Spring XD

Developer Productivity for Data-driven Apps at Scale

Sabby Anandan, Mark Pollack, Mark Fisher

Your Presenters

Mark Fisher

Spring XD
(co-lead)

Mark Pollack

Spring XD
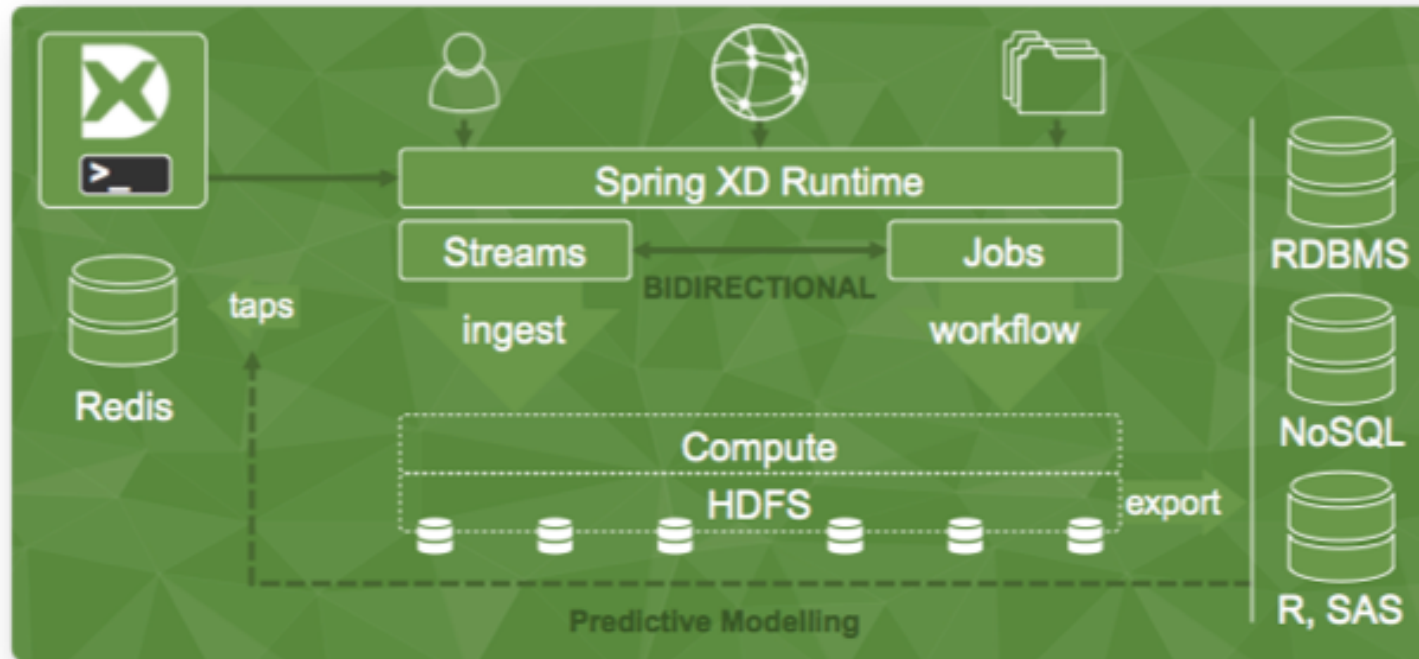*(co-lead)*

Sabby Anandan

Product Manager

spring by Pivotal.

