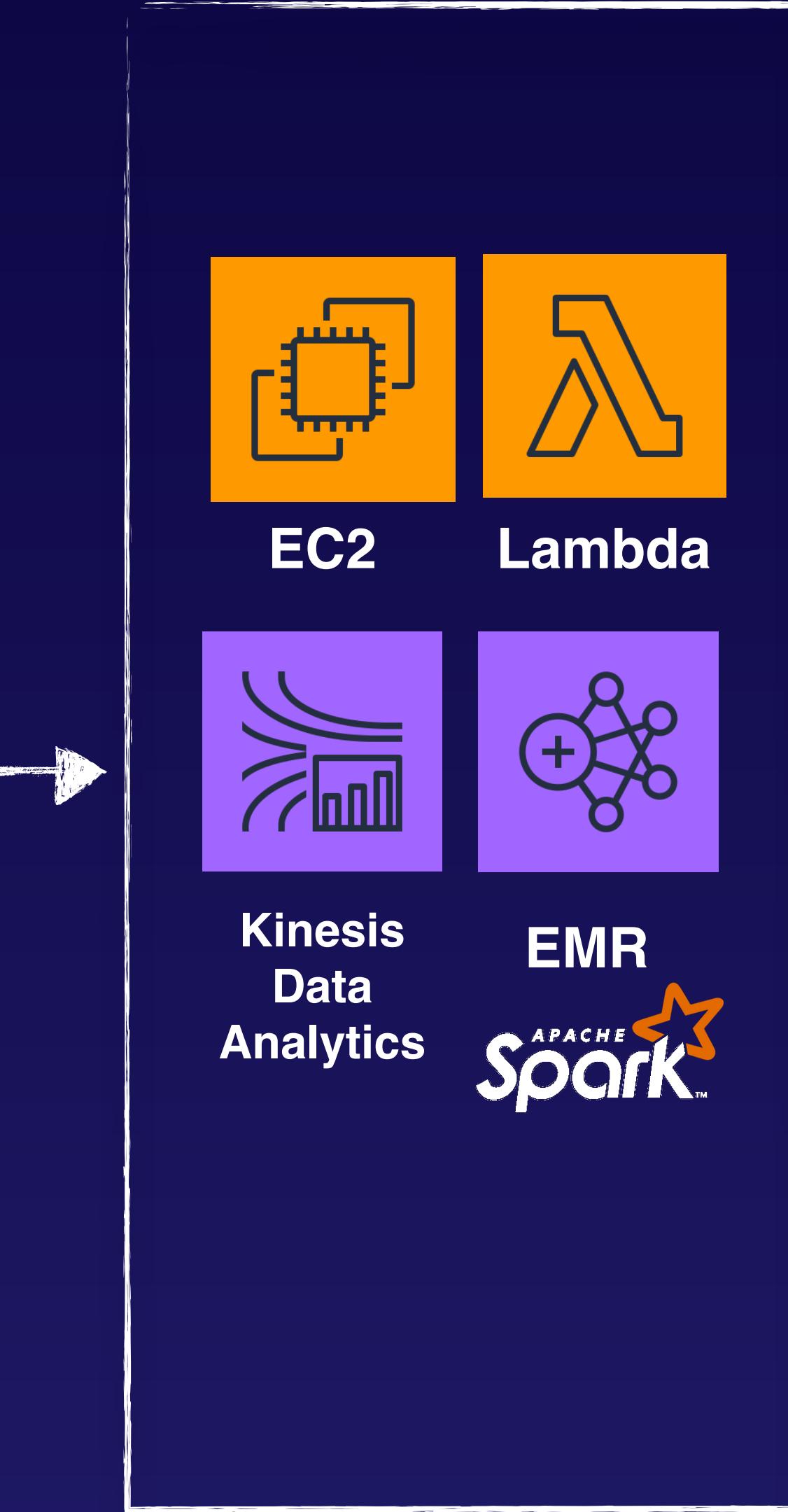
## Data Producers



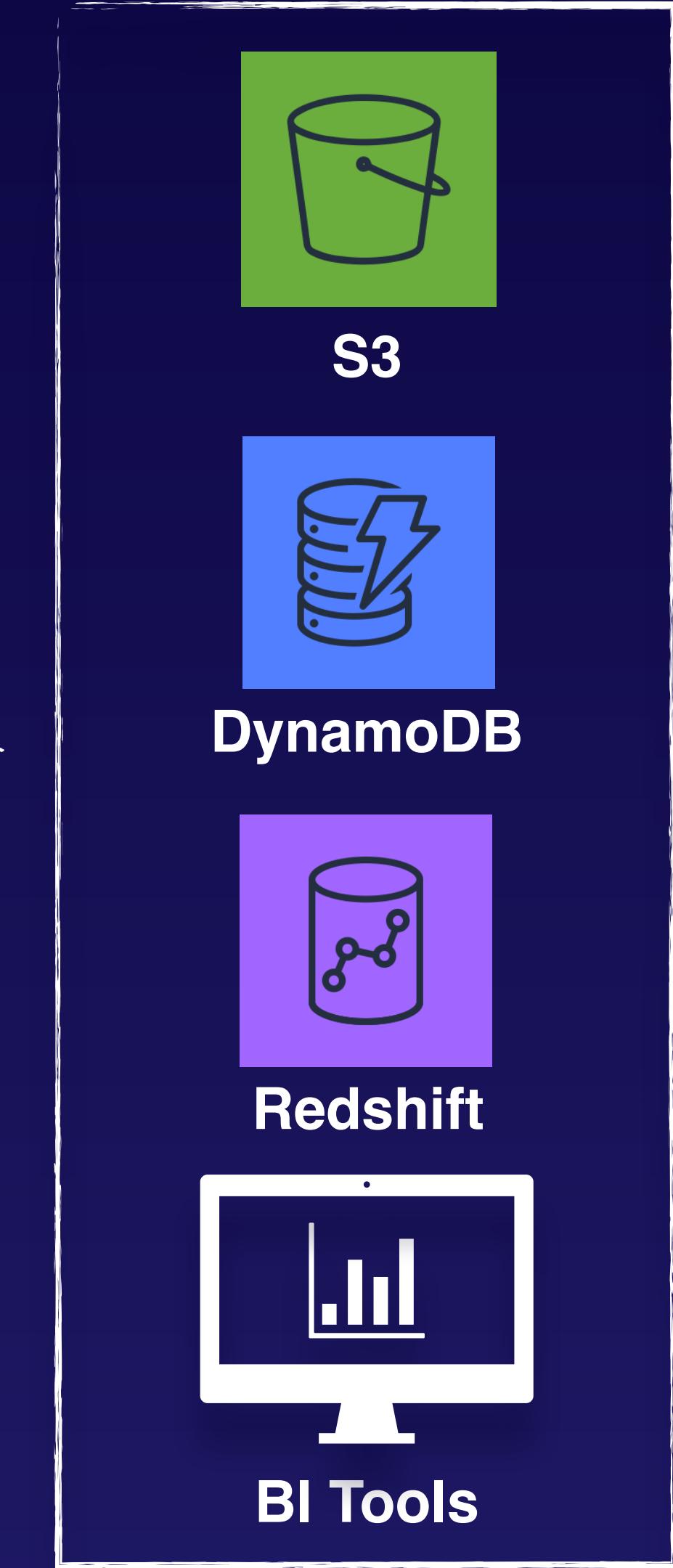
## Kinesis Streams



## Data Consumers (Processing Tools)

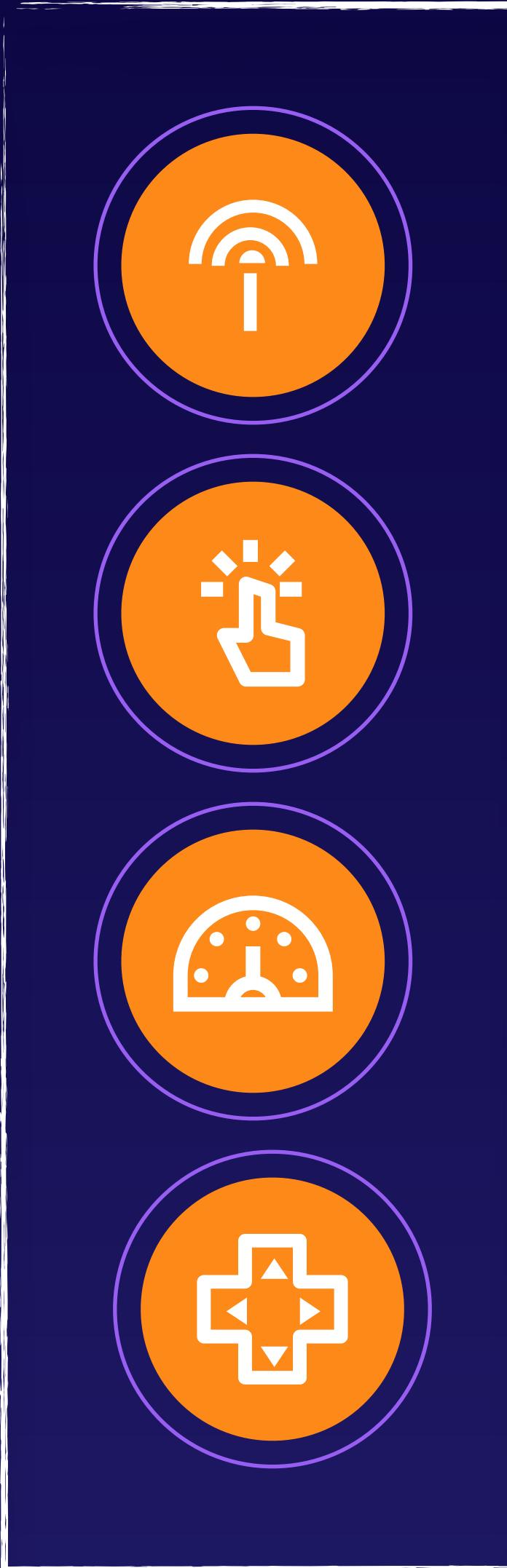


## Storage and Analyzation

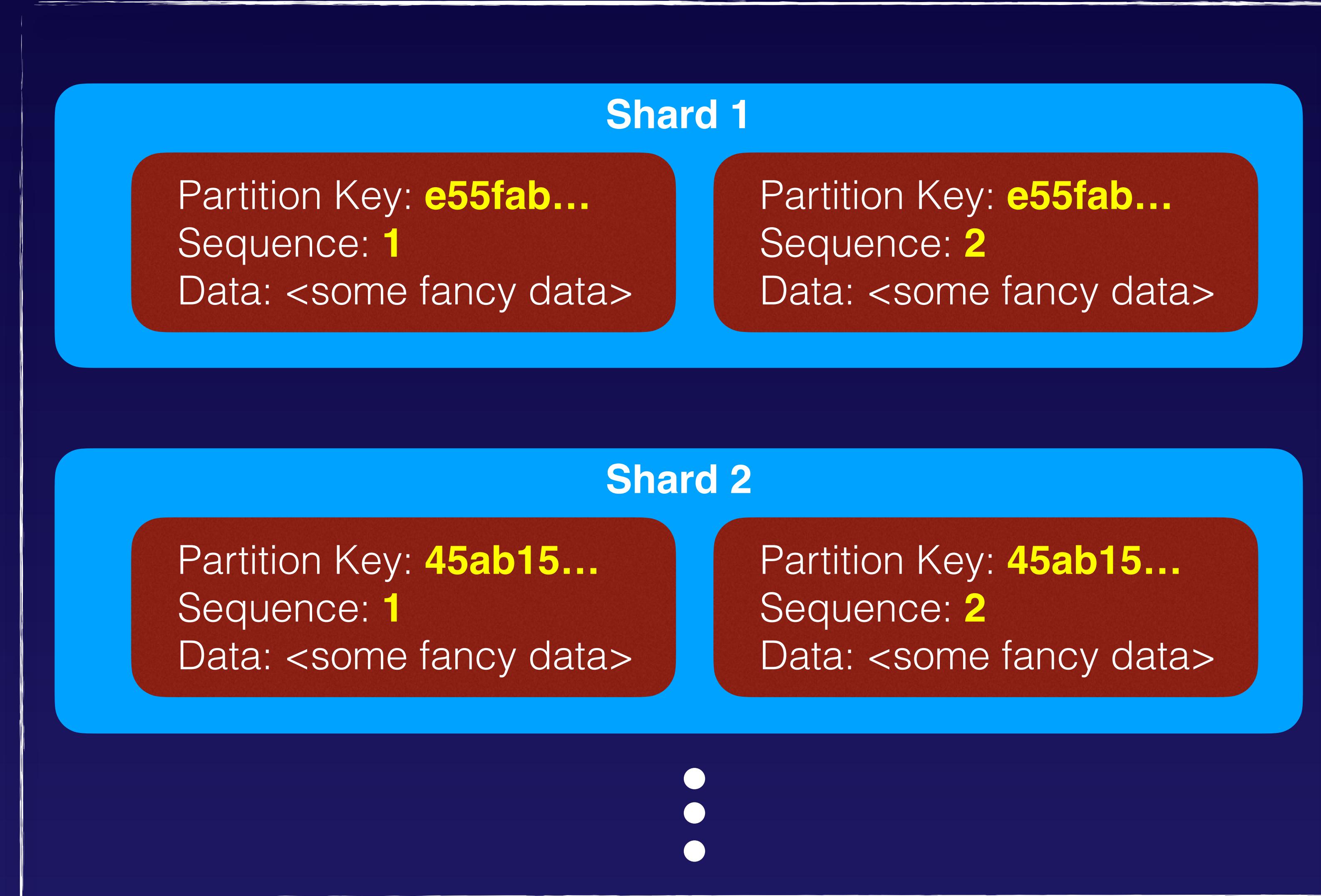


# Kinesis Data Streams

## Data Producers

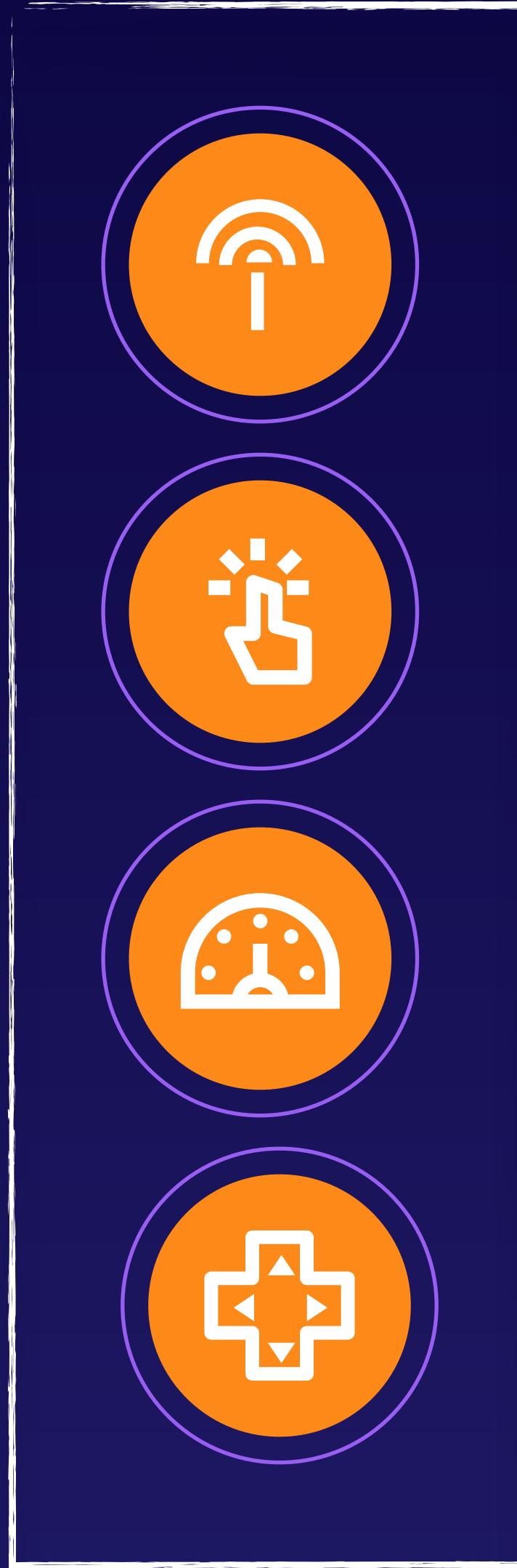


## Kinesis Streams



# Kinesis Data Streams

## Data Producers



**Train Id = Partition Key**



**Train Car = Sequence number**

**Passengers = Data**

- Each shard consists of a sequence of data records. These can be ingested at 1000 records per second.
- Default limit of 500 shards, but you can request increase to unlimited shards.
- A data record is the unit of data captured.
  - sequence number
  - partition key
  - data blob (your payload, up to 1 MB)
