



Simplifying Big Data Analytics: Unifying Batch and Stream Processing

John Fanelli,
VP Product
In-Memory Compute Summit
June 30, 2015

FORRESTER®



Scale-up
Database



Streaming Analytics



Data And
Compute Grid



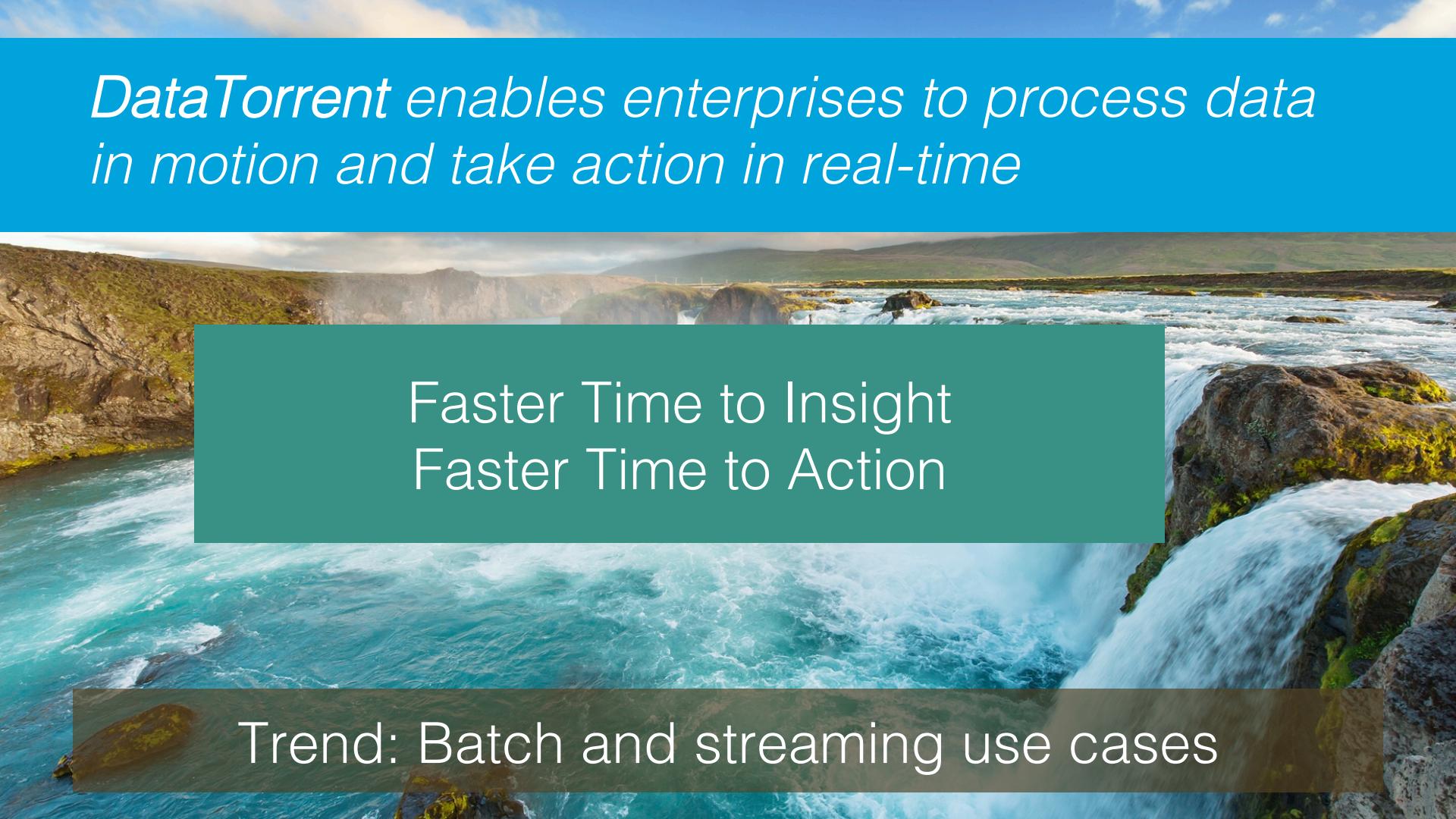
Clustered Database



General-purpose data
processing cluster



*DataTorrent enables enterprises to process data
in motion and take action in real-time*

