

SMART DATA FAST.TM



THE STATE OF STREAMING ANALYTICS: THE NEED FOR SPEED AND SCALE

OUR SPEAKERS



Mike Gaultieri
Principal Analyst
Forrester Research



Peter Vescuso
CMO, VoltDB

OVERVIEW

- The Streaming Analytics Market
- Solution for Analytics and Transactions
- VoltDB -- New Features!
- Use Cases

FORRESTER®

The State Of Streaming Analytics

The Need For Speed and Scale

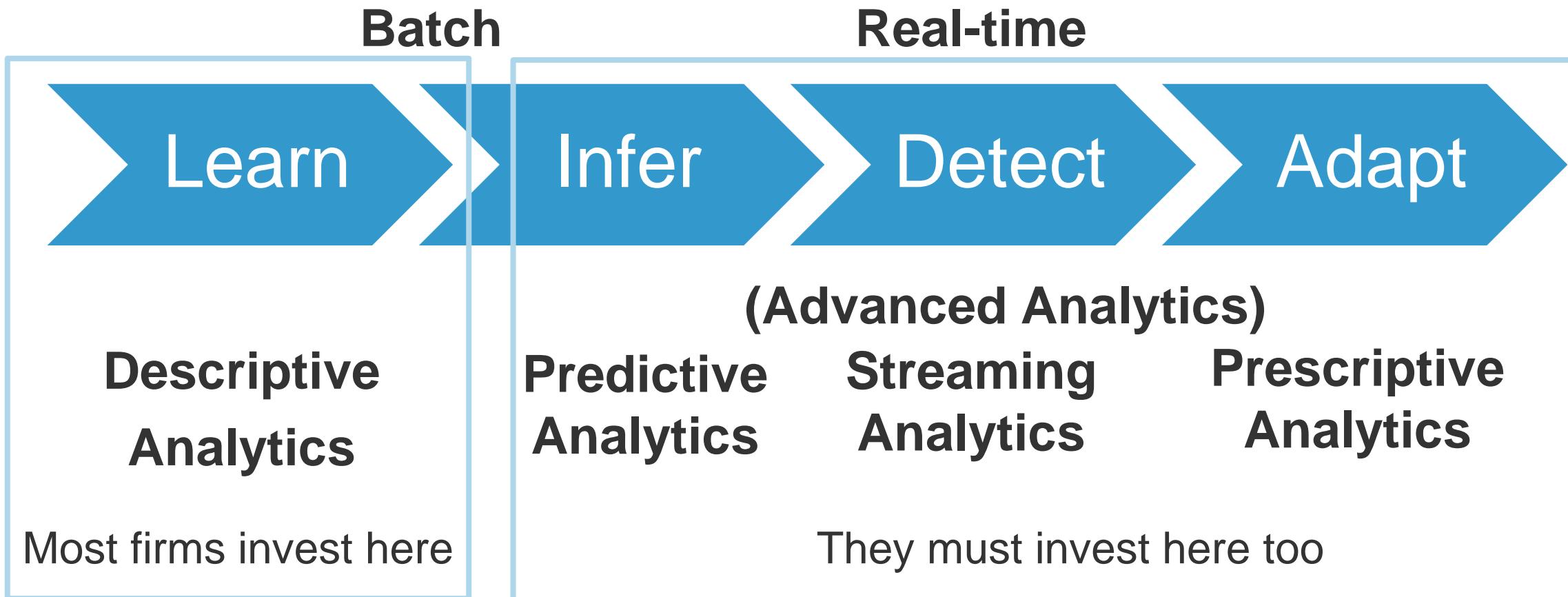
Mike Gualtieri, Principal Analyst

Twitter: @mgualtieri



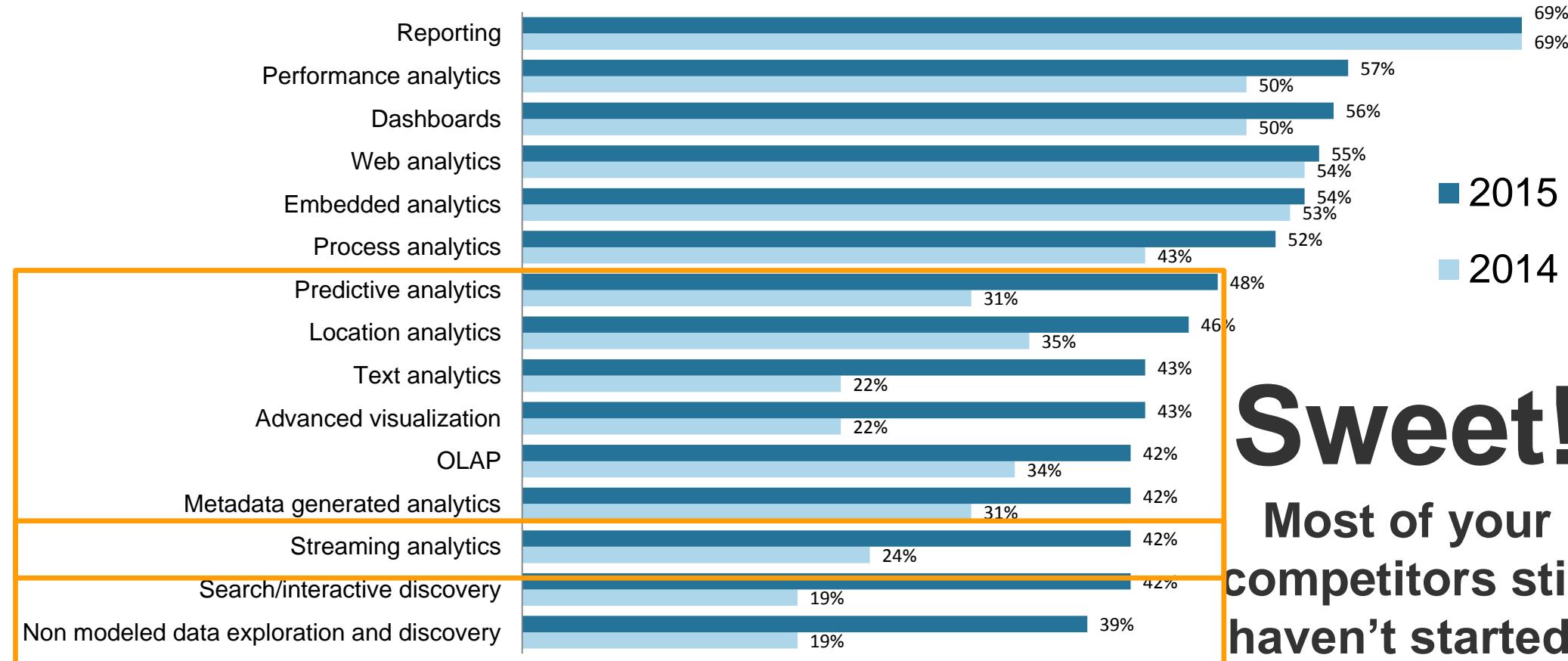
#Analytics

True BI means having all kinds of analytics



Advanced analytics is rightly surging

“What is your firm's/business unit's current use of the following technologies?”



Sweet!
Most of your
competitors still
haven't started!

Source: Forrester Research

Source: Forrester's Global Business Technographics Data And Analytics Survey, 2015 and 2014
Base: 1805 (2015), 1063 (2014)

#FastData



Some say that data is the new oil.



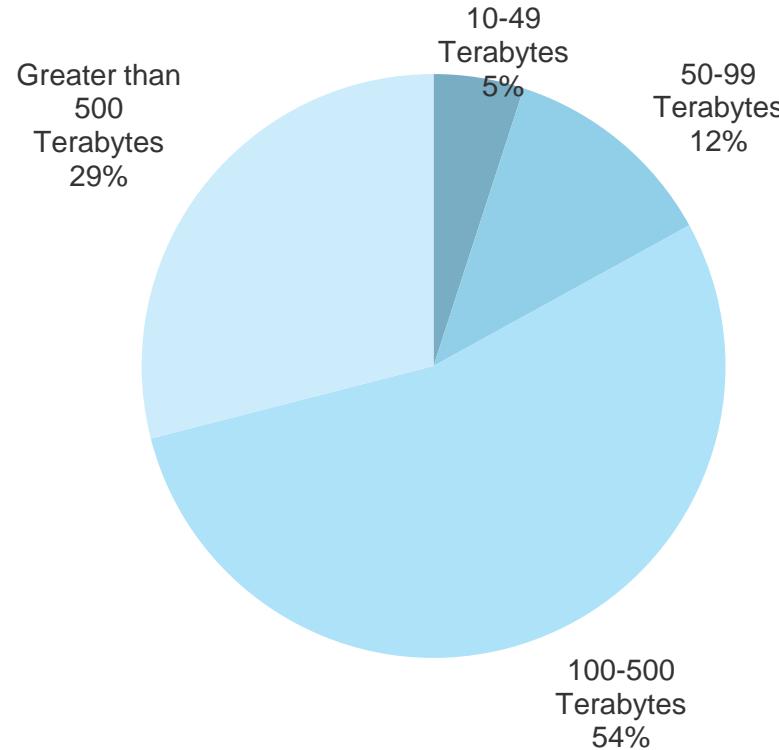
It's more like the Sun – virtually limitless.

Every industry is graced with more data

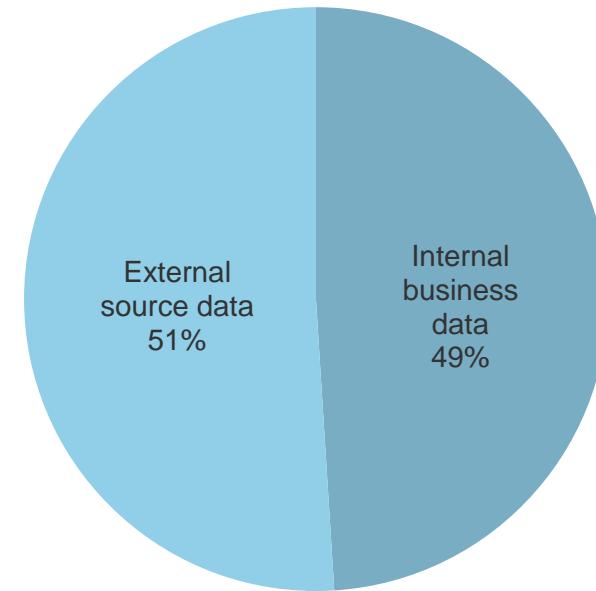
- › Rich transactional data from business applications
- › Usage and behavior data from web and mobile apps
- › Social media data
- › Log data
- › IoT device sensor and event data
- › Data economy – firms buying and selling data

Enterprises have plenty of data from both internal and external sources

Using your best estimate, what is the size of all data stored within your company?



What % of the data available is from internal business applications (ERP and business applications) versus external sources (social, IoT)?



Source: Forrester Research, September 2015

Base: 100 US Managers and above currently using Hadoop for processing and analyzing data.



All data originates fast!

The background image shows a row of five large, cylindrical metal silos standing in a field of tall, dry grass. The silos have dark, conical roofs and are illuminated from within, casting a warm glow and long shadows onto the ground. The sky above is a clear, pale blue.

IoT

110010011011001

Transactions

0100110011011

Customer Data

0100100110

Data Warehouse

0100

But, analytics is usually done much later.

#WhyWait



Perishable insights can have exponentially more value than after-the-fact traditional historical analytics.

#Perishable

A photograph of a young man with long brown hair and black-rimmed glasses, smiling broadly at the camera. He is wearing a dark denim jacket over a blue t-shirt. The background shows shelves filled with books, suggesting a library or bookstore environment.

