Visual Spark

A visual Spark simulator, composing a data *source*, Spark *flow* and data *sink*.

Data Source

Cassandra, Hdfs, Kafka, Text, etc…

Spark Flow

Transformations and Action

Data Sink

Cassandra, Hdfs, Kafka, Text, etc…

Spark Simulator

Executes a visual simulation of a
Data Source ~> Spark Flow ~> Data Sink
pipeline.

Technologies

- Java or Scala
- JavaFX or ScalaFX
- Spark
- Data Source and Sink Technologies

Rationale

- Spark engineers work with Java, Python and Scala.
- Spark simulations can be built with Java, Jython and Scala using JavaFX or ScalaFX.
- JavaFX includes Charts, 2D, 3D and Animations.
- JavaFX apps and controls decorated by CSS.
- Third party JavaFX charts and components exist.
- JavaFX includes a WebView component that allows for the rendering of Javascript content. For instance, D3 could be used to build an exotic chart.
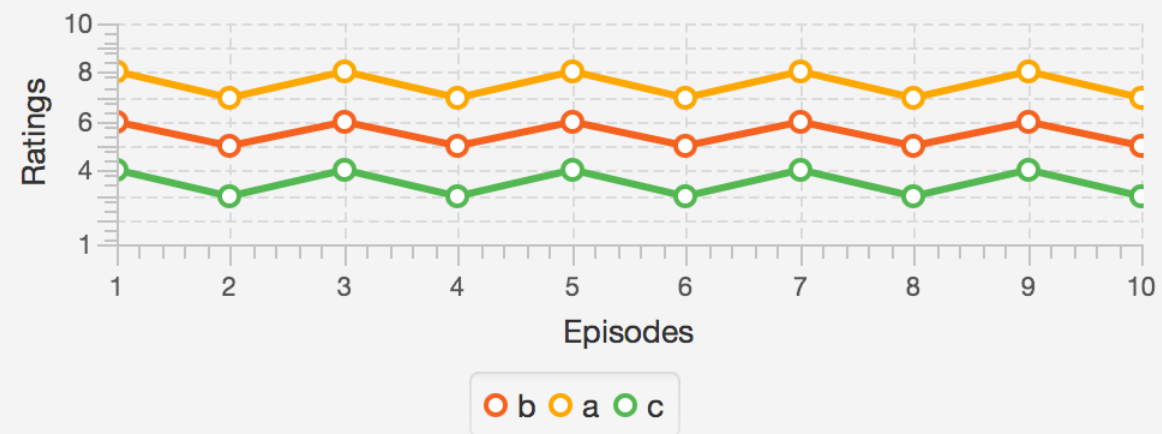
Visual Spark

Play   30 messages processed in 22 seconds.

Source

Program	Season	Episode	Rating	
a	1	1	8	
a	1	2	7	
a	1	3	8	
a	1	4	7	
a	1	5	8	
a	1	6	7	
a	1	7	8	
a	1	8	7	

Flow

Episode Ratings

Ratings

10
8
6
4
2
1

1 2 3 4 5 6 7 8 9 10

Episodes

O b O a O c

Sink

Program Ratings

b(33%)   c(21%)

a(45%)

● a(45%)   ● b(33%)   ● c(21%)
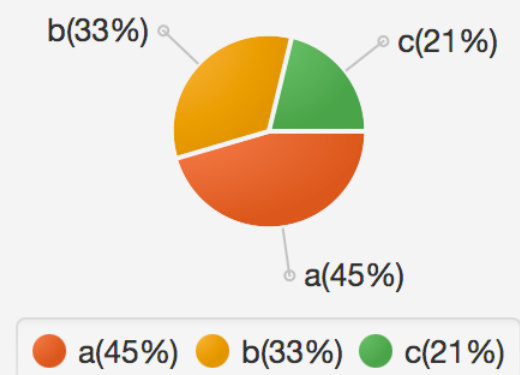
Prototype

- Visual Slick, displayed on the previous slide, details a source ~> flow ~> sink pipeline.
- Source: Kafka Topic Messages
- Flow: Time Series Chart
- Sink: Aggregate Pie Chart