

# X-STREAM ANALYTICS USING URIKA-XA

Spark Streaming for Actionable Insight

Ian Bird

Mercredi 24 juin 2015

# Urika-XA Platform Summary



**Turnkey  
Advanced Analytics  
Platform**

**Next-Generation  
System Architecture**

**Engineered for  
Performance**

- **Hadoop and Spark ecosystem**
- **Emerging analytic workloads**
- **Open platform for current and future frameworks**
- **Single pane of glass for system management**
- **Innovative use of storage technologies**
- **Battle-tested on cutting-edge government/scientific analytic applications**
- **Ready for the enterprise**
- **Dense footprint: over 1,500 cores, 6TB memory**
- **38TB SSD and 120TB POSIX-compliant high-performance storage**
- **InfiniBand**
- **Cloudera Distribution of Hadoop**
- **Scale out to multi-rack configurations**

COMPUTE | STORE | ANALYZE

© CRAY INC. 2015



# Streaming Social Media

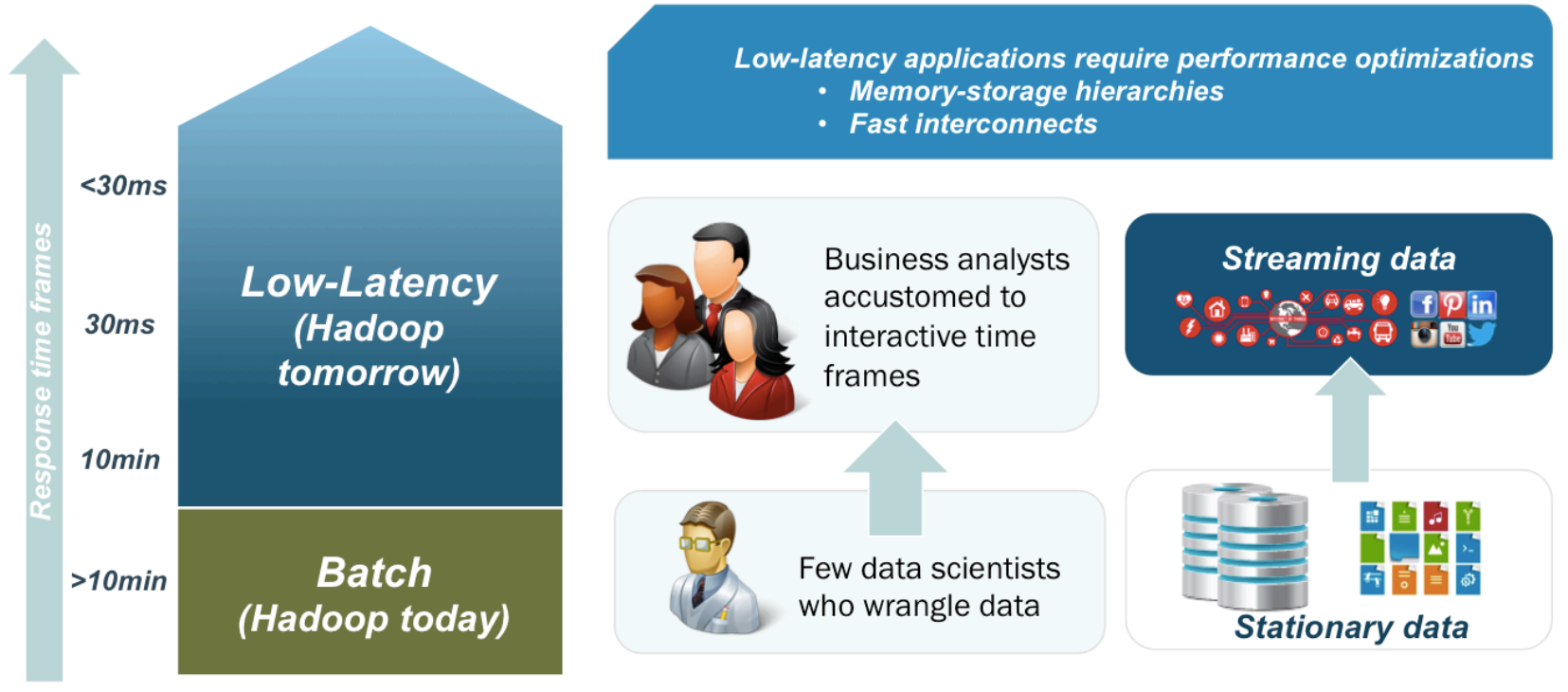
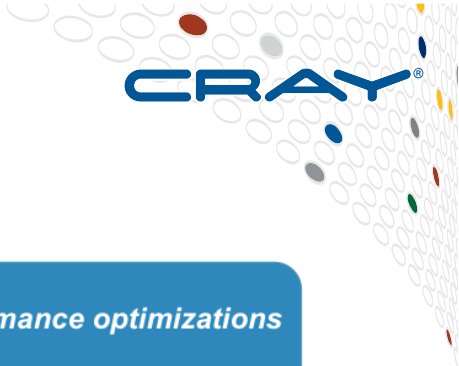
## Spark Streaming Demo

