

SparkSummit 2014

# Apache Spark and the Typesafe Reactive Platform: A Match Made in Heaven



Tuesday, July 1, 14

Copyright © Typesafe, Inc., 2014, Some Rights Reserved.

Photos Copyright © Dean Wampler, 2011-2014, unless otherwise noted. Some Rights Reserved.

The content is free to reuse, but attribution is requested.

<http://creativecommons.org/licenses/by-nc-sa/2.0/legalcode>



# The way we build applications is changing

scroll down, or [Get Started](#)



dean.wampler@typesafe.com  
@deanwampler



# TYPESAFE ACTIVATOR / **TEMP**

## ACTIVATOR TEMPLATES

reactive scala java java8 seed  
starter akka play slick spray

Showing 1 of 146 templates (w

## Spark Workshop



deanwampler



This Activator Spark Wo  
that analyze react



Tuesday, July 1, 14

You can find Spark sample apps (templates) on our community-driven Activator site, [typesafe.com/activator](https://typesafe.com/activator). I have a self-study intro “workshop” for Spark. My demo will be uploaded there next week, once it sucks less ;)



NE Scala Symposium  
@deanwampler  
March 9, 2012

Like a voice  
in the  
wilderness...

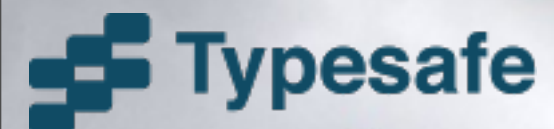
Why Big Data Needs  
to Be Functional



Tuesday, July 1, 14

Since I started Hadoop consulting in 2011, I've wanted functional programming and a functional language for writing applications.  
Title slide from a talk in 2012. The picture is from the Gros Ventres Wilderness, Wyoming.





# Why Spark Is the Next Top (Compute) Model

Philly ETE 2014  
April 22-23, 2014  
@deanwampler  
[polyglotprogramming.com/talks](http://polyglotprogramming.com/talks)

Tuesday, July 1, 14

This is a more recent talk, after Spark had emerged as our savior ;)

Image: Detail of the London Eye

```
public static class LineIndexReducer
    extends MapReduceBase
    implements Reducer<Text, Text,
                        Text, Text> {
    public void reduce(Text key,
        Iterator<Text> values,
        OutputCollector<Text, Text> output,
        Reporter reporter) throws IOException {
        boolean first = true;
        StringBuilder toReturn =
            new StringBuilder();
        while (values.hasNext()) {
            if (!first)
                toReturn.append(", ");
            first=false;
            toReturn.append(
                values.next().toString());
        }
    }
}
```

Awful!

```
val sc = new SparkContext(
  "local", "Inverted Index")

sc.textFile("data/crawl")
  .map { line =>
    val array = line.split("\t", 2)
    (array(0), array(1))
  }
  .flatMap {
    case (path, text) =>
      text.split("""\W+""") map {
        word => (word, path)
      }
  }
  .map {
    case (w, p) => ((w, p), 1)
  }
```

Glorious!



