How Spark is Enabling
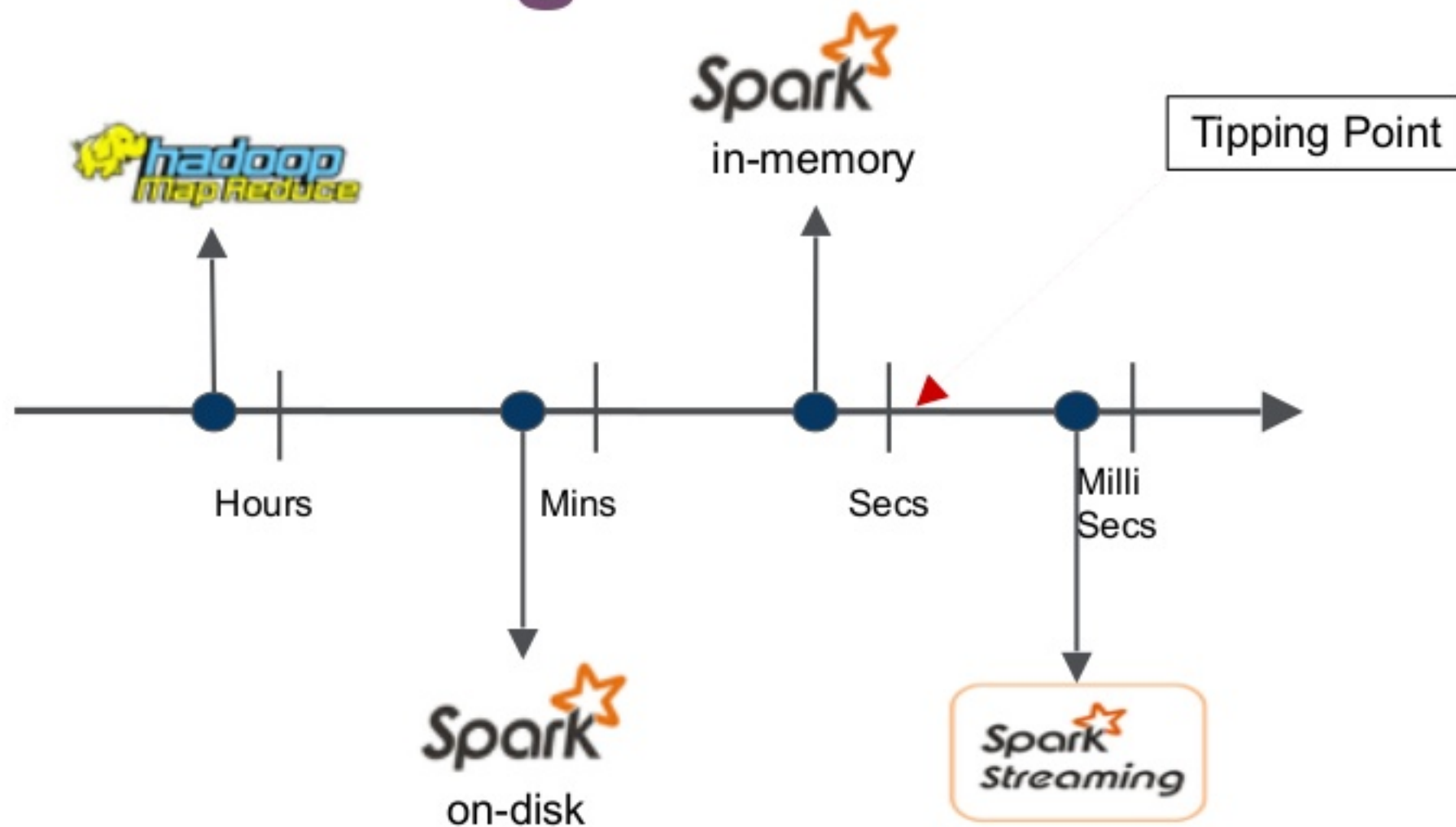# the New Wave of Converged Applications