Agenda

- Architecture

- Functional Stream Processing

- Spark Streaming and Spark Job

- Kafka and Data Partitions

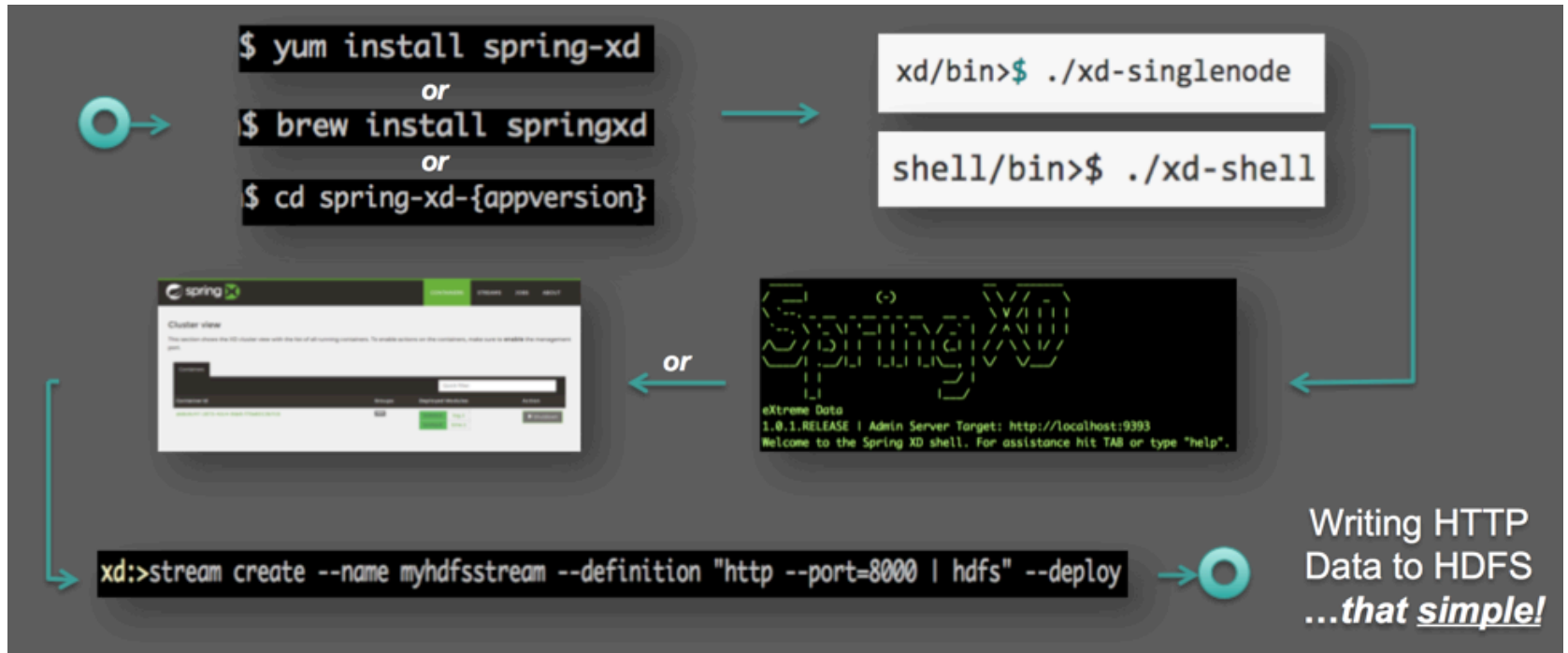spring
by Pivotal.

Purpose

- Unified platform for *stream* and *batch* workloads

- Evolve streams and jobs in *isolation*

- Eliminates technical challenges with distributed computing, fault-tolerance, or high-availability

- Open and Extensible runtime

- Portable: Runs *anywhere* there is a JVM

- Read/write Hadoop without integration headaches

spring
by Pivotal.