**How can you prevent this dude from fleecing
you right now?**



Crowborough CFRs @CrowboroughCFR · Feb 8

For those interested from my last post. flashaholics.co.uk/olight/olight-...

great head torch. #olight

CRB-03

Expand

Reply Retweet Favorite More



ITOURLIGHT @ITOURLIGHT · Jan 30

Small but powerful flashlight #Olight#itourlight

itourlight.com/OLight-S10-Bat...

Expand

Reply Retweet Favorite More



Kristen Williams @TheGunChick · Jan 23

Awesome video! Now that's some extreme testing! #olight fb.me/1cryljaS6

View media

Reply Retweet Favorite More



SOTG @studentofthegun · Jan 19

#View of #LasVegas #Strip for the #Palazzo. Thanks to Uncle Dick for the invitation to the #OLight... instagram.com/p/jXn_7ZsF4o/

Expand

Reply Retweet Favorite More



Followed by Marilyn Terrell

BatteryJunction.com @BatteryJunction · Jan 6

Unique power bank that strips down to 170g. Get up to 100 lumens with a rechargeable AA! Could... instagram.com/p/rIQQn_jta/

Expand

Reply Retweet Favorite More



Hand Of Glory @Handofglory · Dec 21

Genuine night time music, up there with D. Wilson youtube.com/watch?

What are movers and shakers saying about equities that we cover right now?



How can you know if your baby is sleeping soundly or if something is wrong right now?

A photograph of two young women with blonde hair, smiling broadly and laughing. They are standing in front of a colorful, blurred background that suggests a carnival or fairground setting with various attractions and lights.

What offers should you make to your customer if they are within proximity of your store right now?



All data starts off fast!

#Problem

Quiz

What are the technical challenges impeding you from processing and analyzing more data?

- A. Difficulty integrating data from multiple sources.
- B. Data volume is too large.
- C. Difficulty in creating data models and/or preparing data for analytics.
- D. Too many data formats to integrate effectively.
- E. Data is difficult to access from multiple sources

#Streaming

A close-up photograph of a stream of water flowing over dark, textured rocks. A single, crumpled yellow sticky note floats on the surface of the water, appearing to drift downstream. The water is clear enough to see the rocks beneath.

**Streaming analytics can capture and act on
perishable insights.**

FORRESTER

Streaming analytics filter, aggregate, enrich, and analyze a high throughput of data from disparate live data sources to identify patterns, detect urgent situations, and automate immediate actions in real-time.

DEFINITION

Real-time means business time

- › A customer walks into a shopping mall
- › A shopper clicks on an online add
- › A temperature sensor spikes
- › A stock price rises
- › A customer uses a credit card
- › A customer wakes up

Thinking in streams is different...

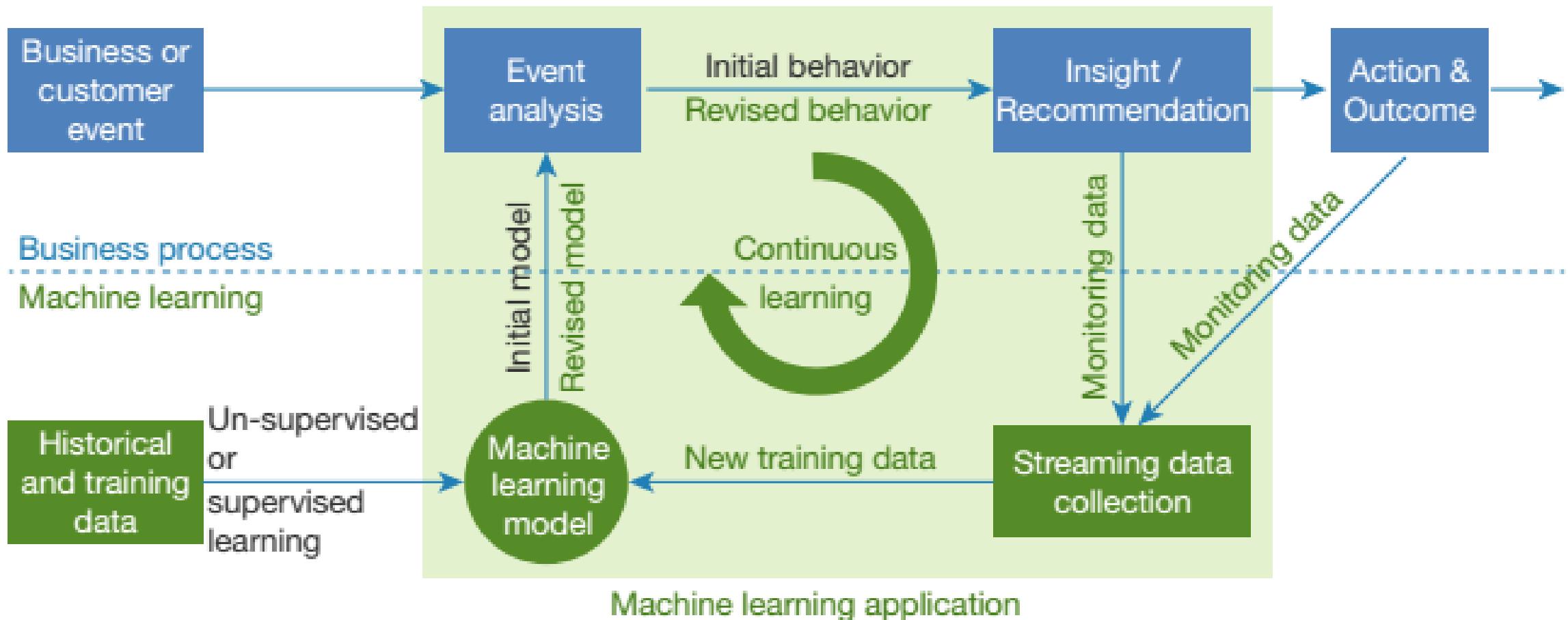
Real-time ETL

- › Ingest
- › Filter
- › Transform
- › Normalize
- › Link
- › Enrich

Real-time Analytics

- › Correlate
- › Location/motion (geofencing)
- › Time windows
- › Temporal pattern detection
- › Business logic/rules execution
- › Action interfaces

Streaming analytics-to-action architecture

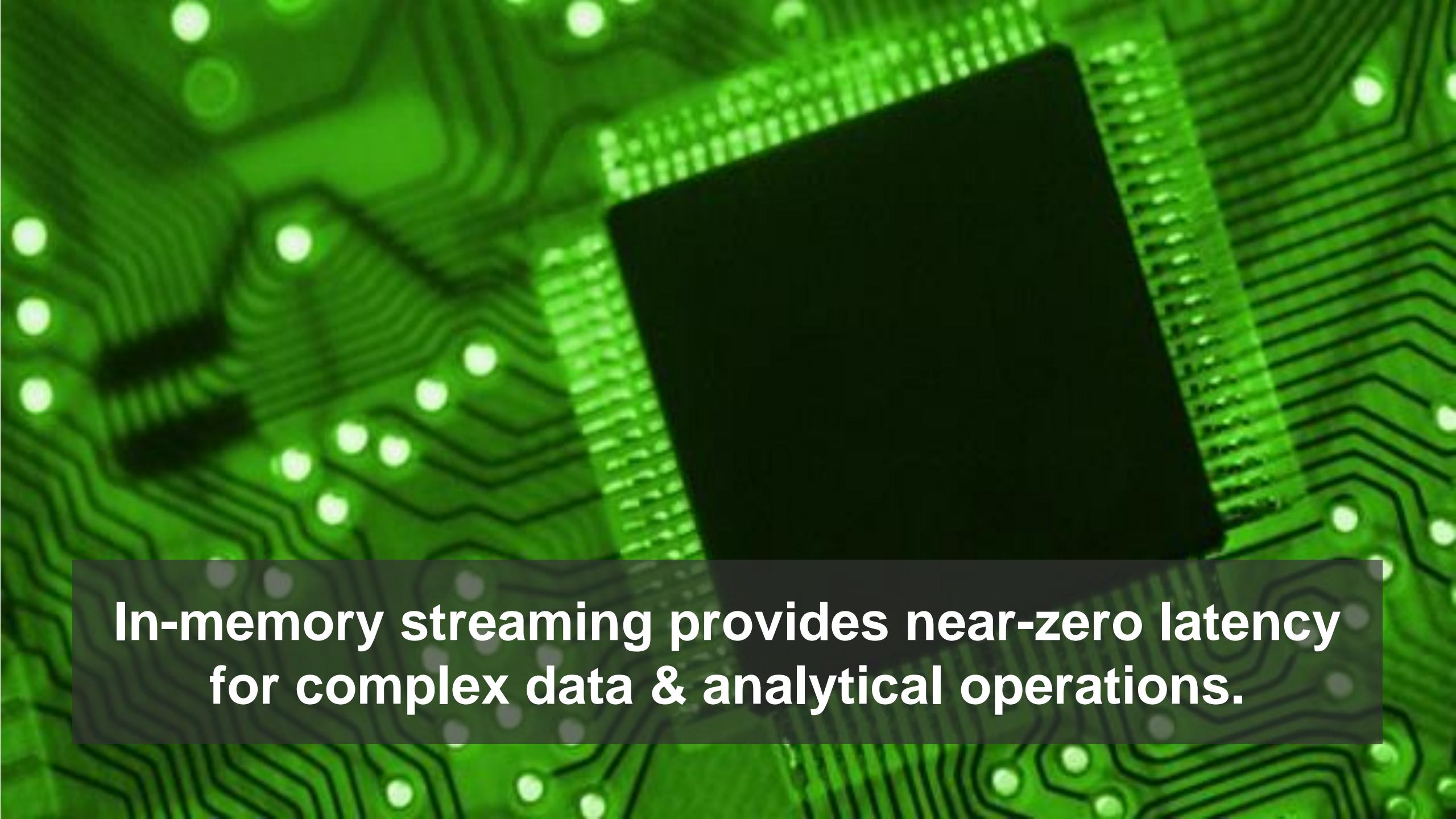


#Translytical

FORRESTER

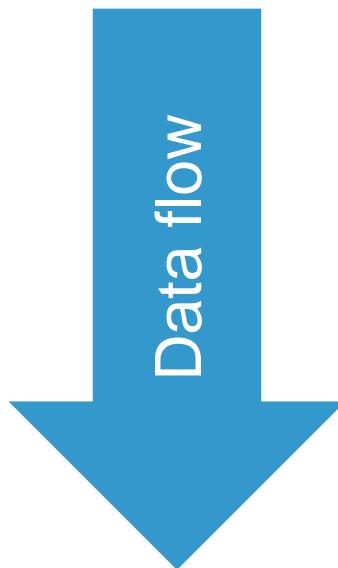
A single, unified database that supports transactions and analytics in real-time without sacrificing transactional integrity, performance, and scale.

