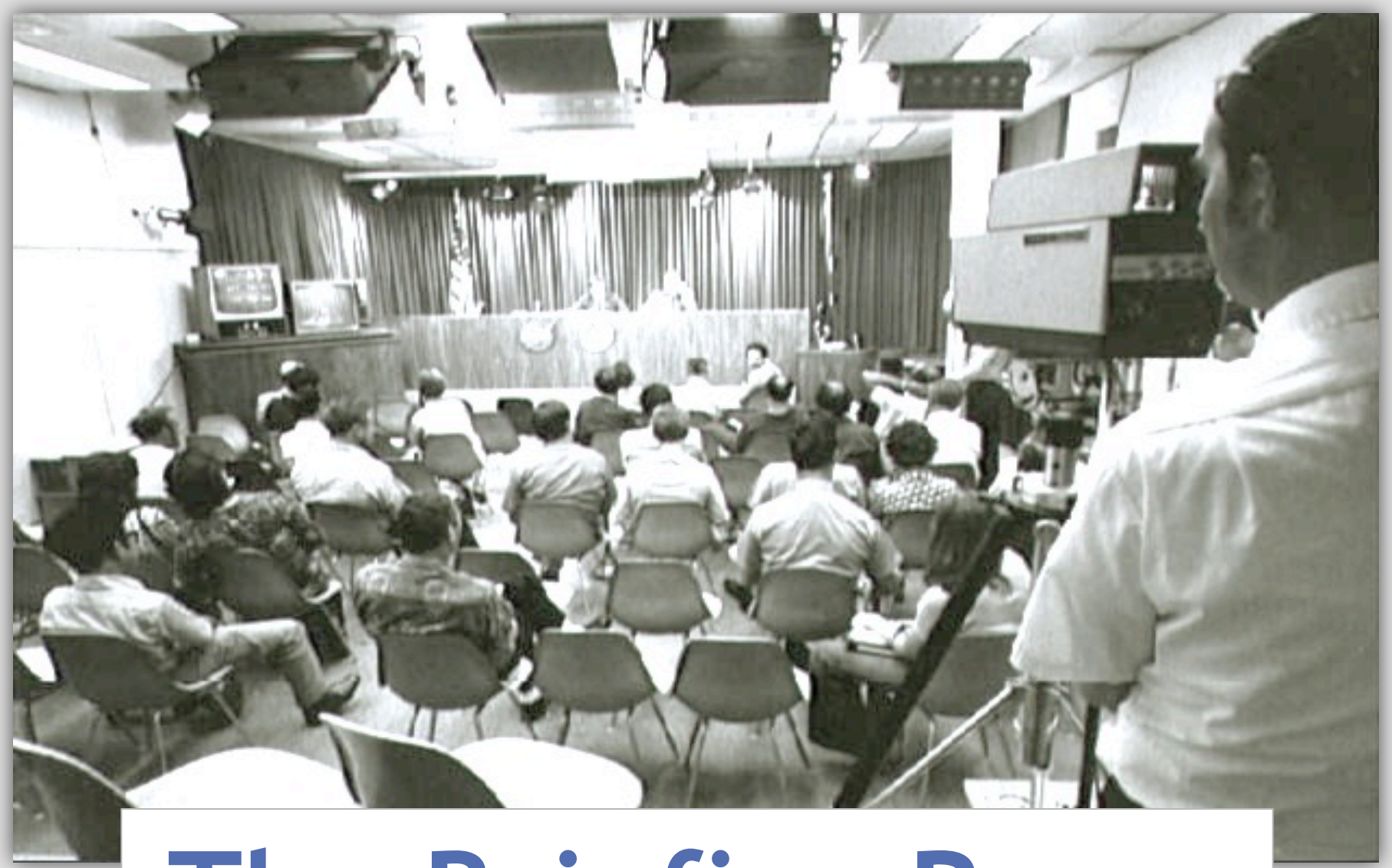
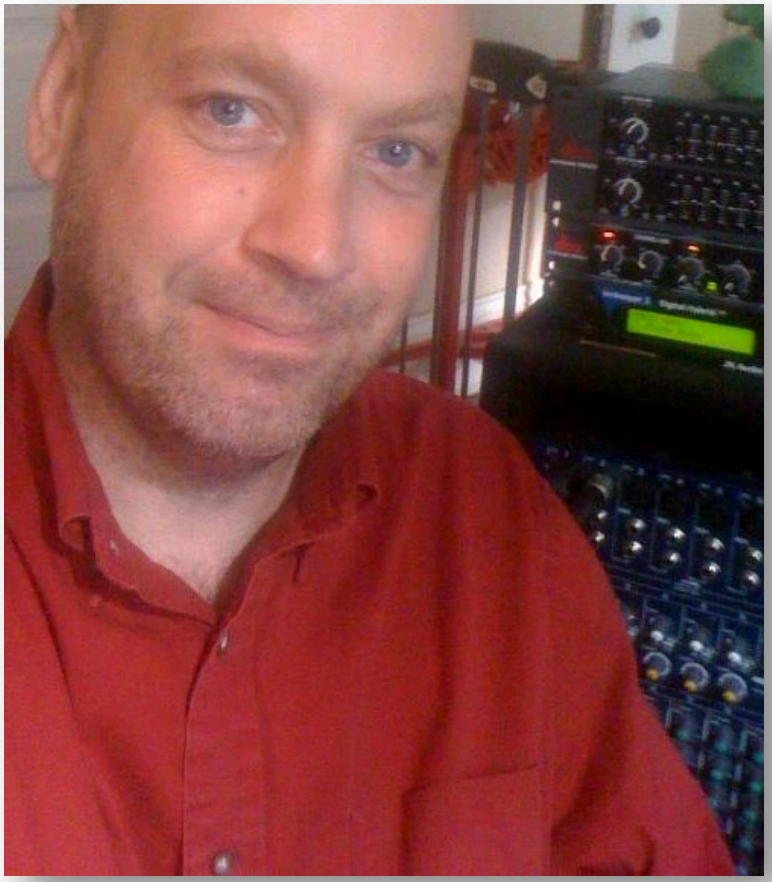
Grab some coffee and enjoy the pre-show banter before the top of the hour!

DISCUSSING THE WAR IN A PARIS CAFE.
SEE PAGE 304.



The Briefing Room

Welcome



Host:
Eric Kavanagh

eric.kavanagh@bloorgroup.com
@eric_kavanagh

Mission

- Reveal the essential characteristics of enterprise software, good and bad
- Provide a forum for detailed analysis of today's innovative technologies
- Give vendors a chance to explain their product to savvy analysts
- Allow audience members to pose serious questions... and get answers!

Topics

October: **DATA MANAGEMENT**

November: **ANALYTICS**

December: **INNOVATORS**

Race Cars Need Professional Race Tracks

- Structural Integrity Matters
- Archaic Architecture Causes Problems
- Once in Place, Enjoy the Show!

Analyst: Mark Madsen

Mark Madsen is president of Third Nature, a technology research and consulting firm focused on business intelligence, data integration and data management. Mark is an award-winning author, architect and CTO whose work has been featured in numerous industry publications. Over the past ten years Mark received awards for his work from the American Productivity & Quality Center, TDWI, and the Smithsonian Institute. He is an international speaker, a contributor to Forbes Online and on the O'Reilly Strata program committee. For more information or to contact Mark, follow @markmadsen on Twitter or visit http://ThirdNature.net

Striim

- Striim, formerly WebAction, is a streaming and intelligence platform specializing in data integration across multiple internal and external sources
- The platform leverages continuous in-memory processing to deliver insights within seconds
- Striim enables data correlation, anomaly detection, alert and workflow triggers and detailed visualizations

Guest: Steve Wilkes

Steve Wilkes, Founder and CTO of Striim, is a life-long technologist, architect, and hands-on development executive. Prior to founding WebAction, Steve was the senior director of the Advanced Technology Group at GoldenGate Software. Here he focused on data integration and continued this role following the acquisition by Oracle, where he also took the lead for Oracle's cloud data integration strategy. His earlier career included Senior Enterprise Architect at The Middleware Company, principal technologist at AltoWeb and a number of product development and consulting roles including Cap Gemini's Advanced Technology Group. Steve has handled every role in the software lifecycle and most roles in a technology company at some point during his career. He still codes in multiple languages, often at the same time.

striim

striim

brought to you by WebAction

Is ETL Now a 4-Letter Word?
Preparing for Streaming Analytics

Steve Wilkes - CTO / Founder WebAction

October 2015

Striim Executive Summary

The Striim Team is uniquely qualified to solve the streaming analytics problem.

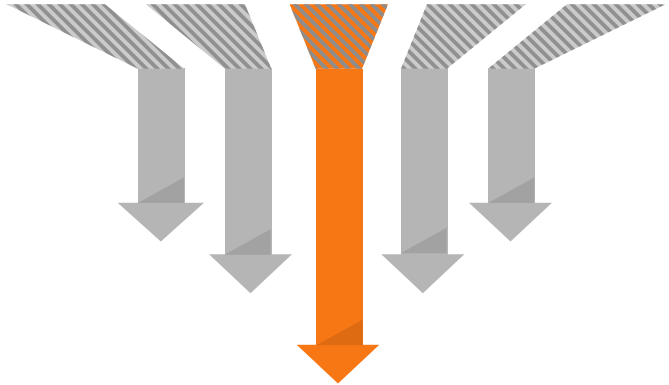
Founded	Founded in 2012 by core team out of GoldenGate Software and WebLogic
Leading investors	Backed by leading investors: Summit Partners and Intel Capital
Technology	Striim Platform - Current release - v. 3.1.4
Customers	Deployments in financial services, telco, retail, gaming, IoT

striim

What is Striim?

Striim provides

Streaming Integration and Intelligence

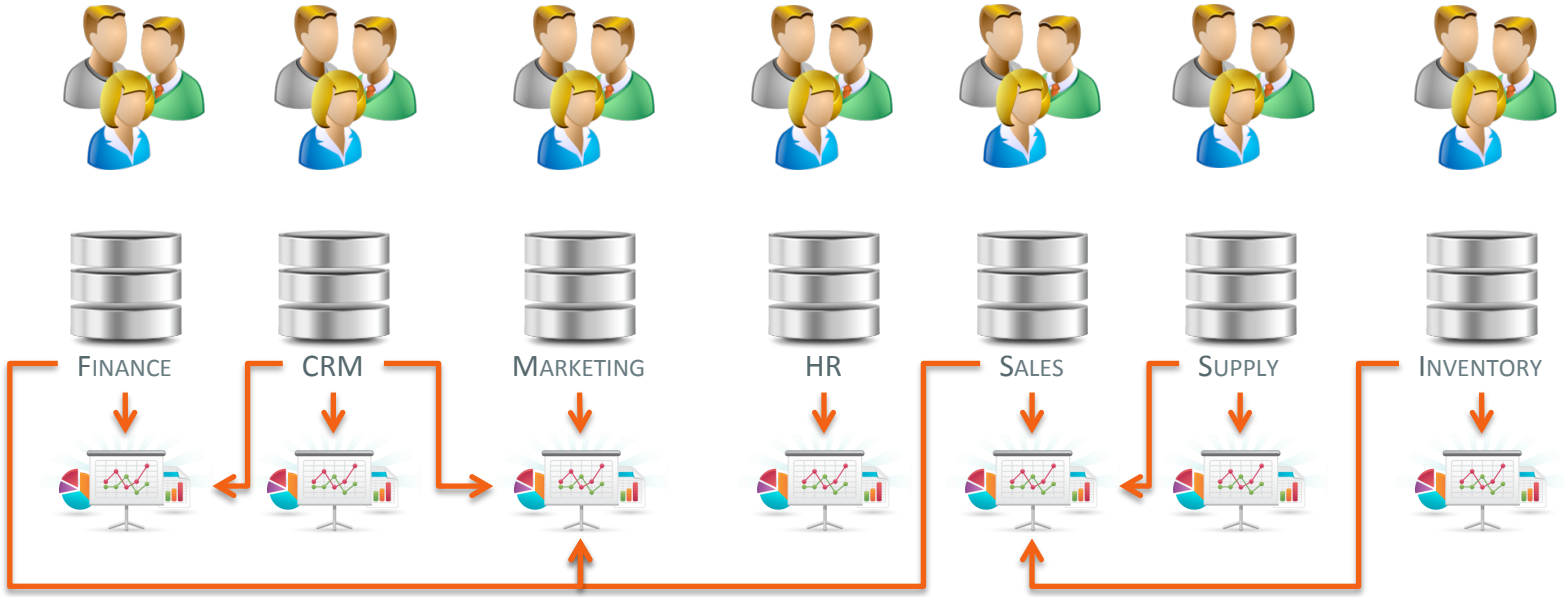


enabling companies to

Make data useful the instant it's born.