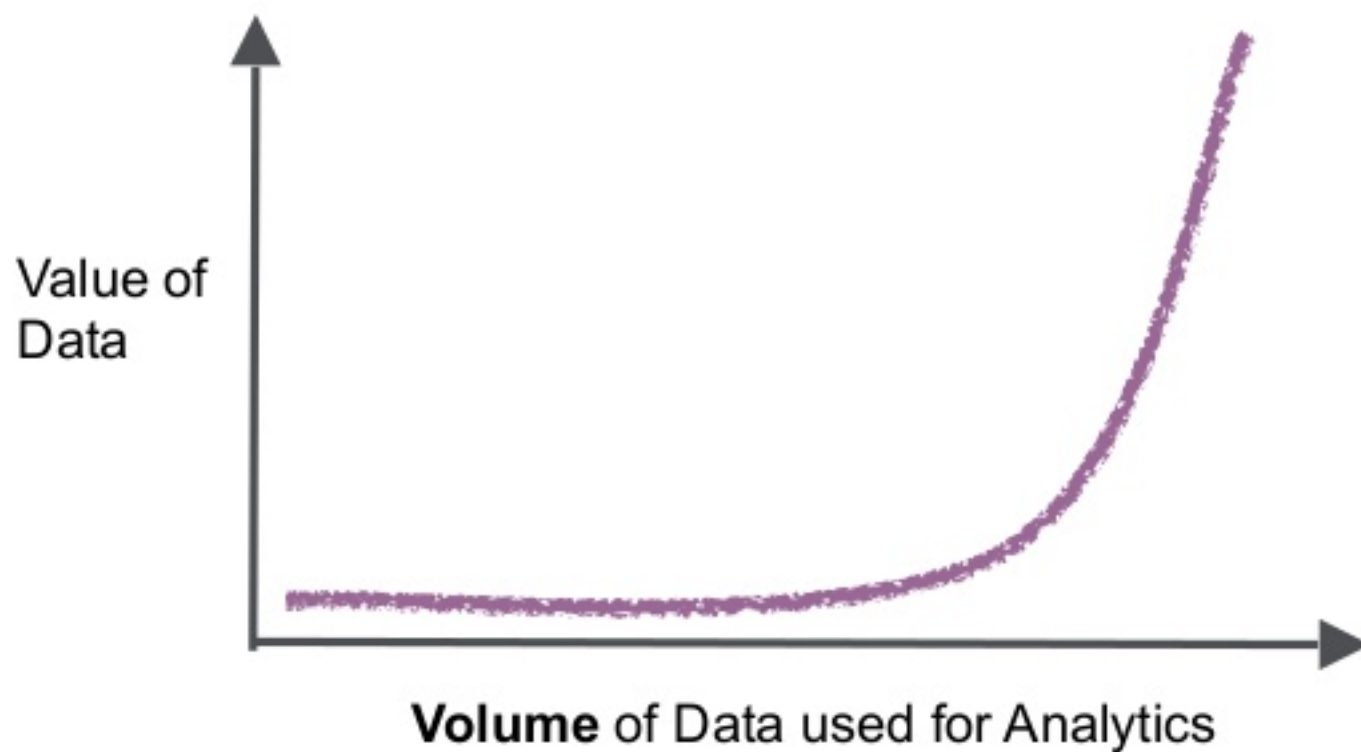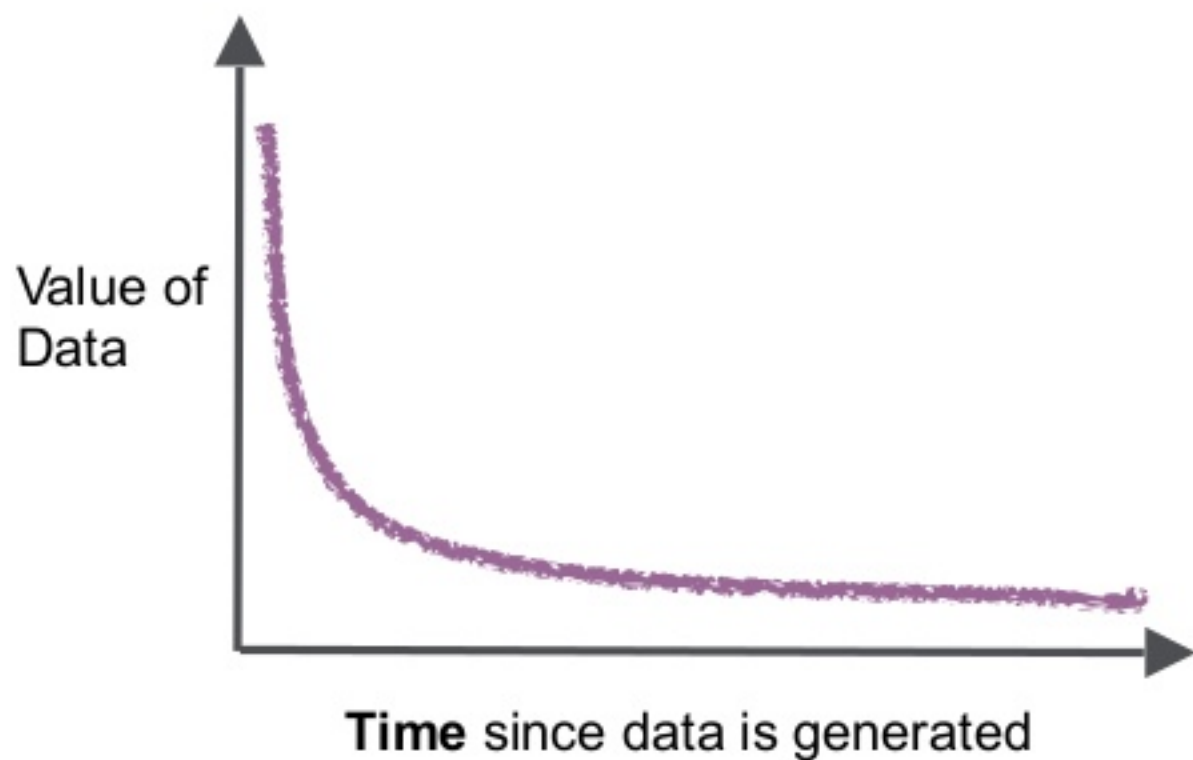
Tugdual Grall

