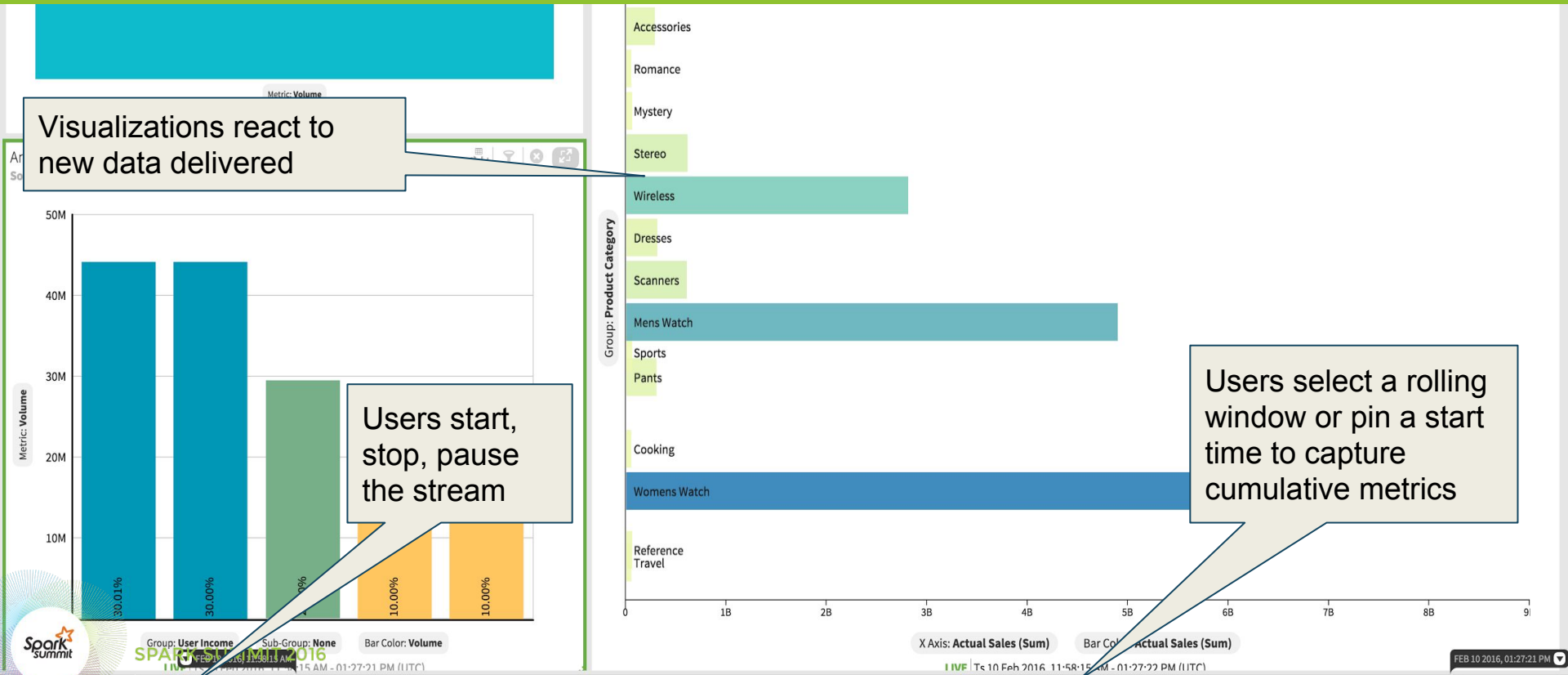
INTERACTIVE VISUALIZATION OF STREAMING DATA POWERED BY SPARK

SPARK SUMMIT 2016
DATA SCIENCE AND ENGINEERING AT SCALE
JUNE 6-8, 2016 SAN FRANCISCO

Streaming @Zoomdata



Drivers for Streaming Data

Data Freshness

Time to Analytic

Business Context

Challenges

- Time
- Frequency
- Retention
- Synchronization
- Order
- Updates

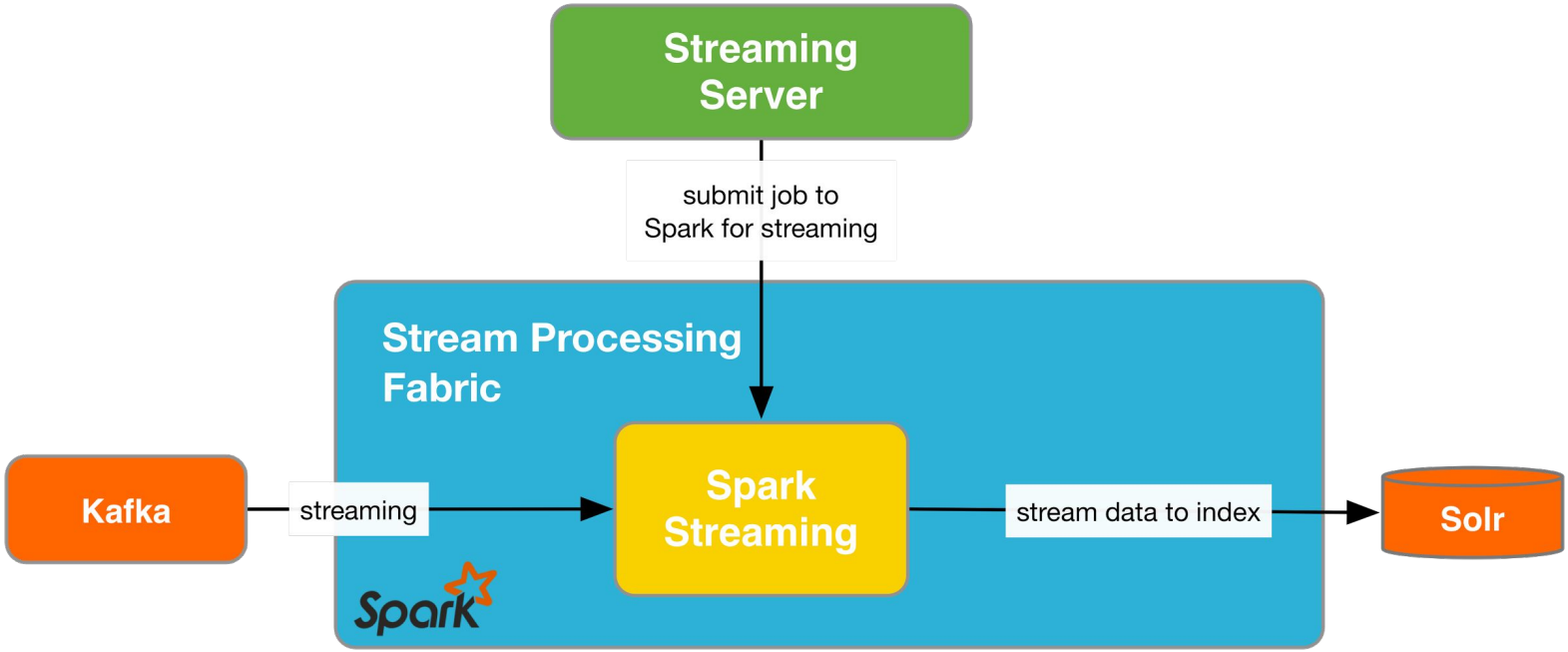