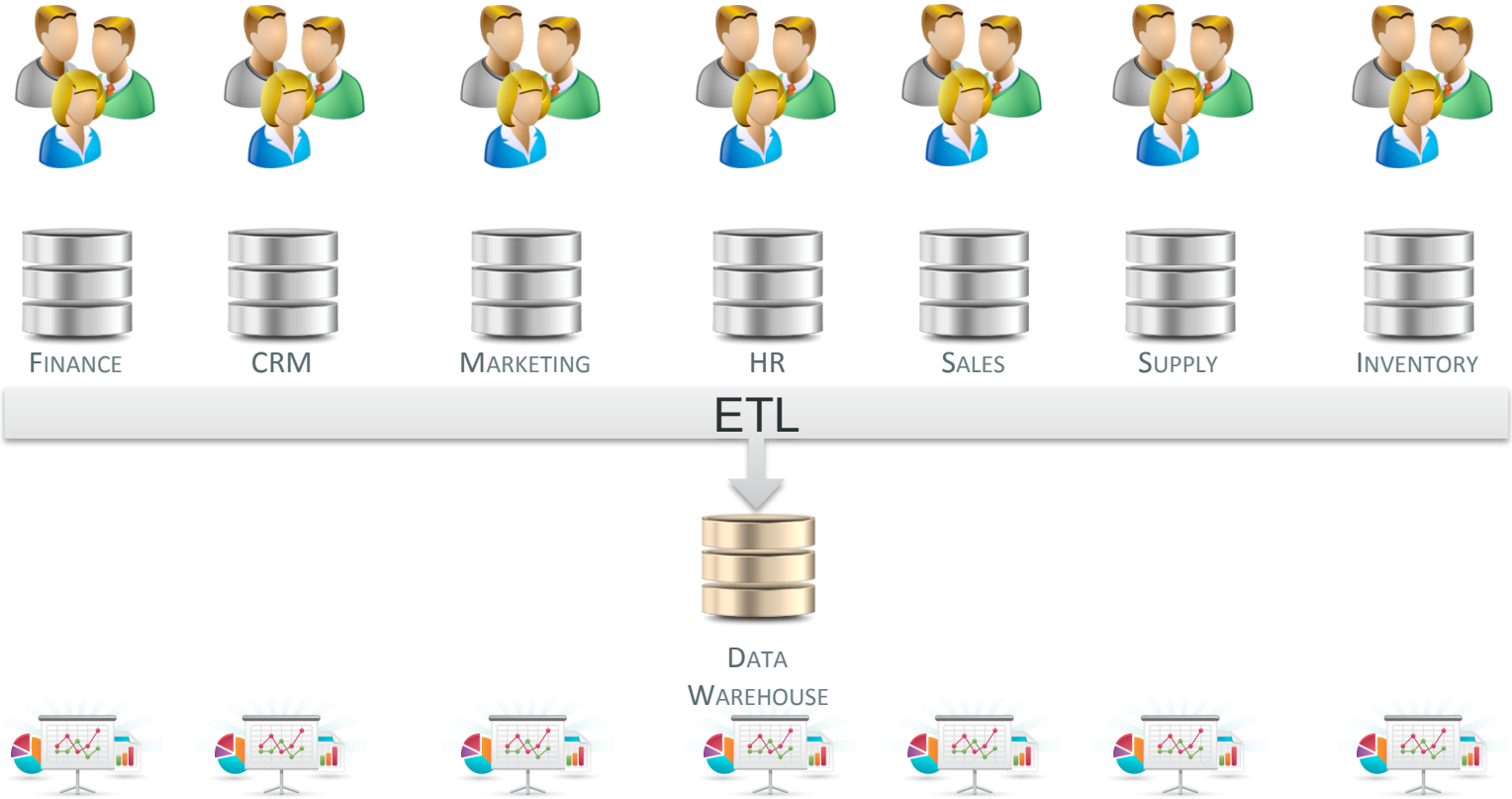
striim

A Brief History of ETL



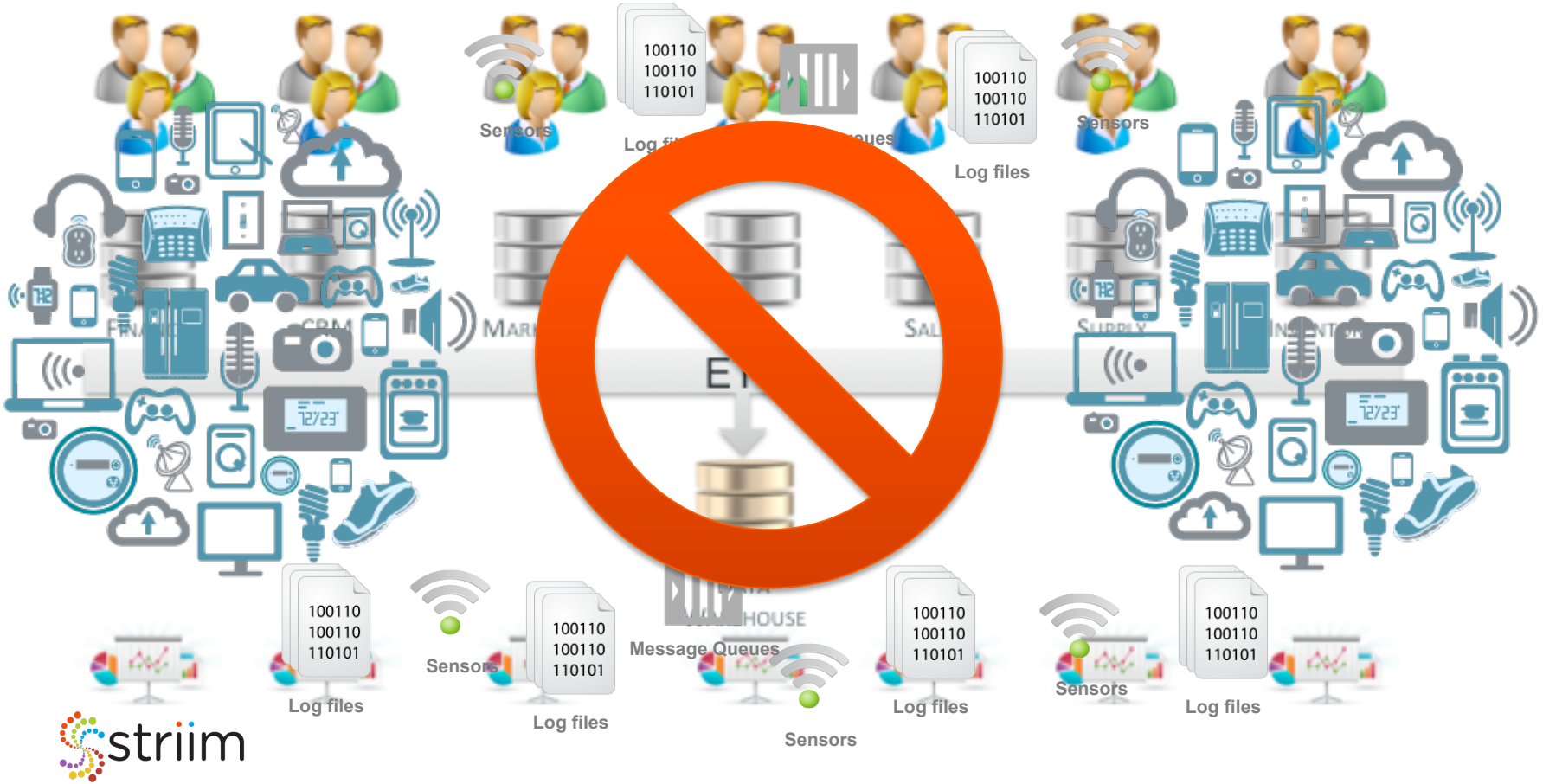
striim

A Brief History of ETL



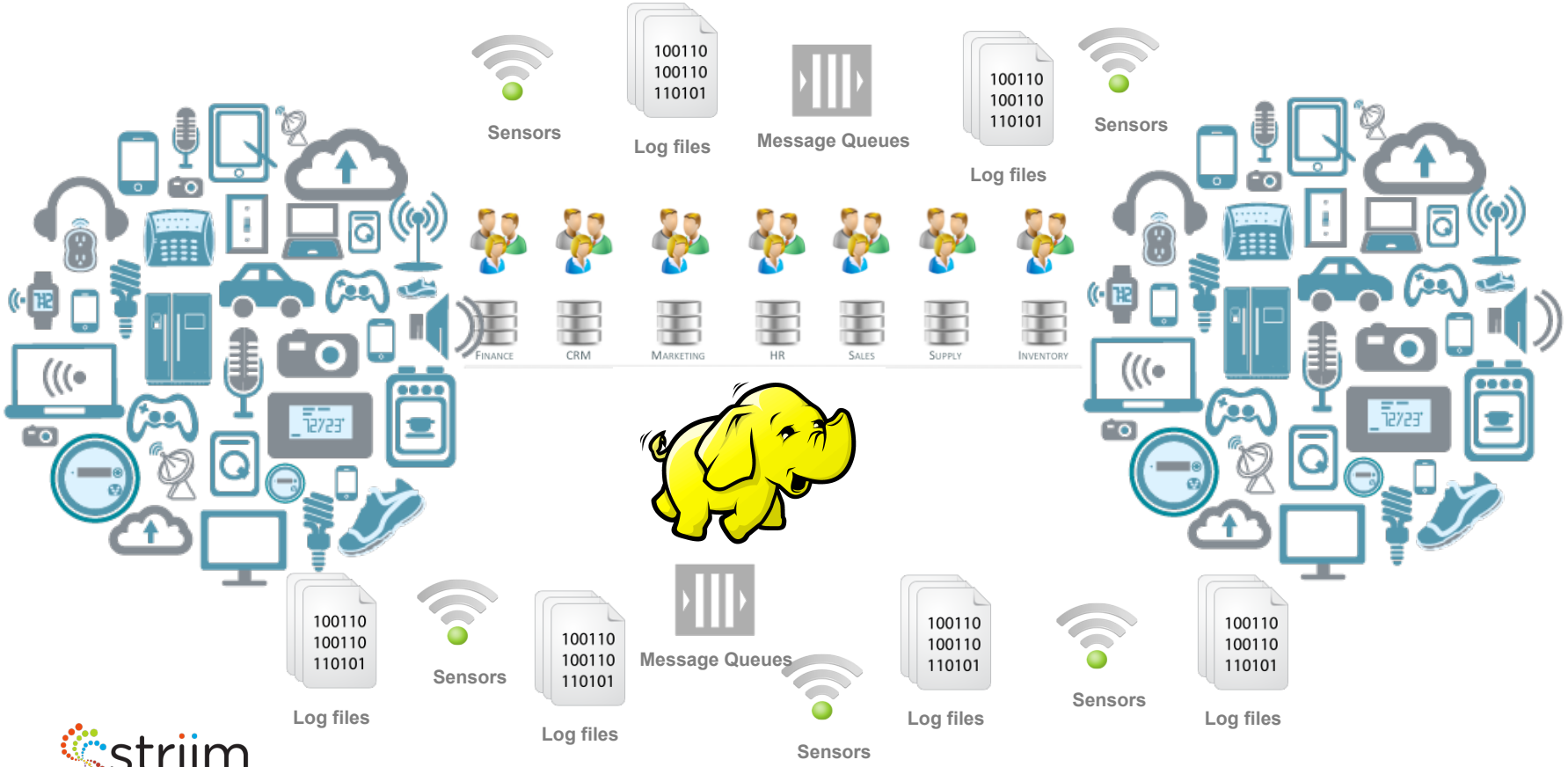
striim

ELTF?


