SMART DATA FAST.™ VOLTDB

# VOLTDB

# FAST DATA – THE NEW BIG DATA

# OVERVIEW

- Trends
- Fast vs Big
- Approaches
- Use Cases

© 2015 VoltDB

# DATA-FICATION OF LIFE

## The 10 Trillion Device World

"Smartness can be embedded everywhere," said Professor Sangiovanni-Vincentelli, EE/CS at University of California at Berkeley.

"The entire environment is going to be full of sensors of all kinds. Chemical sensors, cameras and microphones of all types and shapes. Sensors will check the quality of the air and temperatures. Microphones around your environment will listen to you giving commands.“

Computerworld, September 2015

## Fast Data

## Big Data

All data originates as fast data,
why wait to analyze and act on it?

# FAST = ADVANTAGE

© 2015 VoltDB

OPENET
transactional intelligence

“Real-time” contextual offers

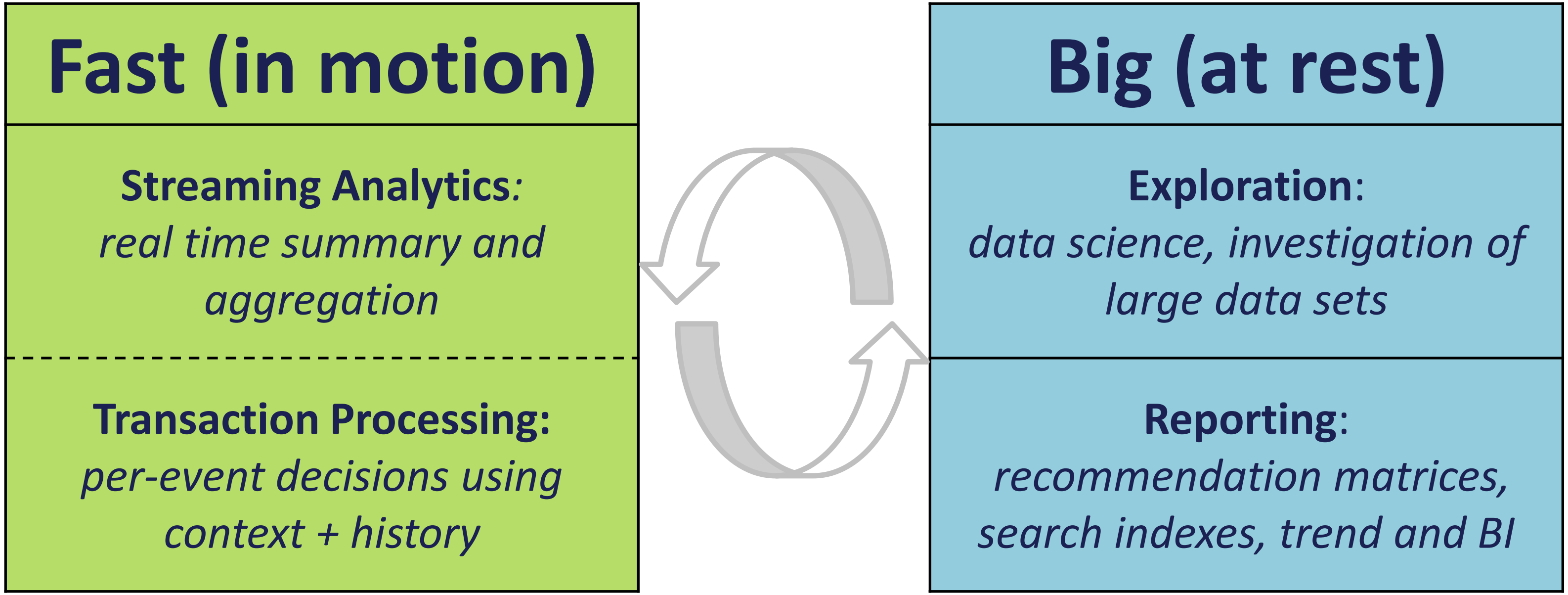
=

↑ offer uptake rates 75%

↑ data revenues by 15%.”