MapR Technologies

SPARK SUMMIT
EUROPE 2016

Decreasing Job Latencies

Analytics & ETL: Batch or Continuous ?

Value of
Data

Time since data is generated

Value of
Data

Volume of Data used for Analytics
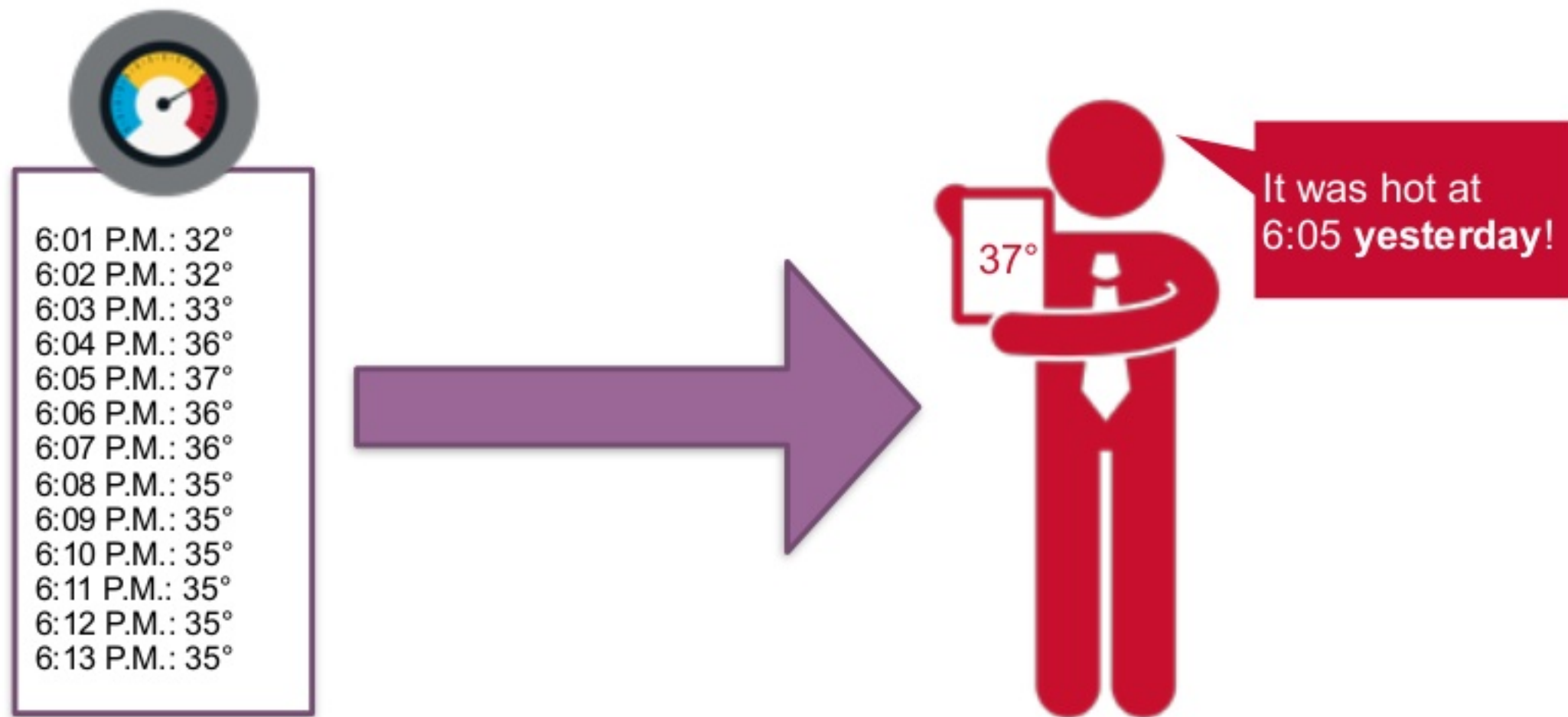
It's not an either or, **you have to do both**

Why Stream Processing?

Batch processing may **be too late** for some events

6:01 P.M.: 32°
6:02 P.M.: 32°
6:03 P.M.: 33°
6:04 P.M.: 36°
6:05 P.M.: 37°
6:06 P.M.: 36°
6:07 P.M.: 36°
6:08 P.M.: 35°
6:09 P.M.: 35°
6:10 P.M.: 35°
6:11 P.M.: 35°
6:12 P.M.: 35°
6:13 P.M.: 35°

37°

It was hot at 6:05 **yesterday**!