---

COMPUTE | STORE | ANALYZE

© CRAY INC. 2015

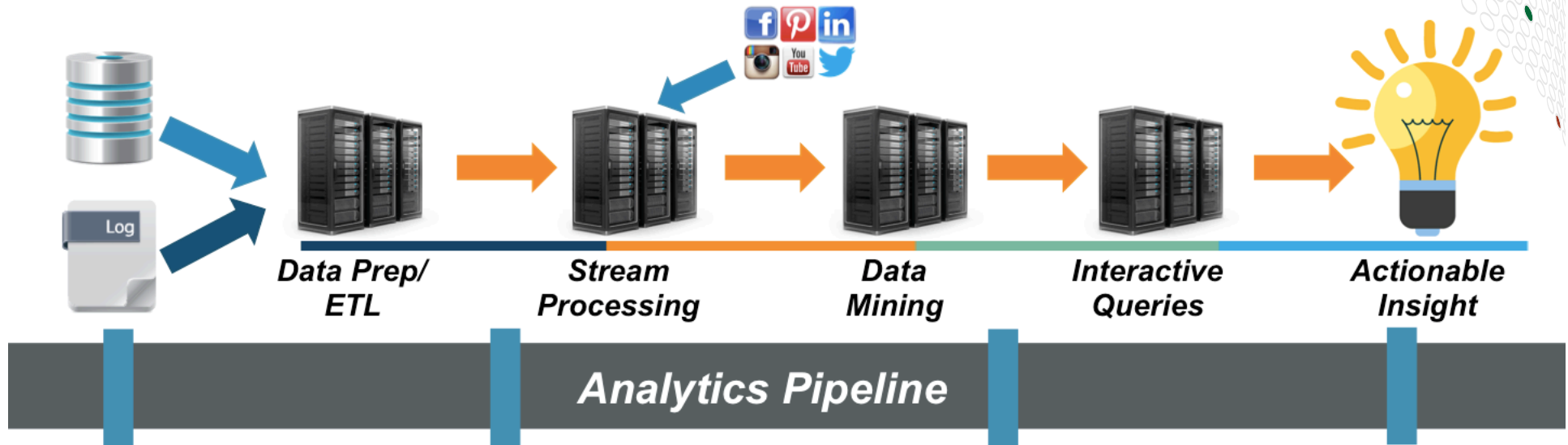
# The quest for in-time Analytics



COMPUTE | STORE | ANALYZE

# Analytics Cluster Sprawl

CRAY®



- Multiple steps of analytics processing
- Each with different computing characteristics
- Requiring separate cluster for each step

- High TCO**
- Large datacenter footprint
  - High management cost
  - Significant data movement

COMPUTE | STORE | ANALYZE



Spark  
SQL

Spark  
Streaming

Mlib  
(machine  
learning)

GraphX  
(graph)