DEFINITION

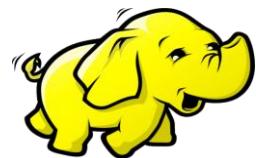


**In-memory streaming provides near-zero latency
for complex data & analytical operations.**

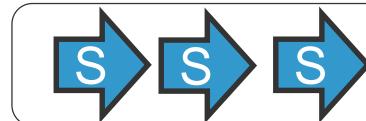
Real-time



Batch



Streaming Analytics



Scale-up Database



Data And Compute Grid



Clustered Database

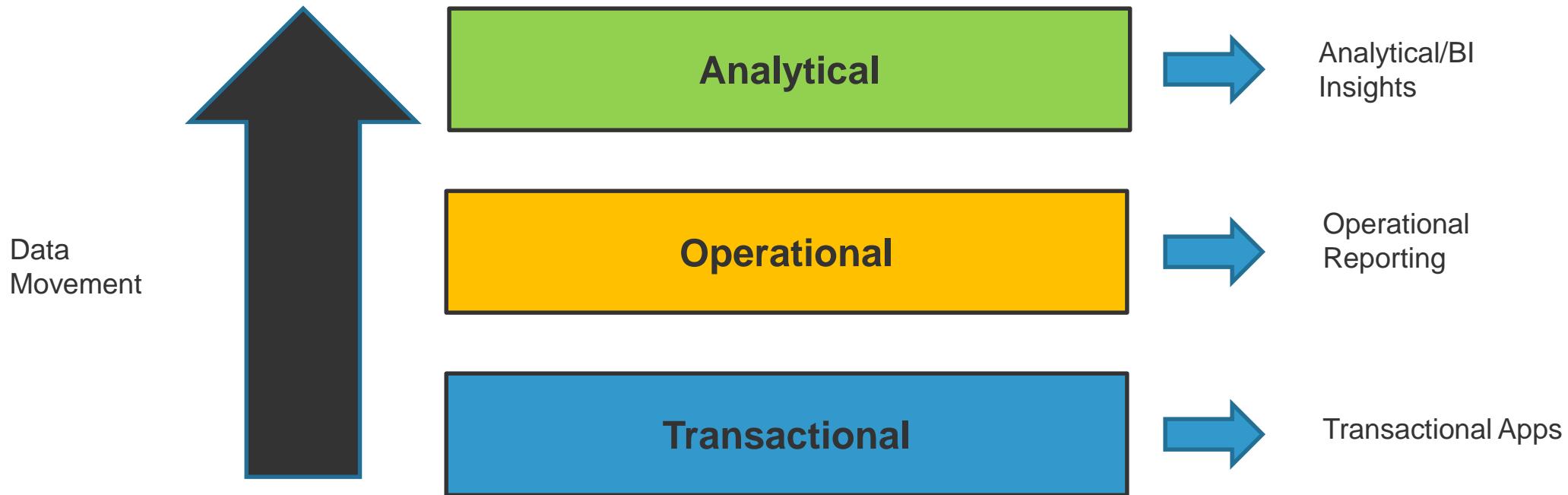


General-purpose data processing cluster

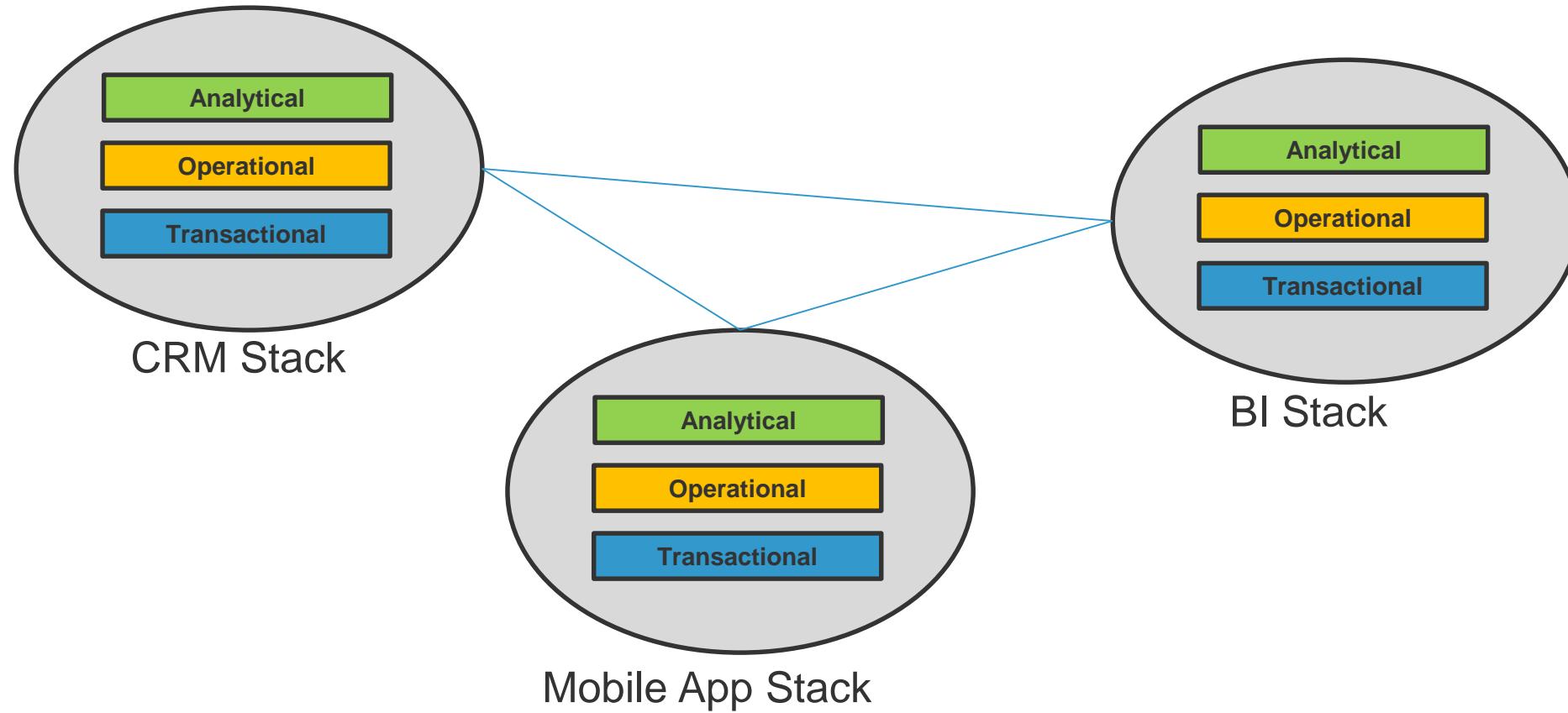


Spark

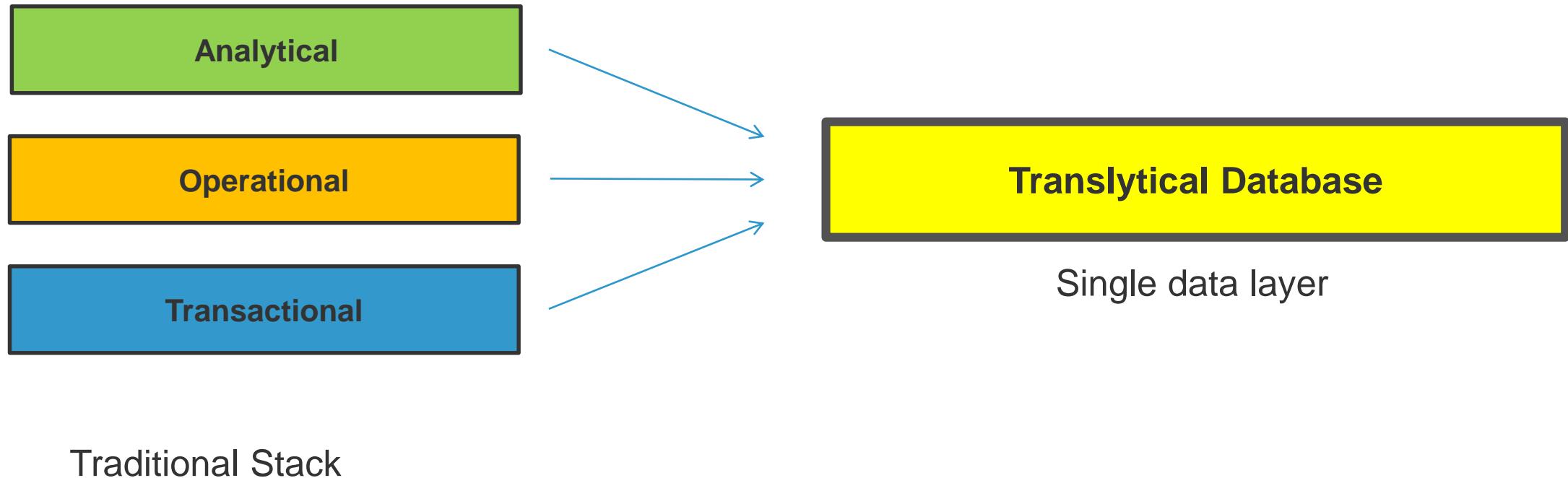
Traditional analytic architectures must move data to analyze it



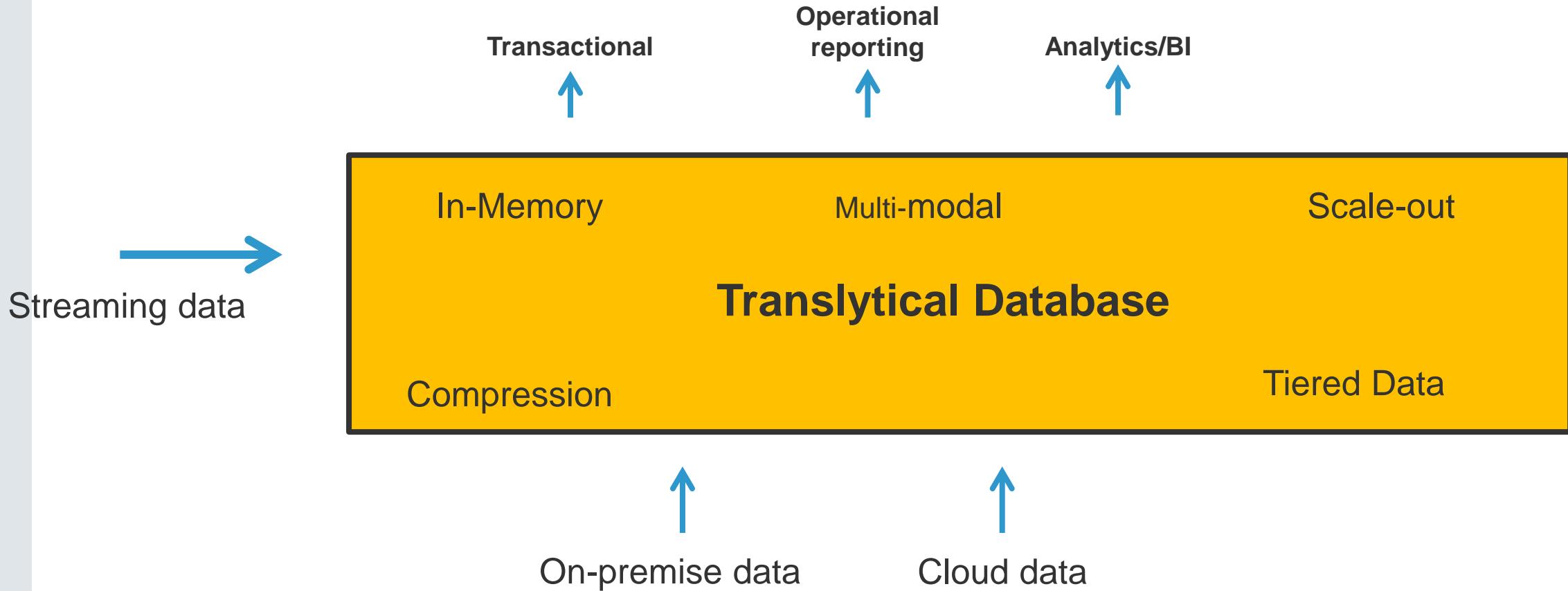
Analytical silos have proliferated



Translytical databases collapse the stack to reduce complexity and deliver analytics in real-time