Sensors
100110
100110
110101
Log files
Sensors
Log files
Message Queues
Sensors
Log files
Sensors
Log files
Sensors
Log files

striim

Hadoop to the Rescue?



striim

Not Quite!

- Exponential increase in volume/velocity/variety of data
- Increasing business need to address issues while you can affect the outcome
- Customer / Employee expectations for instant response
- Know what the customer knows as soon or before

- Big Data and ETL are designed for batch processing
- Enterprises aren’t getting value out of Big Data investments
  - data dumped into data lakes with no organization/filtering
  - by the time it’s processed/analyzed, it’s too late
  - can't integrate database change into Big Data

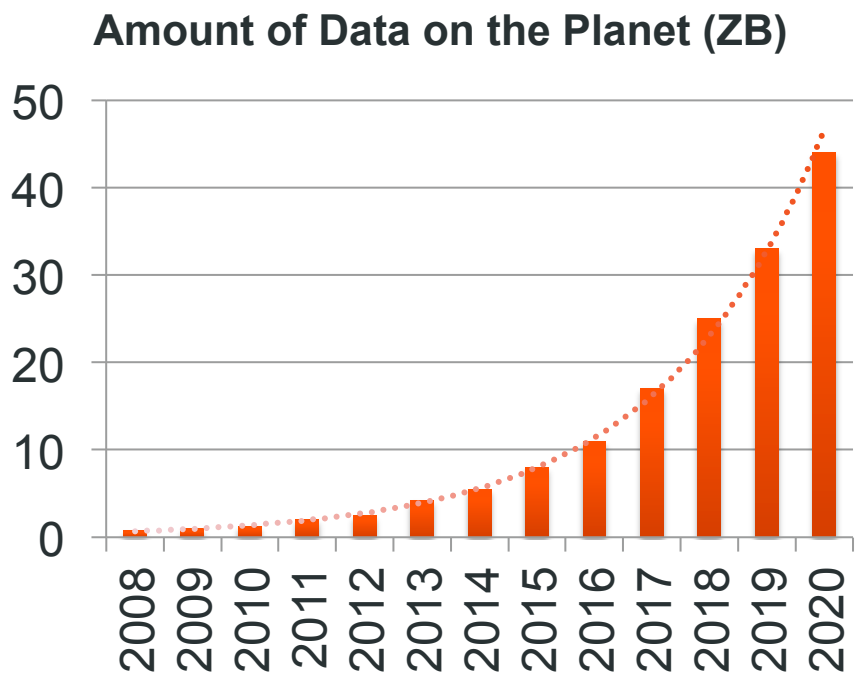
striim

Data Growth Challenges

- Data Growth is Exponential
- More data created in the next 2 years than ever created before

- Storage cannot keep up
  - In 2014 we could store 1/3 of the data
  - By 2020 it will be 1/6

- Only half of the data that should be secured is secured



* According to IDC's 2014 Digital Universe Study

striim

The Data Lake Fallacy

Gartner 2014 (Casonato, Heudecker, Beyer, Adrian)

“Through 2018, 90% of deployed data lakes will be useless as they are overwhelmed with information assets captured for uncertain use cases. ”

Bill Schmarzo 2015 - CTO of EMC Global Services

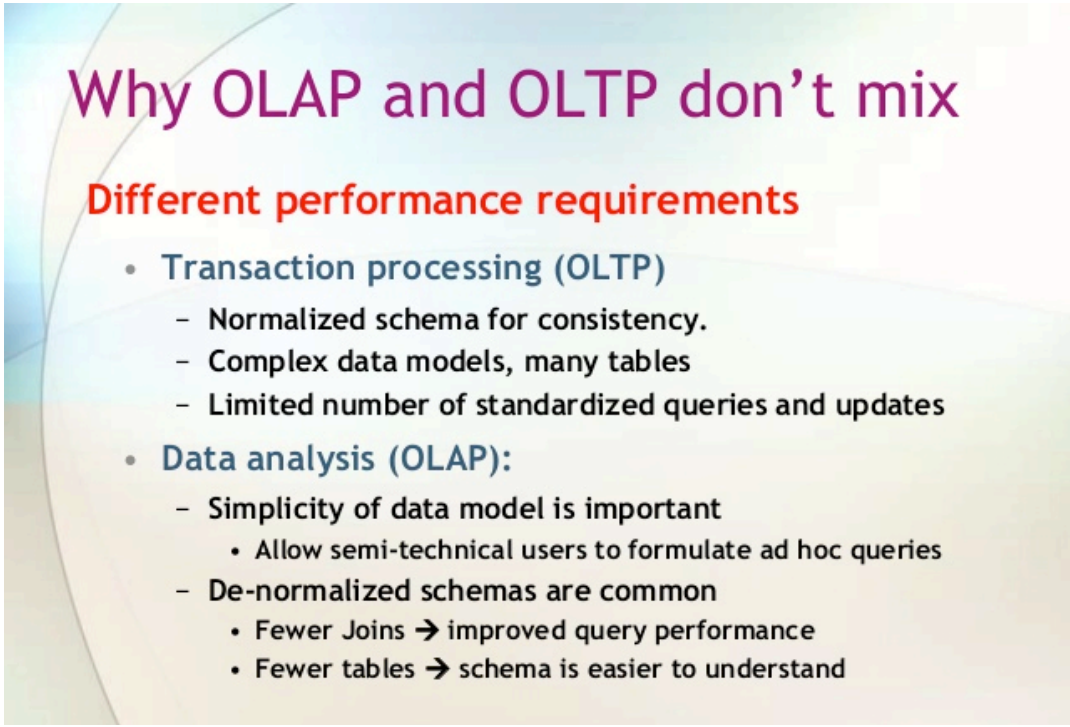
“The problem is that most customers are still tackling this whole big data conversation. You start with your technology, bring in some Hadoop, throw some data in there and you kind of hope magic stuff happens. It’s really a process fraught with all kind of misdirection.”

striim

Avoid Repeating History

You Don't Query OLTP

- You need to:
  - simplify
  - filter
  - denormalize
- Data dumped into Hadoop is not designed for querying



* OBIEE 11g Training Online Class

striim

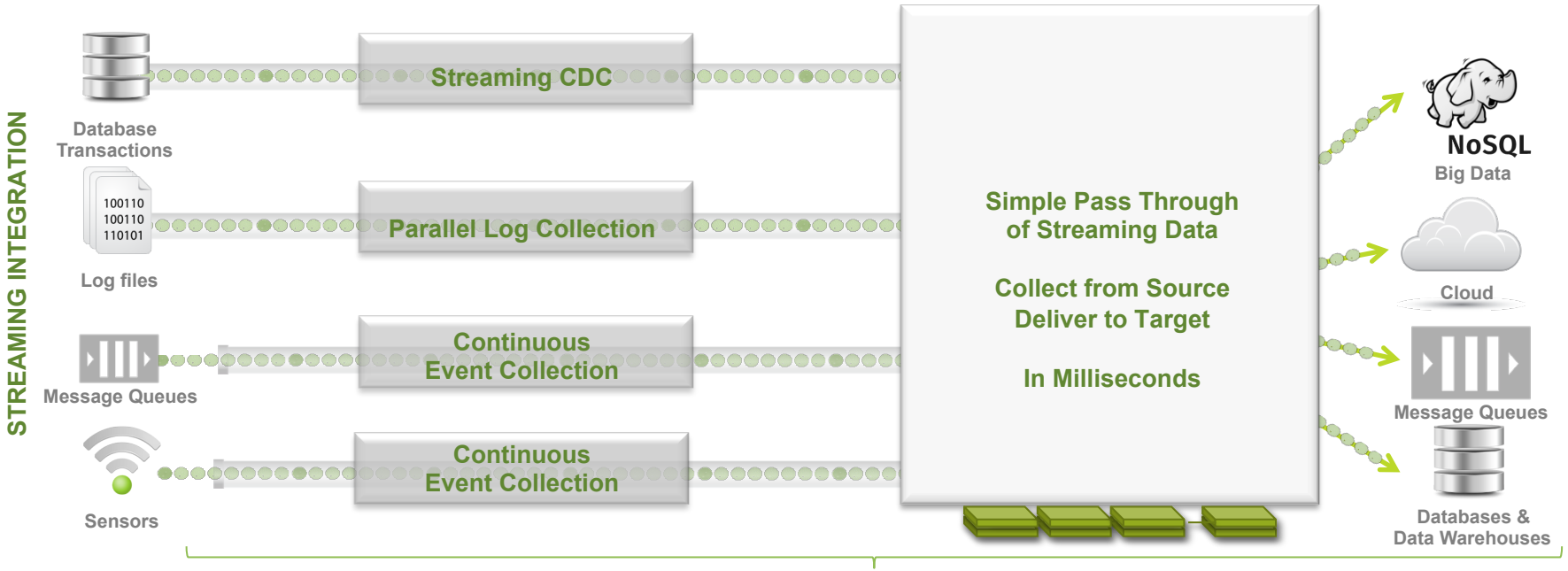
So How Do You Handle This?

- Access to data
  - CDC for databases
  - Parallel collection of log files
  - Edge processing for sensor data
- In-flight processing
  - Transformation / Filtering
  - Enrichment / Denormalization
  - Aggregation / Removal of redundant data
- Scale-out
  - Handle huge and increasing volume
  - Handle incremental processing requirements
- Security and Governance
  - Know what data you have
  - Being able to protect it granularly