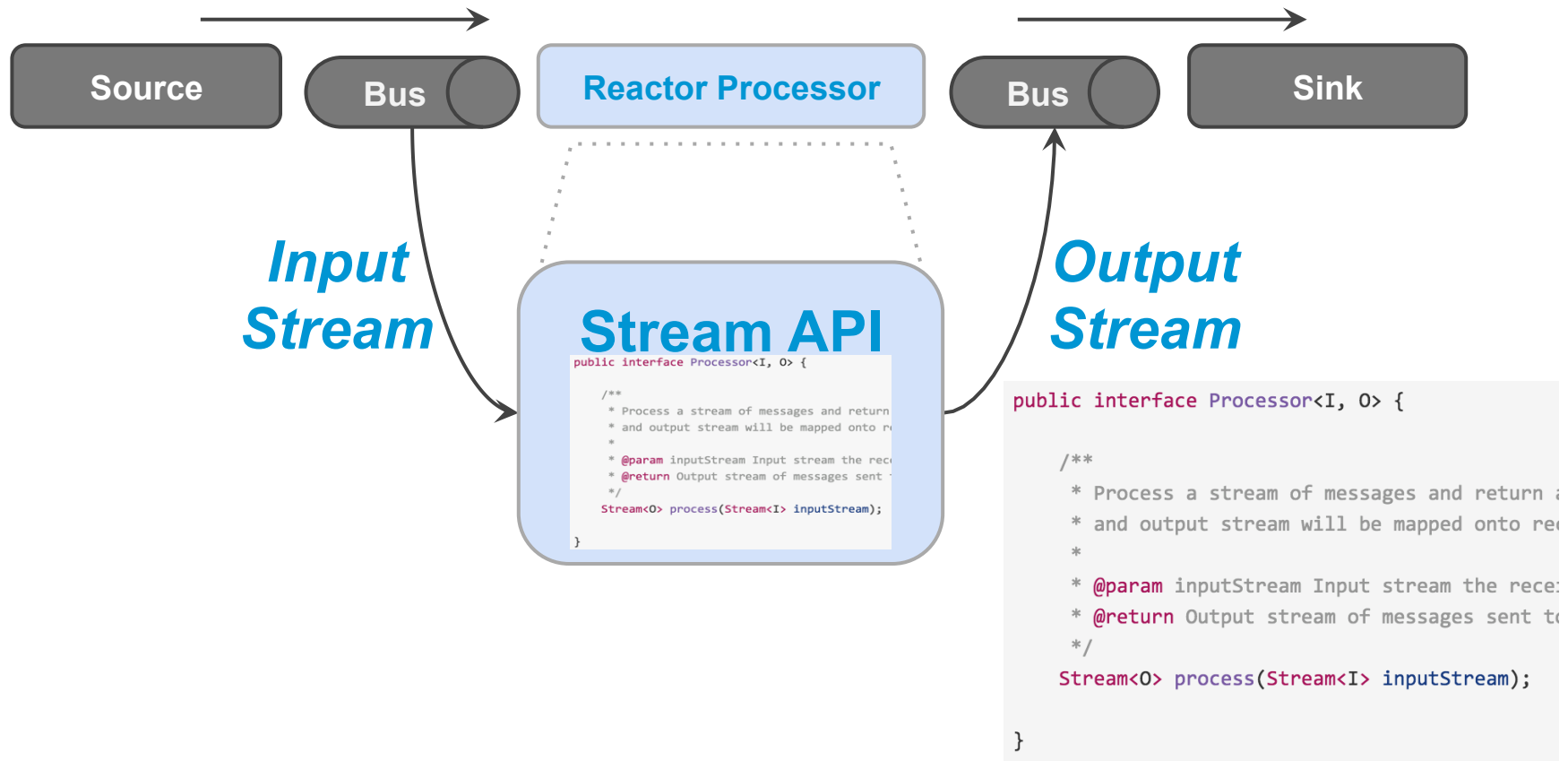# Distributed Runtime

Hadoop without integration headaches..

spring
by Pivotal.