Why Stream Processing?

It's becoming important to **process events as they arrive**

Advanced Analytics

Descriptive

- What Happened
- Why did it happen
- Discovery in nature
- **Batch** Analytics

Predictive

- What will happen
- Combines historical data with rules and algorithms
- **ML** (Batch + Real Time)

Streaming

- Analyse data as it happens
- Triggers and Alarms.
- Anomaly detection
- **Continuous** ETL and Analytics

Prescriptive

- What + When + Why
- Suggestions
  to take advantage of future opportunity or mitigate risks
- **Agility** is key to success.

There is a need to **converged** these Analytics

Converged Computing

	Offline	Real Time
Programmatic	Spark & ML	Spark Streaming
SQL	Spark SQL	Spark Structured Streaming

The Many "Convergences" In Progress
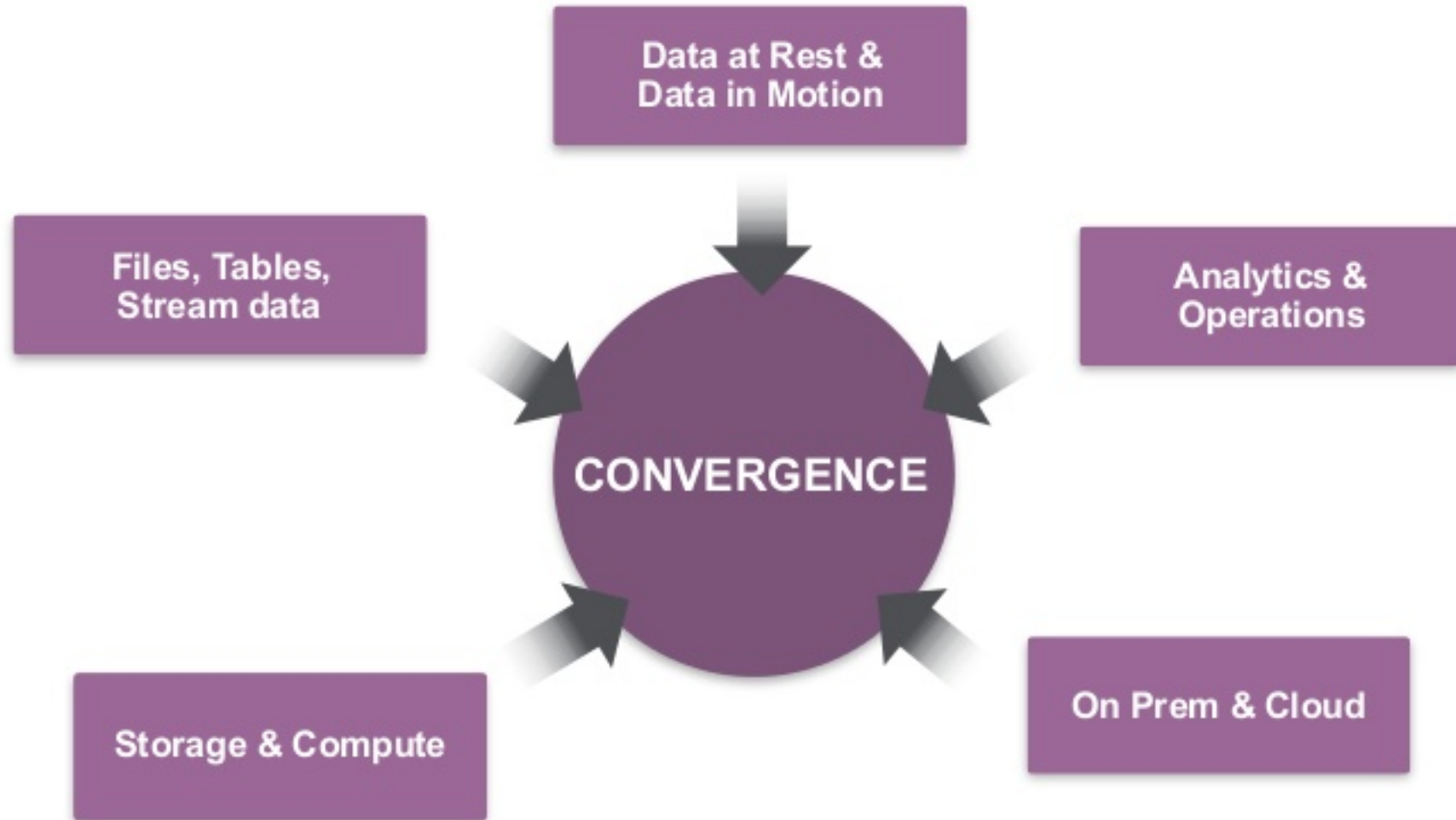
Spark on Non-Converged Platform
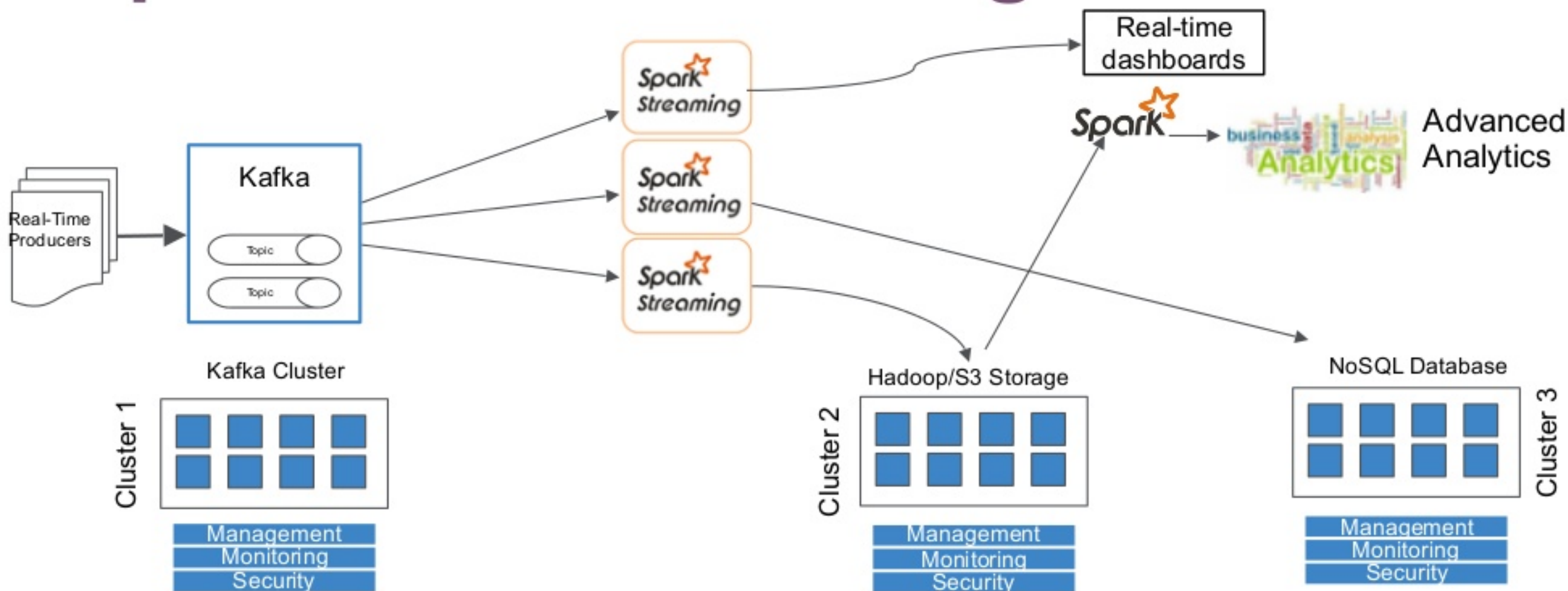
- Redundant 3x **Management, Monitoring and Security**
- Redundant 3x **Data Storage**