Source: Openet 2014 survey of 87 mobile operators

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



## Fast (in motion)

**Streaming Analytics***:*
*real time summary and aggregation*

**Transaction Processing:**
*per-event decisions using context + history*

## Big (at rest)

**Exploration**:
*data science, investigation of large data sets*

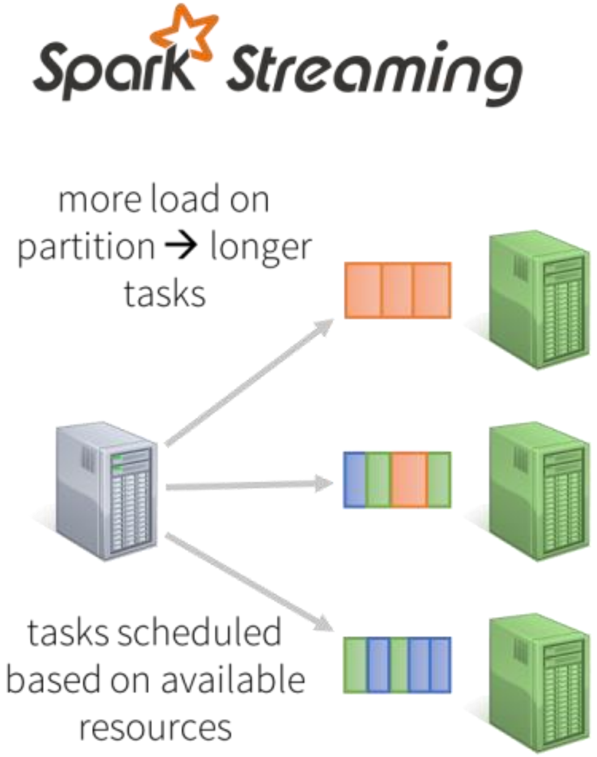
**Reporting**:
*recommendation matrices, search indexes, trend and BI*

SMART DATA FAST.™ VOLTDB

# APPROACHES

# IN THE BEGINNING THERE WAS BATCH….

- Collect data, process it (used to be overnight), produce a report (output)
  - If batch job fails, delete the data, and start over
- Distributed systems made this better, more efficient
- Challenges
  - Response time (latency)
  - Processing events in order

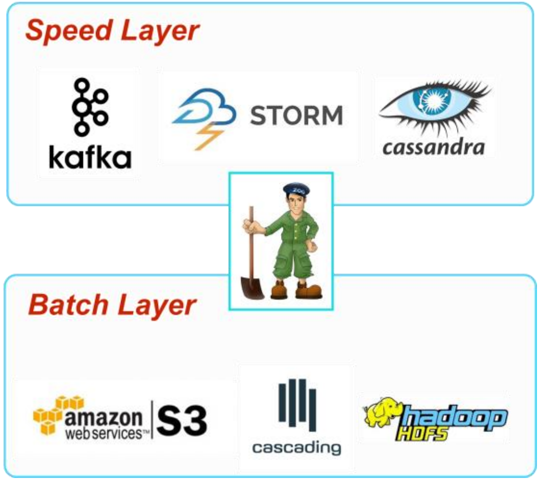

Spark Streaming
more load on partition → longer tasks
tasks scheduled based on available resources

© 2015 VoltDB

# NOSQL AND “EVENTUALLY CONSISTENT” SOLUTIONS

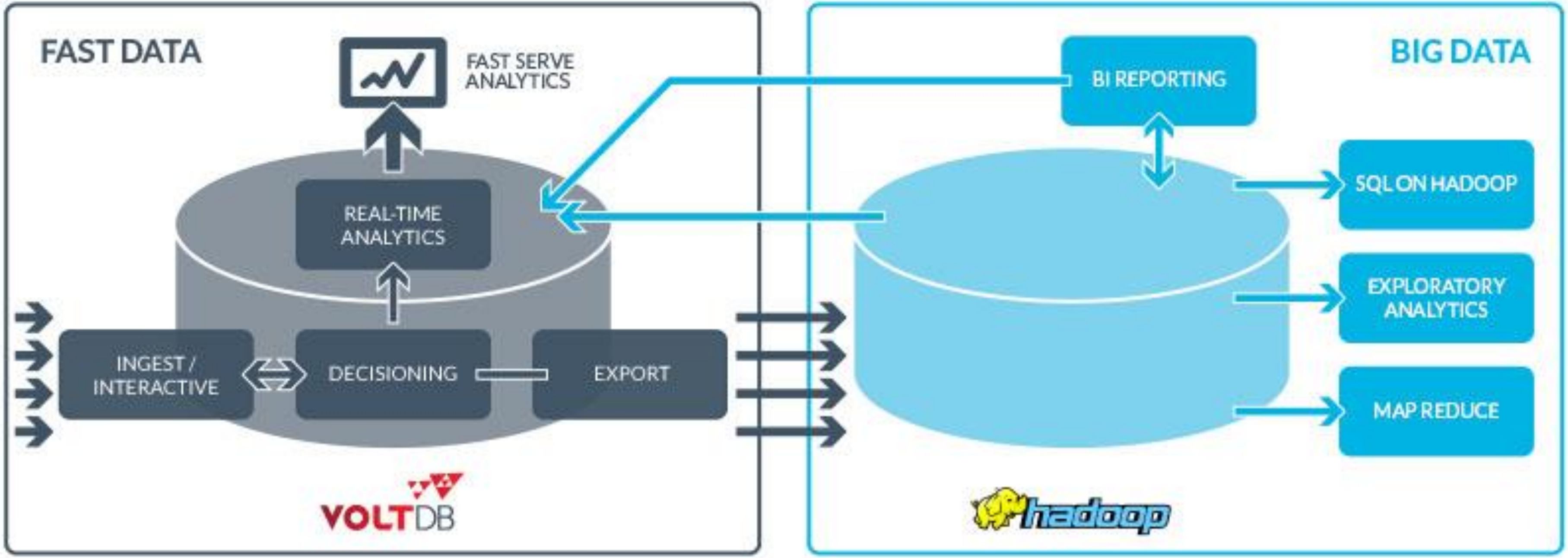
- Combine stream processing frameworks with NoSQL DBs
- Challenges
  - DiY requires building in reliability, code for ‘book keeping’ to ensure accuracy
  - Response time/latency goes up as components are added
  - Failure modes