Addressing streaming @Zoomdata

	Historical	Revised
Receive Data	JMS	Kafka
Manipulate Stream	Single JVM in Memory	Spark Streaming
Hold Data in Buffer	MongoDB	Pluggable
Interact with Data	Custom Code	Pluggable

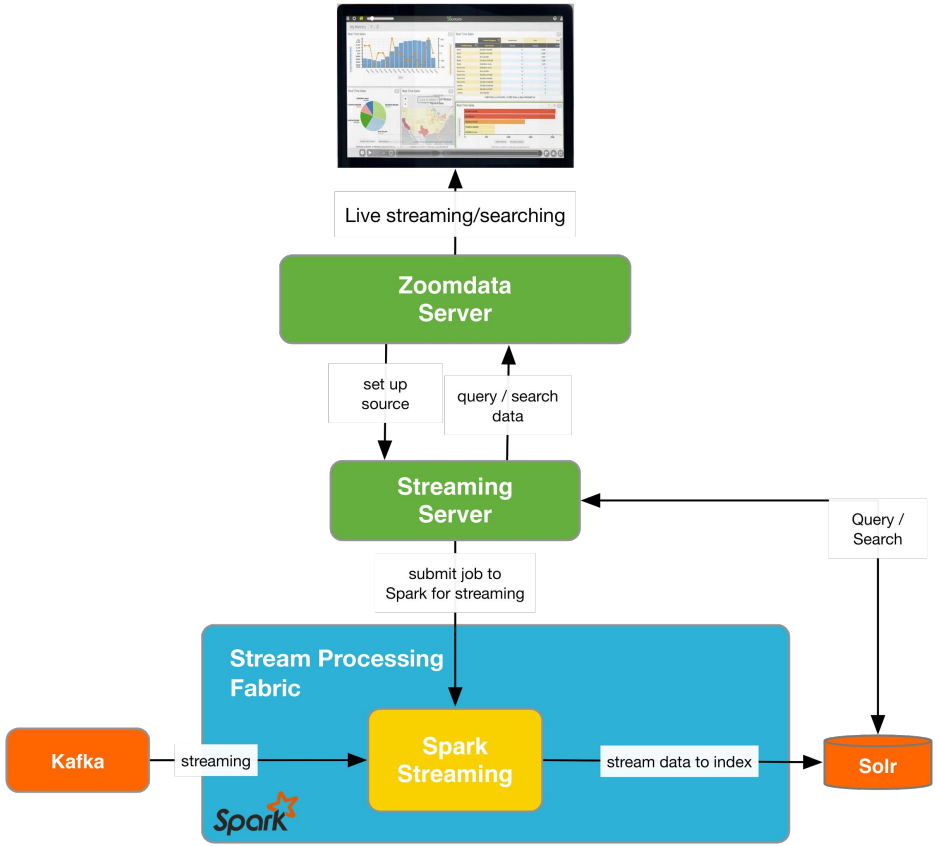
Technology Cast

- The Stream - Kafka, Kinesis, JMS
- Processing Fabric - Spark Streaming
- Landing Area - MemSQL, Solr, Kudu, Others