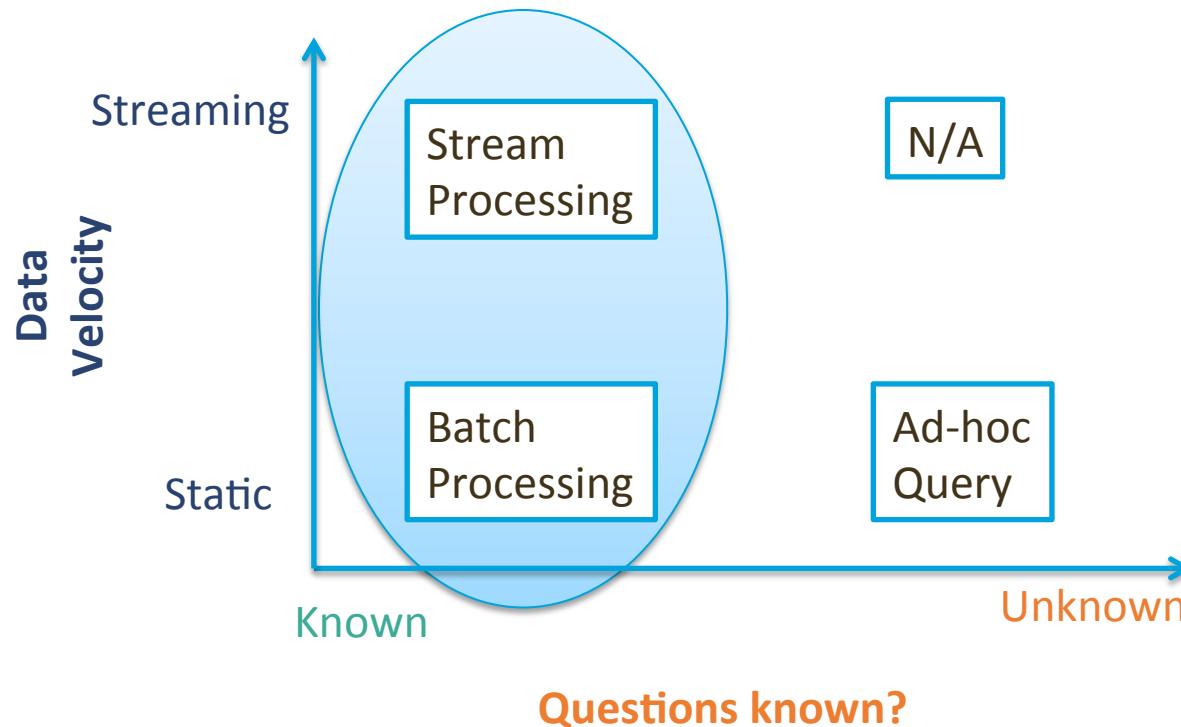
A wide-angle photograph of a waterfall cascading over mossy rocks into a turbulent blue river. A large teal rectangular overlay is positioned in the center of the image. Inside this overlay, the text "Faster Time to Insight" and "Faster Time to Action" is displayed in white, sans-serif font.

Faster Time to Insight
Faster Time to Action

A close-up photograph of water falling over a dark, textured rock. A dark brown rectangular overlay is positioned at the bottom of the image. Inside this overlay, the text "Trend: Batch and streaming use cases" is displayed in white, sans-serif font.