# Spark Deployments...

```
graph TD; Spark[Spark] --- YARN[YARN]; Spark --- Mesos[Mesos];
```

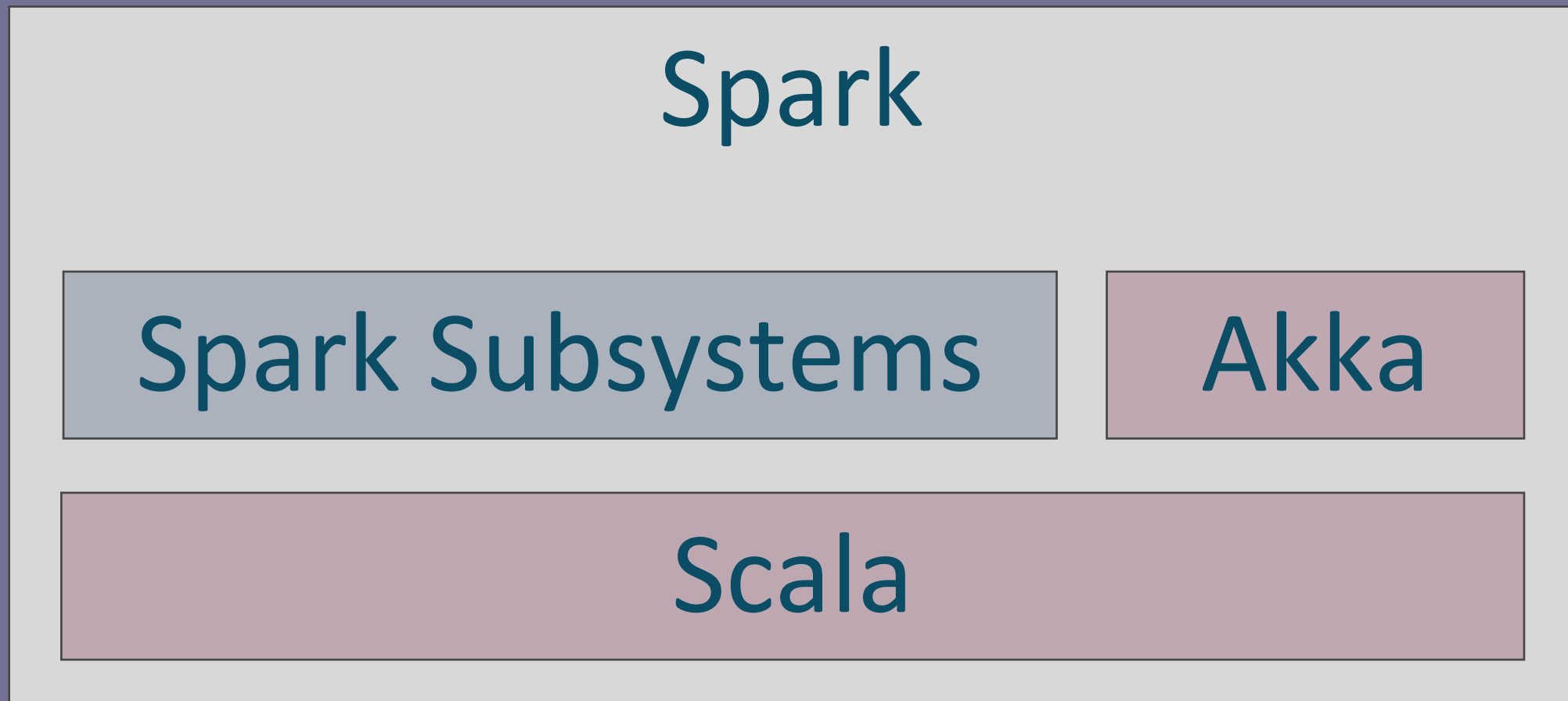
Spark

YARN

Mesos



# Inside Spark

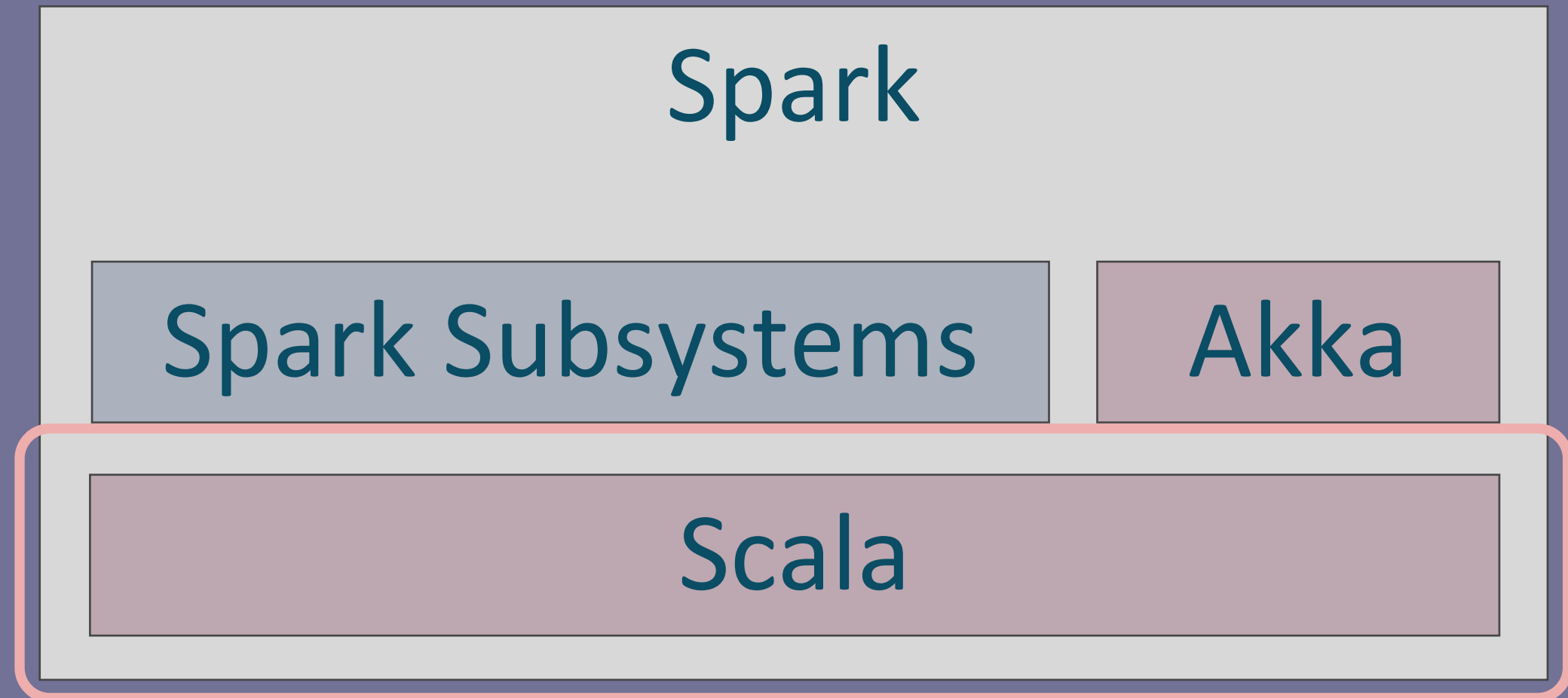


Tuesday, July 1, 14

Typesafe has a special relationship with Spark, unusual compared to the other partners. Spark is *written* in our tool: Scala. Spark also uses Akka, our actor-based middleware, in some parts of Spark.