Lambda Architecture



NEW
PARADIGM
AHEAD

# NEW ENTERPRISE ARCHITECTURE: FAST + BIG



# ARCHITECTURE IS IMPORTANT….

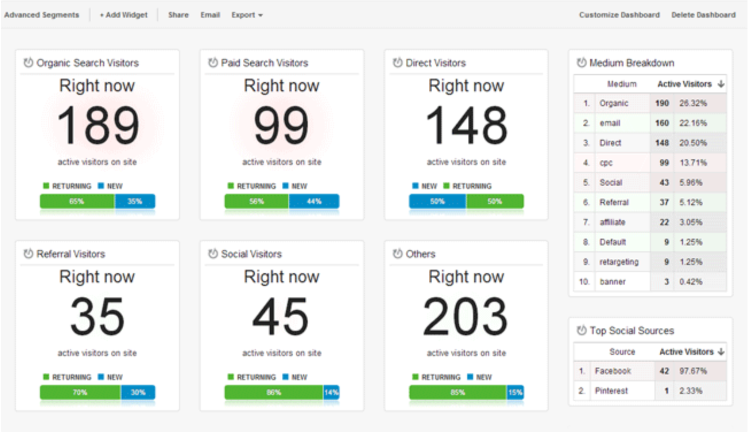
Fast data requires a different architecture.

© 2015 VoltDB

# STREAMING ANALYTICS

**What:**
Filter, aggregate, enrich, and analyze a high throughput of data from live data sources

**Why:**
To identify patterns, detect urgent situations, and automate immediate actions in real-time



Organic Search Visitors
Right now
189
active visitors on site
🟩 RETURNING 🟦 NEW
65% 35%

Paid Search Visitors
Right now
99
active visitors on site
🟩 RETURNING 🟦 NEW
56% 44%

Direct Visitors
Right now
148
active visitors on site
🟦 NEW 🟩 RETURNING
50% 50%

Referral Visitors
Right now
35
active visitors on site
🟩 RETURNING 🟦 NEW
70% 30%

Social Visitors
Right now
45
active visitors on site
🟩 RETURNING 🟦 NEW
86% 14%

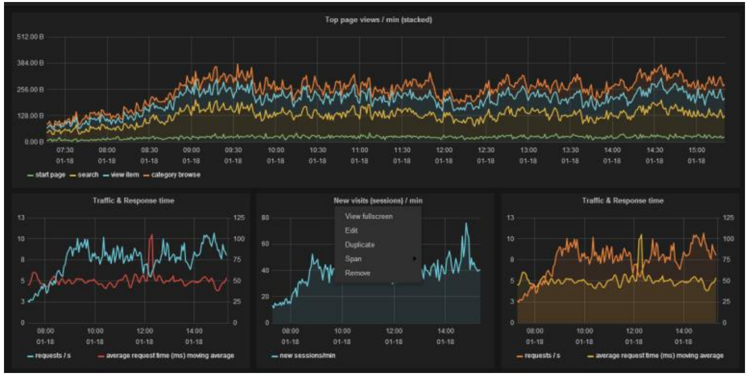
Others
Right now
203
active visitors on site
🟩 RETURNING 🟦 NEW
85% 15%

Medium Breakdown

| | Medium | Active Visitors ↓ | |
|---|---|---|---|
| 1. | Organic | 190 | 26.32% |
| 2. | email | 160 | 22.16% |
| 3. | Direct | 148 | 20.50% |
| 4. | cpc | 99 | 13.71% |
| 5. | Social | 43 | 5.96% |
| 6. | Referral | 37 | 5.12% |
| 7. | affiliate | 22 | 3.05% |
| 8. | Default | 9 | 1.25% |
| 9. | retargeting | 9 | 1.25% |
| 10. | banner | 3 | 0.42% |

Top Social Sources

| | Source | Active Visitors ↓ | |
|---|---|---|---|
| 1. | Facebook | 42 | 97.67% |
| 2. | Pinterest | 1 | 2.33% |