Trend: Batch and streaming use cases

Data Processing Categories in Big Data Use Cases



Data Sources

Transactional Data

Web Click Stream

Mobile Devices

Operational Log Files

Public Data

Sensor Data



Customer Uses

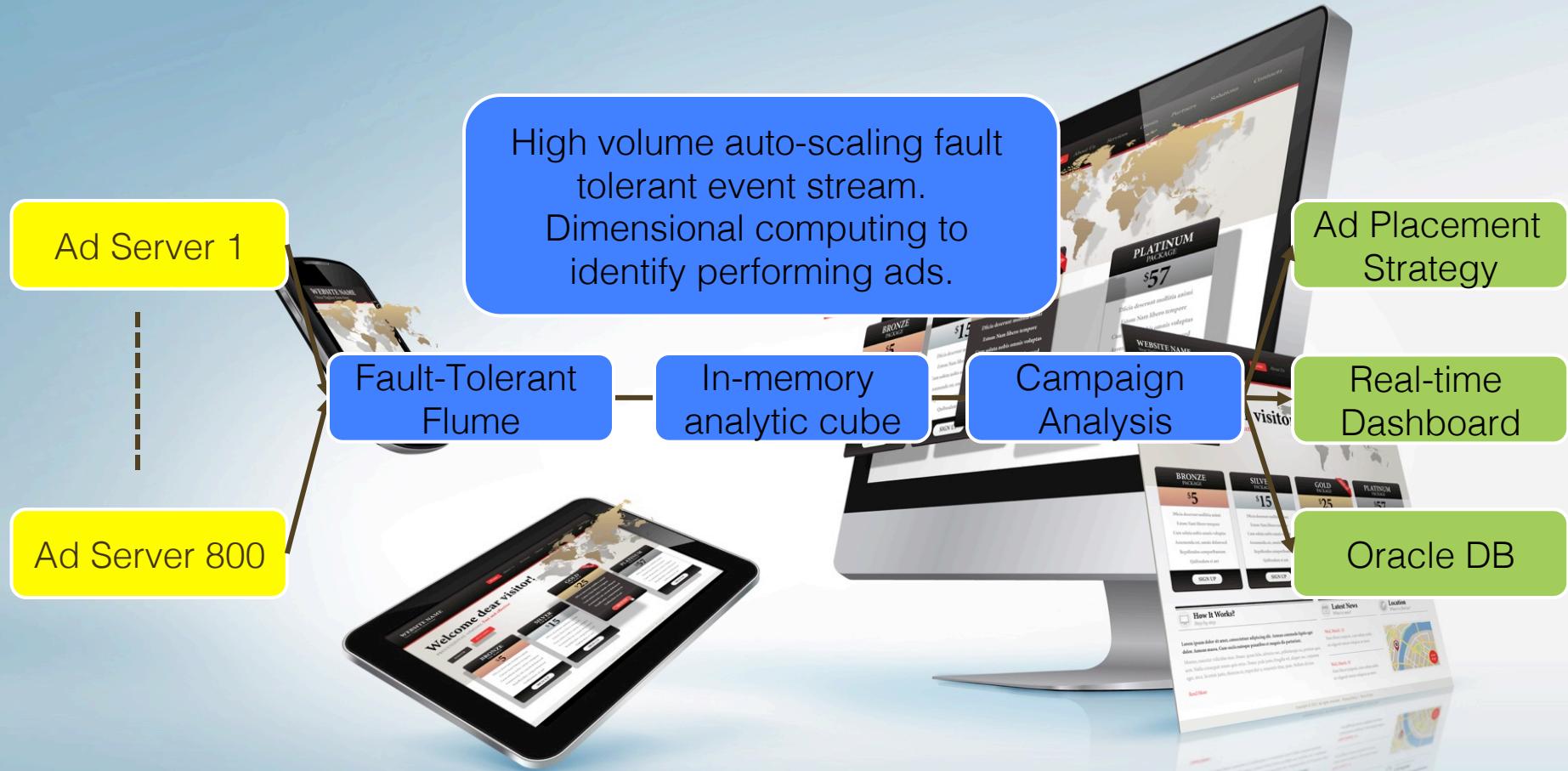


Real-Time Advertising
Customer Service
Operational
Fraud Detection
Predictive Maintenance