# We're Working Together



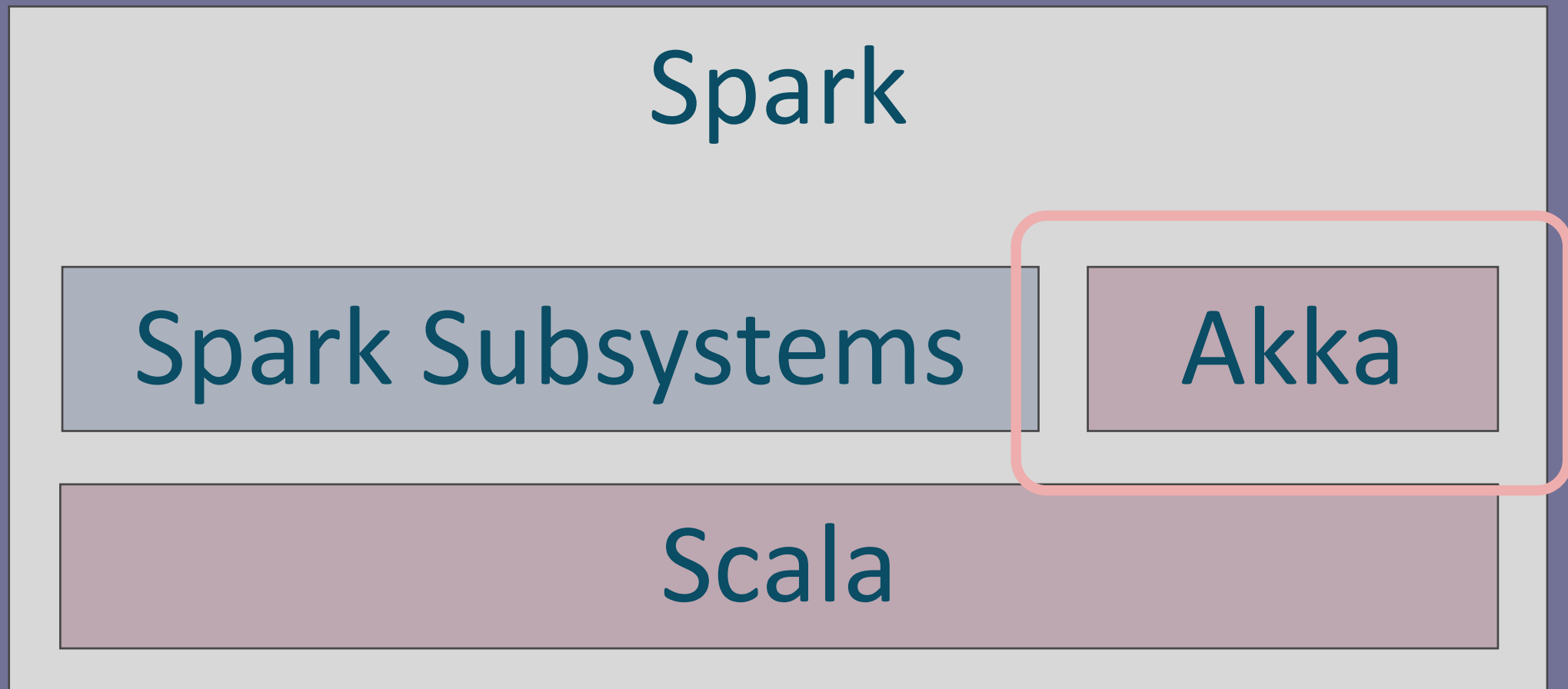
## Scala Improvements

Tuesday, July 1, 14

We're working with Databricks to improve Scala and the libraries to meet their needs, such as improved techniques for specifying and transmitting anonymous functions around a cluster.