Easy to Setup and Run

Functional Stream Processing

spring
by Pivotal

Reactor Stream API

Source → Bus

Reactor Processor

Bus → Sink

Input Stream

Output Stream

Stream API

```
public interface Processor<I, O> {

    /**
     * Process a stream of messages and return
     * and output stream will be mapped onto re
     *
     * @param inputStream Input stream the rece
     * @return Output stream of messages sent
     */
    Stream<O> process(Stream<I> inputStream);

}
```

```
public interface Processor<I, O> {

    /**
     * Process a stream of messages and return
     * and output stream will be mapped onto re
     *
     * @param inputStream Input stream the rece
     * @return Output stream of messages sent t
     */
    Stream<O> process(Stream<I> inputStream);

}
```

spring
by Pivotal

RxJava Observable API

Source → Bus → **RxJava Processor** → Bus → Sink

Input Stream

Observable API

Output Stream

```
public interface Processor<I,O> {

    /**
     * Process a stream of messages and return an outp
     * and output stream will be mapped onto receive/s
     *
     * @param inputStream Input stream the receives me
     * @return Output stream of messages sent to the m
     */
    Observable<O> process(Observable<I> inputStream);
}
```

spring
by Pivotal

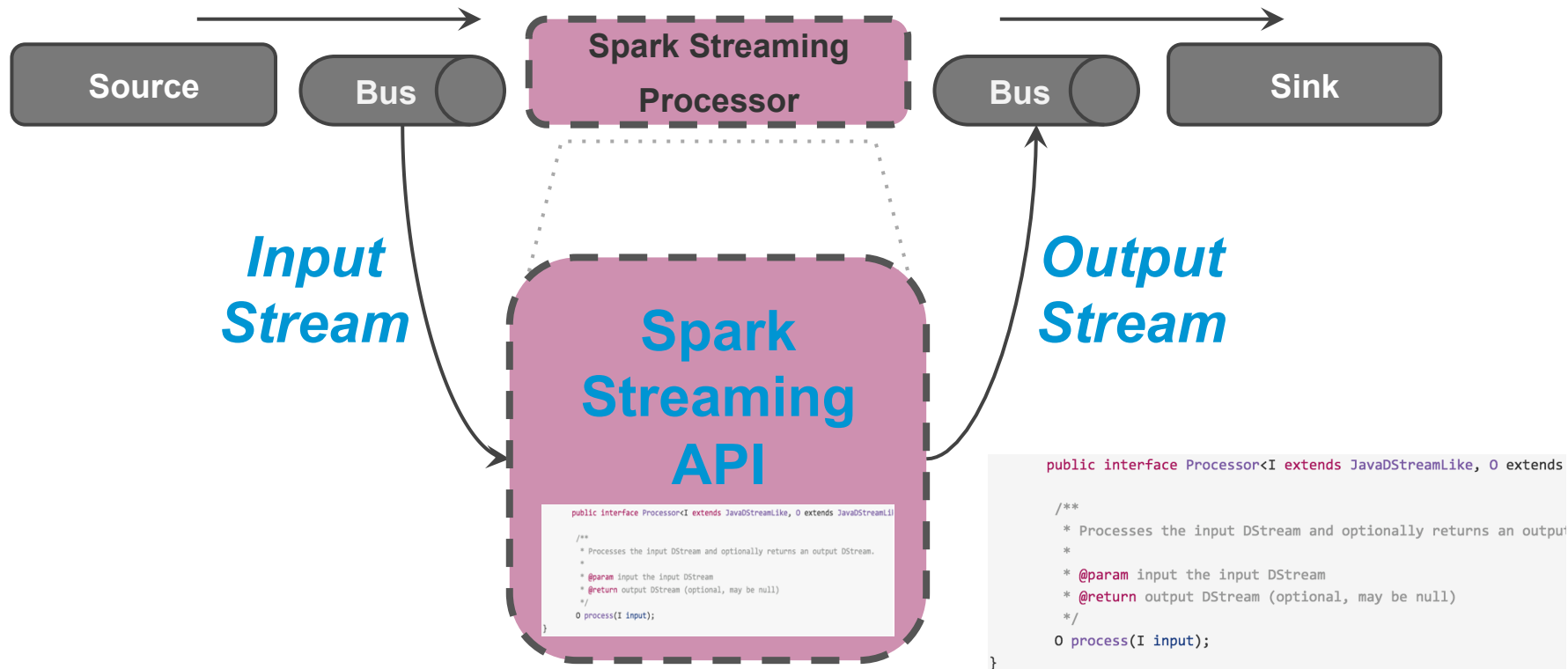# Spring XD + Spark

spring by Pivotal.