= Streaming Integration

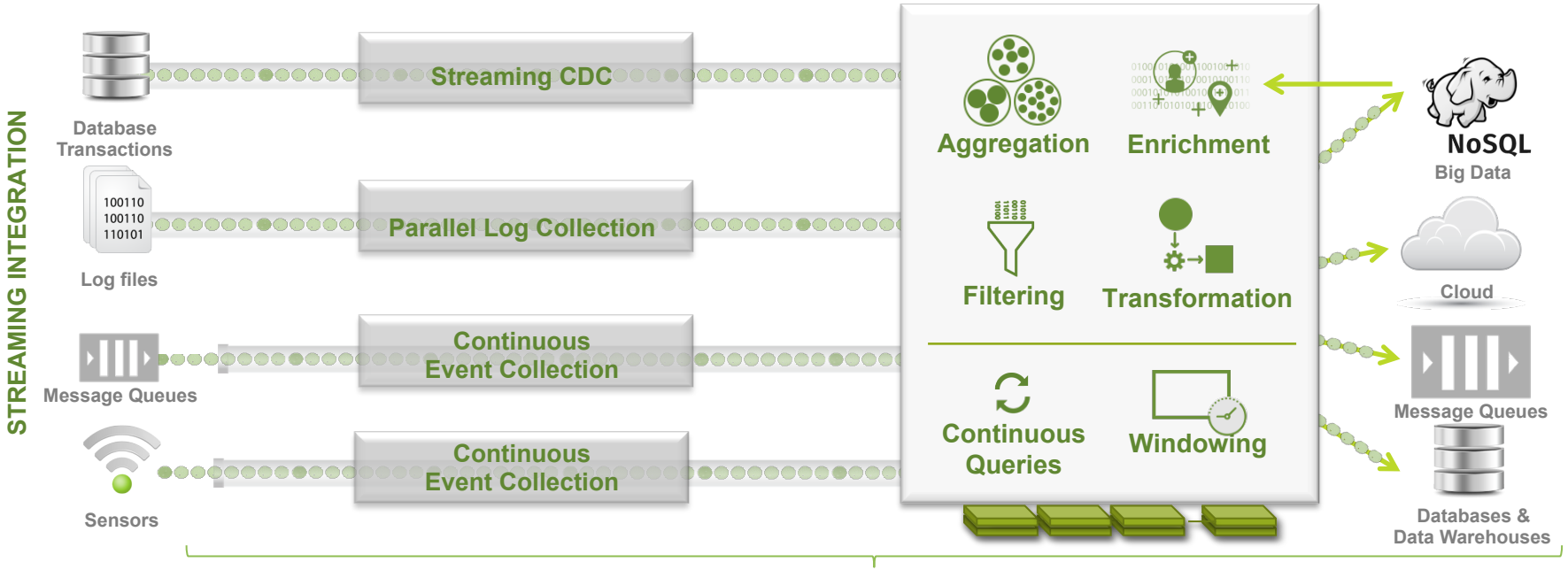
striim

Streaming Integration



striim

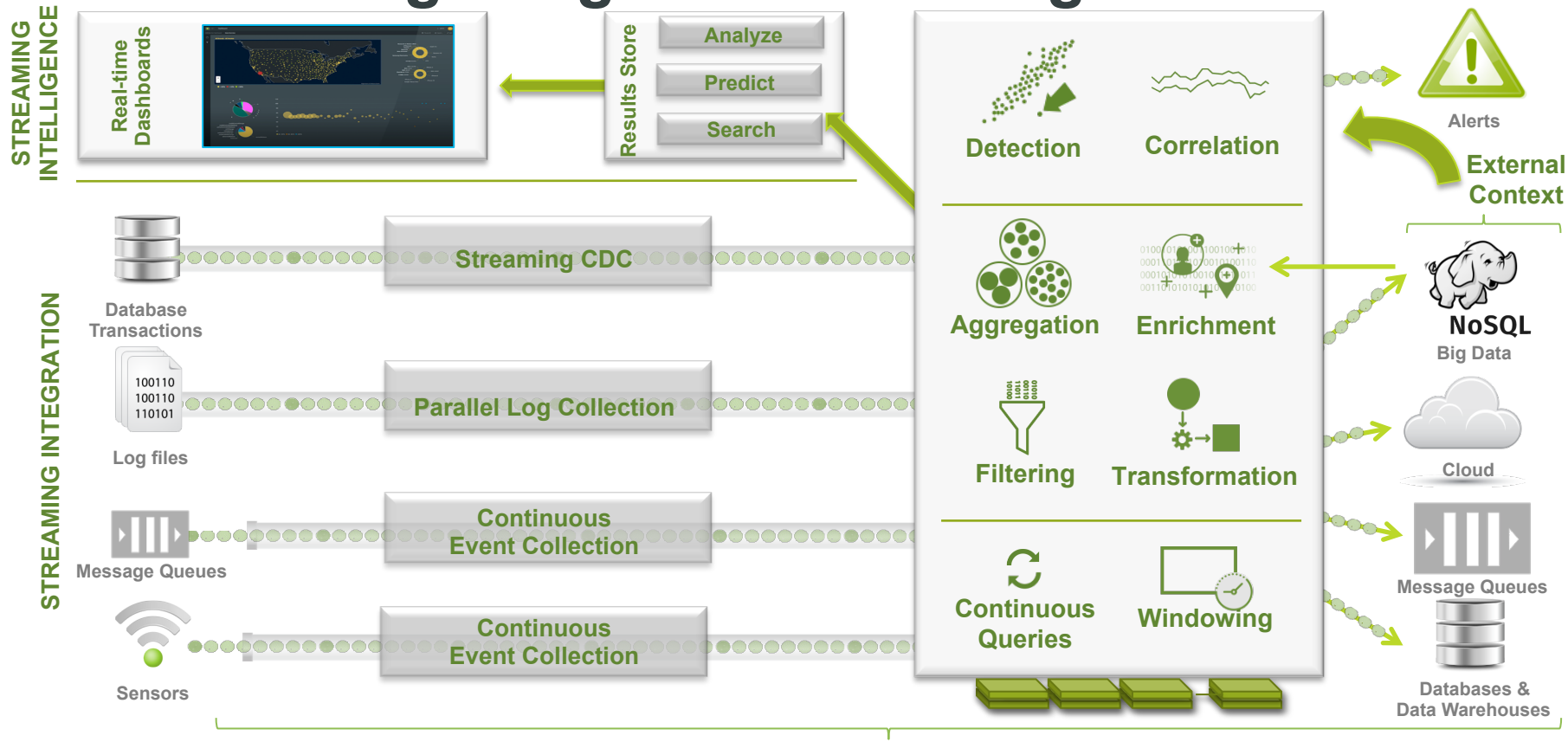
Streaming Integration



striim

EASY TO IMPLEMENT | COST EFFECTIVE | REALTIME CONTINUOUS PROCESSING

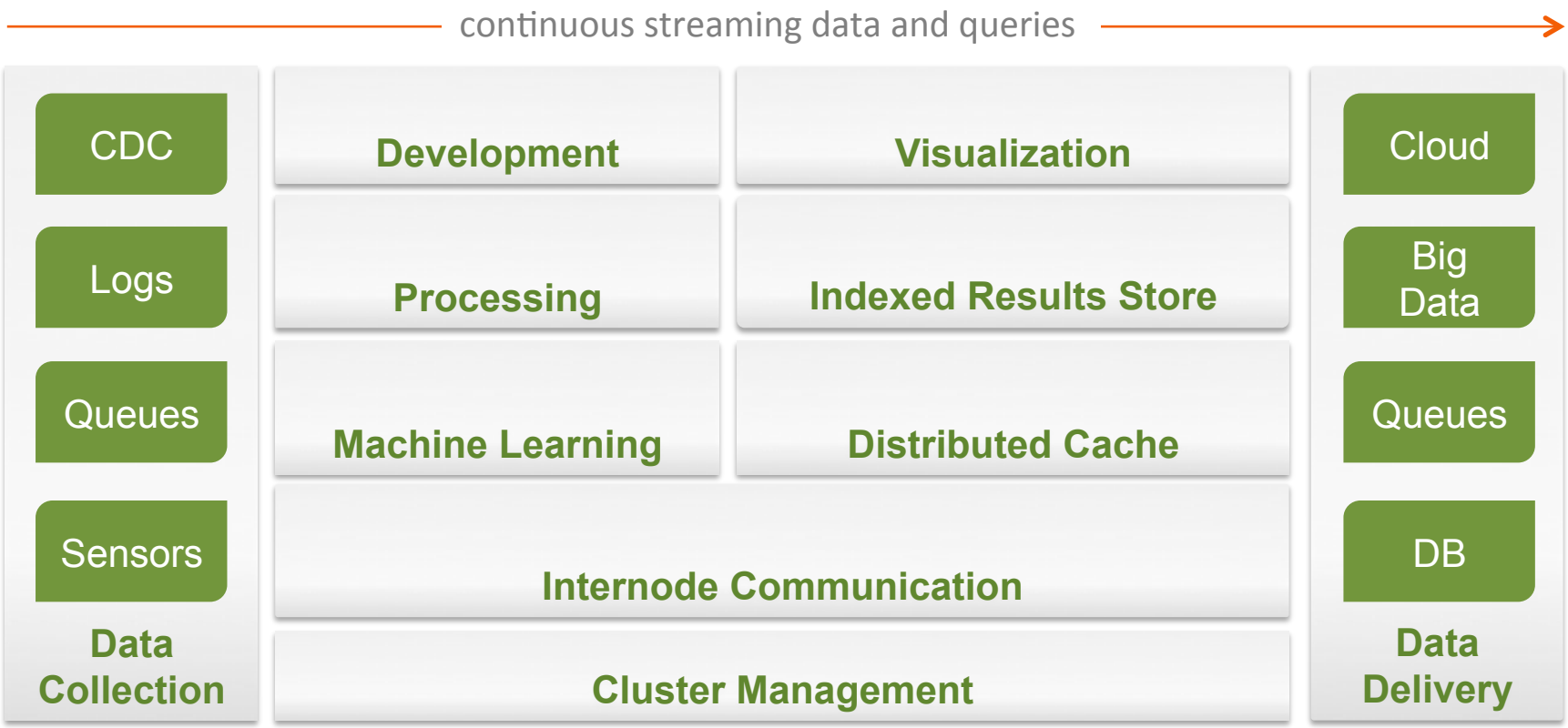
Streaming Integration & Intelligence



striim

EASY TO IMPLEMENT | COST EFFECTIVE | REALTIME CONTINUOUS PROCESSING

Striim Platform Components



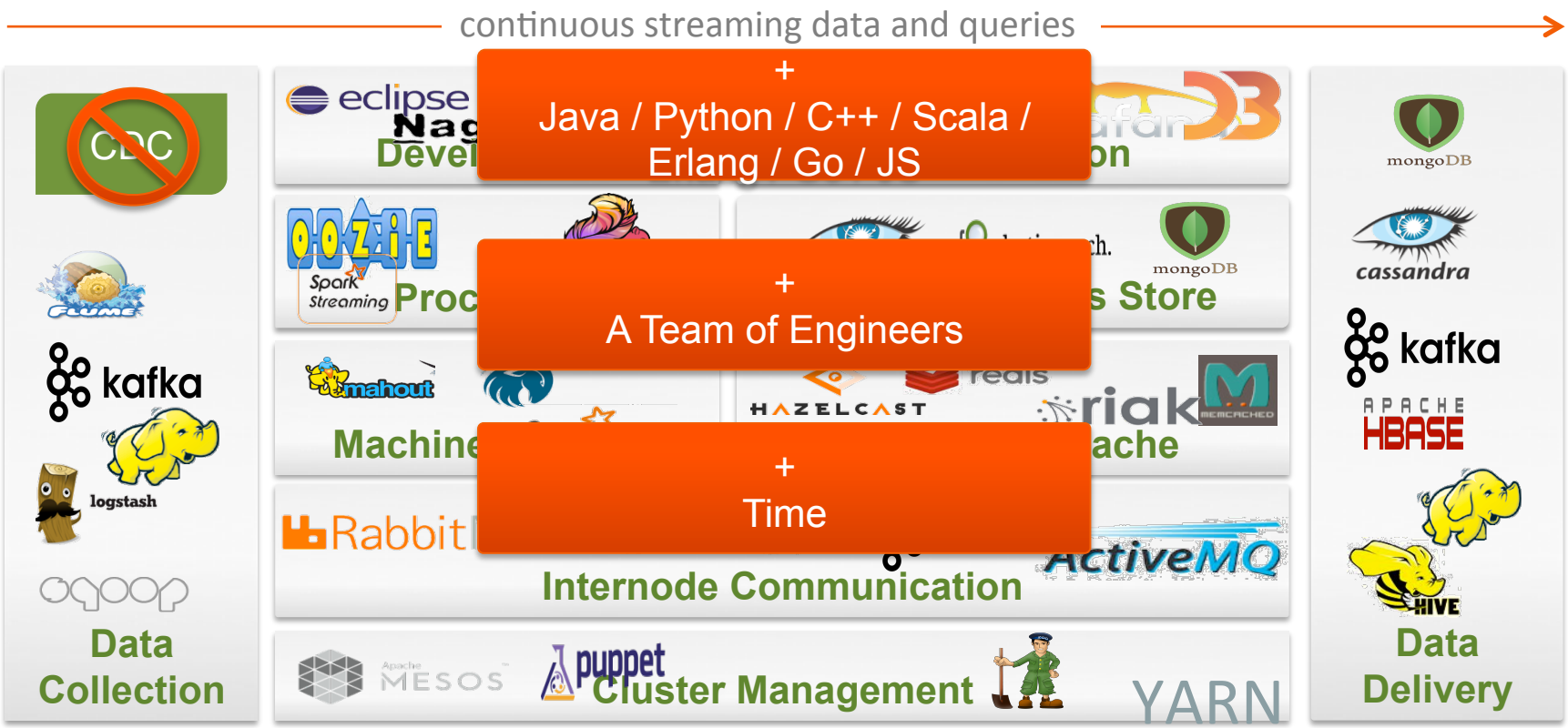
striim

With Open Source?