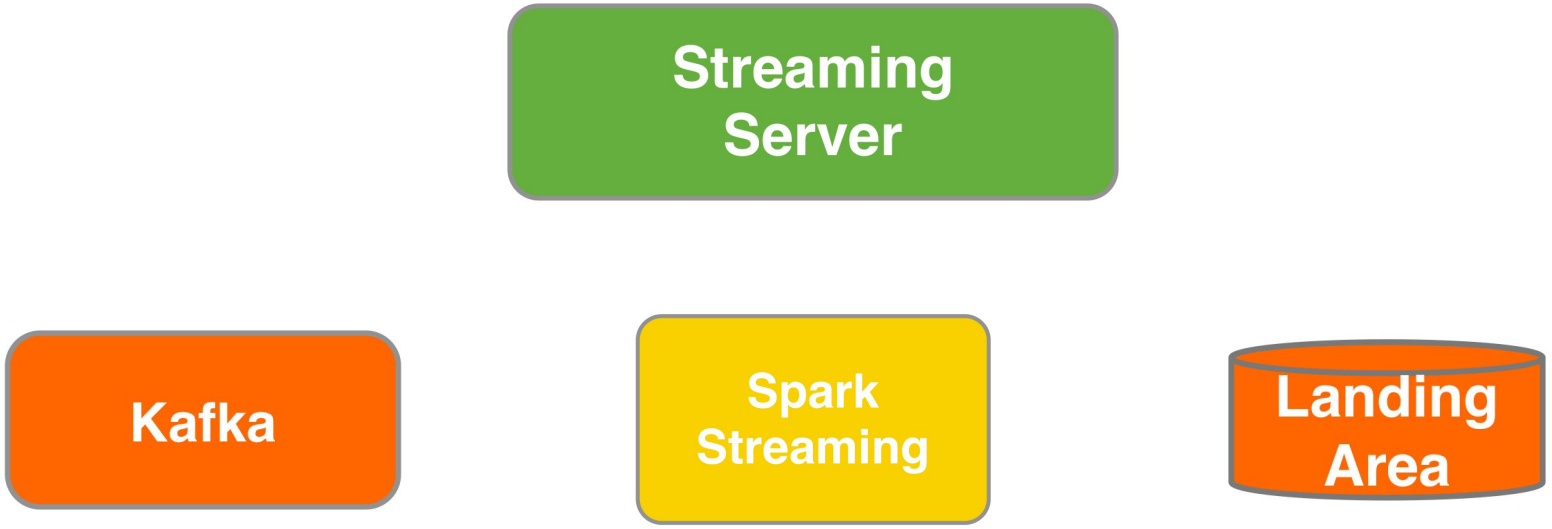
How it looks


Streaming Server
submit job to Spark for streaming
Stream Processing Fabric
Kafka
streaming
Spark Streaming
stream data to index
Solr
Spark

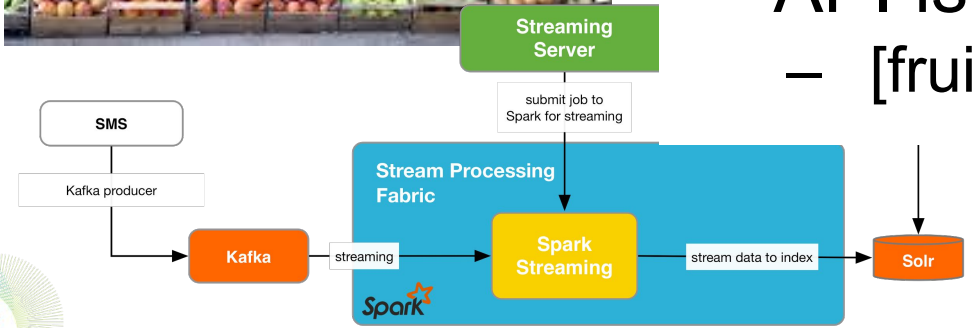
With the rest of the app



Scale Out



Let's put it all together



- Process fruit orders from all over
- Buyers make orders via an API
- API is SMS
  - [fruit] [quantity] e.g., apple 10

Demo

Benefits

- Contextual Expressiveness with Streaming Data
- Independent scalability (scale-up, scale-around)
- Expressiveness powered by Spark -- using Windowing (dataframe API with stream)
- DR COOP, other Data management concerns

Future Work

- Cross stream synchronization & fusion
- On-demand scale out and resource management via Mesos
- Schema evolution
- More extensible landing strategies

Questions

For more information contact:

ruhollah@zoomdata.com

Come visit Zoomdata at our booth H2!

Thank You