Converged Computing & Converged Data Management

Notebook*	Workflow Management	Commercial Apps	QSS

Installer, Management, and Monitoring

Spark Streaming	Spark Machine Learning (MLlib)	Spark SQL	GraphX	SparkR

Spark Core

MapR-FS	MapR-DB	MapR Streams
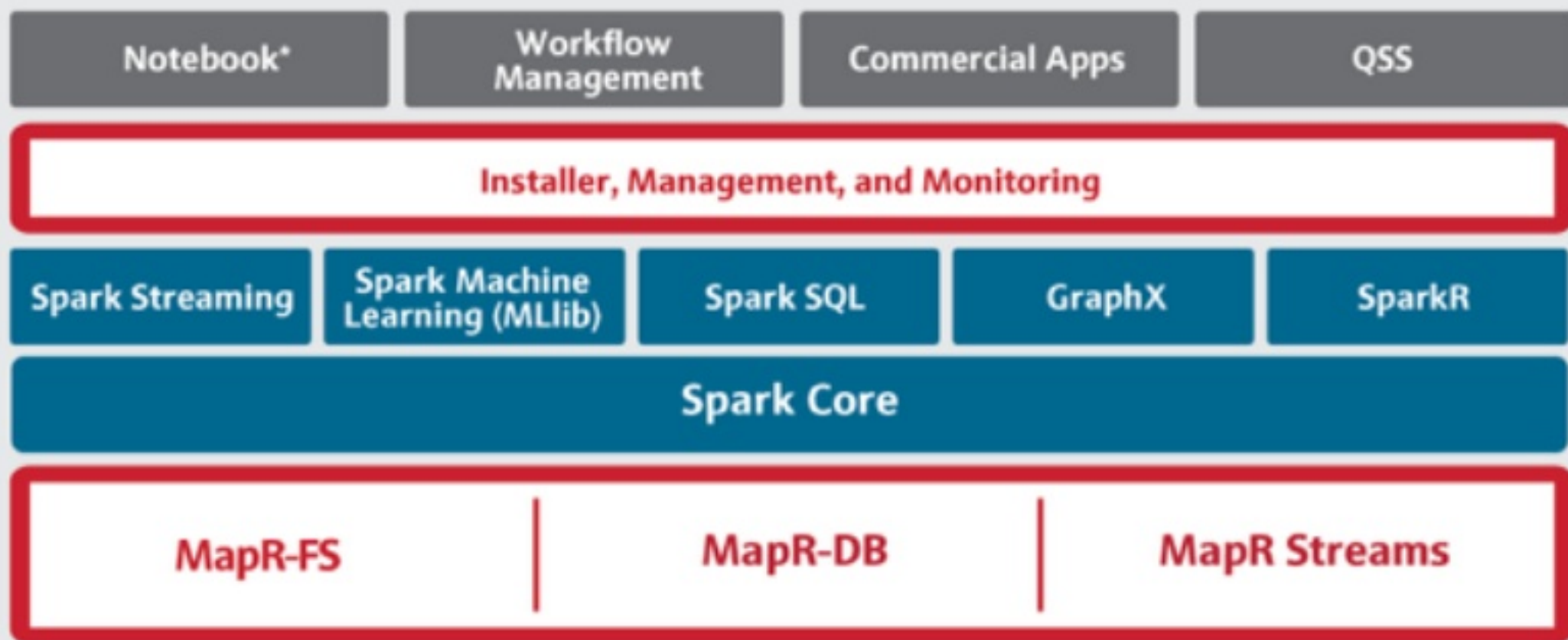
MapR Converged Data Platform

