

# RxKinesis

IN4389 Reactive Programming

Alex Garella

# Rx interface for Amazon Kinesis

- Amazon Kinesis
- RxKinesis
- Demo
  - RxKinesisChat
  - TwitterAnalytics
- Tips
- Questions

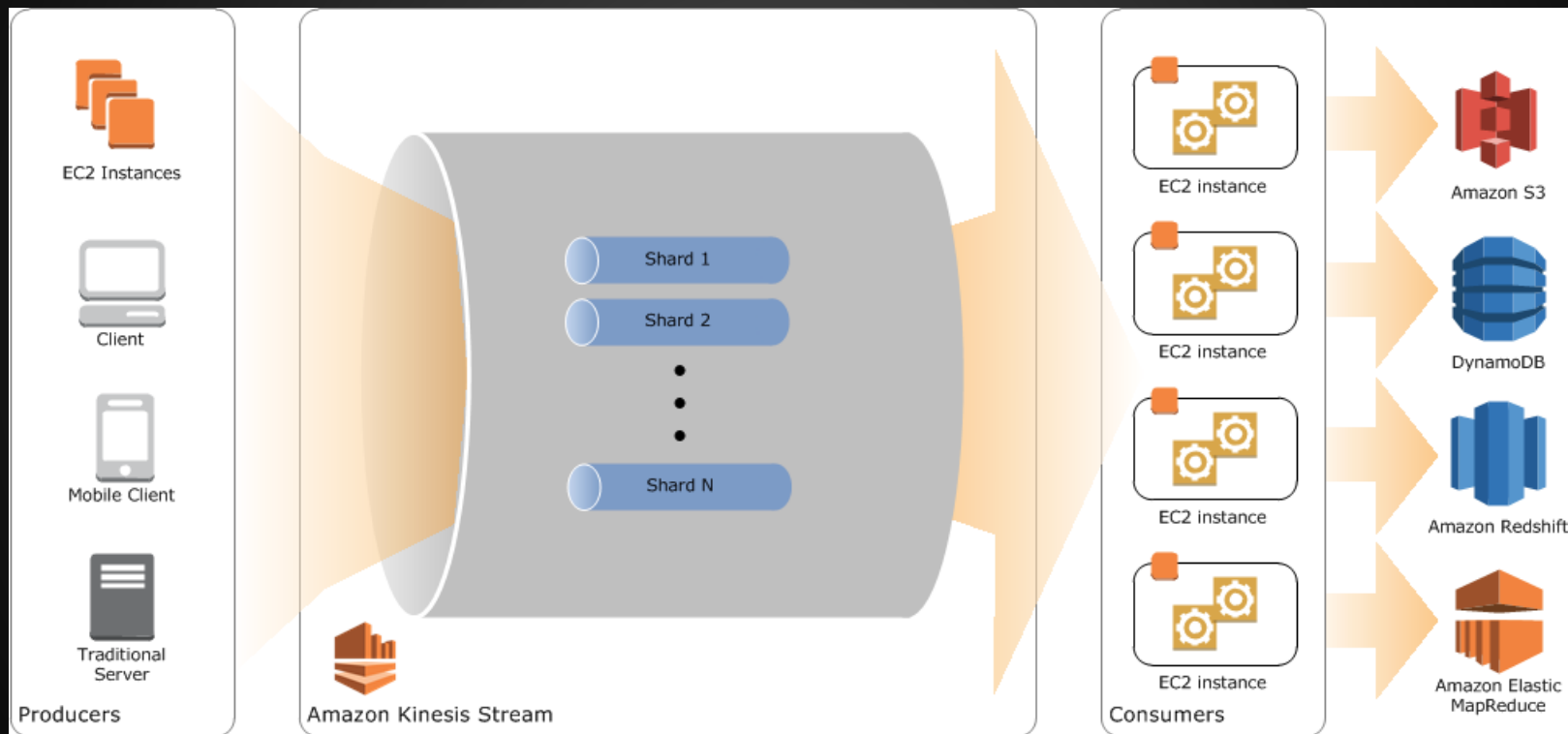
# Amazon Kinesis

- Part of the AWS `Big Data` solutions
- Collect & process large data streams
- Real time
- Scenarios
  - Real-time metrics & reporting
  - Real-time data analysis

# Amazon Kinesis Concepts

- Streams
- Data records
- Sequence number
- Shards / Partition Keys
- Producers / Consumers
- Application name

# Amazon Kinesis Overview



# Amazon Kinesis API

- AWS Java SDK
  - AmazonKinesisClient
- amazon-kinesis-client
  - IRecordProcessor
  - IRecordProcessorFactory
  - Worker
- Cloudwatch for metrics
- DynamoDB to persist application state

# RxKinesis

- Provide Rx interface to Amazon Kinesis
- Publish data to Kinesis stream:  
RxKinesisPublisher
  - Create Kinesis stream
- Consume data from Kinesis stream:  
RxKinesisConsumer
- Written in Scala

# RxKinesisConsumer[T]

- Takes:
  - ConsumerConfiguration case class
  - Optional: parser: String => T
- Provides:
  - Observable[T]
- ReplaySubject under the hood
  - Provides Observable interface to the outside
  - Processes data in the RecordProcessor on the inside



# RxKinesisPublisher[T]

- Takes:
  - PublisherConfiguration & Observable[T]
  - Optional serializer:  $T \Rightarrow \text{String}$
- Internal Observer[T] serializes and publishes data to Amazon Kinesis

# RxKinesisChat

- RxKinesisPublisher[Message]
  - Takes Observable[Message]
  - Serialize Message to JSON
- RxKinesisConsumer[Message]:
  - Parse JSON to Message
  - Provides Observable[Message] with data from Kinesis
- JSON parsing / serialization with spray-json library

# RxKinesisChat

Demo

# Twitter Analytics

- Twitter Streaming API
- HoseBird client (hbc-core)
- Incoming stream of tweets in JSON
- Tweets => RxKinesisPublisher => Kinesis
- Kinesis => RxConsumer => Analyze
- Sort hashtags by their frequency
- Real-time

# Twitter Analytics

Demo

# Tips for using Amazon Kinesis

- Non-comprehensive documentation but plenty of code examples!
- Beware of Kinesis streams in different regions! \$\$\$
- Delete shards & DynamoDB tables if you don't use them! \$

**Questions?**