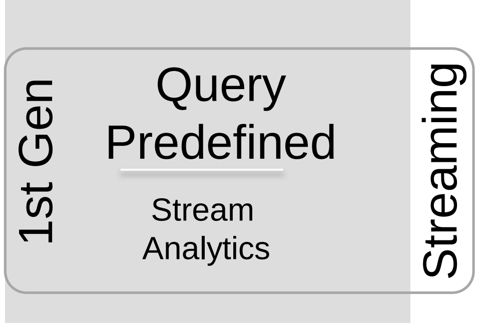


# 1ST GENERATION FAST DATA: STREAMING ANALYTICS

- 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

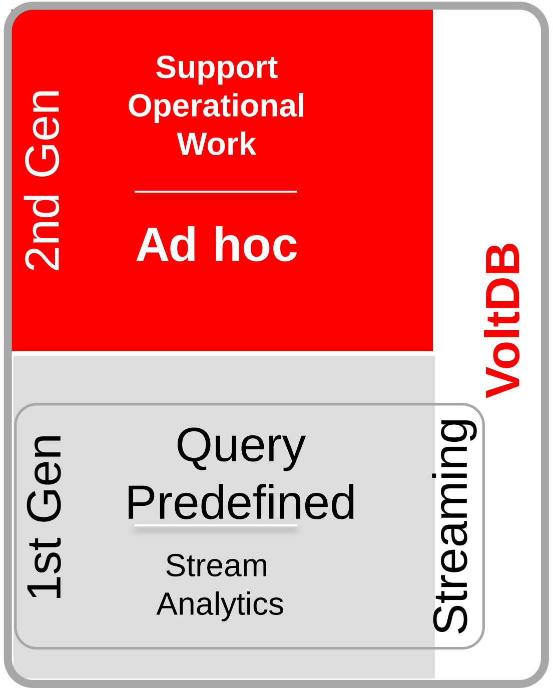
- Functional:
  - 1st Gen only offers streaming analytics
  - Separate database required for any meaningful work
  - Proprietary interface is inconsistent with the rest of the data pipeline
  - Does not support applications requirement for interaction



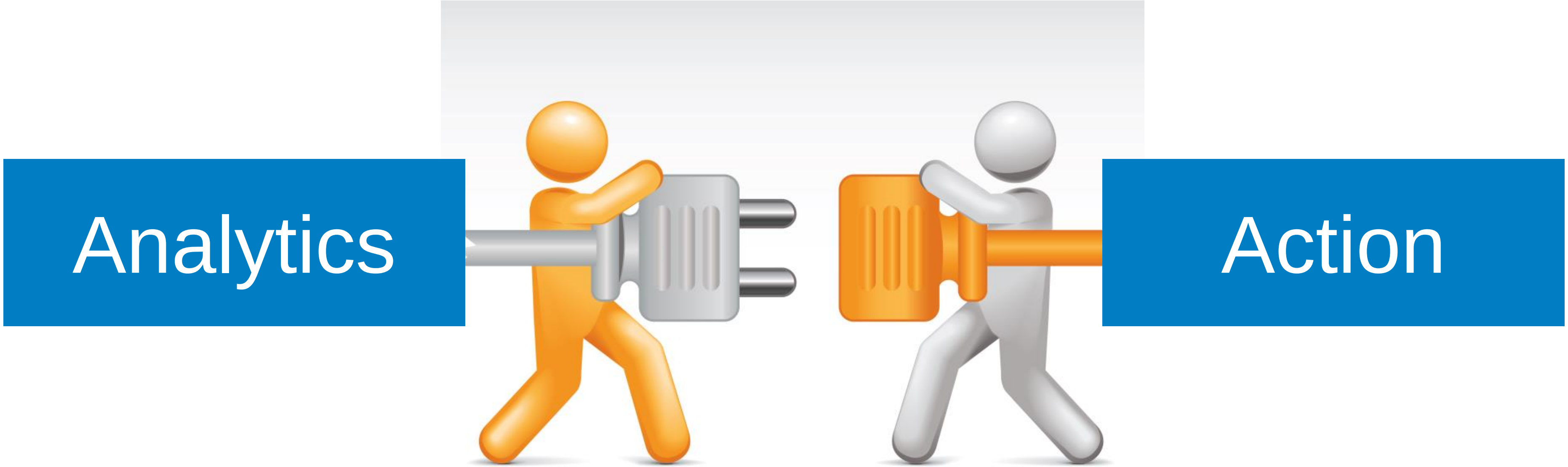
# 2ND GENERATION FAST DATA: STREAMING ANALYTICS & OPERATIONAL WORK

- 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*



# WHAT’S NEW HERE?



Analytics

Action

Combining streaming analytics and transactions allows you to act at the rate that you learn.