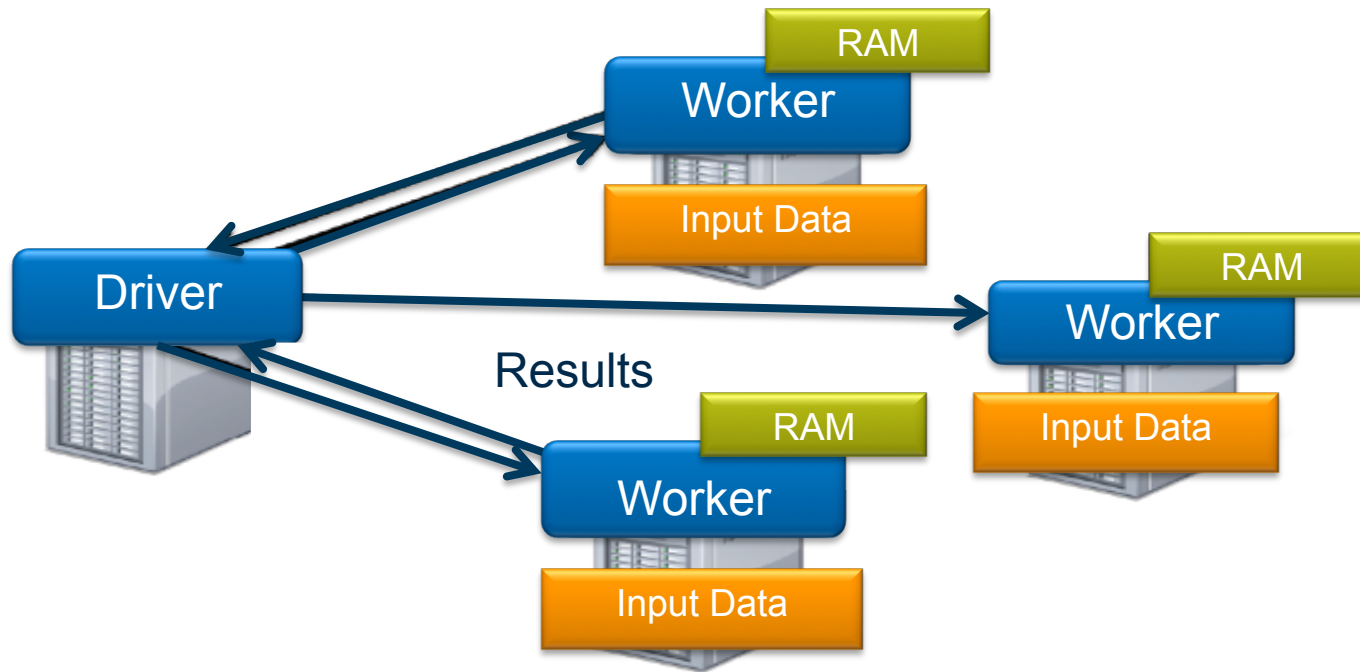
Apache Spark

---

COMPUTE | STORE | ANALYZE

© CRAY INC. 2015

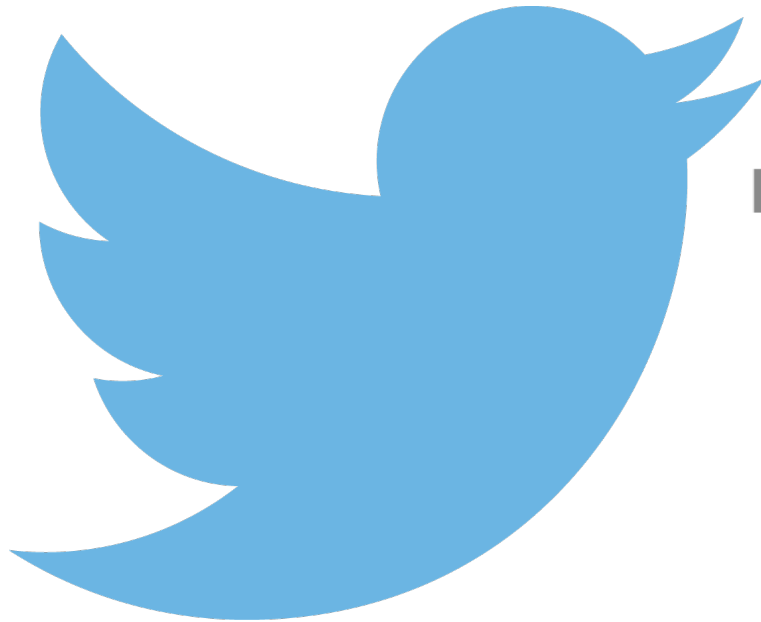
# Spark Runtime



COMPUTE | STORE | ANALYZE

© CRAY INC. 2015

# Hashtags, Mentions, Hullabaloo



UKLabour  
eureferendum  
ScotToriesNicolaSturgeon  
SLATUKIP Grexit BetterOffOut  
labour Brexit EUreferendumYES  
UKIP eeas CameronConservatives  
Nigel Farage eu EUreferendumNO  
David scottishlabour  
euobs  
EUobserver  
ukipmeps Commission  
referendum  
LibDems MailOnline  
labourpress