Open Source Engines & Tools

Commercial Engines & Applications

Processing

| | | | Search and Others | Cloud and Managed Services | VERTICA SAP MySQL | Custom Apps |

Unified Management and Monitoring

Data

HDFS API   POSIX, NFS   HBase API   JSON API   Kafka API

Web-Scale Storage
MapR-FS

Database
MapR-DB

Event Streaming
MapR Streams

High Availability   **Real Time**   **Unified Security**   **Multi-tenancy**   **Disaster Recovery**   **Global Namespace**
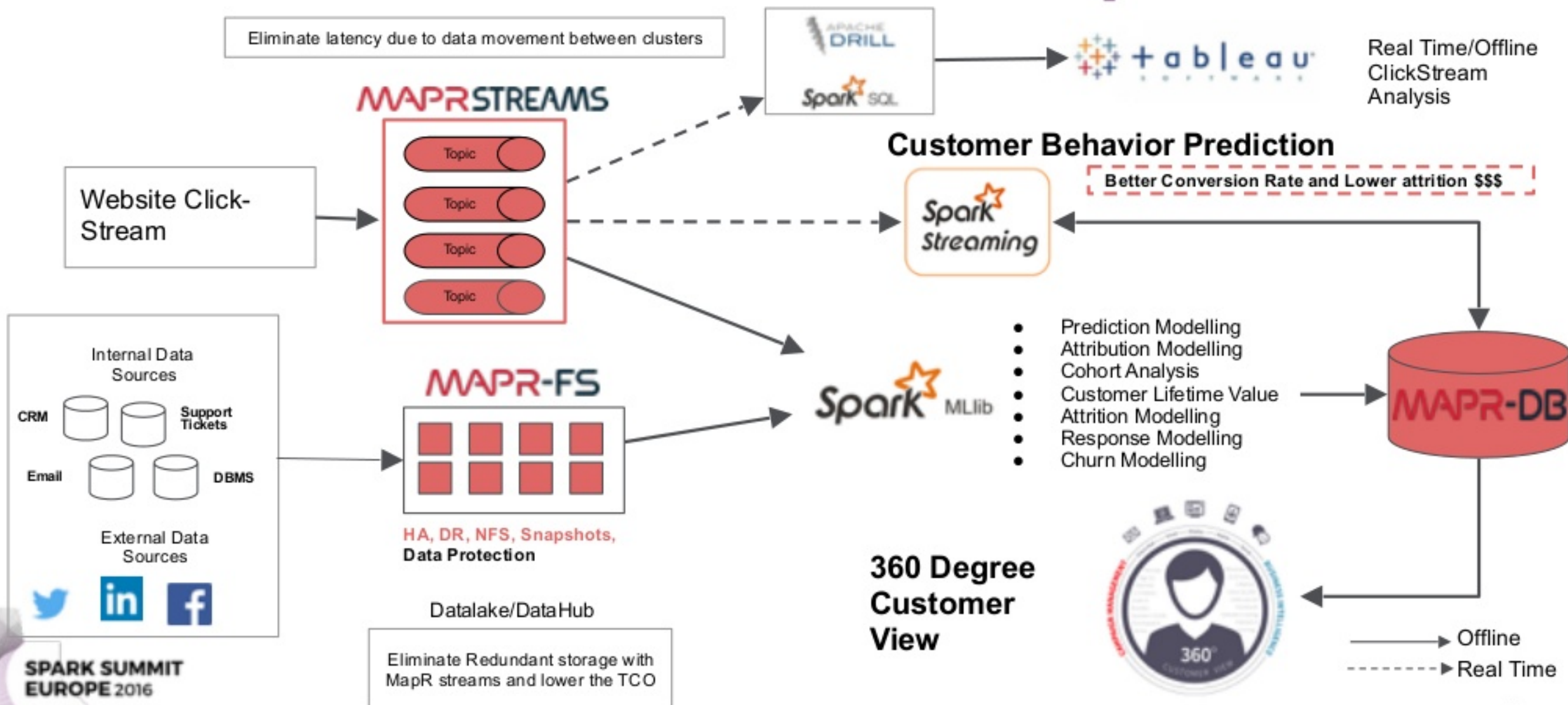
Enterprise-Grade Platform Services