# TRANSLYTICAL DATABASES

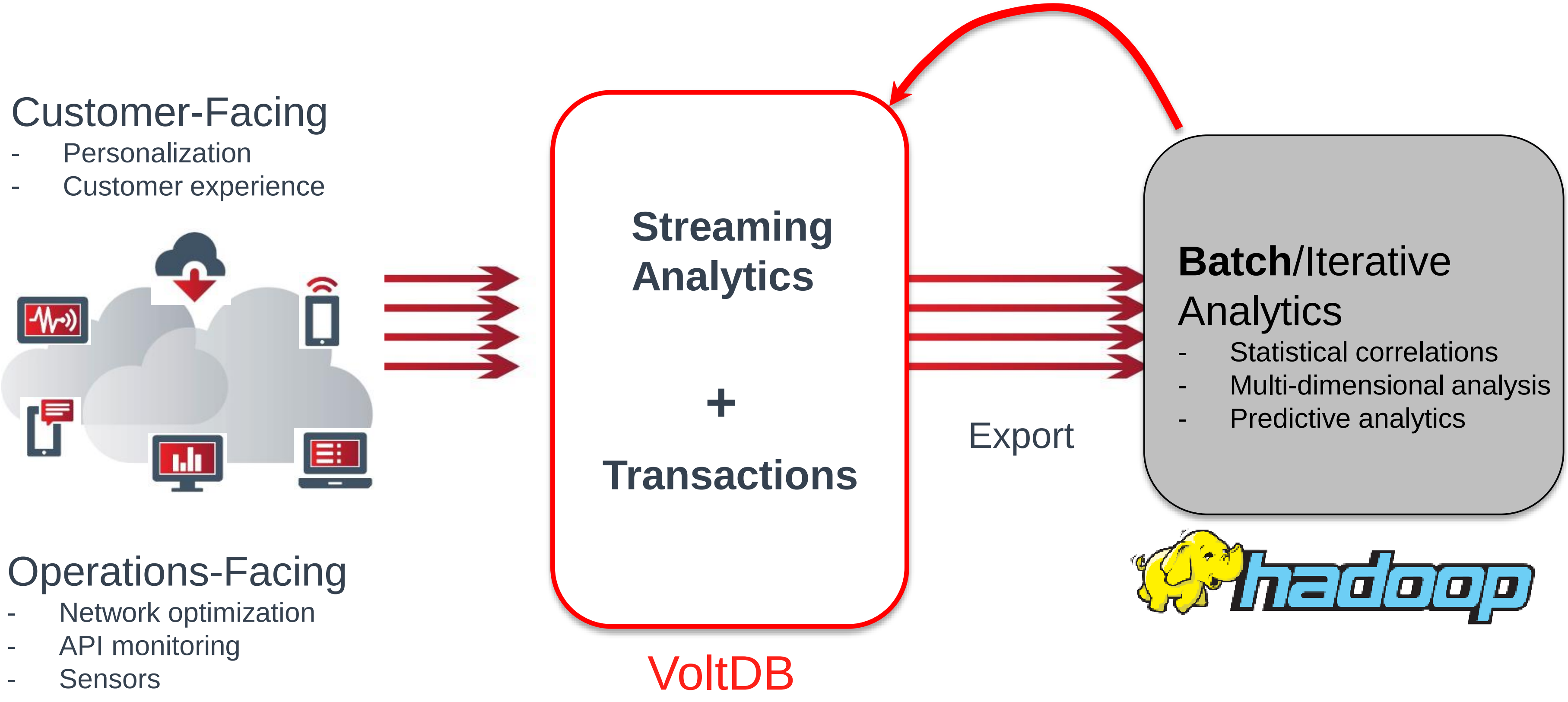
FORRESTER®

“By definition the only way to do streaming analytics is to do it **in-memory**. Don’t make the mistake of thinking that streaming is just about ingestion. Streaming analytics is about *analytics* more than it is about *ingestion*.”

“Spark Streaming is micro batch processing. That’s still batch processing but it does it in micro batches. I don’t consider that a true real-time streaming platform because it’s geared more for batch processing.”

A new category of databases is emerging we call **translytical databases**: streaming analytics with transactions in a single database.