# We're Working Together

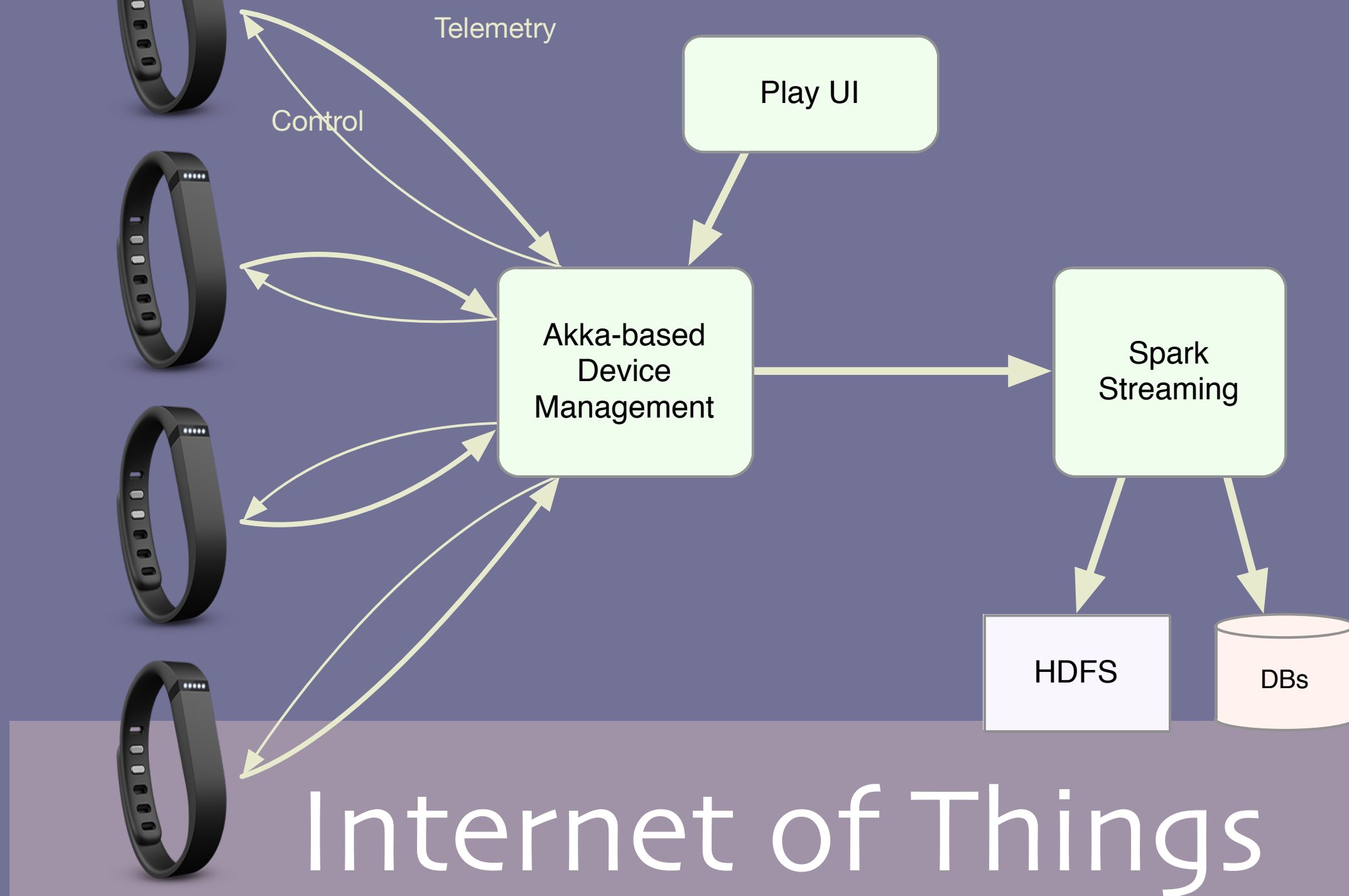


## Libraries: Akka Streams

Tuesday, July 1, 14

We're building new libraries, like Akka Streams, based on the open-source specification for Reactive Streams, that Spark might use. (Reactive Streams have a back-pressure channel, so that no receiver is ever overwhelmed.)