SAMPLE CUSTOMER USE CASE

Customer 360 & Behavior prediction

Eliminate latency due to data movement between clusters

APACHE DRILL

Spark SQL

tableau SOFTWARE

Real Time/Offline ClickStream Analysis

MAPR STREAMS

Topic
Topic
Topic
Topic

Customer Behavior Prediction

Spark Streaming

Better Conversion Rate and Lower attrition $$$

Website Click-Stream

Internal Data Sources

CRM   Support Tickets

Email   DBMS

External Data Sources

MAPR-FS

Spark MLlib

- Prediction Modelling
- Attribution Modelling
- Cohort Analysis
- Customer Lifetime Value
- Attrition Modelling
- Response Modelling
- Churn Modelling

MAPR-DB

HA, DR, NFS, Snapshots, Data Protection

360 Degree Customer View

360 CUSTOMER VIEW

Datalake/DataHub

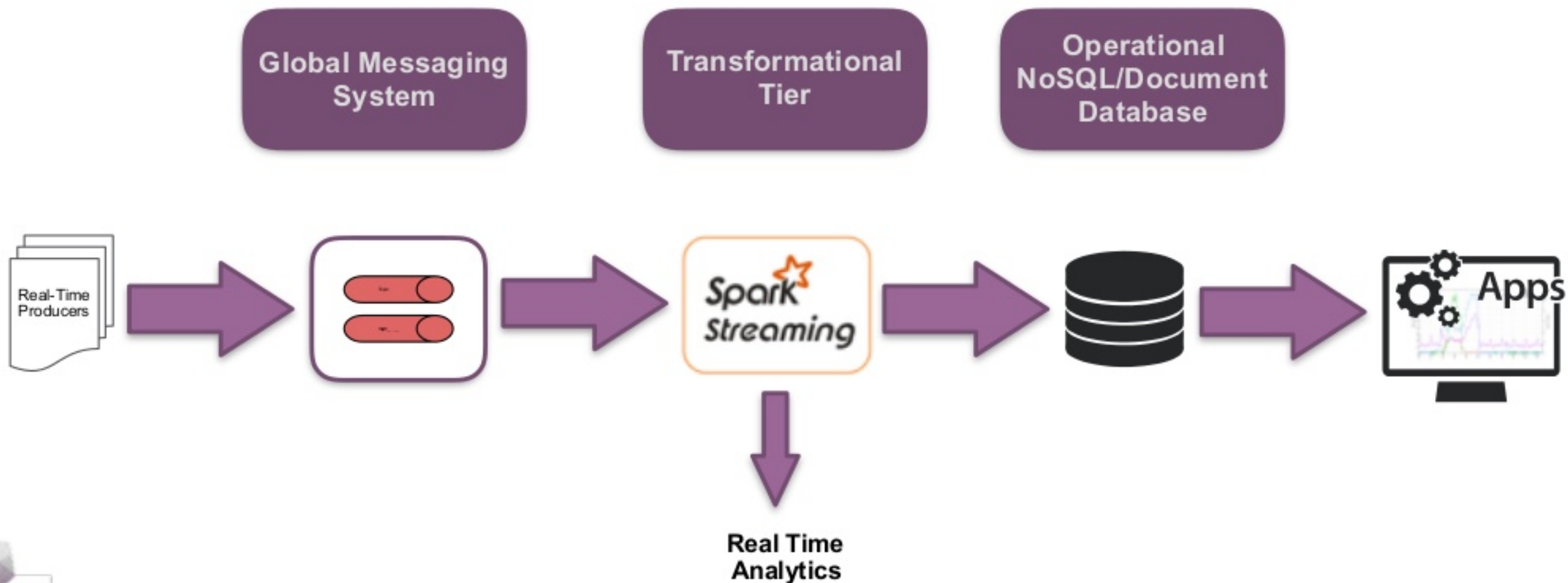Eliminate Redundant storage with MapR streams and lower the TCO