# FAST DATA REQUIRES ANALYTICS WITH (TRANS)ACTIONS



Customer-Facing

- Personalization
- Customer experience

Operations-Facing

- Network optimization
- API monitoring
- Sensors

**Streaming Analytics**

**+**

**Transactions**

VoltDB

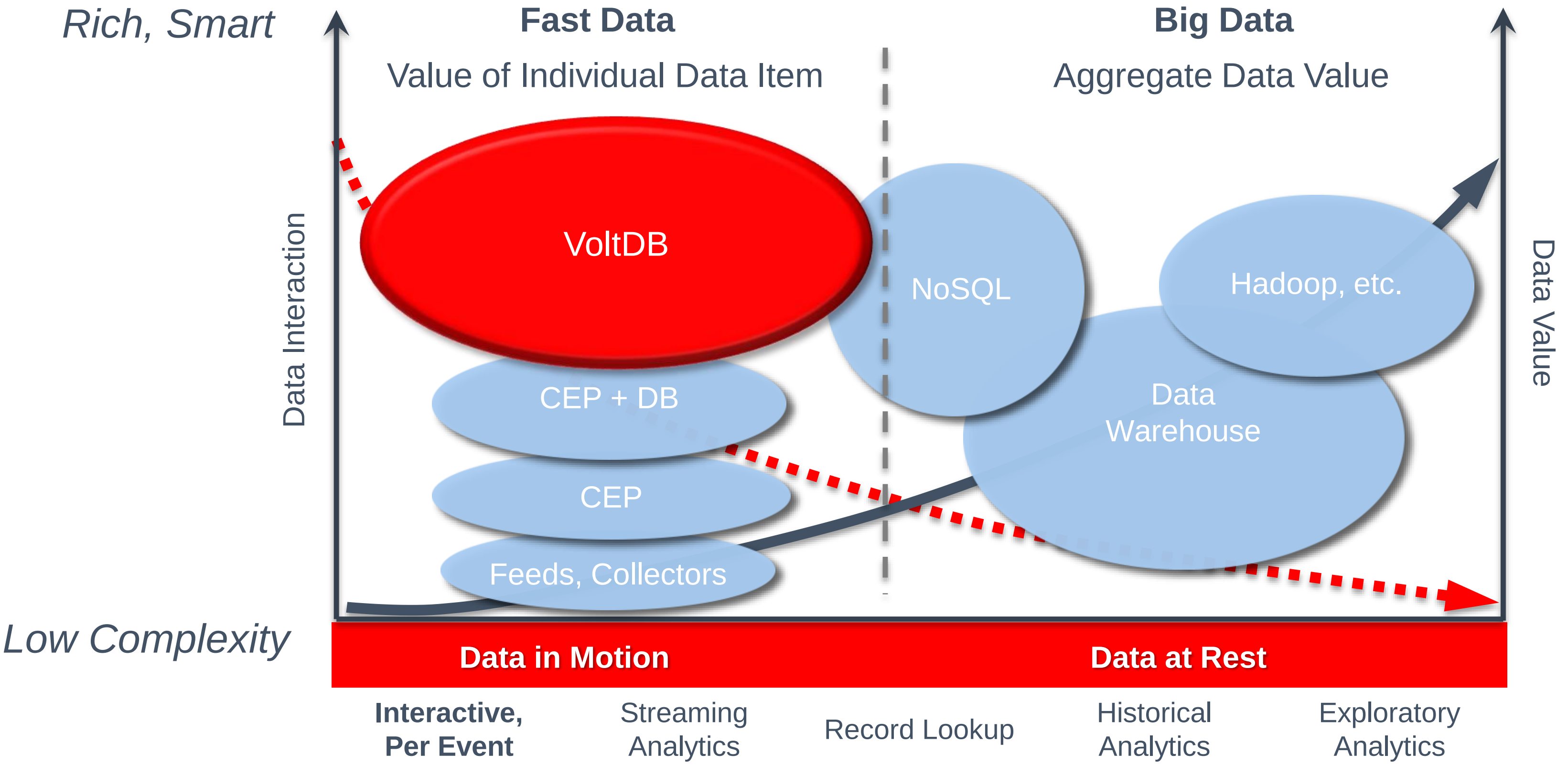
Export

**Batch**/Iterative Analytics

- Statistical correlations
- Multi-dimensional analysis
- Predictive analytics

© 2015 VoltDB

# THE TIME VALUE OF DATA



© 2015 VoltDB

# VOLTDB: A SUPERIOR ARCHITECTURE FOR FAST DATA

- ✓ In-Memory performance
- ✓ Scale-out, shared nothing
- ✓ ACID & SQL & Java
- ✓ Continuous, per event
- ✓ Reliability and fault tolerance
- ✓ Hadoop ecosystem integration

VoltDB is really different than everything else

# THE SO WHAT

VoltDB allows companies to act on data in real-time, enabling new levels of application functionality and performance that drive new revenue streams while reducing infrastructure costs

© 2015 VoltDB

# USE CASES

# USE CASE EXAMPLES: ANALYTICS + (TRANS)ACTIONS

| | Streaming Analytics<br>(Stream Proc. or OLTP) | (Trans)Actions<br>(OLTP) |
|---|---|---|
| Mobile Usage | Count current usage minutes | Will current usage plus previous balance cause the customer to exceed his quota? |
| Gaming | Real-time stats on player effectiveness | Change game interaction to increase engagement of the player |
| Real-time Risk | Determine position values as prices and positions change | Does a new trade violate the defined risk tolerance? If “no,” place trade |
| Ad placement | With which segment is this user identified | Identify ad, check vendor quota balance, determine best network and place ad |
| Content Delivery Service | Count content views | Update log records in real time for accurate billing based on content views |

© 2015 VoltDB

# USE CASES

## Telco

- Subscriber Management
- Session Management
- OSS/BSS – policy, billing, routing
- SLA Management

## Financial Services

- Risk Management (portfolio, trading)
- Fraud Detection
- Compliance (BB&O)
- Customer Engagement

## Media and Entertainment

- Personalization
- Digital Advertising
- Content Delivery
- Gaming

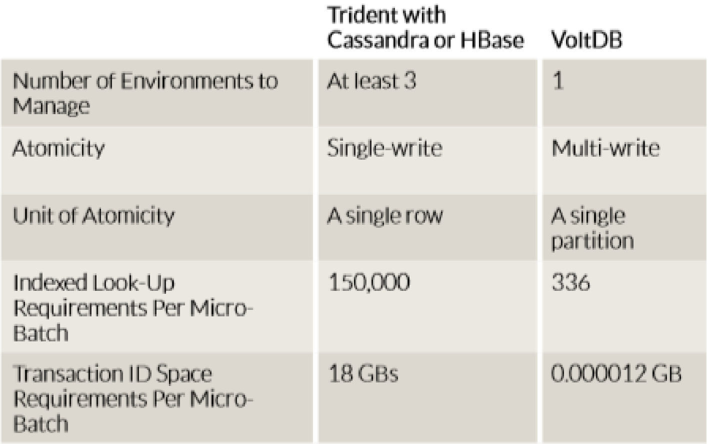
## IoT/Sensors

- Smart Energy
- Connected Home
- Patient Monitoring

# SIMPLIFYING THE LAMBDA ARCHITECTURE

maxcdn

Content delivery network service provider



| | Trident with Cassandra or HBase | VoltDB |
|---|---|---|
| Number of Environments to Manage | At least 3 | 1 |
| Atomicity | Single-write | Multi-write |
| Unit of Atomicity | A single row | A single partition |
| Indexed Look-Up Requirements Per Micro-Batch | 150,000 | 336 |
| Transaction ID Space Requirements Per Micro-Batch | 18 GBs | 0.000012 GB |