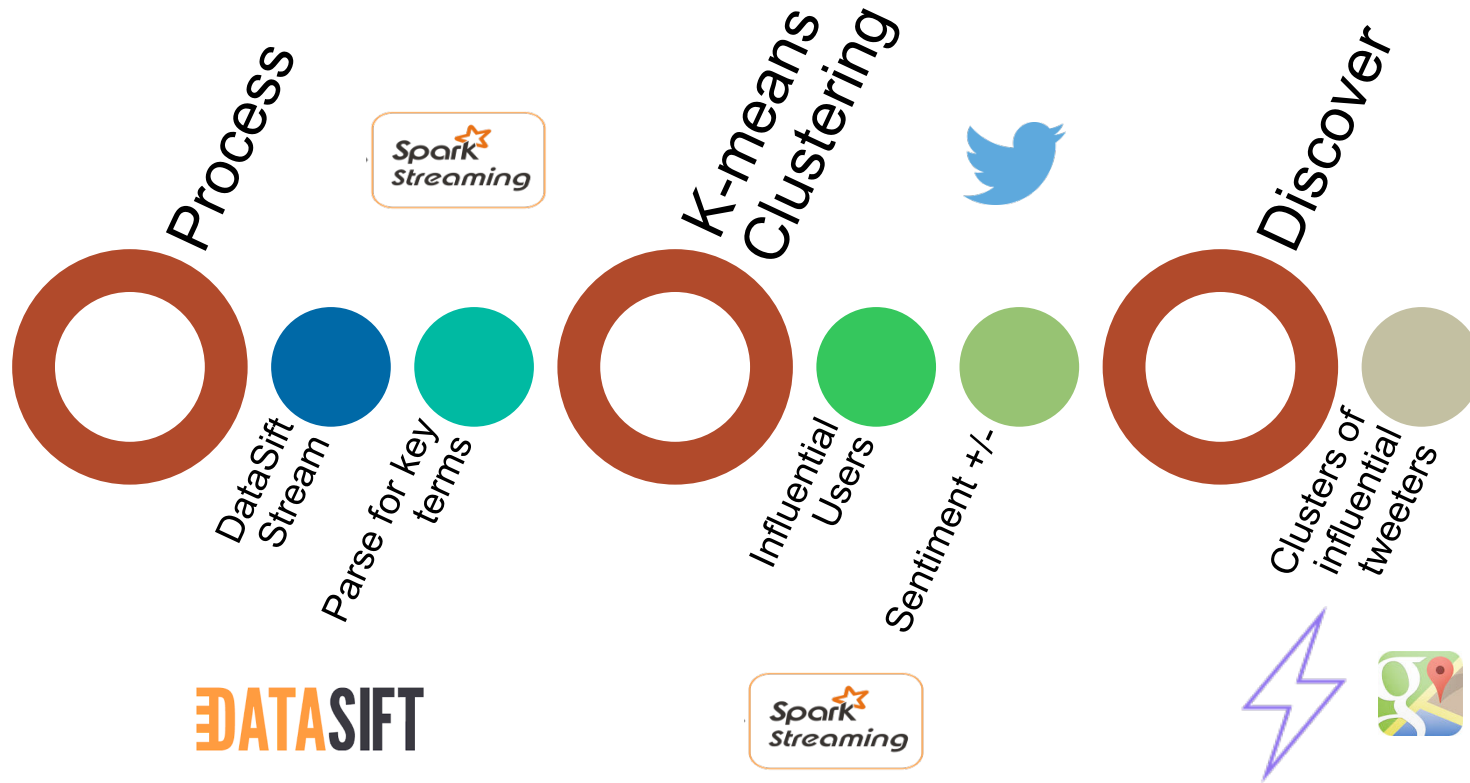


# 1 – UK EU exit, Greek EU exit, Hullabaloo



COMPUTE | STORE | ANALYZE

# 2 – Clusters of Influential Tweeters

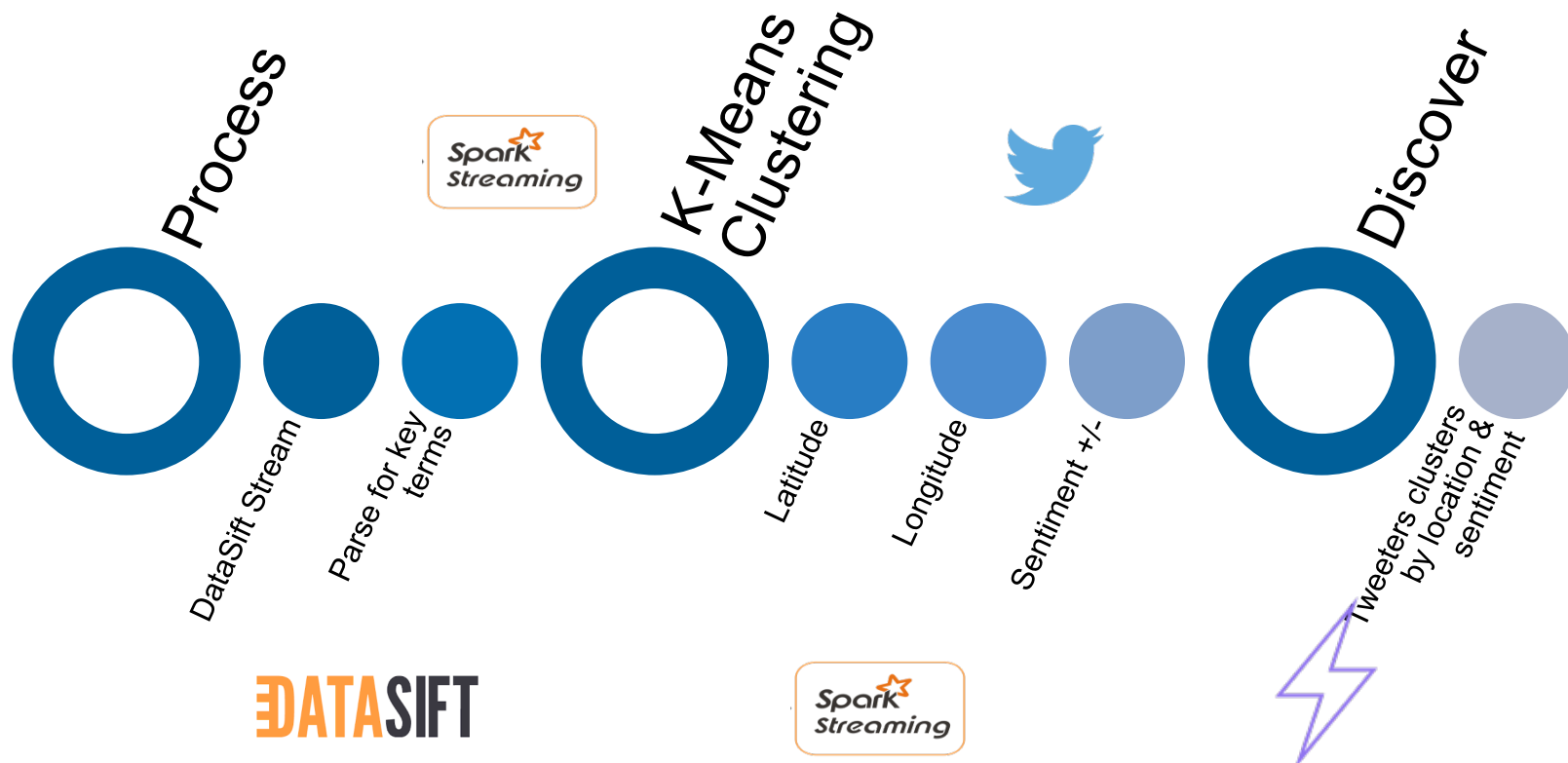


**DATA SIFT**

COMPUTE | STORE | ANALYZE

© CRAY INC. 2015