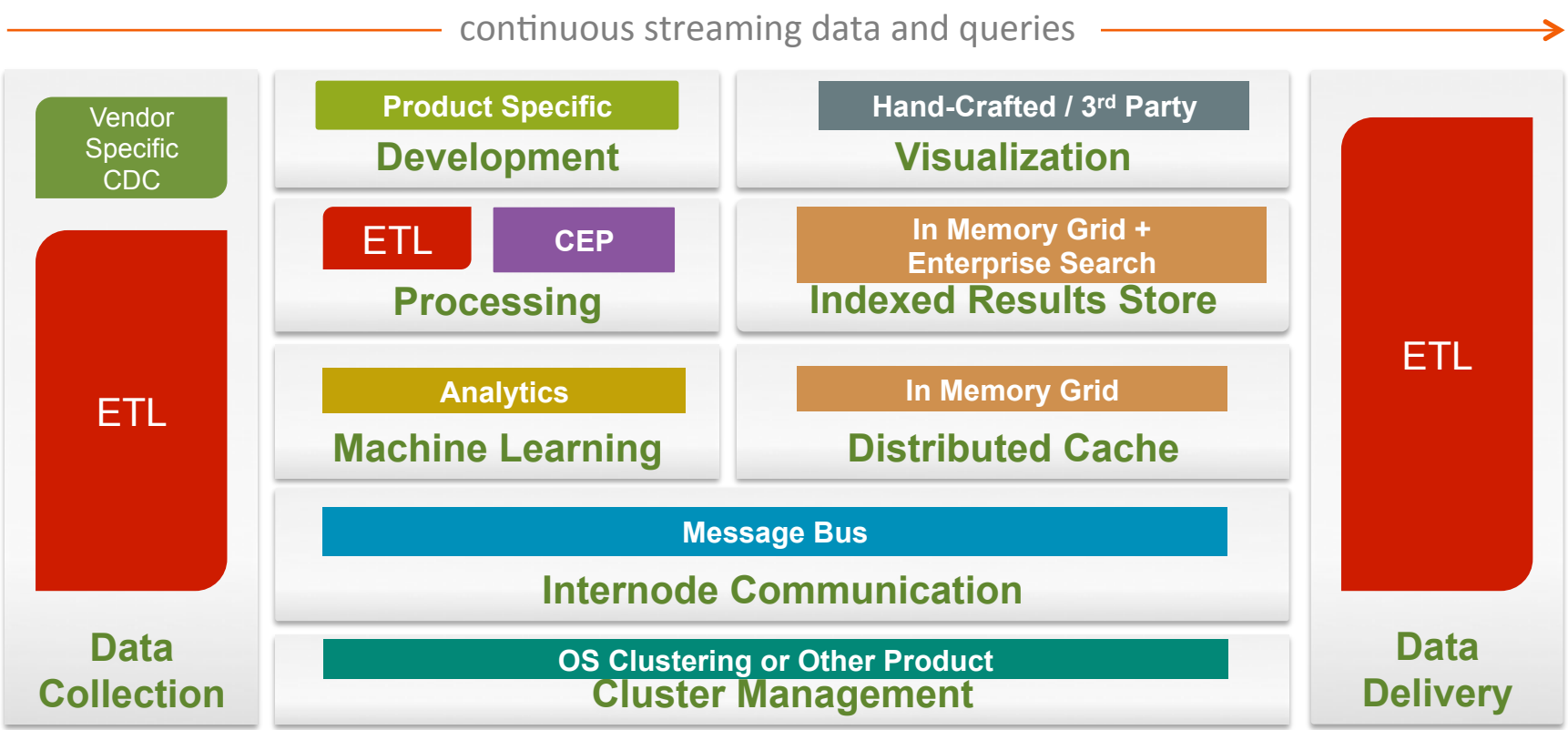
continuous streaming data and queries

striim

With Open Source?

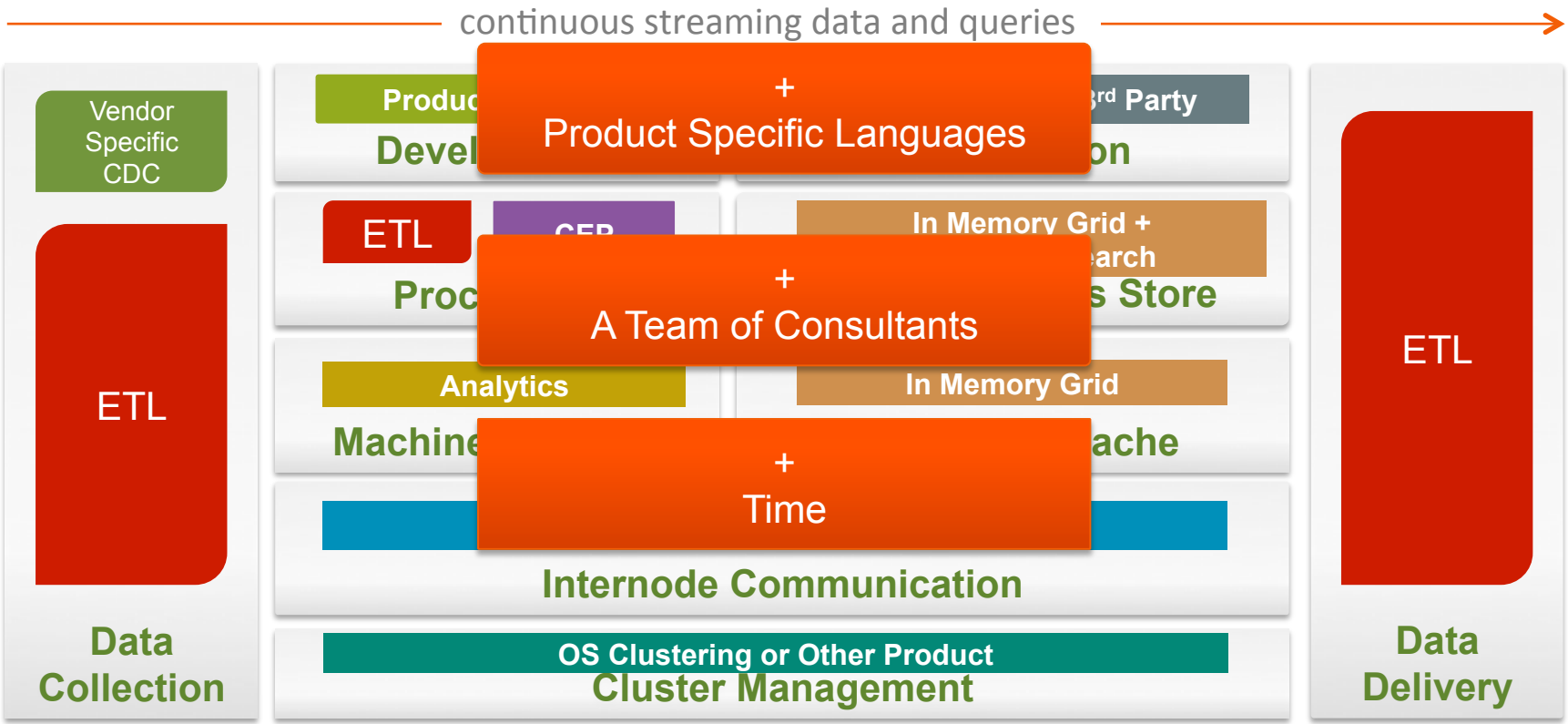


From Legacy Vendors?



striim

From Legacy Vendors?



striim

With Striim


continuous streaming data and queries
CDC
Logs
Sensors
Data Collection
Cloud
Big Data
DB
Data Delivery
Cluster Management
Single End-To-End Platform
Enterprise Ready Secure and Robust
striim
Scale-Out Architecture
Develop in UI with SQL
striim

Unmatched

An **end-to-end** streaming integration and intelligence solution.

Streaming Integration

- Capture **data the instant it's born** from structured and unstructured sources
- **Non-intrusive streaming CDC** for transactional data sources
- **Correlate** data across multiple streams in real time
- **Enrich** streaming data with reference and historical data **for continuous integration**
- Build streaming integration pipelines **quickly, using a SQL-like language**
- **Granular security framework** for streaming data

Streaming Intelligence

- Analyze data and trigger **alerts and workflows the instant data is born**
- **Continuously** updating **visualizations** of streaming data
- **Enrich** streaming data with reference and historical data **for immediate context**
- Process events **once-and-only-once**
- **Enterprise-strength and enterprise-scale leveraging low-cost compute**.

striim

Solutions and Use Cases






Cloud Application Control

Security Event Processing

Risk & Fraud Alerting

Quality of Service Management

Consumer Analytics

Device (IOT), Datacenter Analytics

Anti-Money Laundering
Anomaly Detection
Predictive Device Maintenance
VIP Customer Quality of Service Monitoring
Geo-targeted Mobile Marketing
Connected Cars
Cross-platform Attack Monitoring
Point of Sale Monitoring

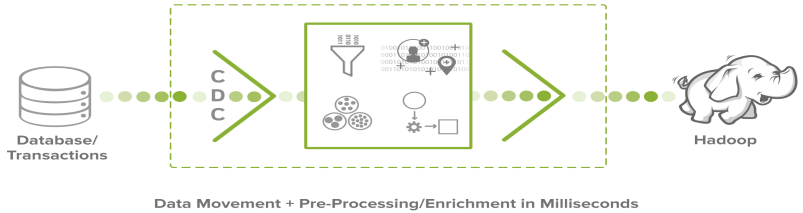
Fraud Detection and Prevention
Risk Management
Credential Monitoring
Multi-log Correlation
Retail Trends and Anomalies
API Usage Monitoring
SLA Monitoring
Partner Activity Monitor

striim

Integration Use Case: CDC to Hadoop

Scenario: You need to get activity in the form of change from relational databases into Hadoop, processing it on the way

- User and application activity in enterprise databases is an essential information asset but it can be difficult and expensive to access
- To get a complete picture of what's happening in your enterprise, you need to include the realtime change in corporate databases
- Upfront stream processing such as filtering and enrichment, optimizes data storage footprint and stored data become more actionable
- Change data can be used for
  - maintaining an audit trail
  - understanding activity patterns
  - cross referencing with other sources like websites



striim

Intelligence Use Case: Fraud Prevention

Scenario #1: A credit or debit card is used from two different ATM devices in distant locations within a short span of time.

- Fraud case over distance
- 10 km minimum distance between two ATM locations
- 5 minute window
- Tens of Thousands of ATMs

Scenario #2: Account balances of multiple cards have been requested from the same ATM device within a short span of time without completing a transaction.