*Implementing VoltDB and micro-batching with multi-write atomicity within the Lambda framework simplified management and improved performance, storage, scalability, and operations.*

## Use Case

- Counting “content” views in real time for billing and reporting

## Why VoltDB?

- Real-time analytics + transactions w/scale
- Need for accuracy – chose VoltDB over Trident/Storm+Cassandra combination for real-time streaming aggregations with “exactly once” semantic

maxcdn

MaxCDN uses 1/10th compute resources of alternate solutions.

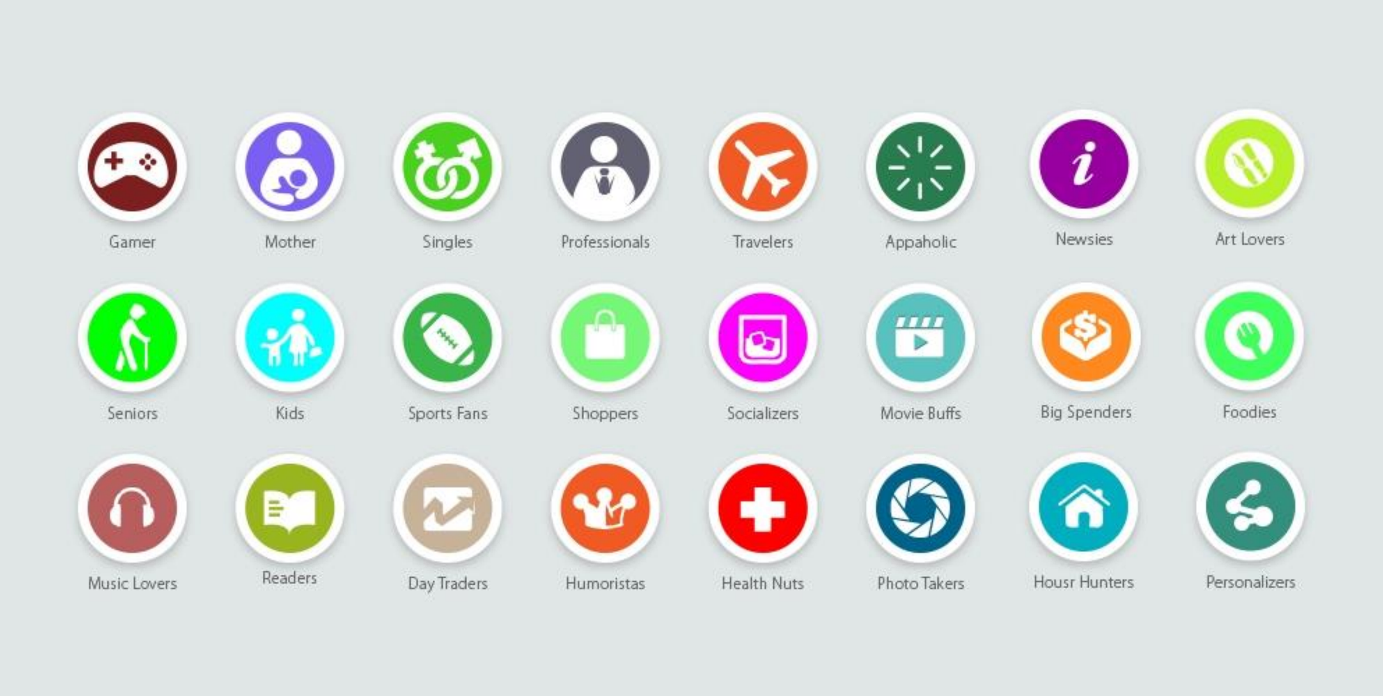
Behzad Pirvali
Performance Architect

airpush

- Mobile advertising service
- Managing over 150,000 applications

## HYPER**TARGET**

Real-Time targeting = f(persona, interests, behaviors)



Requirement: Hundreds of thousands of concurrent connections with round-trip latencies in milliseconds