- Transient Data Store - retention period for the data records are 24 hours to 7 days.

**1**

## Kinesis Producer Library (KPL)

Easy to use library that allows you to write to a Kinesis Data Stream.

**2**

## Kinesis Client Library (KCL)

Integrated directly with KPL for consumer applications to consume and process data from Kinesis Data Stream.

**3**

## Kinesis API (AWS SDK)

Used for low level API operations to send records to a Kinesis Data Stream.

## Kinesis Producer Library (KPL)

- Provides a layer of abstraction specifically for ingesting data
- Automatic and configurable retry mechanism
- Additional processing delay can occur for higher packing efficiencies and better performance
- Java wrapper

## Kinesis API

- Low-level API calls (PutRecords and GetRecords)
- Stream creations, resharding, and putting and getting records are manually handled
- No delays in processing
- Any AWS SDK

Amazon Kinesis ◀ Create Kinesis stream ?

Dashboard

**Data Streams**

- Data Firehose
- Data Analytics
- Video Streams

External resources

What's new

Kinesis stream name\*

Acceptable characters are uppercase and lowercase letters, numbers, underscores, hyphens, and periods.

**Shards**

A shard is a unit of throughput capacity. Each shard ingests up to 1MB/sec and 1000 records/sec, and emits up to 2MB/sec. To accommodate for higher or lower throughput, the number of shards can be modified after the Kinesis stream is created using the API. [Learn more](#)

▶ Estimate the number of shards you'll need

Number of shards\*

You can provision up to 498 more shards before hitting your account limit of 500.  
[Learn more](#) or [request a shard limit increase for this account](#)