Translytical databases are transactional and analytical





Translytic databases scale-out



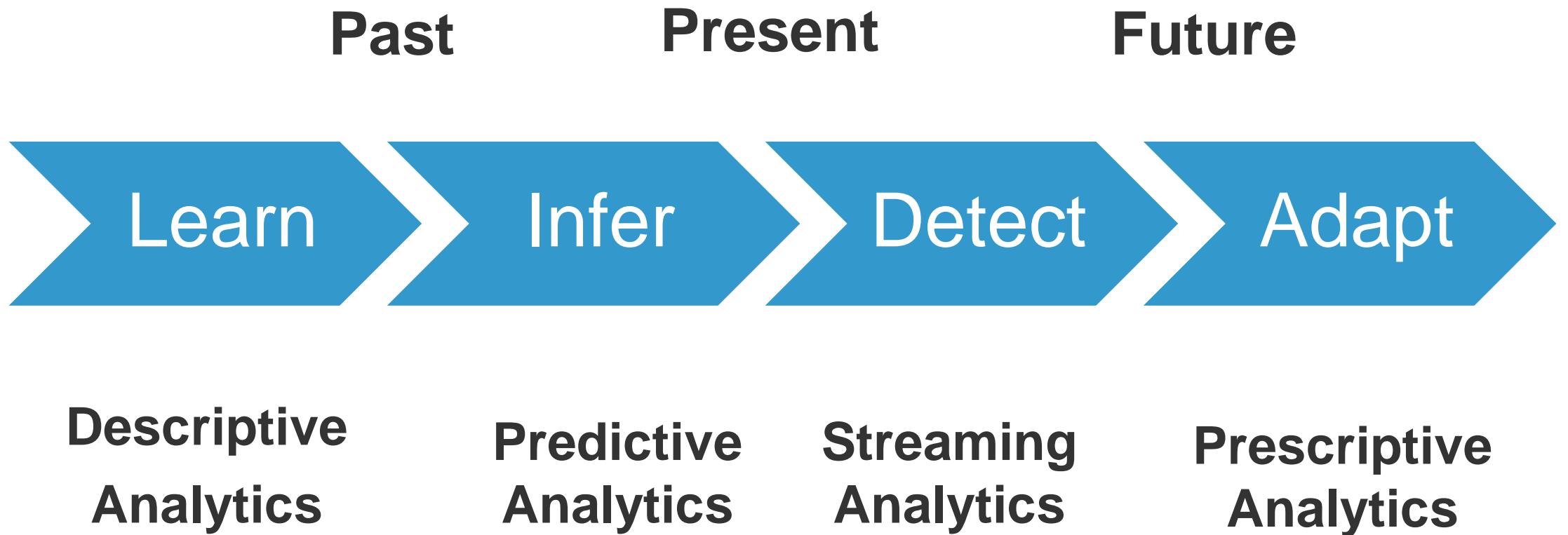
Translytic databases use RAM to be real-time.



Transactions + Analytics = Translytical

#Opportunity

Start by adopting translytic databases for real-time applications



Thank you

Mike Gualtieri

mgualtieri@forrester.com

Twitter: @mgualtieri



FORRESTER®

forrester.com

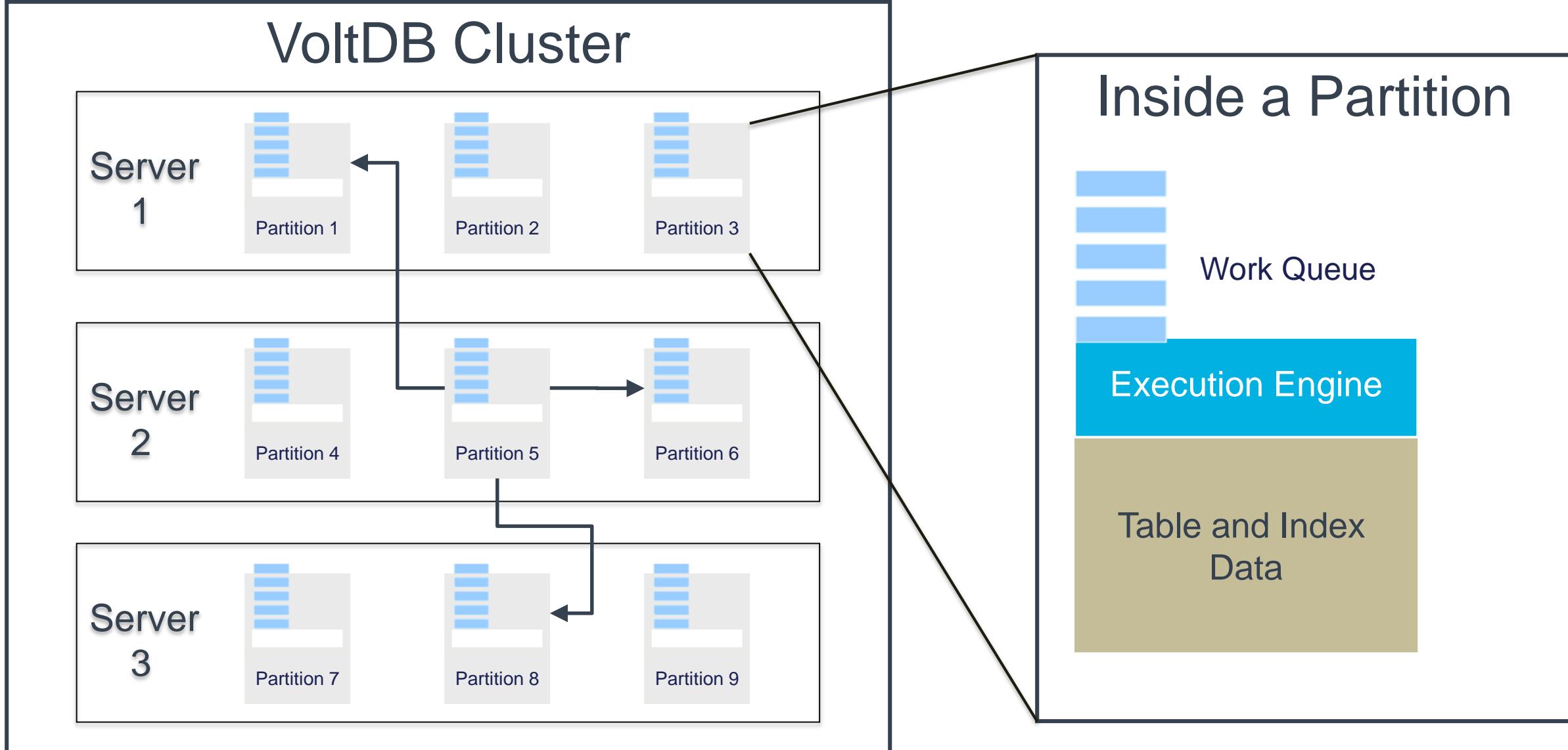
THE VOLTDB ARCHITECTURE FOR FAST DATA

- ✓ In-Memory performance (never lose data)
- ✓ Scale-out, shared nothing
- ✓ ACID & SQL & Java
- ✓ Real-time analytics
- ✓ Reliability and fault tolerance
- ✓ Hadoop ecosystem integration