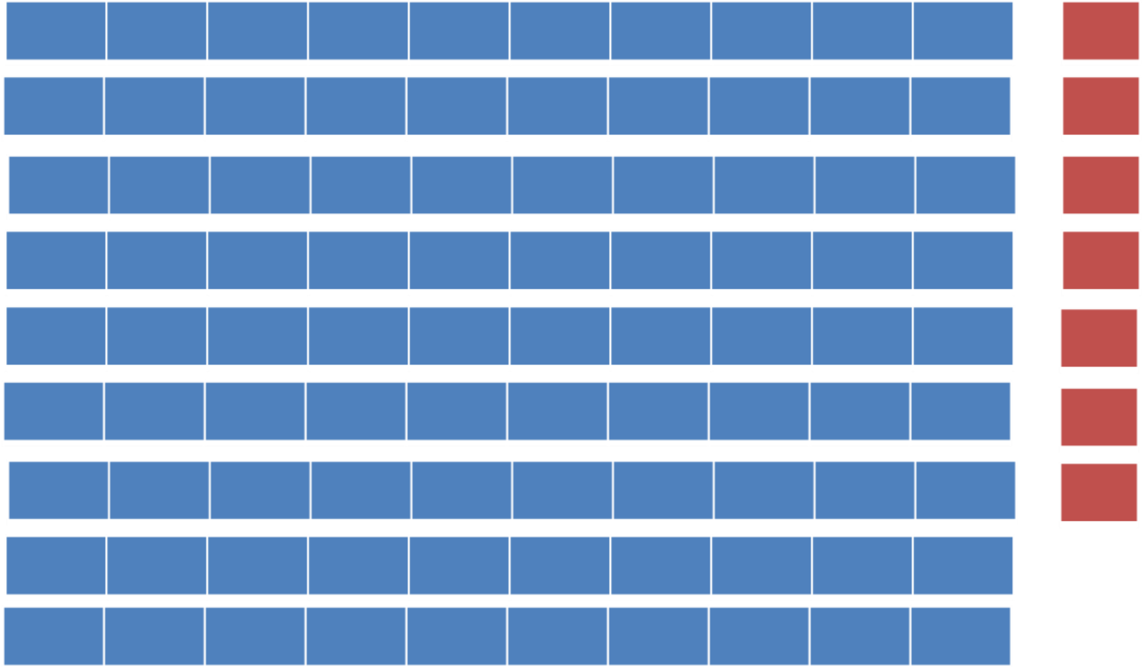
airpush

## Before (MySQL)

100 servers

## After (VoltDB)

7 servers



airpush

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

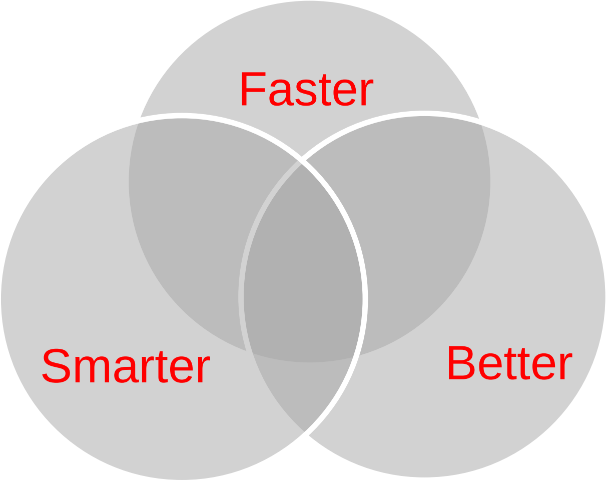
Dan Khasis
Chief Technology Officer

© 2015 VoltDB

# APPLICATIONS BUILT WITH VOLTDB ARE:

- ✓ **Faster, more performant**
  - tps, latency
- ✓ **Simpler**
  - Fraction of components and coding vs. alternatives
  - Lower maintenance and support
- ✓ **Better**
  - Lower system risk
  - Correct results
  - Higher availability and reliability

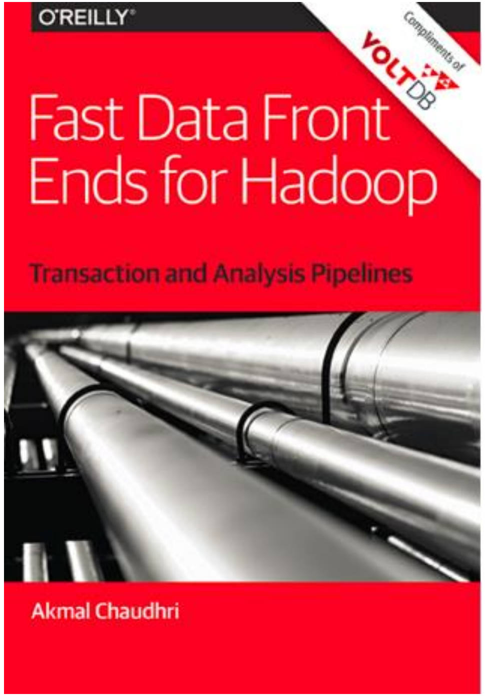
# WHY VOLTDB? Our customers realize exceptional business value



SMART DATA FAST.™

# QUESTIONS?

- Use the chat window to type in your questions
- Try VoltDB yourself:
  - ➢ Free trial of the Enterprise Edition:
    - www.voltdb.com/Download
  - ➢ Open source version is available on github.com
- Use the chat tab to ask your questions.
- Join the conversation on Twitter #VoltDBFastData
- Download our latest report from O'Reilly in the resources window


O'REILLY
Compliments of VOLTDB
Fast Data Front Ends for Hadoop
Transaction and Analysis Pipelines
Akmal Chaudhri