- At least three or more cards used
- 2 minute window
- Tens of Thousands of ATMs

striim

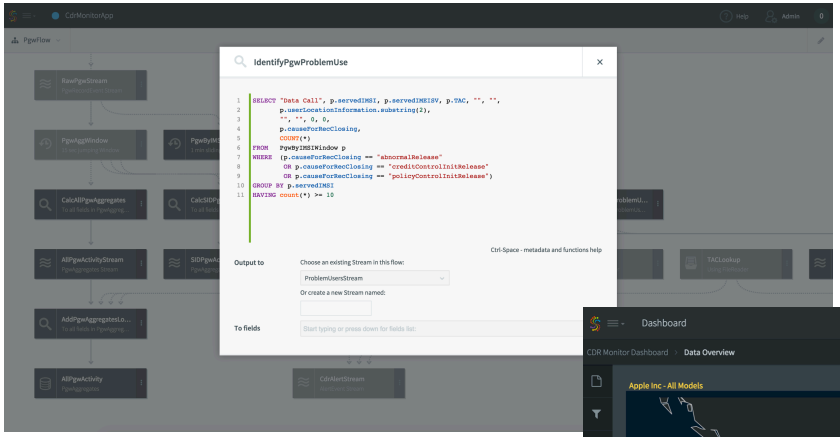
Intelligence Use Case: Customer Experience

Scenario: Monitor activity of customers in realtime to ensure best possible customer experience

- Compare behavior to historical records and flag anomalies
- Realtime risk profiling, alerting and workflow automation
- Identify issues affecting customers before they experience them
- Utilize customer context to determine actions
  - deal with VIP customers differently
  - use customer models for realtime scoring
- Detect and react to patterns that may indicate customer churn

striim

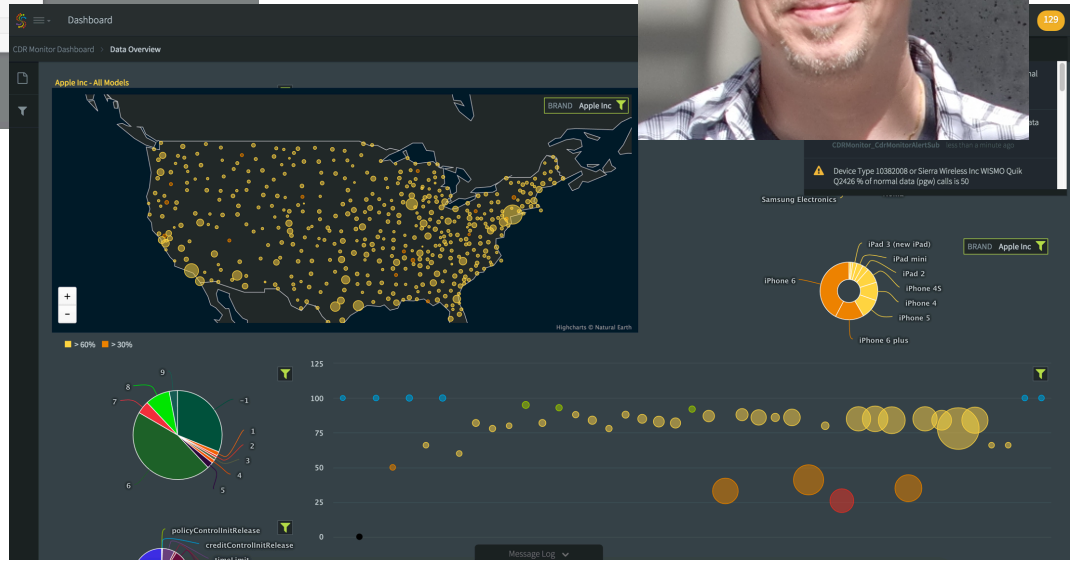
Want To Know More?



steve@striim.com

striim.com

@striimteam



striim

Perceptions & Questions

Analyst:
Mark Madsen

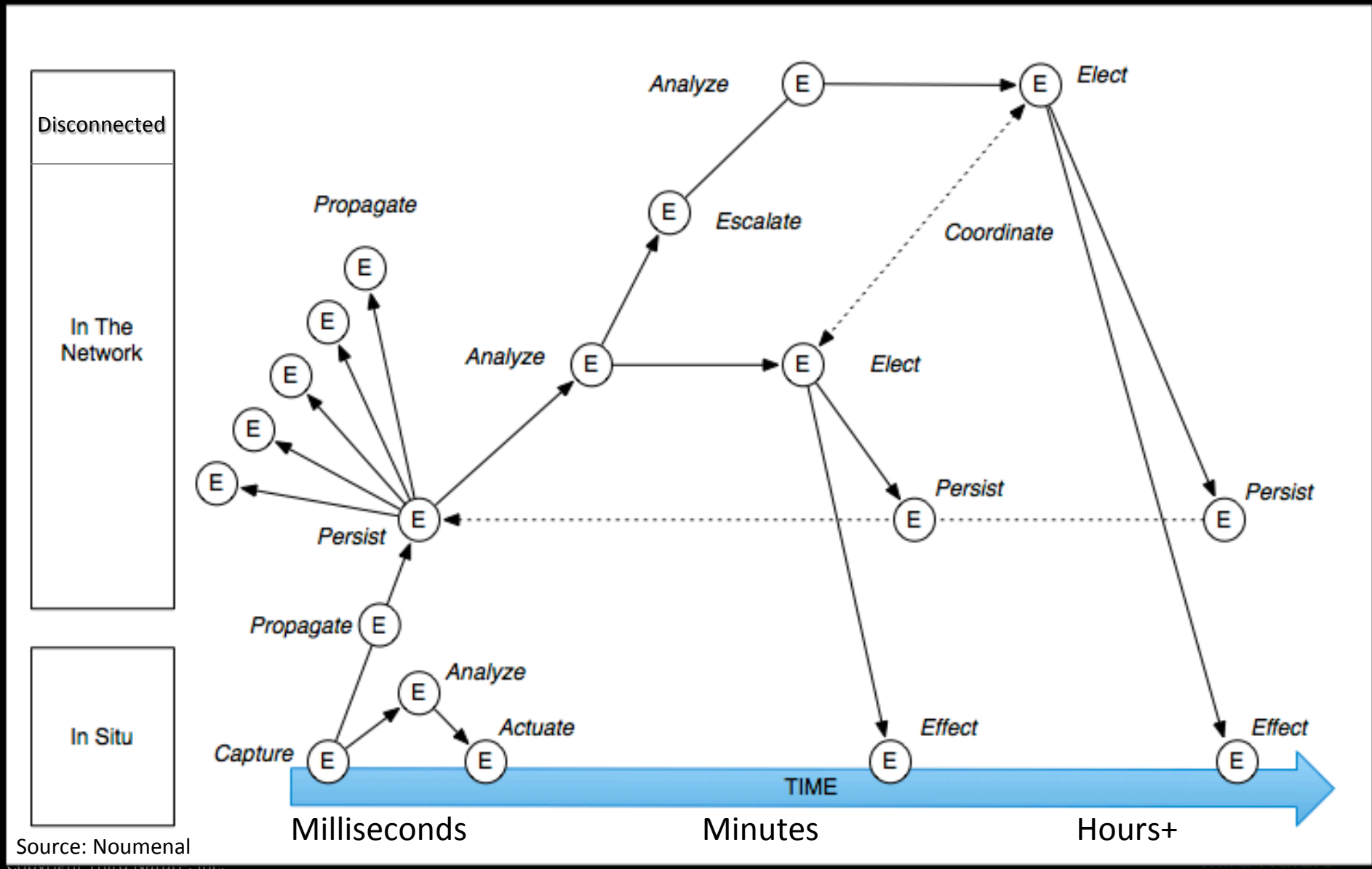
Is ETL Now a 4-Letter Word? Preparing for Streaming Analytics

Analyst commentary

October, 2015
Mark Madsen
Third Nature
@markmadsen

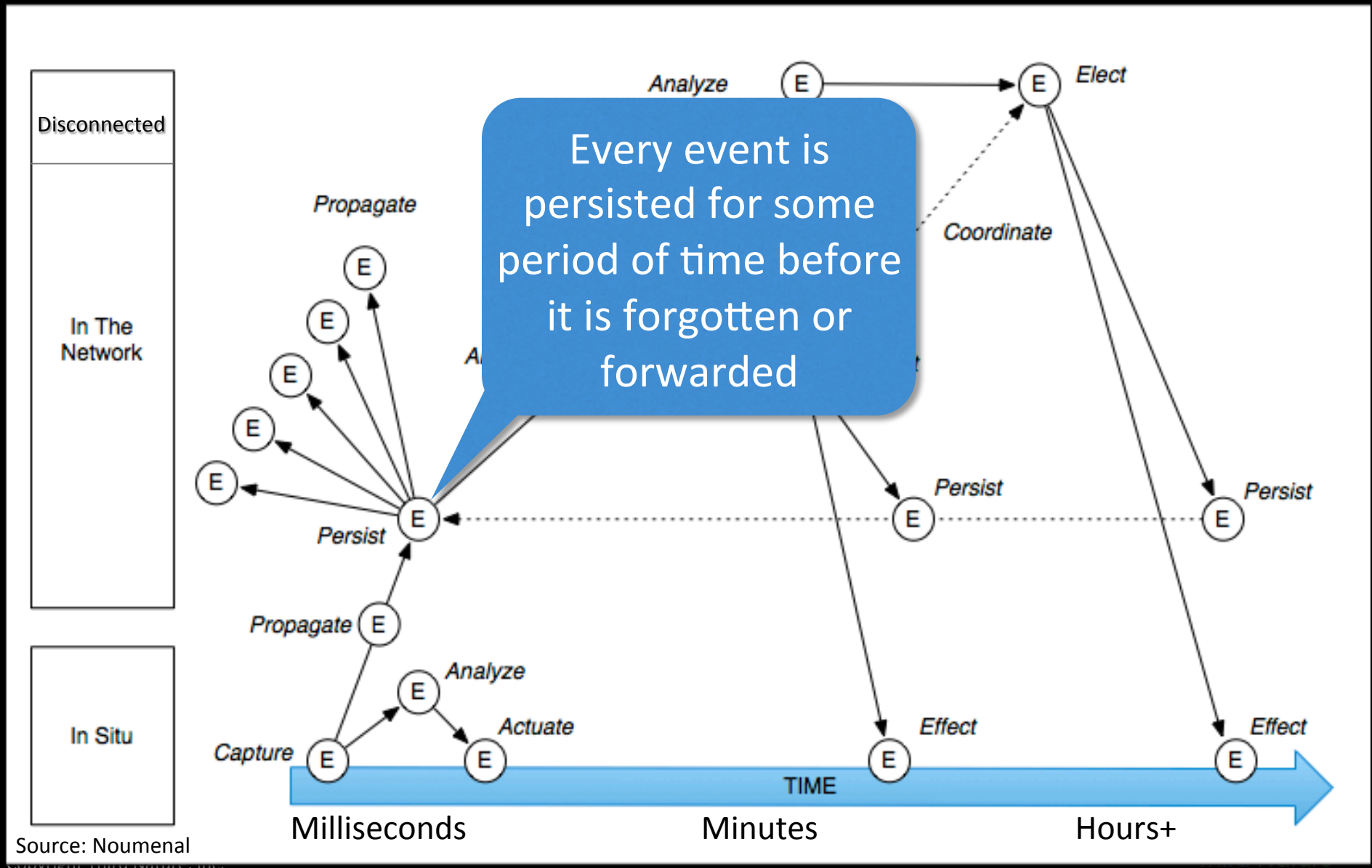
Third Nature

In a mostly-connected world, events occur in different time frames, follow different cycles of use