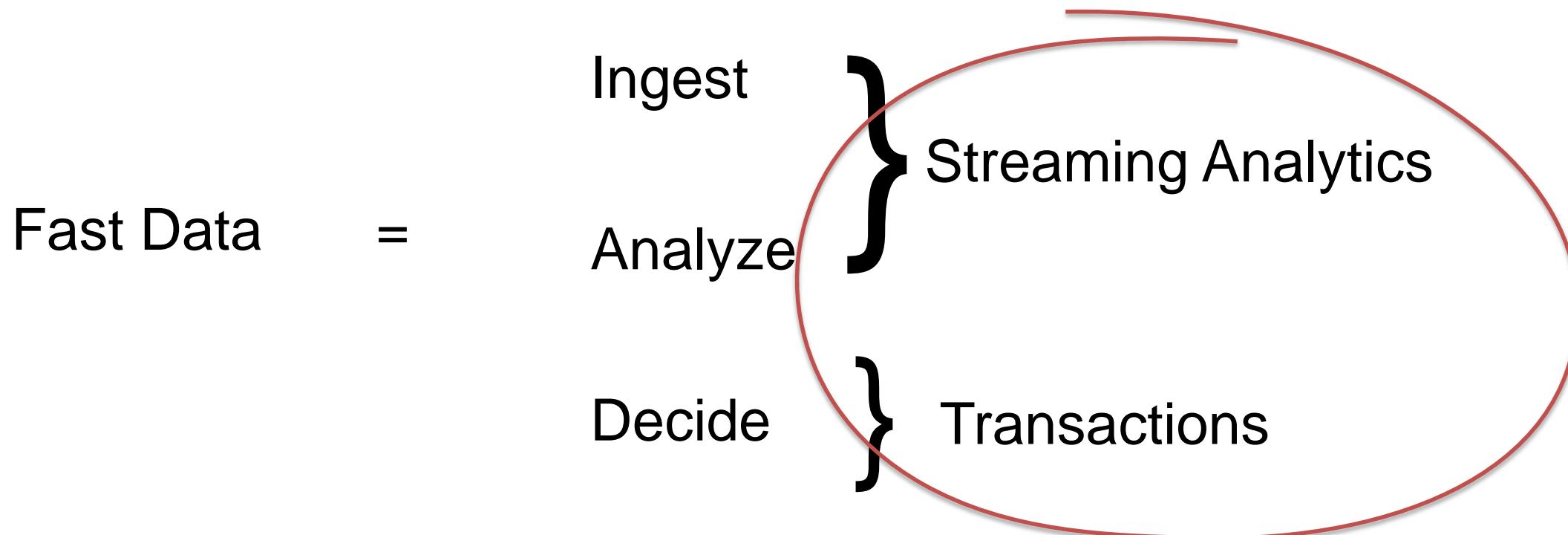


VoltDB is really different than everything else

VOLTDB: A BEAUTIFUL ARCHITECTURE

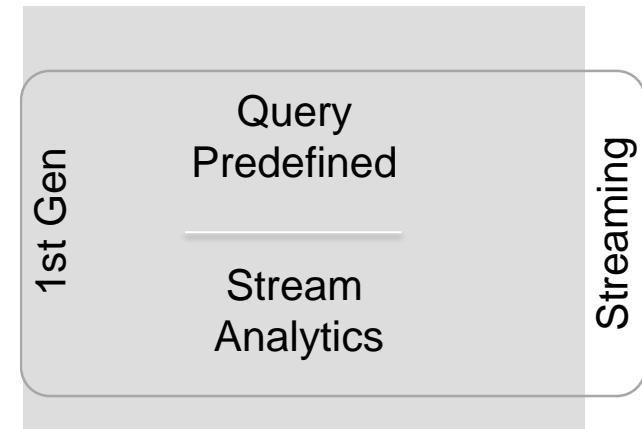


THE MODERN OPERATIONAL APPLICATION



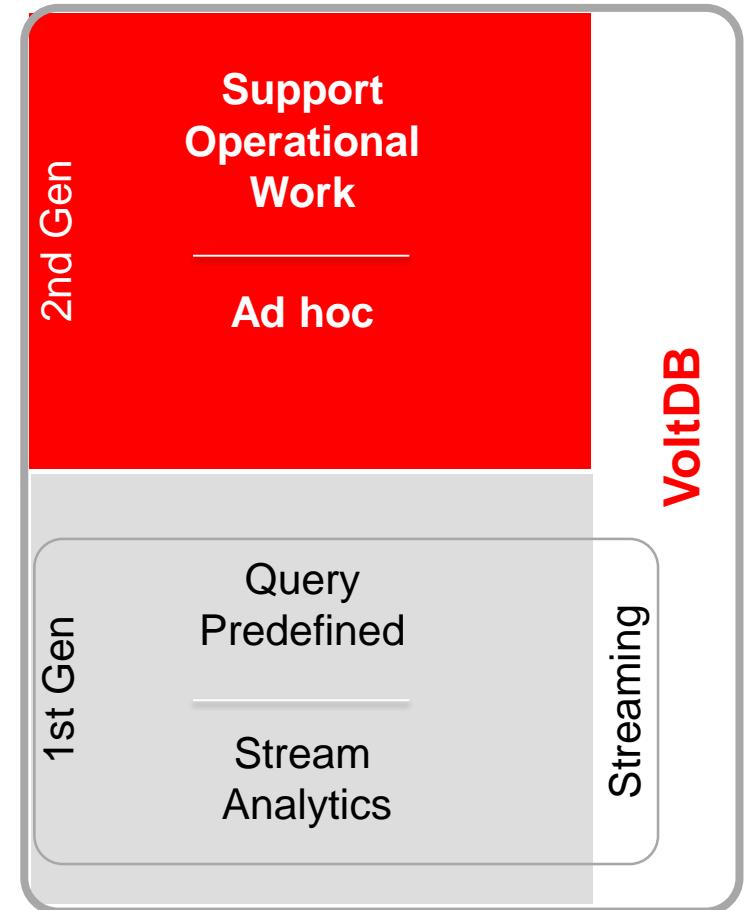
1ST GENERATION FAST DATA: STREAMING ANALYTICS

- Examples: Spark Streaming, Storm, Kinesis, Tibco Streambase, et al
- Technical:
 - Lack “state” for transaction processing (operational)
 - Complex programming model
 - No ability to do ad hoc queries
- Functional:
 - 1st Gen only offers streaming analytics
 - Separate database required for any meaningful work
 - Proprietary interface is inconsistent with the rest of the data pipeline
 - Does not support applications requirement for interaction



2ND GENERATION FAST DATA: STREAMING ANALYTICS & OPERATIONAL WORK

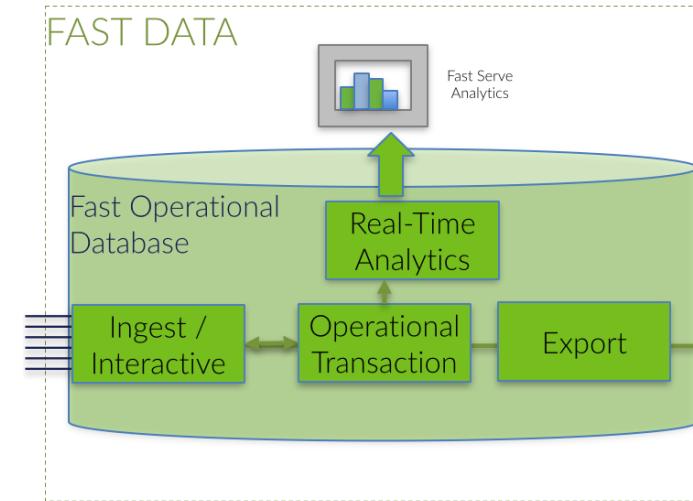
- Streaming Analytics converges with the operational applications
 - Convergence is necessary to use data in real-time
 - Automated application interactions are informed by data
 - Brings the application into the “data analytics” world
- Streaming Analytics alone is *passive*, Fast Data is *interactive*



VOLTDB'S SOLUTION FOR FAST DATA

Single System

- Operational DB (high speed)
- Streaming Analytics

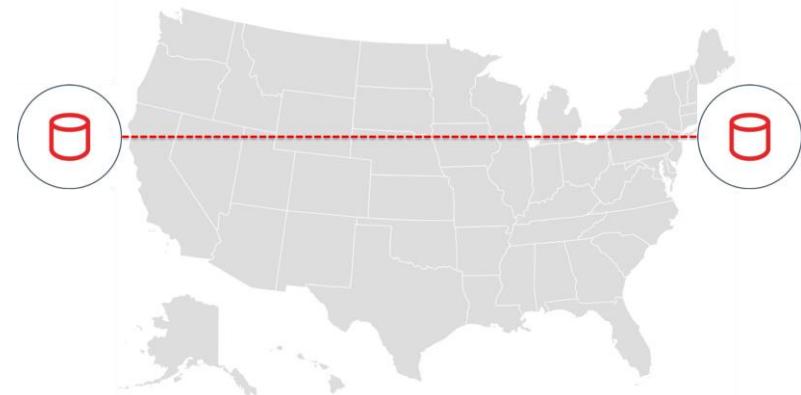
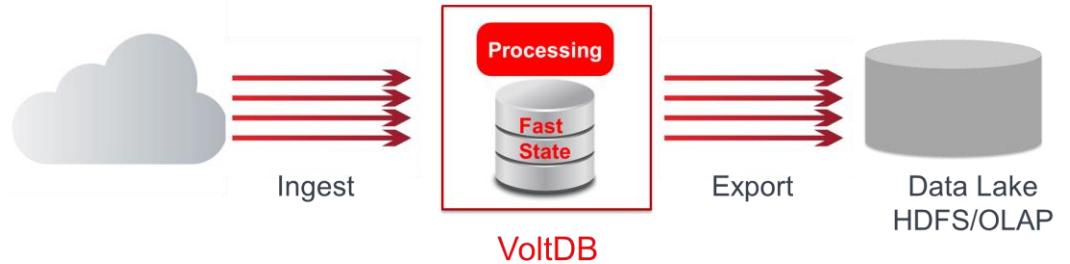


Approached the problem from a database perspective

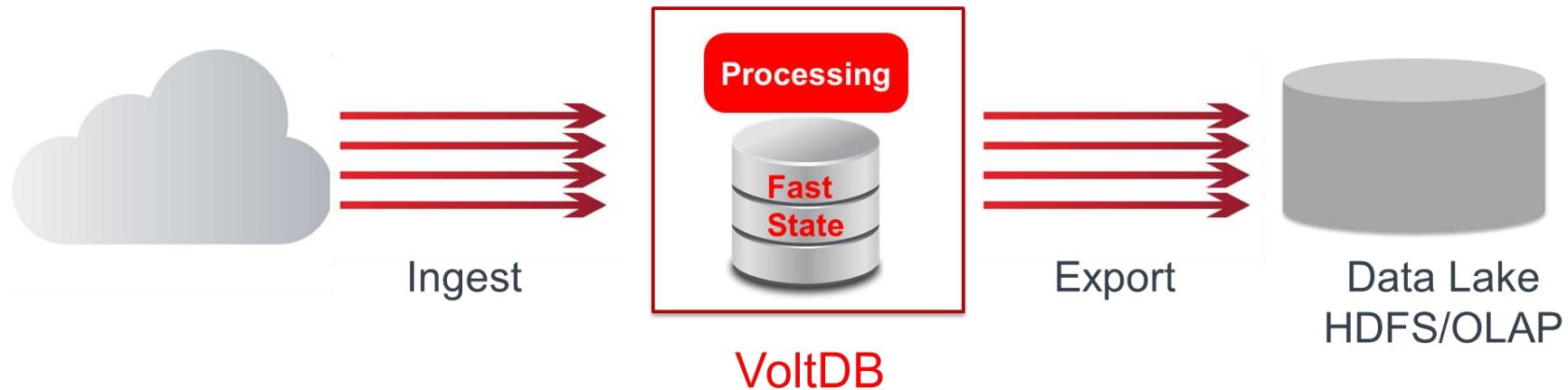
- Same way low speed apps have always solved the problem
- Performance scales to streaming levels (millions of tps)
- Fully Integrated with Big Data Ecosystem
- Import and Export are core functionalities

VOLTDB 5.6

- Fast Data Pipeline
- Distributed Database Replication

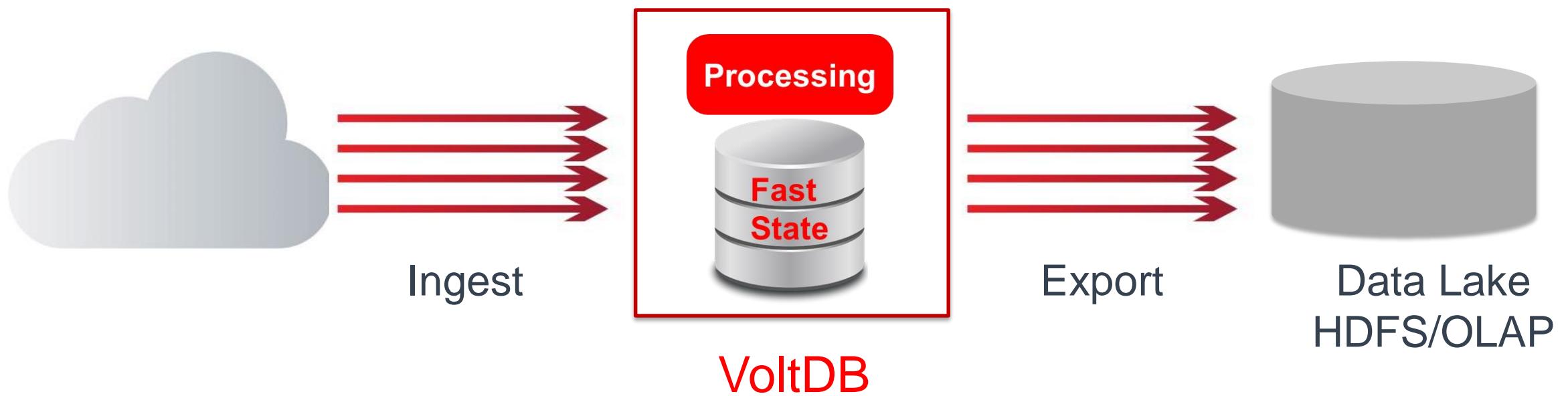


FAST DATA PIPELINE



VoltDB takes over the distributed system complexity of building high-speed ingestion fast data pipelines