Offline

Real Time

13

STREAMING FIRST ARCHITECTURE

What Do We Exactly Need to Do ?
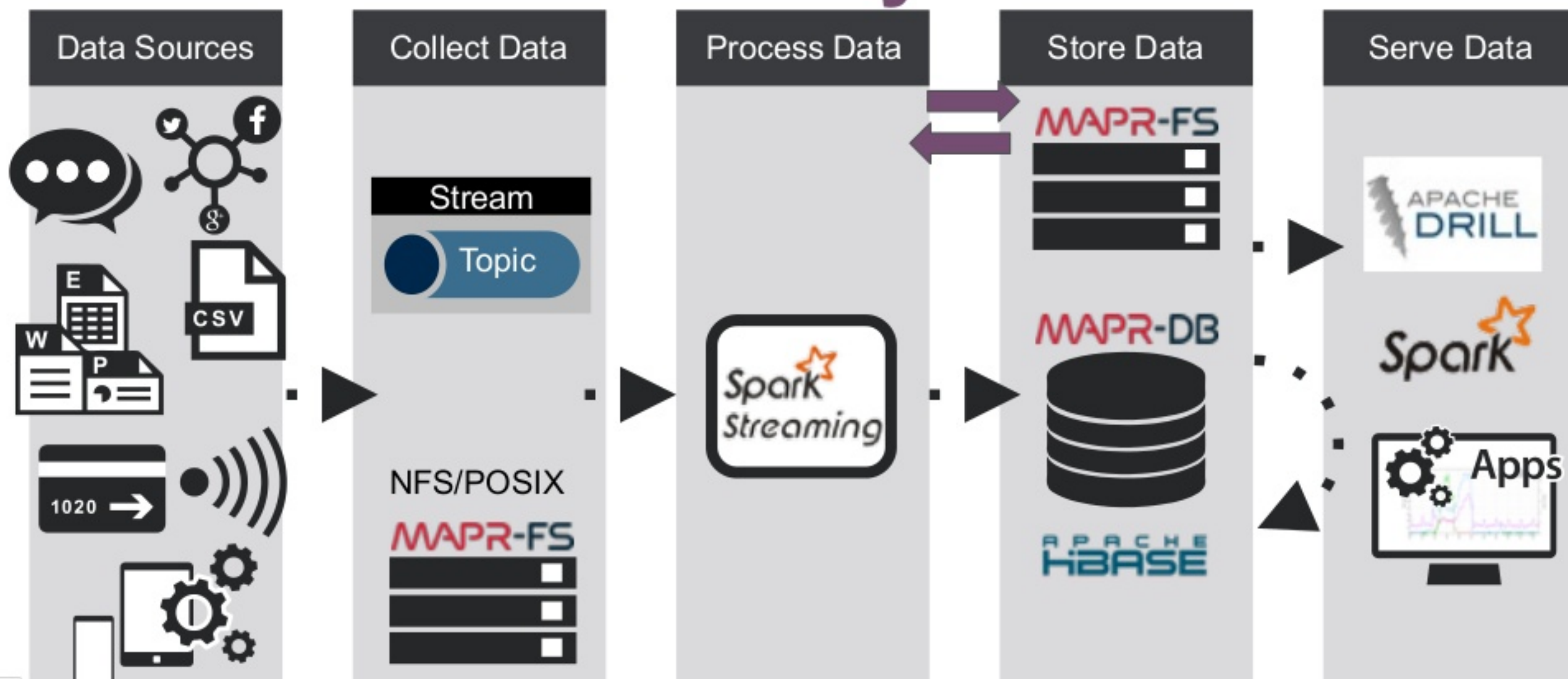
Trinity of Real Time

Continuous Streaming ETL & Computed Analytics

Application

DB

MAPRSTREAMS

- Topic
- Topic
- Topic
- Topic

- 60 events/sec
- 10 MB/event
- Tabled based topics

Spark Streaming

Delta Aggregates

Pre-compute analytics with Spark Streaming on Data-in-motion

Apache **Solr**

Search Application

MAPR-FS

Multi-Tier Data Archival

MAPR-DB

Pre-Computed

Level 1 Aggregates

Level 2 Aggregates

On-Demand

Level 3 Aggregates

Advanced ML Analytics

Q&A
Engage with us!

1. Read explanation of and Download code
   - https://www.mapr.com/blog/fast-scalable-streaming-applications-mapr-streams-spark-streaming-and-mapr-db
   - https://www.mapr.com/blog/spark-streaming-hbase

2. **Get Started:** MapR Converged Data Platform
   https://www.mapr.com/get-started-with-mapr

3. **Get Answers:** MapR Converge Community
   https://community.mapr.com/community/answers

4. **Get Trained:** MapR On-Demand Training
   https://learn.mapr.com

DEV 360 - Apache Spark Essentials

FREE

DEV 361 - Build and Monitor Apache Spark Applications

FREE

DEV 362 - Create Data Pipelines Using Apache Spark

FREE

THANK YOU.

Contact information or call to action goes here.

SPARK SUMMIT
EUROPE 2016