# We're Using Spark



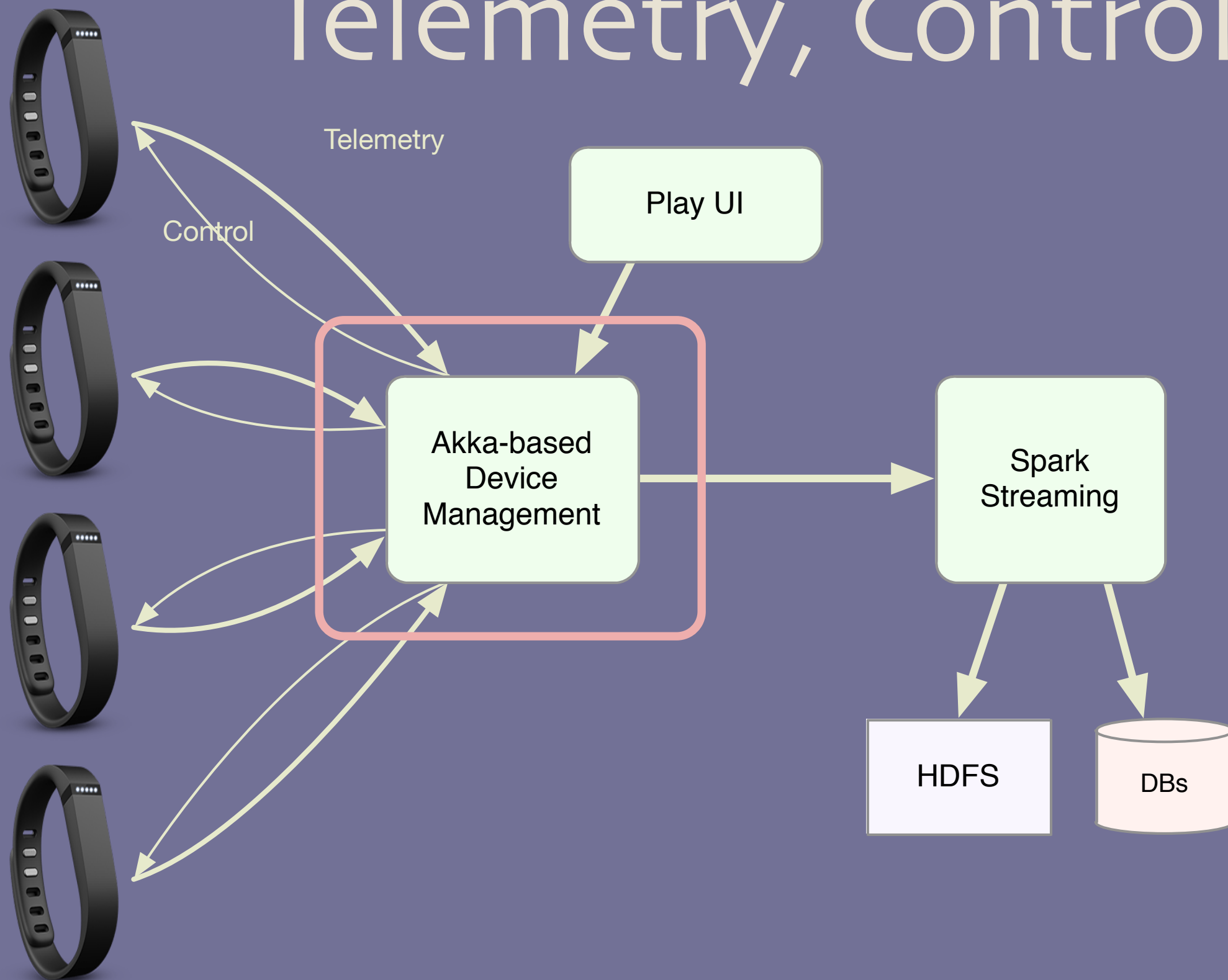
Tuesday, July 1, 14

We're helping clients build new applications. Some of them are building "Internet of Things" (IoT) systems. I'm working on reference designs and tools for IoT. This is a sketch of the demo I'll do. (The Play UI not included...)