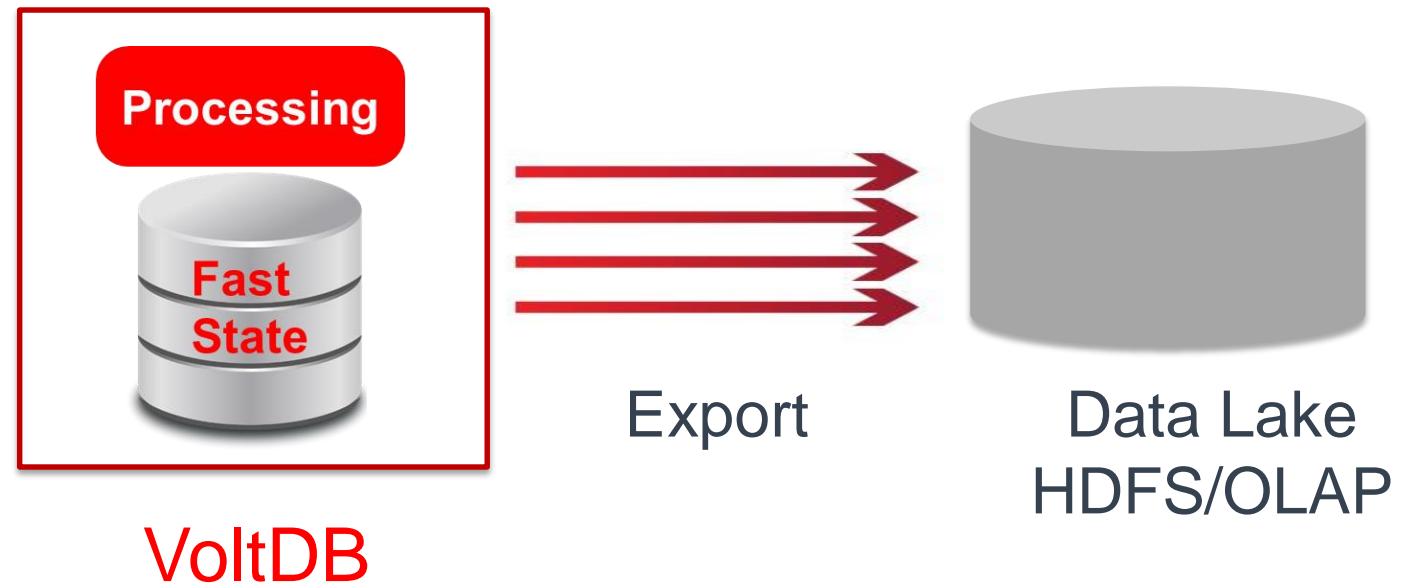
FAST DATA PIPELINE



FAST DATA PIPELINE

Customer-Facing

- Personalization
- Customer experience



Operations-Facing

- Network optimization
- API monitoring
- Sensors

FAST DATA PIPELINE

Customer-Facing

- Personalization
- Customer experience



Operations-Facing

- Network optimization
- API monitoring
- Sensors

Streaming Analytics

- Filtering
- Windowing
- Aggregation
- Enrichment
- Correlations

+

Transactions

- Context-aware
- Personal
- Real-time

VoltDB

Export



Data Lake
HDFS/OLAP

FAST DATA PIPELINE

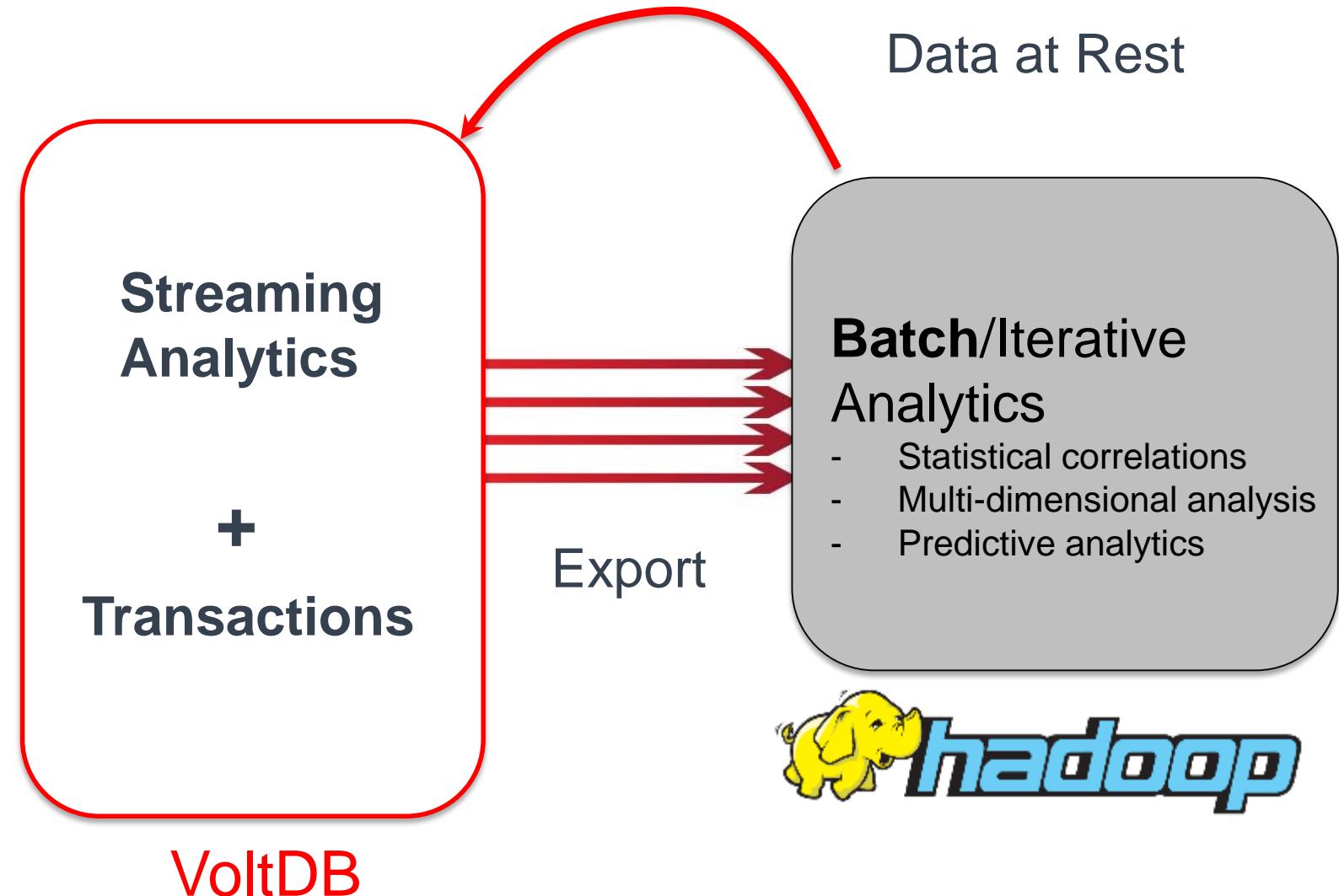
Customer-Facing

- Personalization
- Customer experience



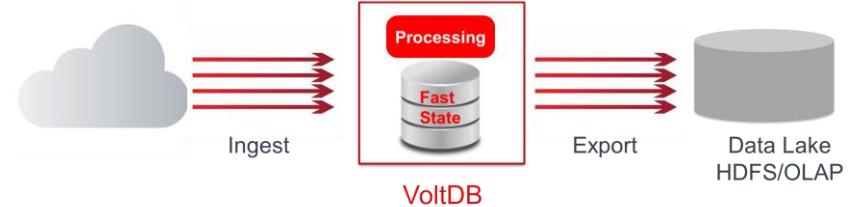
Operations-Facing

- Network optimization
- API monitoring
- Sensors



FAST DATA PIPELINE

What's New?

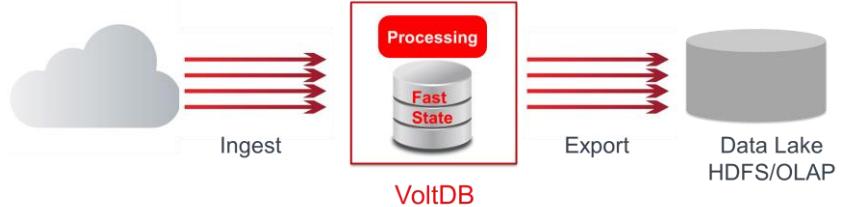


Simpler, faster connections to your data pipeline infrastructure

- New Kafka Importer
 - Without writing code, import multiple Apache Kafka streams; built-in, distributed, and fault tolerant.
 - Import different Kafka topics from a single set of brokers, or from different brokers, to separate tables or stored procedures.
- Fast, synchronized export
 - Supporting the latest Kafka release, synchronized export avoids the connector hanging if the Kafka infrastructure becomes unresponsive

FAST DATA PIPELINE

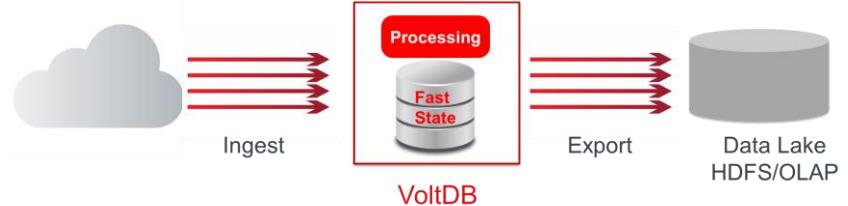
What's New?



Simpler, faster connections to your data pipeline infrastructure

- **Multi-stream import *and* export**
 - **Import** multiple streams to different tables or stored procedures
 - Import different Kafka topics from a single set of brokers, or from different brokers, to separate tables
 - **Export** data to multiple targets simultaneously, e.g.,
 - Export de-duped sensor data to Hadoop once it has been processed
 - “Export” alerts regarding unusual events to HTTP for distribution via SMS, email, or other notification service

FAST DATA PIPELINE



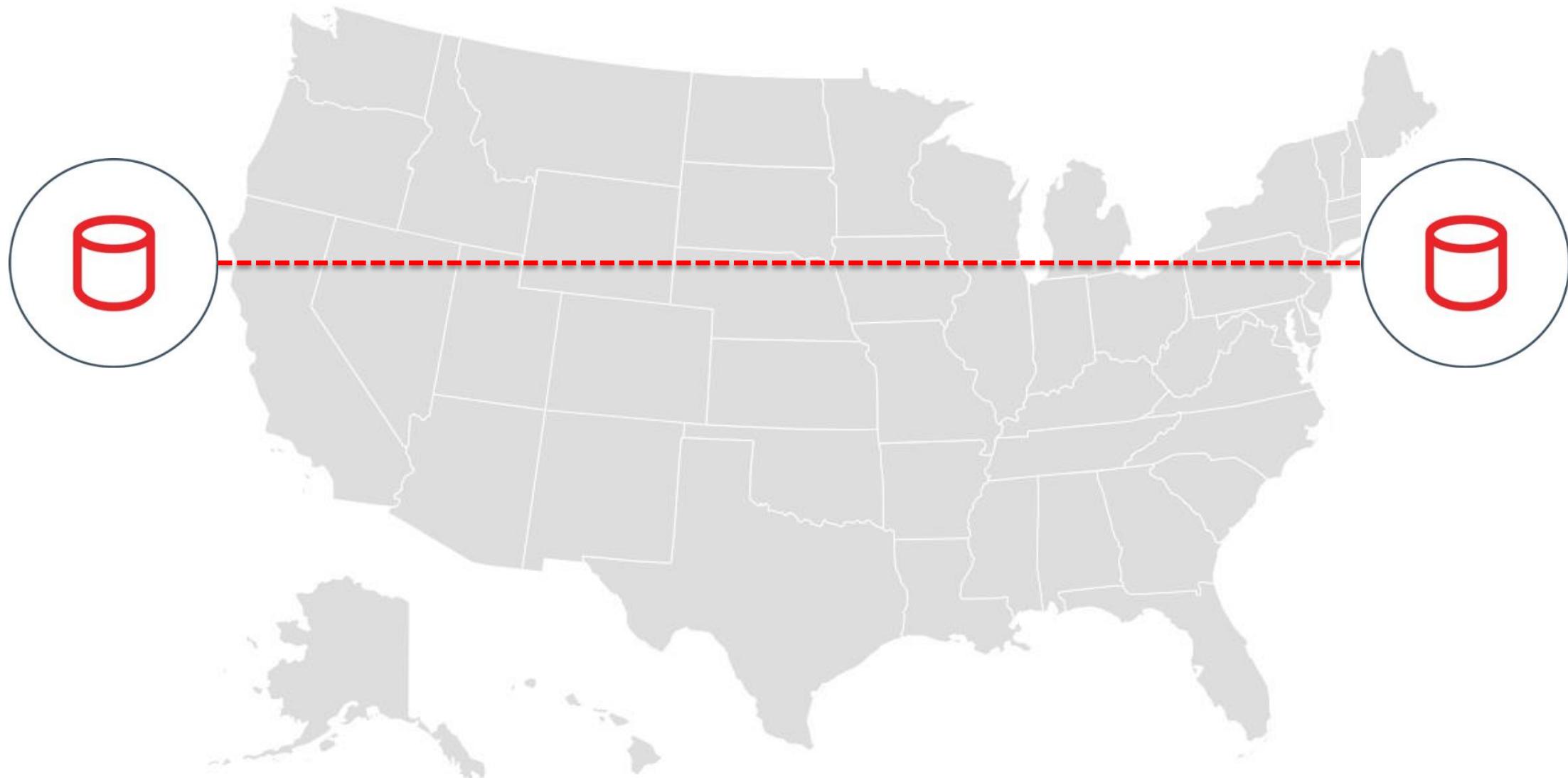
What's New?