Processing Data In Motion



Online advertising dynamic inventory purchases



SmartGrid Connected Home



Batch Data

Customer Information

Historical Sales

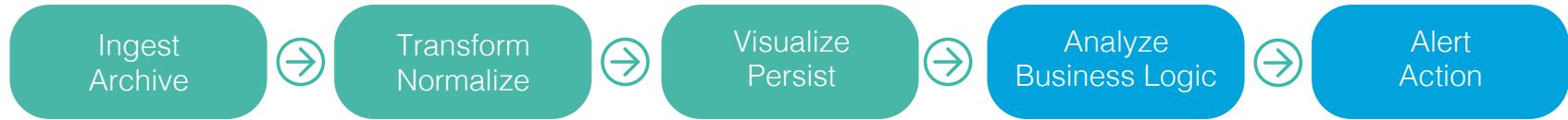
Support Data

Product Configuration

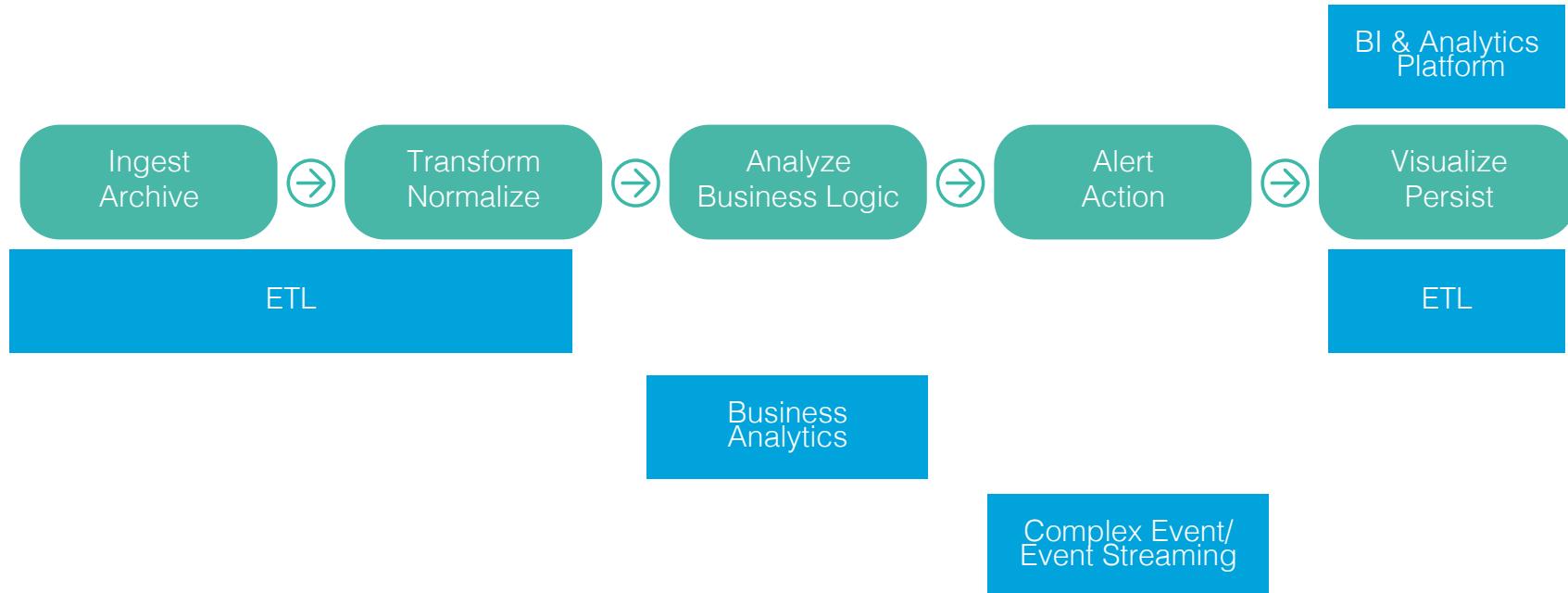
Corporate Info



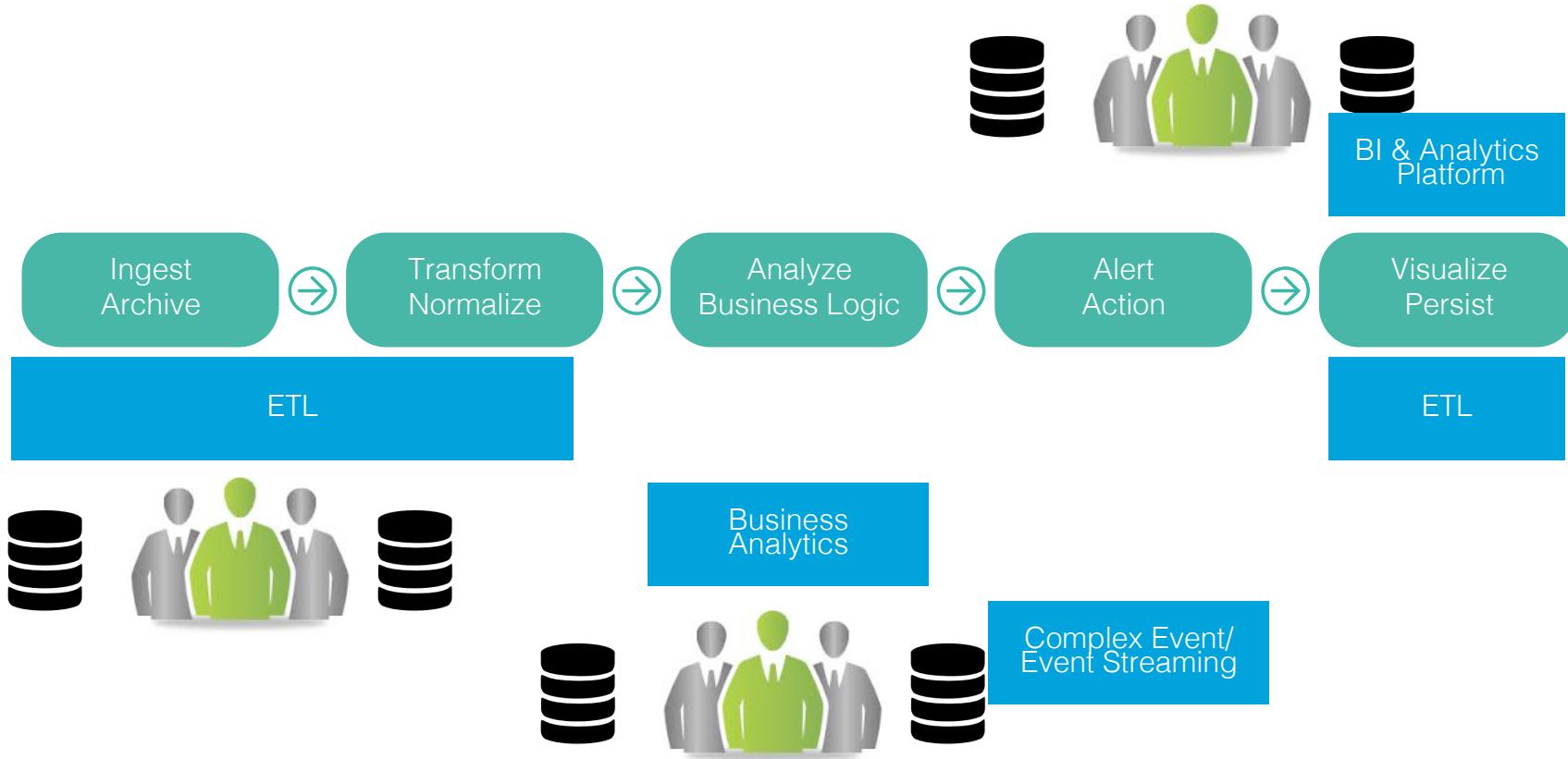
Batch Processing Data



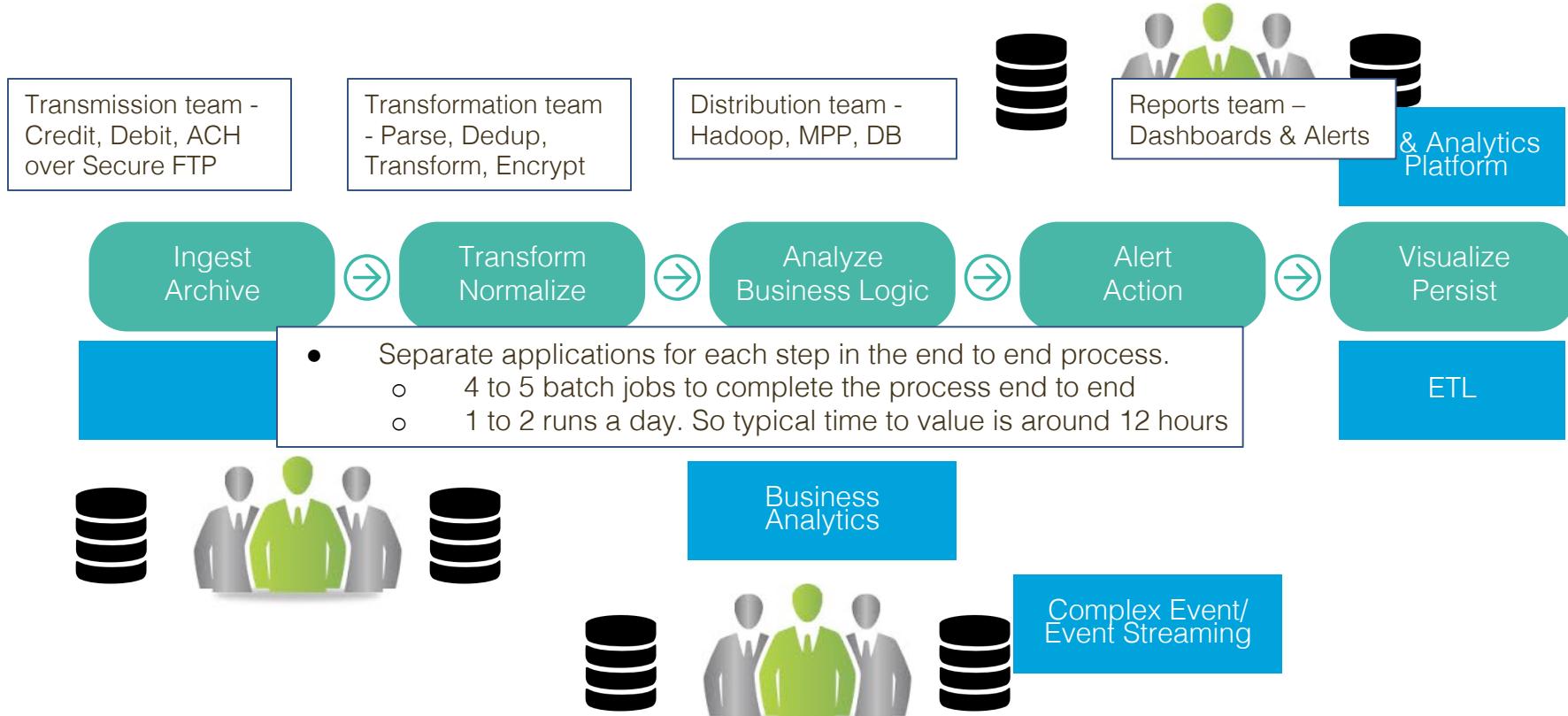