Fitbit photo from fitbit.com



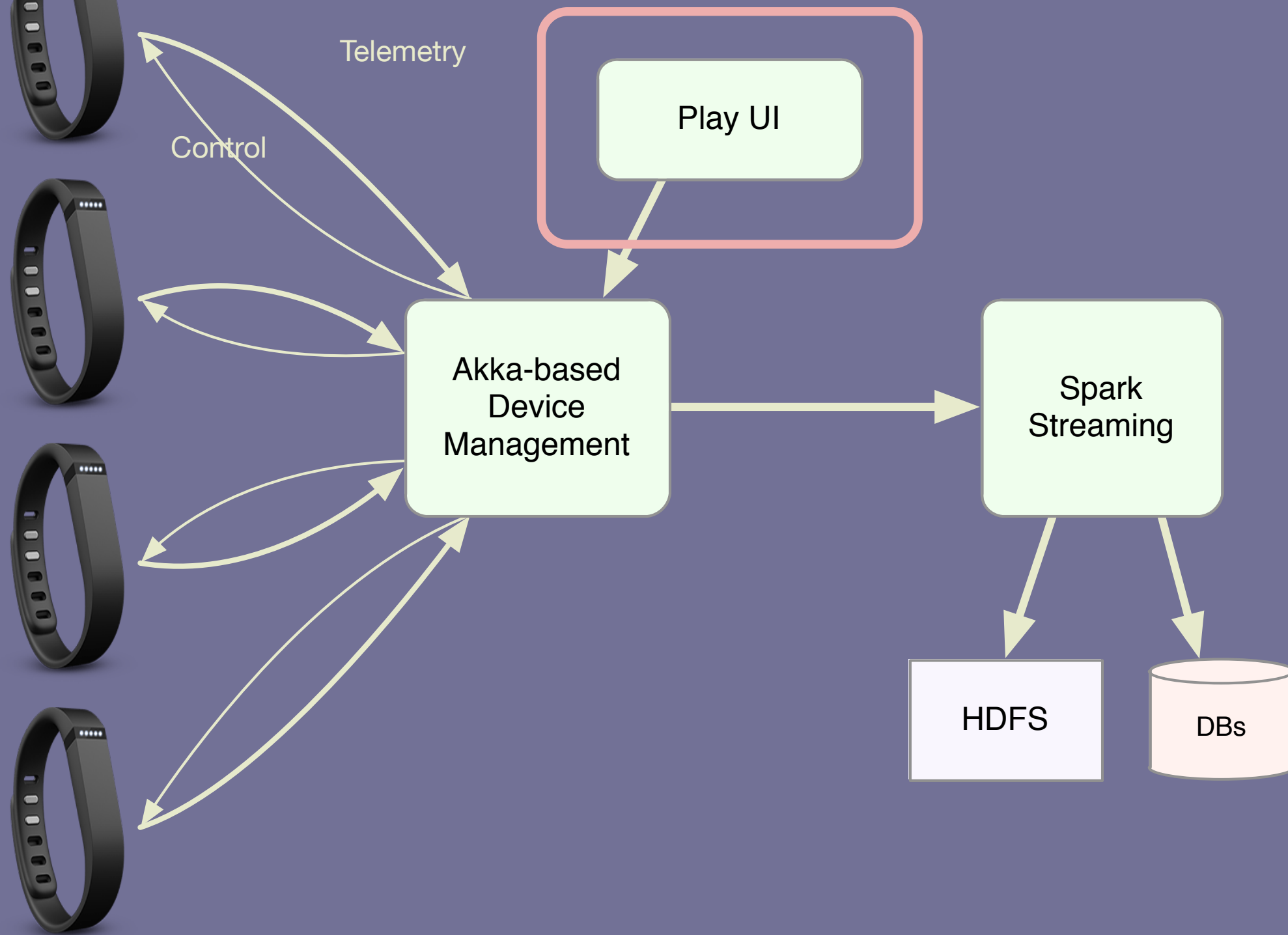
# Telemetry, Control



Tuesday, July 1, 14

You can use Akka for ingesting data and controlling the devices.