Simpler, faster connections to your data pipeline infrastructure

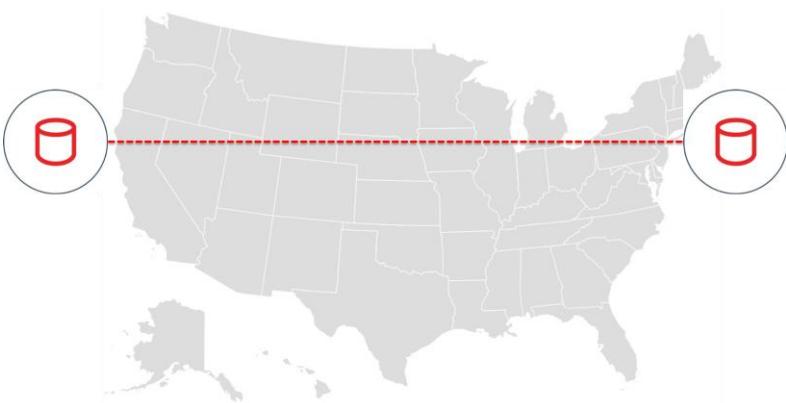
- Elasticsearch Export Connector
 - Receives serialized data from the export tables and inserts it into an Elasticsearch server or cluster
- Export data to Elasticsearch to perform full-text searches on VoltDB data more flexibly and faster.

DATABASE REPLICATION



DATABASE REPLICATION

What's New?



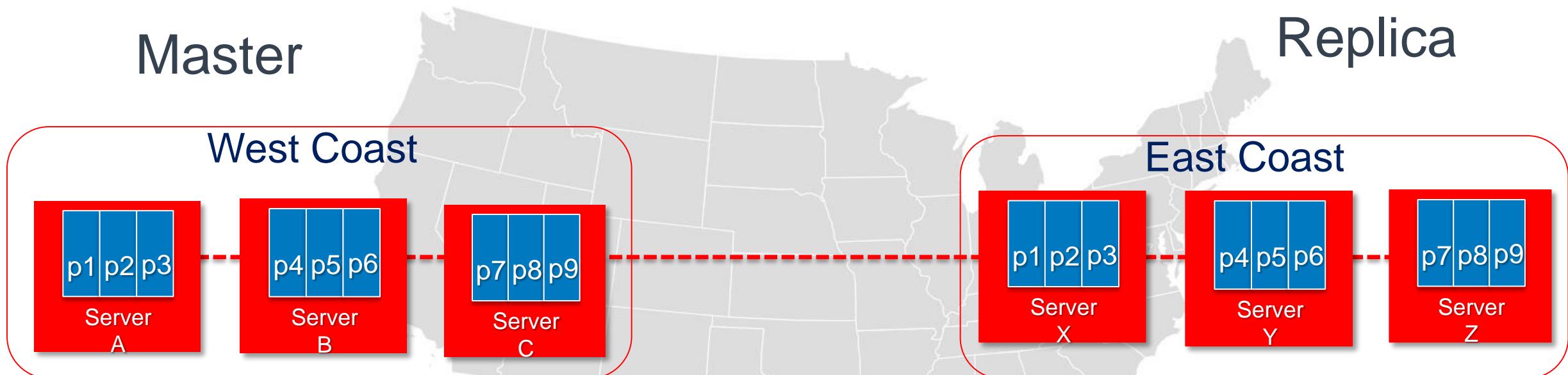
Active/Active database replication for mission critical applications

- Enables multiple database instances, including geographically distributed instances, to support the same application
- Scalable, with no single point of failure, and supporting binary log-based disaster recovery (DR)

Uses

- “Hot standby” to improve uptime in case of failure
- Protecting against catastrophic events -> disaster recovery
- Offload read-only workloads, such as reporting

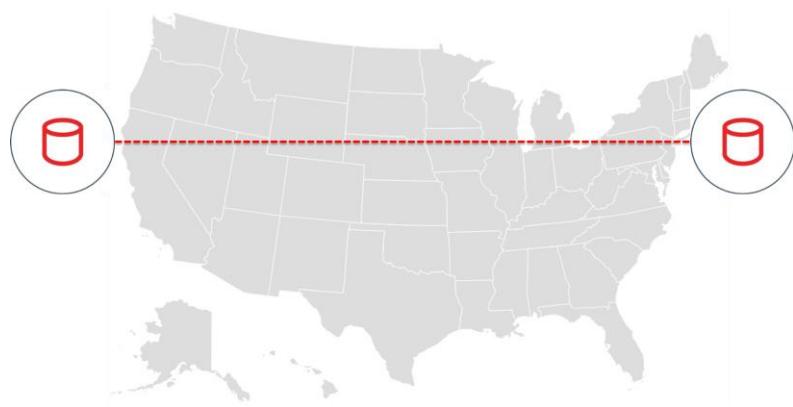
DATABASE REPLICATION



The VoltDB advantage: **Replication by partition**

- Faster - parallel processing improves performance
- More durable – individual partition streams can be redirected without interfering with the replication of the other partitions

DATABASE REPLICATION

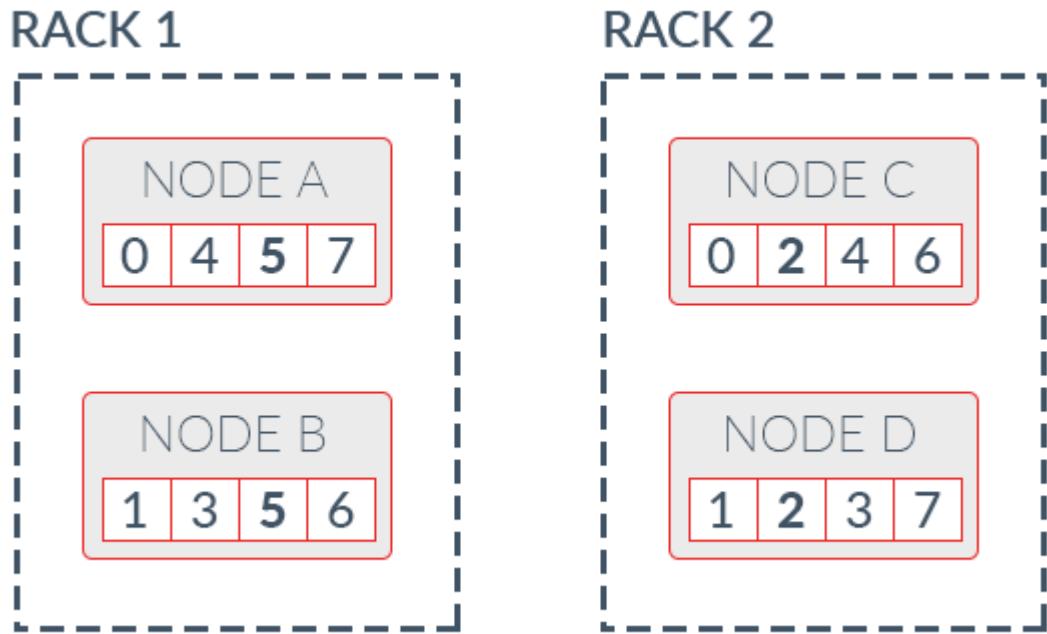


What's New?

Rack-Aware Partitioning to improve availability

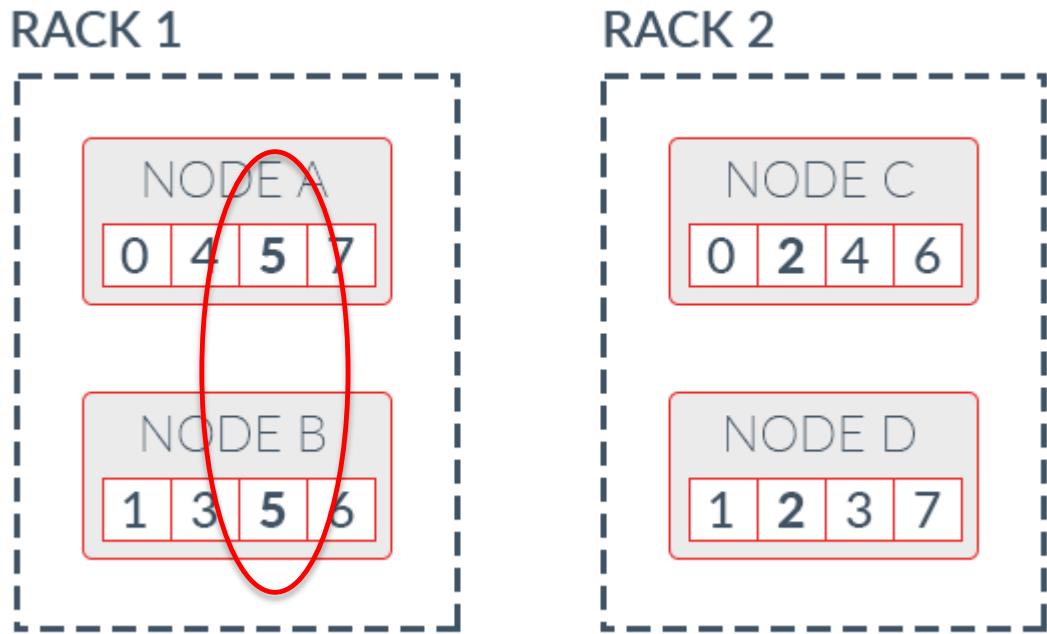
- Partition database replicas with rack-awareness to prevent having replicas on the same physical rack, thereby improving availability in the event of a rack, chassis or power failure.

RACK-AWARE PARTITIONING



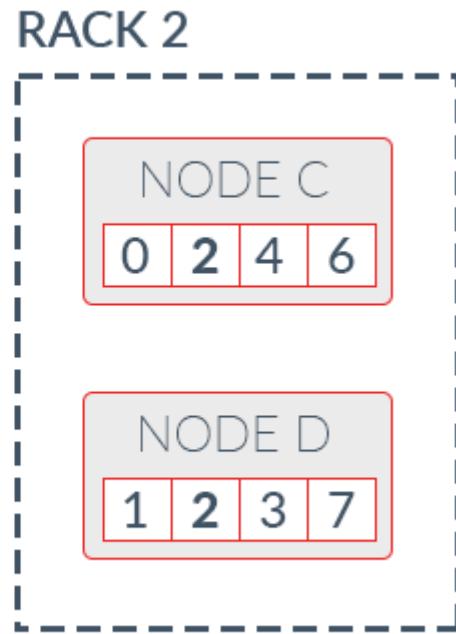
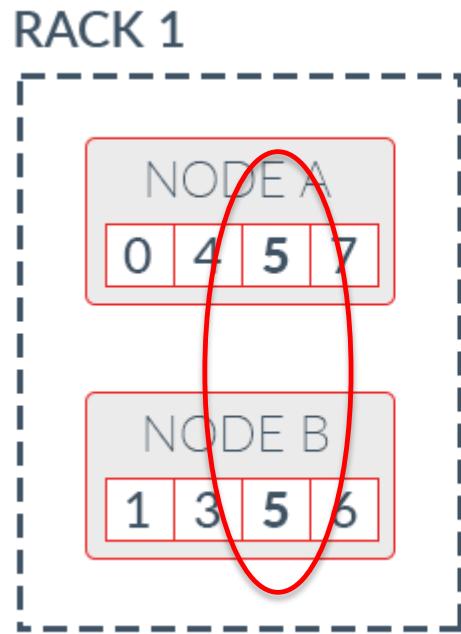
Old
“All Nodes are Equal”