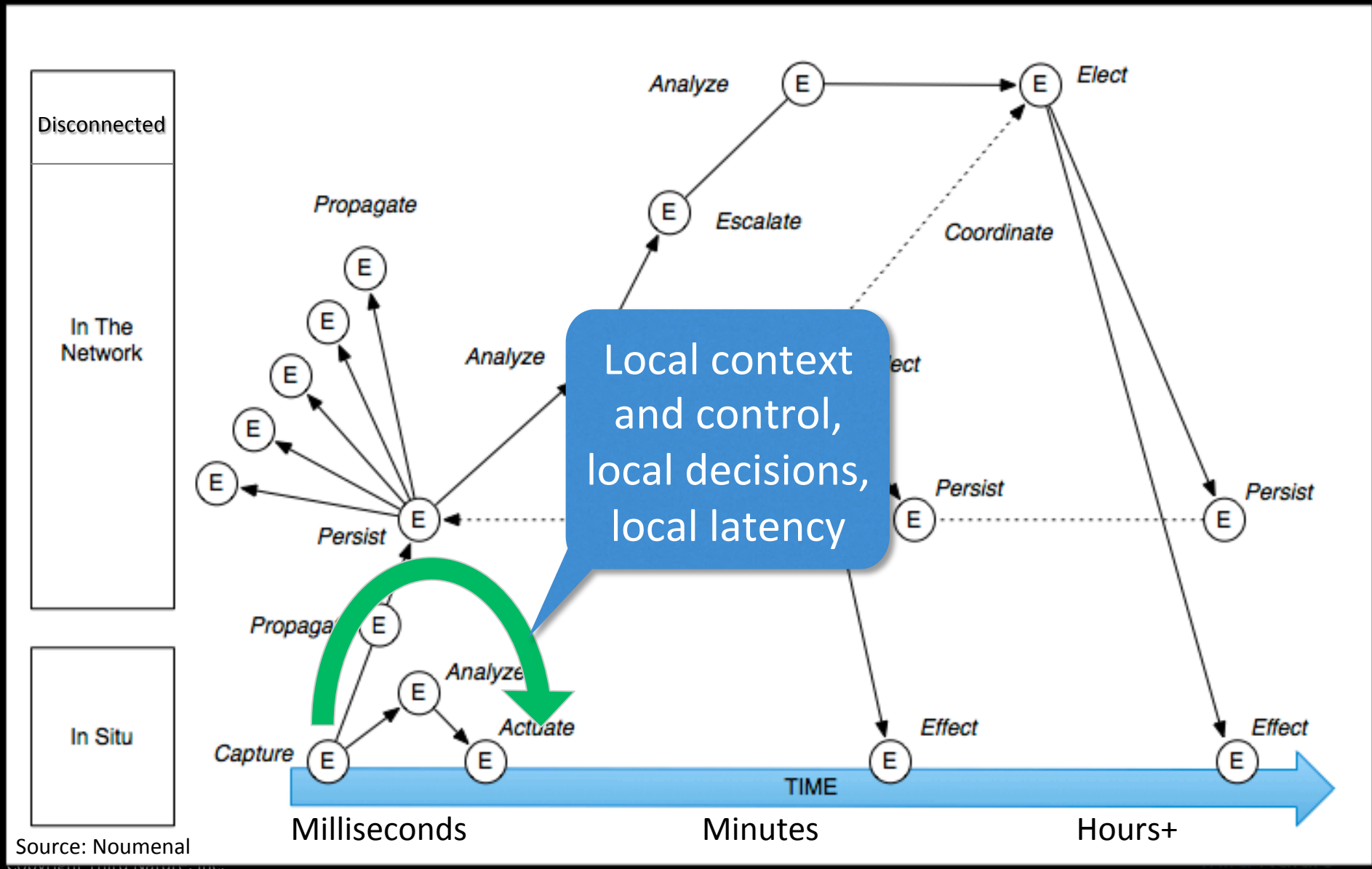
Source: Noumenal

In a mostly-connected world, events occur in different time frames, follow different cycles of use



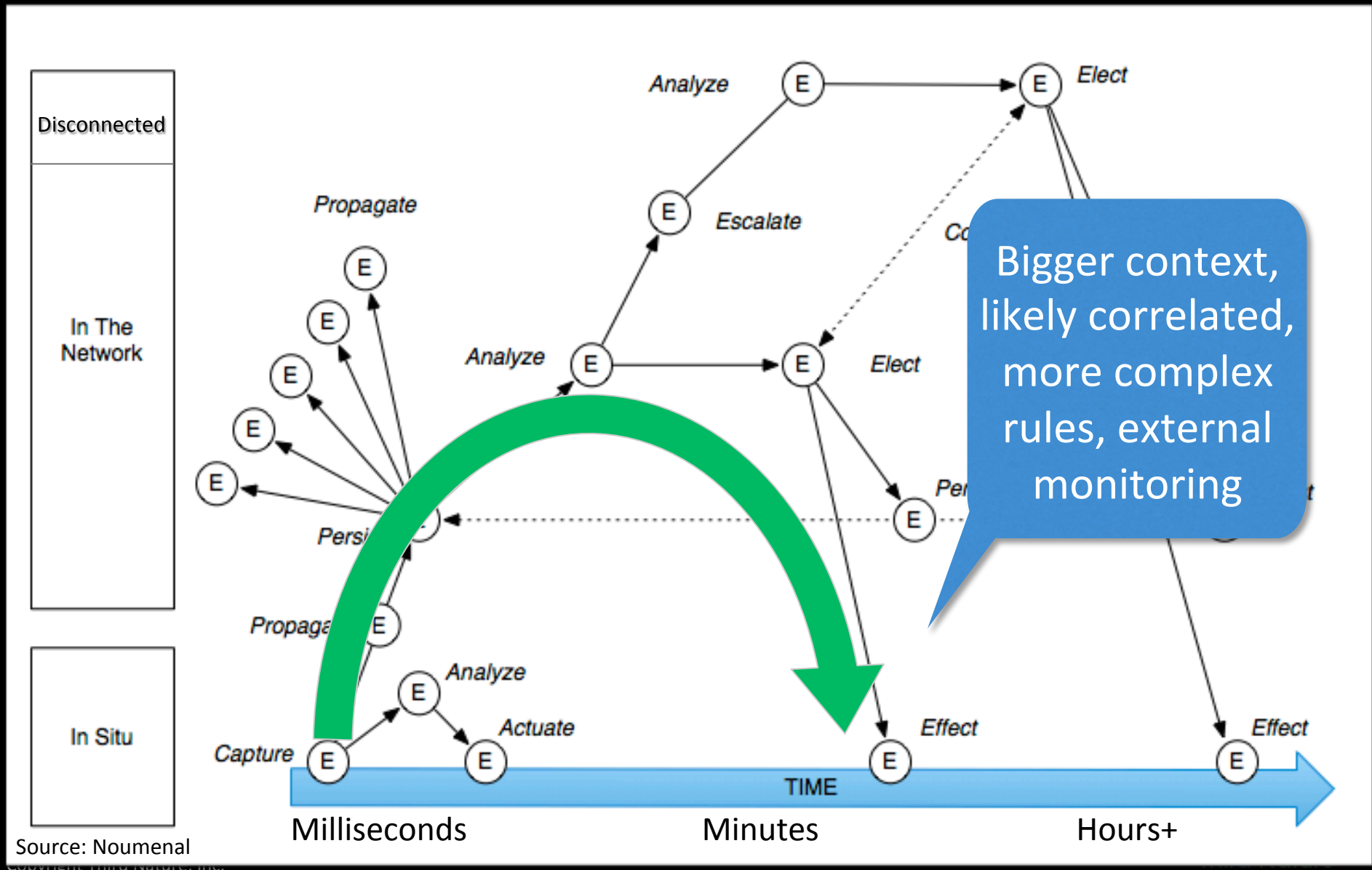
Source: Noumenal

In a mostly-connected world, events occur in different time frames, follow different cycles of use



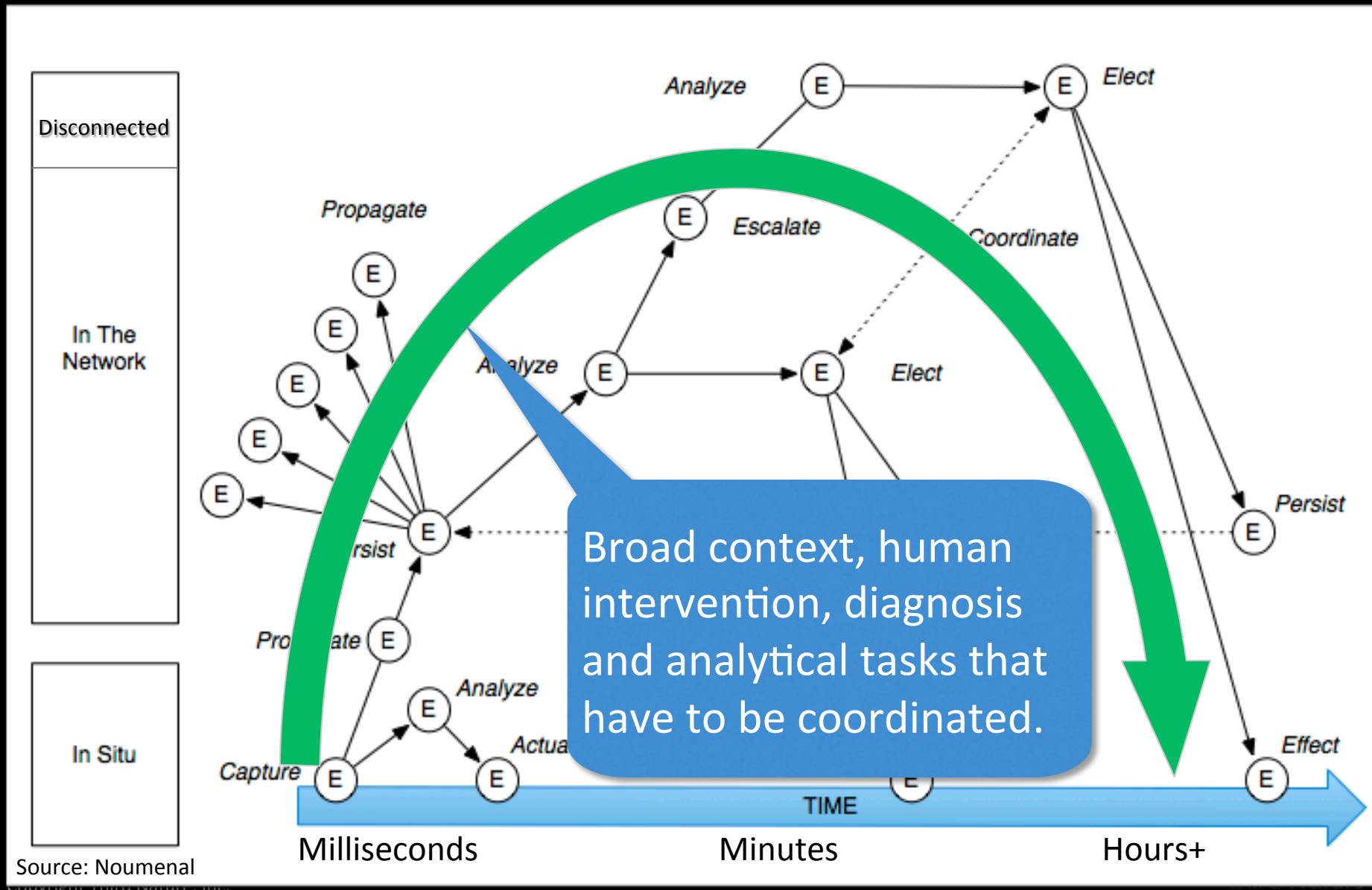
Source: Noumenal

In a mostly-connected world, events occur in different time frames, follow different cycles of use



Source: Noumenal

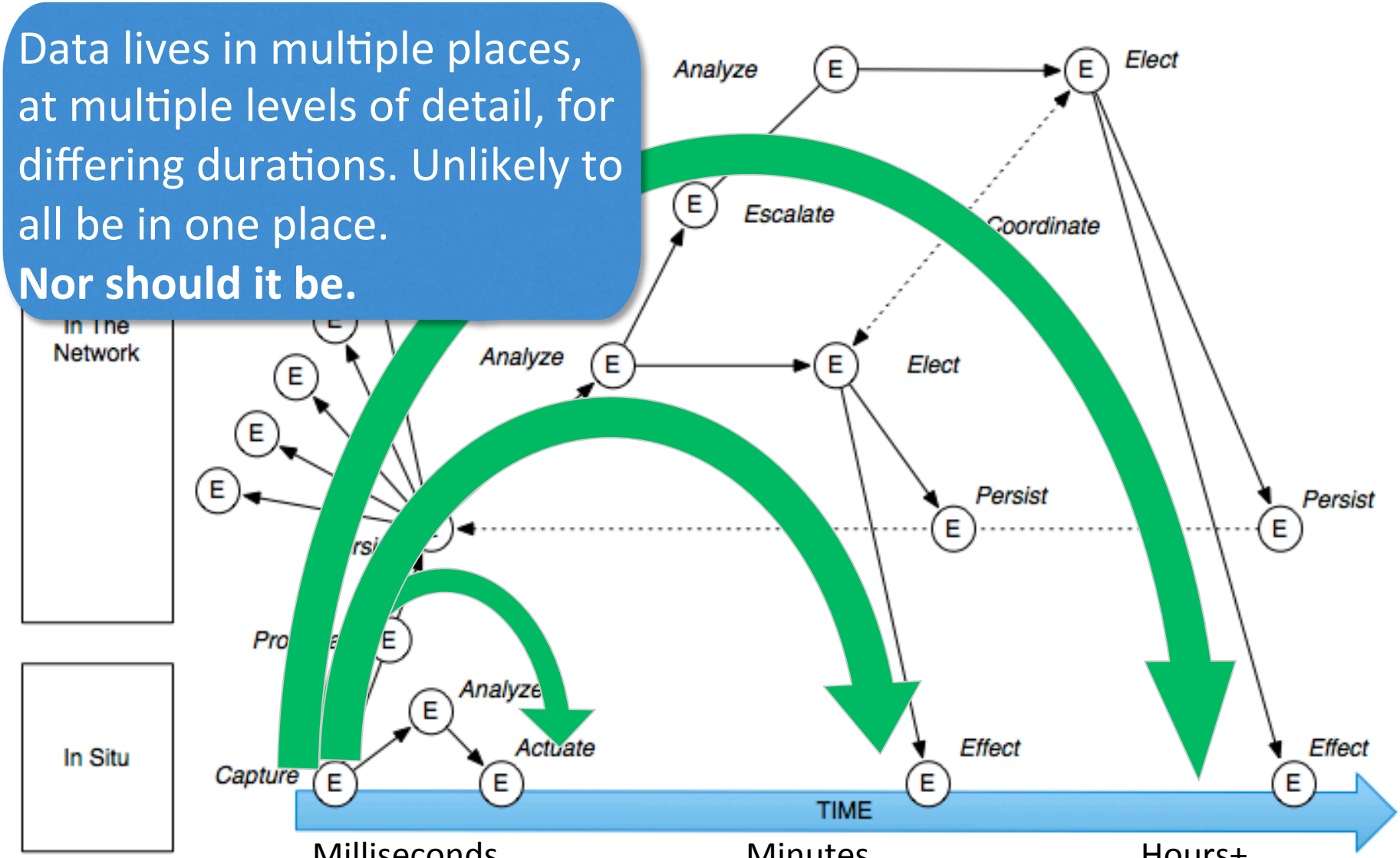
In a mostly-connected world, events occur in different time frames, follow different cycles of use