# 3 – Clusters of Tweet by location



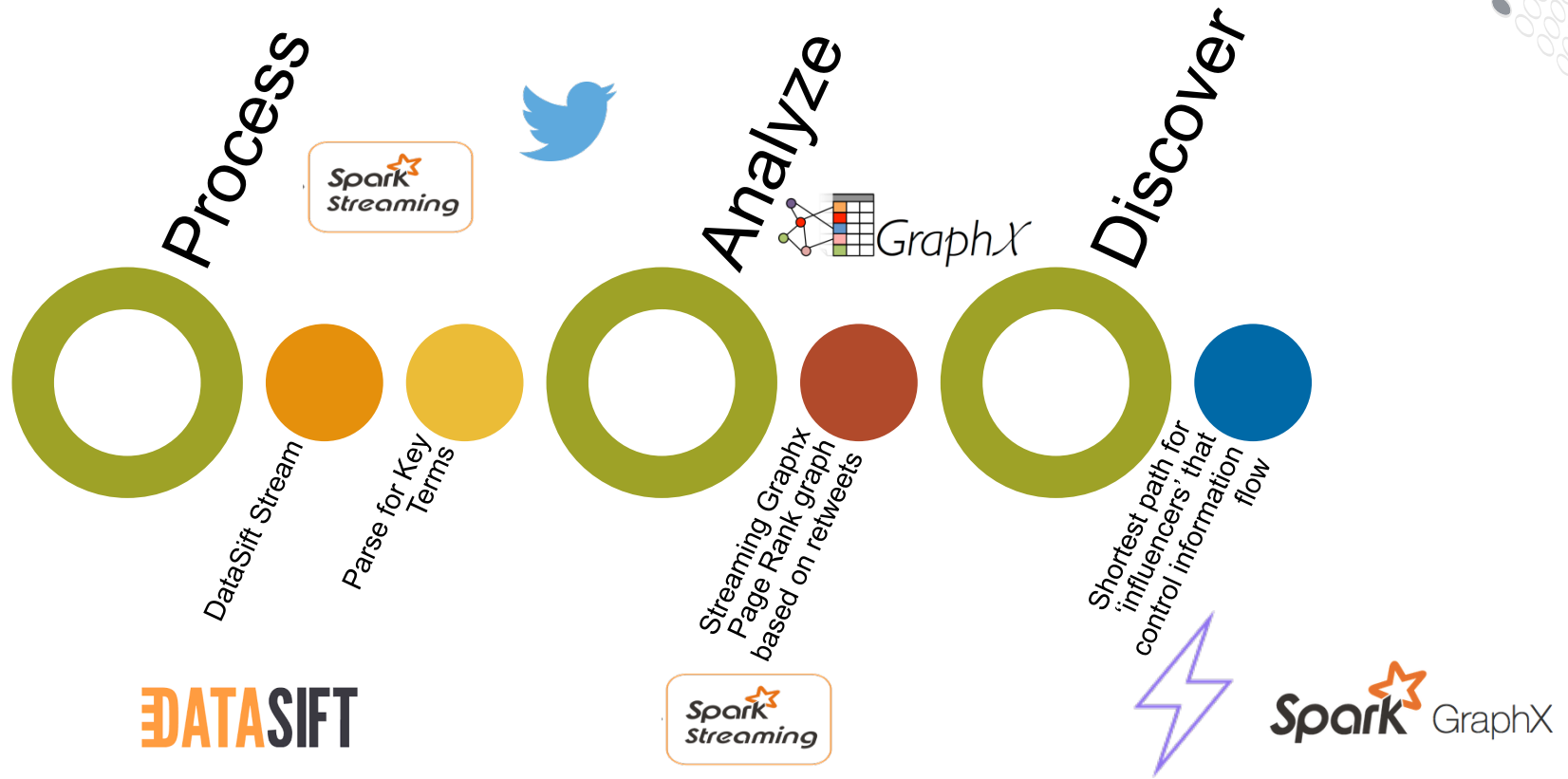
**DATA SIFT**

**Spark Streaming**

COMPUTE | STORE | ANALYZE

© CRAY INC. 2015

# 4 – Clusters of Tweets from original mention



COMPUTE | STORE | ANALYZE

# Cray Urika-XA™ Extreme Analytics Platform

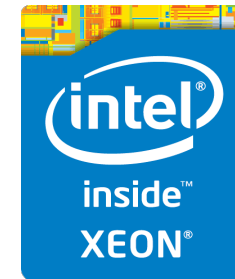
*Advanced Analytics at Lower TCO*



**CRAY**  
**URIKA-XA**

- Pre-Integrated, Open Platform
- Hadoop and Spark ecosystem
- High Performance storage technologies
- Multi-rack configurations

Based on the Intel® Xeon® processor



COMPUTE | STORE | ANALYZE

© CRAY INC. 2015