# Management

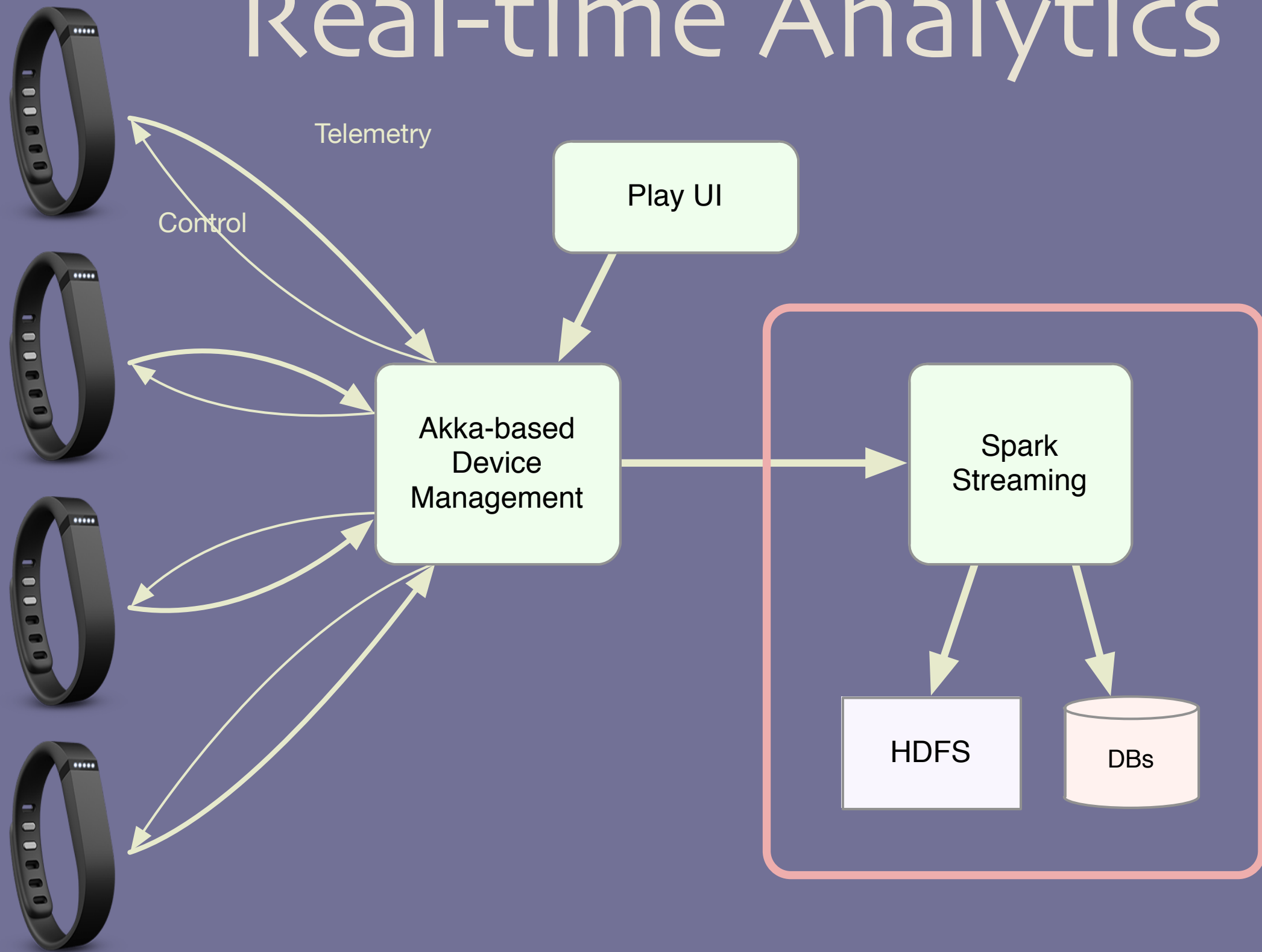


Tuesday, July 1, 14

You can use Play Web UIs for management, data viz.