Source: Noumenal

In a mostly-connected world, events occur in different time frames, follow different cycles of use

Data lives in multiple places, at multiple levels of detail, for differing durations. Unlikely to all be in one place.
Nor should it be.



Source: Noumenal

We have a model for the persisted portion only

The DW can't handle real time ingest

- One of the original DW design assumptions: solve for conflicting workloads by using a different database
- Workload management has limits
- Scalability problem for event streams
- Spiky flow patterns and dynamic scaling

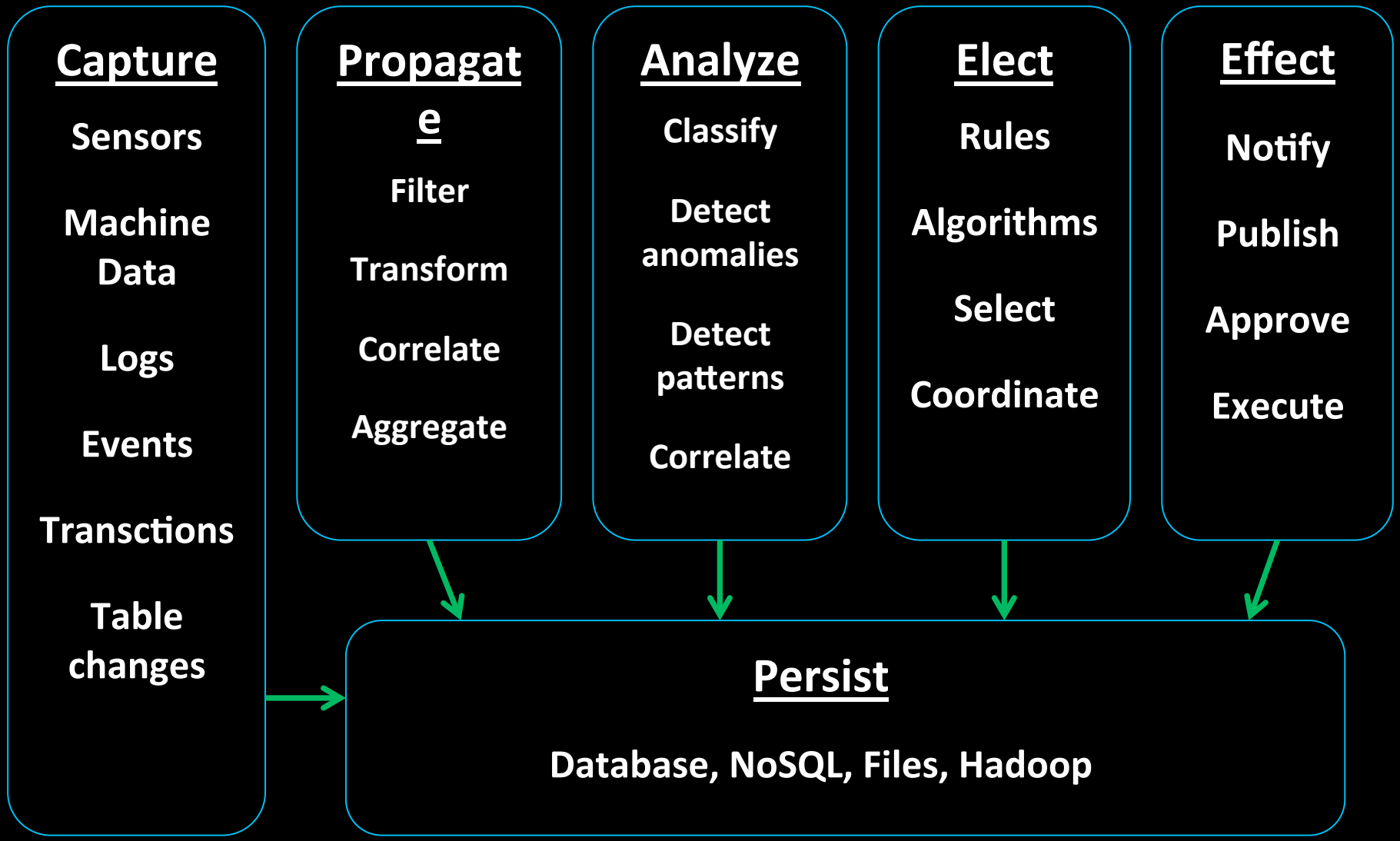
Static schema:

- Reaction time - shapes, holes, dropped packets
- What happens first, upstream change or data model change?

Polling architectures do not work well for streaming

- Introduces latency
- Polling creates performance and scaling problems

Third Nature



Third Nature

Streaming isn't either-or, it's part of core architecture

Sliding window of "now"

Persisted but not yet loaded into a platform

Queryable history

Event streams, in-mem stores, CEP streaming SQL can be used for these

A DB or ETL can get you to within minutes (at large scale) but it won't be easy or cheap; mainly lives in the realm of history

Real time monitoring doesn't use only real time data: windows, restarts, detecting deviation, so the above boundaries are crossed.

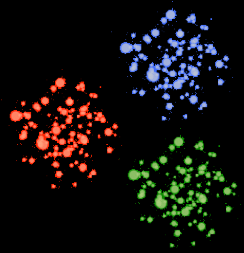
Third Nature

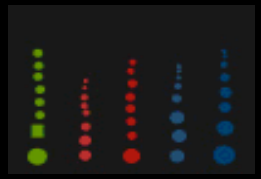
If you want to do realtime and still manage your data effectively then you need to think about data architecture

Microservices

Metadata & reuse?

Metadata





Stream Collect Refine Manage Deliver

Flowing | Persisted | Managed history

Flow, persisted, managed define different access, processing, storage and retrieval requirements

Third Nature

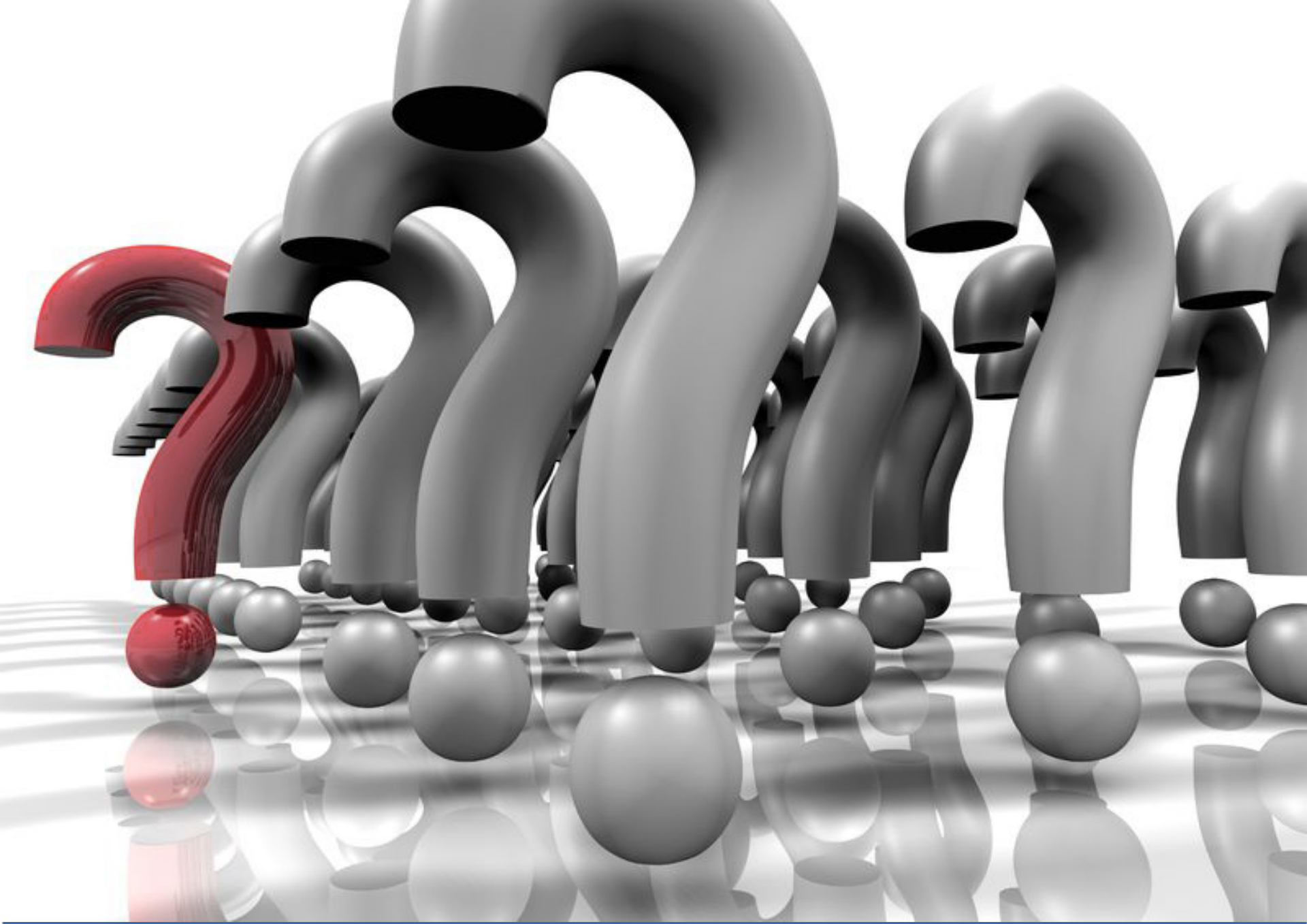
About Third Nature

Third Nature is a research and consulting firm focused on new and emerging technology and practices in analytics, business intelligence, and performance management. If your question is related to data, analytics, information strategy and technology infrastructure then you're at the right place.

Our goal is to help companies take advantage of information-driven management practices and applications. We offer education, consulting and research services to support business and IT organizations as well as technology vendors.

We fill the gap between what the industry analyst firms cover and what IT needs. We specialize in product and technology analysis, so we look at emerging technologies and markets, evaluating technology and hw it is applied rather than vendor market positions.

Third Nature



Upcoming Topics

October: **DATA MANAGEMENT**

November: **ANALYTICS**

December: **INNOVATORS**

www.insideanalysis.com

THANK YOU
for your
ATTENTION!

Some images provided courtesy of Wikimedia Commons