The Enterprise Data Processing Problem



The Enterprise Data Processing Problem



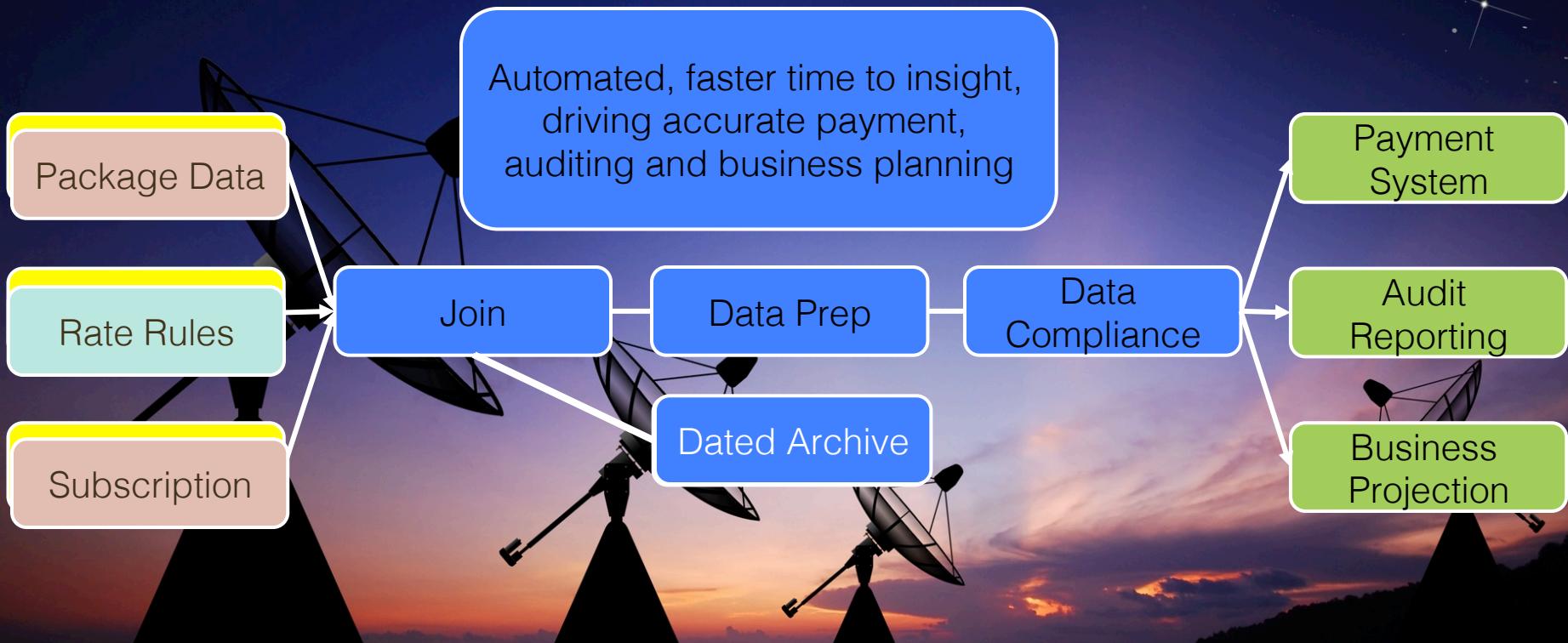
The Enterprise Data Processing Problem



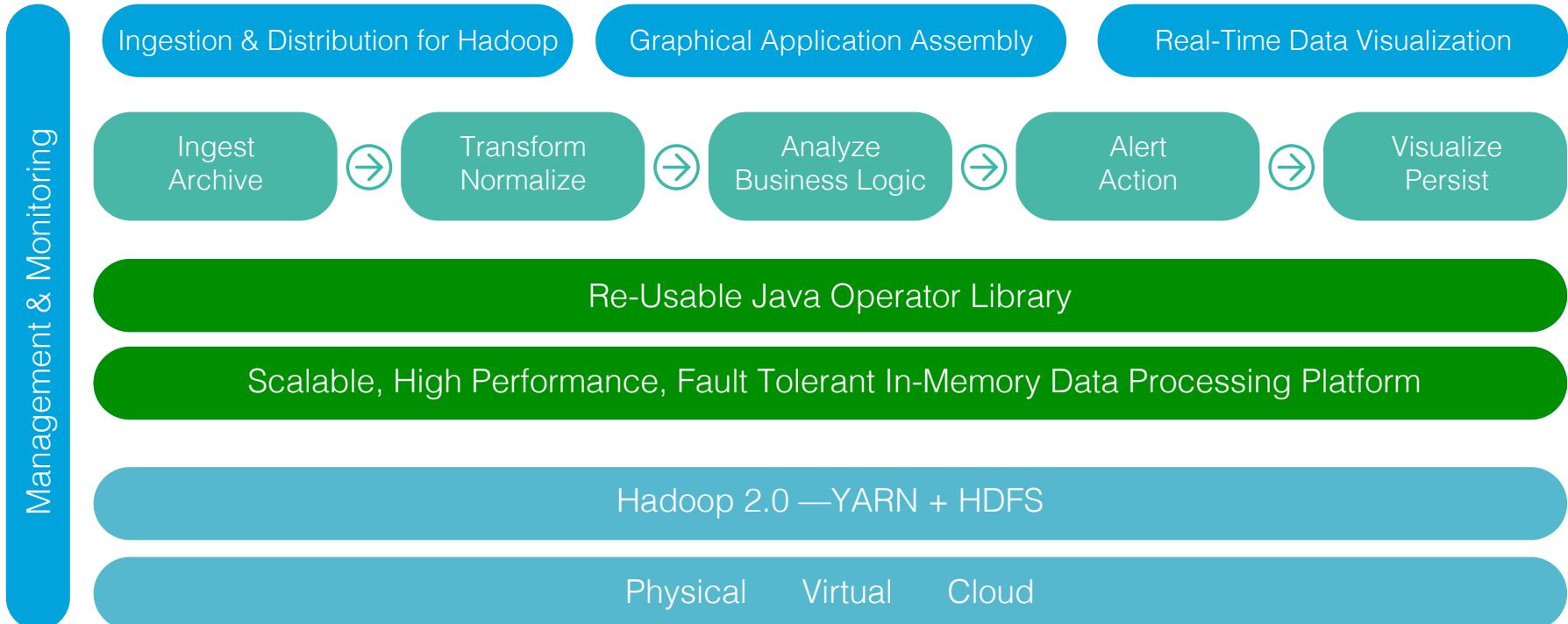
Financial services big data fabric



Satellite Television Provider



DataTorrent RTS Architectural Overview



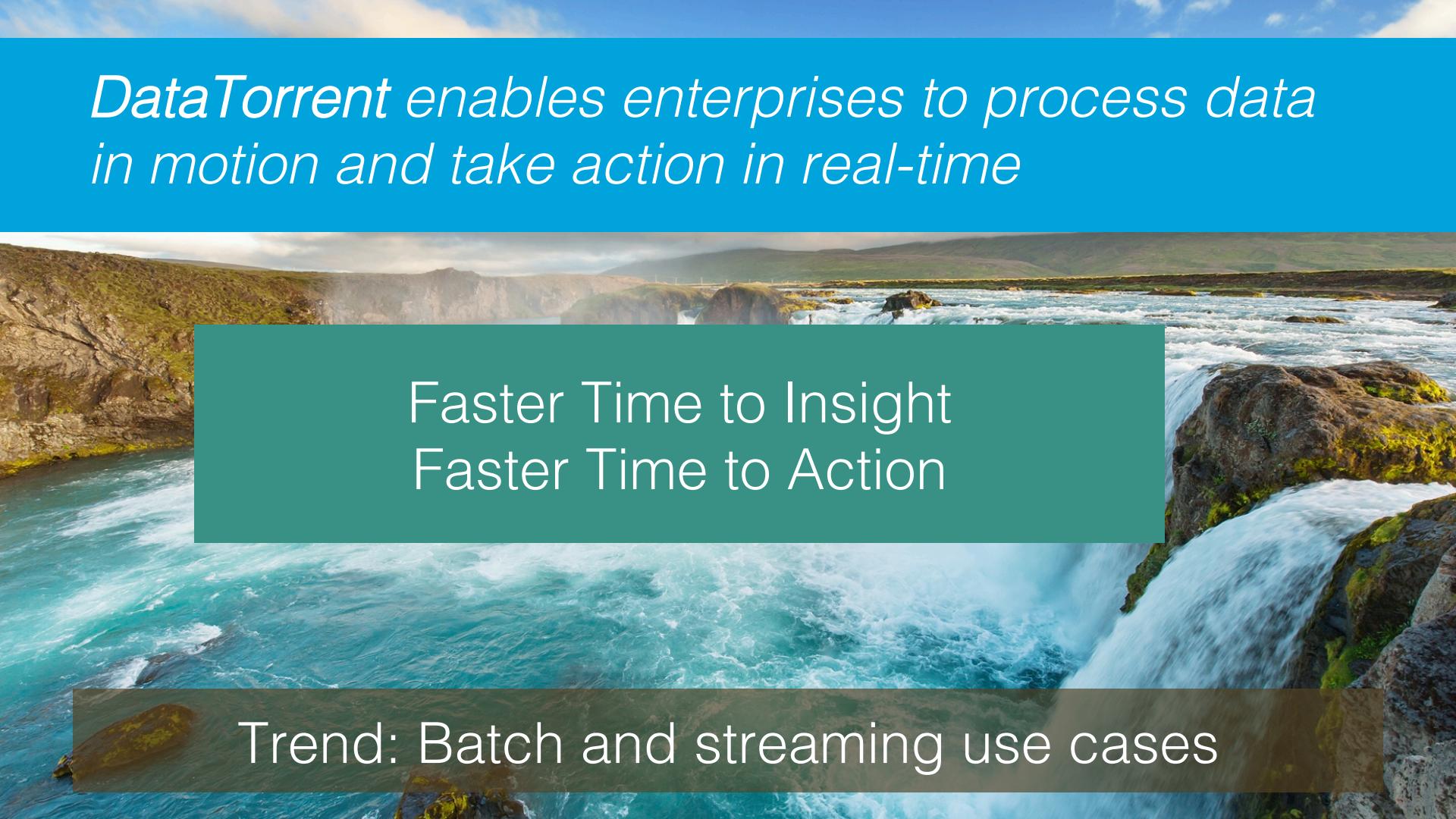
Available July 31

DataTorrent - Project Apex

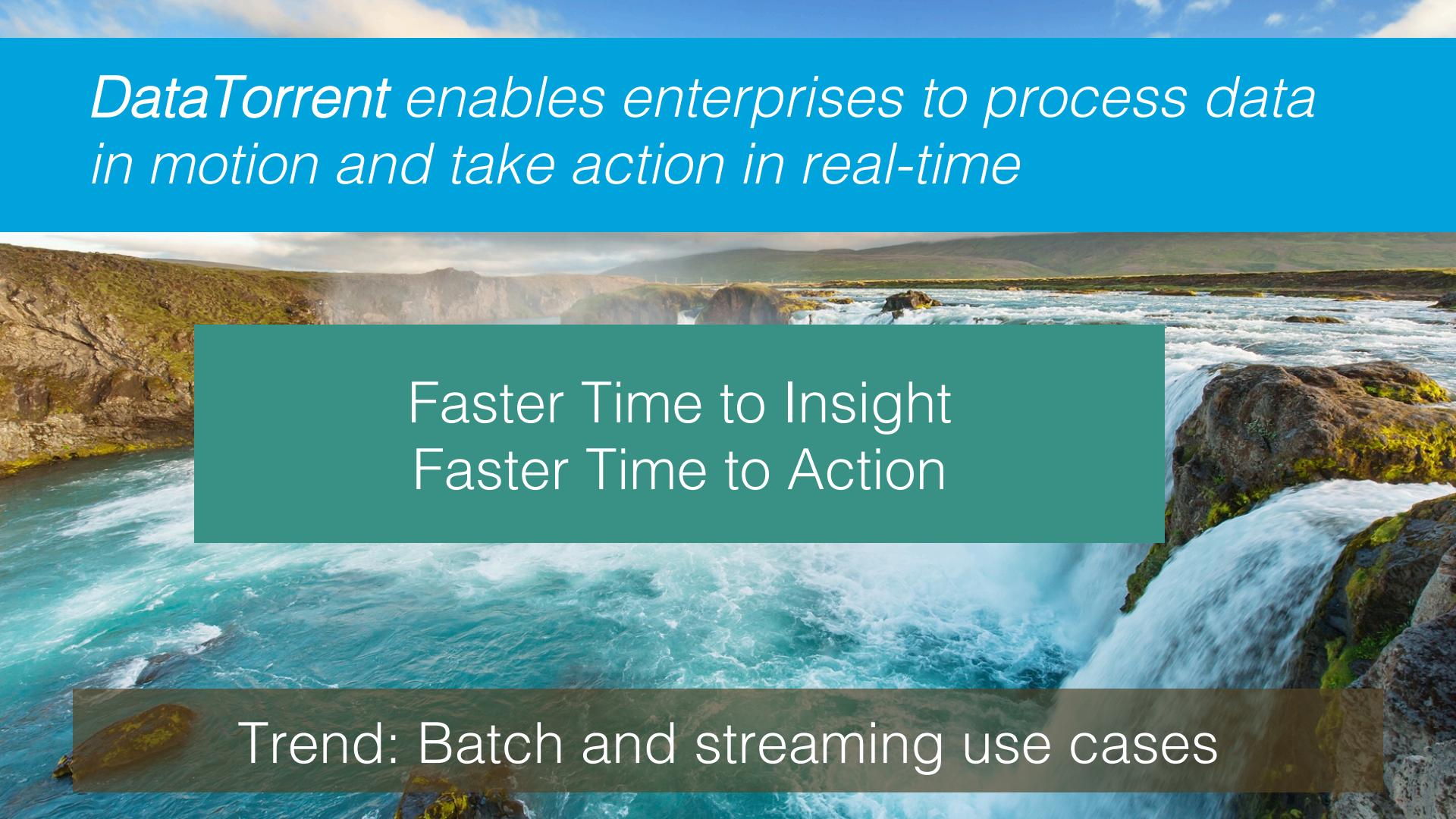
- Industry's first *open source* enterprise-class unified stream and batch processing platform
- DataTorrent RTS 3 Core engine
- Key features requested by open source developer community
 - Event processing guarantees
 - In-memory performance & scalability
 - Fault tolerance and state management
 - Native rolling and static window support
 - Hadoop-native YARN & HDFS implementation
- Apache 2.0 License
- Complemented by open source Malhar operator library

<https://www.datatorrent.com/product/project-apex/>

*DataTorrent enables enterprises to process data
in motion and take action in real-time*

A wide-angle photograph of a waterfall cascading over mossy rocks into a turbulent blue river. A large teal rectangular overlay is positioned in the center of the image. Inside this overlay, the text "Faster Time to Insight" and "Faster Time to Action" is displayed in white, sans-serif font.

Faster Time to Insight
Faster Time to Action

A close-up photograph of water falling over a dark, textured rock. A dark brown rectangular overlay is positioned at the bottom of the image. Inside this overlay, the text "Trend: Batch and streaming use cases" is displayed in white, sans-serif font.

Trend: Batch and streaming use cases