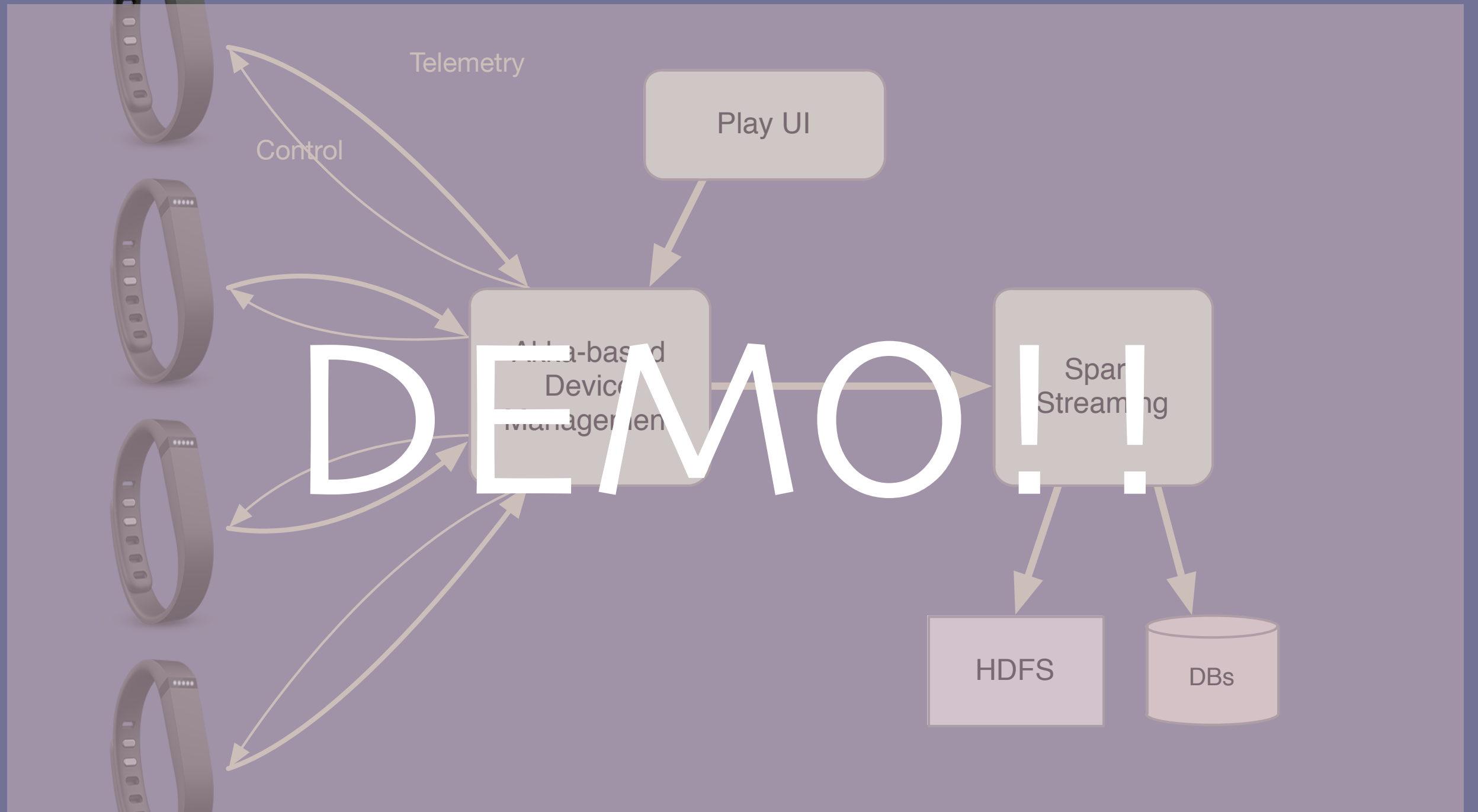


# Real-time Analytics



Tuesday, July 1, 14

You can use Spark streaming for ingesting data for analytics, both batch-mode and real-time.



Tuesday, July 1, 14

The demo shows a simulated client sending data over an Akka Stream to an Akka manager, which then forwards the data to Spark Streaming for analytics.



# TYPESAFE ACTIVATOR / **TEMP**

## ACTIVATOR TEMPLATES

reactive scala java java8 seed  
starter akka play slick spray

Showing 1 of 146 templates (v

## Spark Workshop



deanwampler



This Activator Spark Wo  
that analyze real



Tuesday, July 1, 14

Check out Spark sample apps (templates) on our community-driven Activator site, [typesafe.com/activator](https://typesafe.com/activator). I have a self-study intro “workshop” for Spark. My demo will be uploaded there next week.



 Typesafe

Whew!

dean.wampler@typesafe.com  
@deanwampler  
[polyglotprogramming.com/talks](http://polyglotprogramming.com/talks)

Tuesday, July 1, 14

Copyright © Typesafe, Inc., 2014, Some Rights Reserved.

Photos Copyright © Dean Wampler, 2011-2014, unless otherwise noted. Some Rights Reserved.

The content is free to reuse, but attribution is requested.

<http://creativecommons.org/licenses/by-nc-sa/2.0/legalcode>

Image: My cat Oberon, enjoying the morning sun...