Spark

Fast and general engine for large-scale data processing.

What is Spark in Spring XD?

- Leverage OOTB source/sinks from XD and delegate computation logic to Spark cluster
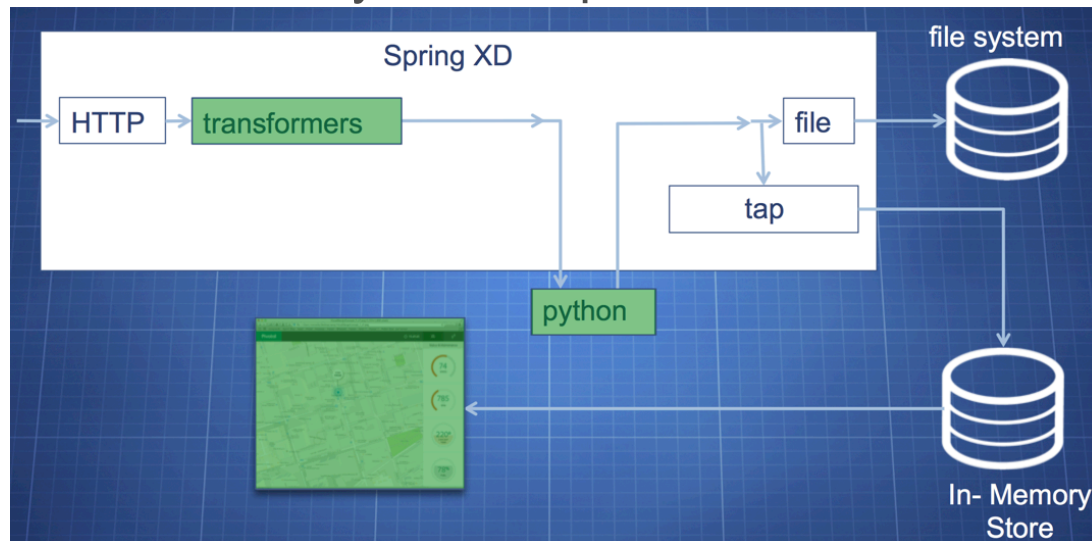- Automatic failover capabilities in case of Spark Driver failure

spring
by Pivotal.

Spark Streaming API

Spring XD + Python

spring
by Pivotal.

Python

What is Python in Spring XD?
● OOTB 'shell' as *processor* and *sink* to execute external shell
  commands such as a Python script

Spring XD + Kafka

spring
by Pivotal.

Kafka

Apache Kafka is publish-subscribe messaging rethought as a distributed commit log.

What is Kafka in Spring XD?

- As a *message bus* to send data from the output of one module to the input of the next module

  - As a *source* module to ingest data from Kafka topics

  - As a *sink* module to ingest data into Kafka topics

spring
by Pivotal.

Looking Ahead

spring
by Pivotal

Roadmap

- Custom Modules and HDFS

- UI Canvas:

  - *create, visualize and deploy data pipelines*

  - *metrics and monitoring capabilities*

- Enrichment: OOTB modules/jobs

- Apache Ambari Support

- Security Enhancements

- Spring XD on Pivotal Cloud Foundry (GA Release: May 2015)

spring
by Pivotal.

Resources

- [Project Site](#)

  - [GitHub](#)

  - [Wiki](#)

- [Spring XD Samples](#)

- [JIRA: Features and Issues](#)

spring by Pivotal

Learn More. Stay Connected.

- **Project Site**

  - GitHub

  - Wiki

- **Spring XD Samples**

- **JIRA: Features and Issues**

Twitter: twitter.com/springcentral

YouTube: spring.io/video

LinkedIn: spring.io/linkedin

Google Plus: spring.io/gplus