RACK-AWARE PARTITIONING

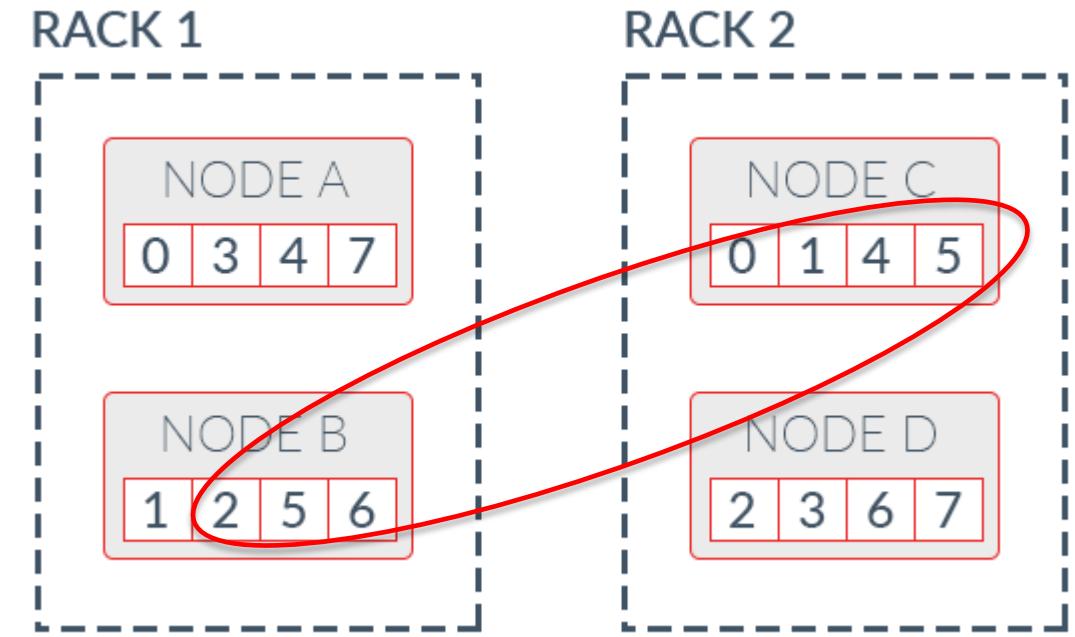


Old
“All Nodes are Equal”

RACK-AWARE PARTITIONING



Old
“All Nodes are Equal”



New
“Rack-aware”
Replicas assigned to different racks

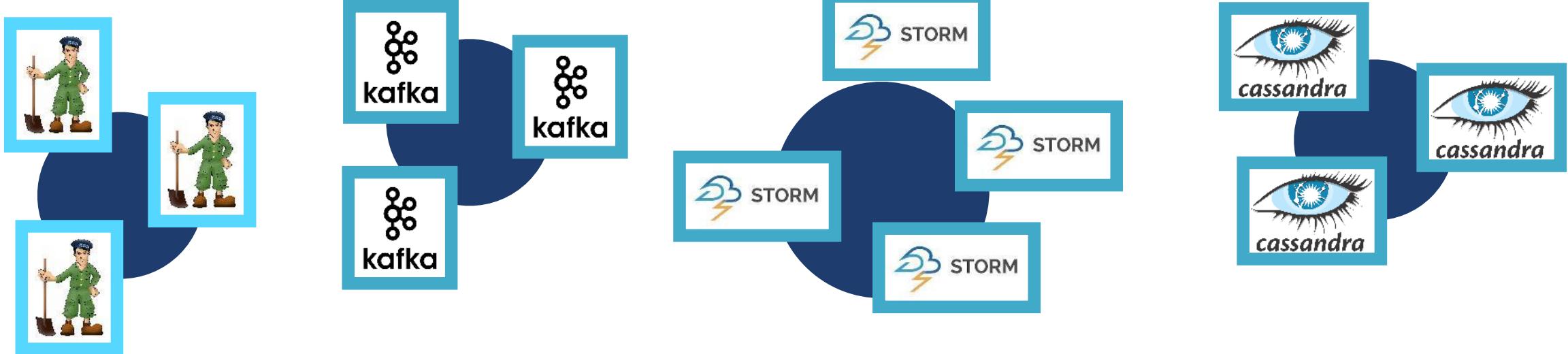


SMART DATA FAST.™



USE CASES

DIY DISTRIBUTED DATA INFRASTRUCTURE?



How to ensure data consistency, accuracy?

SIMPLIFYING THE LAMBDA ARCHITECTURE



Use Case

- Content delivery network service provider
- Counting content views

Why VoltDB?

- Real-time analytics+ transactions w/scale
- Replaced Storm, Cassandra with VoltDB for real-time streaming aggregations with “exactly once” semantic

Bottom line

- Accurate – guaranteed correct results with VoltDB’s ‘exactly-once’ semantics
- Faster time to market
- 32 TB of data processed with 7 servers
- 1/10th the resources of the alternatives

Real-Time Event Decisioning



Use Case

- The Emagine real-time event decision making platform for Communications Service Providers (CSPs)

Why VoltDB?

- Real-time analytics+ transactions
- Scale - billions of network events per day, analyzing hundreds of thousands of transactions simultaneously, and then intelligently interacting with customers

Bottom line

- 3 ms system response time
- 253% increase in offer purchases

EAGLE INVESTMENT SYSTEMS, BNY MELON

The screenshot shows the homepage of Eagle Investment Systems. At the top, there's a navigation bar with links for LOGIN, UNITED STATES - ENGLISH, and SEARCH EAGLE. Below the navigation is a banner with the text "EAGLE INVESTMENT SYSTEMS" and a "WINNER" badge. The main title "Eagle Investment Systems" is prominently displayed. A banner for the "data management summit Awards 2014" highlights that Eagle Wins "Best Fund Accounting and Data Platform". Below the banner, there are sections for "YOUR ROLE...", "YOUR MARKET...", and "YOUR ASSETS UNDER MANAGEMENT". A quote from Mal Cullen, Head of the Americas and Eagle ACCESS™, emphasizes flexibility, scalability, efficiency, accuracy, integrity, and cost effectiveness. A photo of Mal Cullen is also present.

Use Case

- Financial portfolio management and performance measurement
- Over 160 clients, supporting over \$21 trillion in assets
- Performance tracking and risk management

Why VoltDB?

- Performance, scalability to meet application workloads and client SLAs
- High speed data cache for risk calculations with large and rapidly changing data sets

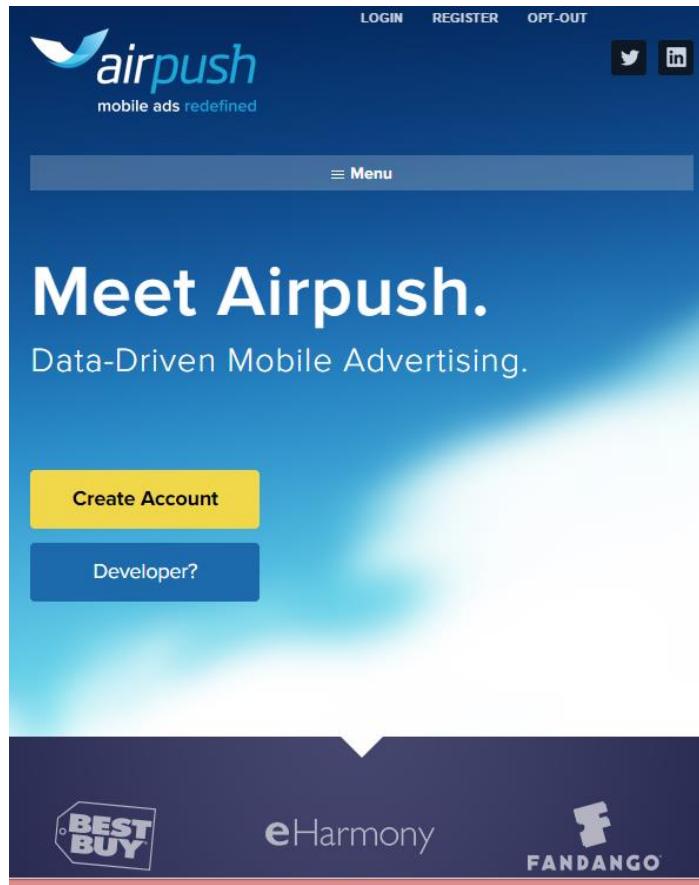
Bottom Line

- Lower TCO

"We plumb Eagle data with third party analytic engines with VoltDB as part of our overall architecture to collect, manufacture, validate and then distribute this information as fast as possible, as fast as the operations could support it."

Marc Firenze, CTO

AIRPUSH



Use Case

- Managing online mobile advertising
- Over 120,000 live applications

Why VoltDB?

- Replaced costly MySQL infrastructure with scalable VoltDB cluster

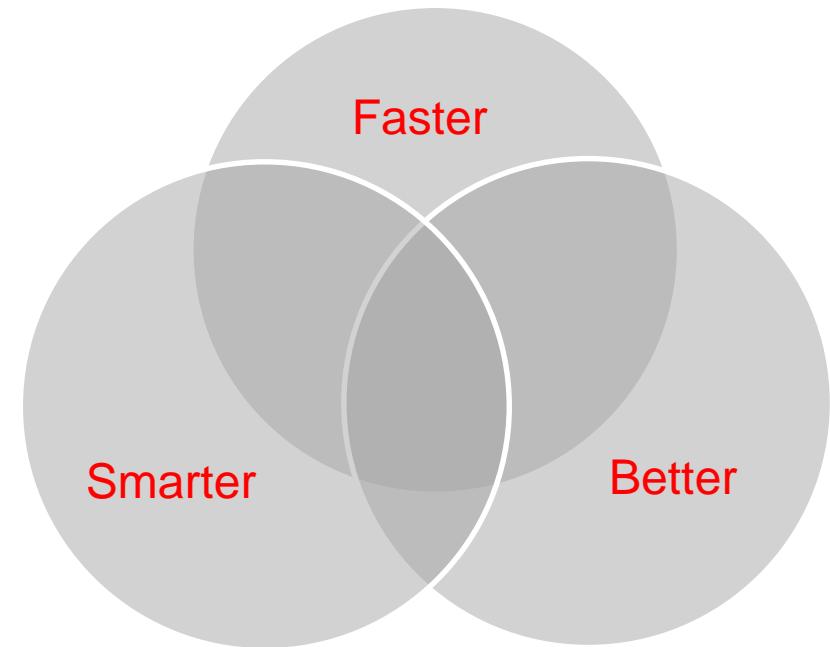
Bottom line

- Enabled accurate ad-campaign balance tracking, dramatically improving “last-dollar” decisions
- Eliminated the opportunity cost of placing wrong ads
- Reduced infrastructure cost by 93% (7 servers vs. 100)

“Achieved a previously impossible level of ad budget management accuracy”

WHY VOLTDB?

- Superior architecture for fast data/translytics
 - In-Memory, Scale-out, ACID, SQL+JSON
 - Rapid data ingestion with transactions
 - Data durability and HA
- VoltDB customers realize exceptional business value



QUESTIONS?

- Use the chat window to type in your questions
- Try VoltDB yourself
 - www.voltdb.com/Download
- Download our new ‘recipes’ eBook
 - <https://voltdb.com/ebook/fast-data-recipes>
- Email us your questions:
askanengineer@voltdb.com

**WE'RE
HIRING!**

careers@voltdb.com