Total stream capacity Values are calculated based on the number of shards entered above.

Write  MB per second

Records per second

Read  MB per second

Amazon Kinesis ◀ Create Kinesis stream ?

Dashboard

**Data Streams**

- Data Firehose
- Data Analytics
- Video Streams

External resources

What's new

Kinesis stream name\*

Acceptable characters are uppercase and lowercase letters, numbers, underscores, hyphens, and periods.

**Shards**

A shard is a unit of throughput capacity. Each shard ingests up to 1MB/sec and 1000 records/sec, and emits up to 2MB/sec. To accommodate for higher or lower throughput, the number of shards can be modified after the Kinesis stream is created using the API. [Learn more](#)

▶ Estimate the number of shards you'll need

Number of shards\*

You can provision up to 498 more shards before hitting your account limit of 500. [Learn more](#) or [request a shard limit increase for this account](#)

Total stream capacity Values are calculated based on the number of shards entered above.

Write  MB per second

Records per second

Read  MB per second



# Kinesis Data Streams

## When should you use Kinesis Data Streams?

- Needs to be processed by consumers.
- Real time analytics.
- Feed into other services in real time.
- Some action needs to occur on your data.
- Storing data is optional.
- Data retention is important.

- **Process and evaluate logs immediately**

**Example:** Analyze system and application logs continuously and process within seconds.

- **Real-time data analytics**

**Example:** Run real-time analytics on click stream data and process it within seconds.