The Internet of Analytics
Discovering actionable insights from high-velocity streams of real-time IoT data

Sami Akbay, Founder and EVP, WebAction
In-Memory Computing Summit | San Francisco, CA | June 2015
Enabling Internet of Things

Cheap and Efficient Data Capture
- Affordable sensors, RFID, antennas, aggregators, cameras
- Smaller footprint
- Low energy consumption

Continuous Connectivity from Everywhere
- Wired networks / wireless networks
- Reliable, high bandwidth connectivity
- Ubiquitous access virtually from anywhere

Abundant Compute Power and Storage
- Faster chips
- Cheaper Memory
- Hadoop / big data stores
IoT Data Is Not Useful Until Analyzed

Traditional
• Capture only events of interest

• ERP, CRM, Billing …
• Human generated/captured
• Stored in Databases
• Data inherently valuable
• Data already useful for operations before analytics

Internet of Things
• Capture Everything

• Sensor, log, location, …
• Machine generated/captured
• Stored in big data frameworks
• Most of the data has little inherent value
• Value of data unknown until after it is analyzed
The Internet of Analytics

- Continuous Connectivity from Everywhere
- Cheap and efficient Data Capture
- Abundant Compute Power and Storage
IoT Data is Fast, Big, and Different

- IoT generates large **volumes** of data
  - Expensive to store in traditional data stores
  - Much of it is not useful

- Data comes in spikes or **high-velocity** continuous streams
  - Requires adequate connectivity
  - Uses significant network resources

- Data arrives in a **variety** of different formats
  - Data transformation is required
  - Data models need to facilitate analytics

- Data contains **perishable insights**
  - Requires platforms that can perform sophisticated Stream analytics

Source: *Internet Of Things Applications Hunger for Hadoop and Real-Time Analytics in the Cloud* by Mike Gualtieri and Rowan Curran, Forrester
Paradigm Shift in Computing – 3Vs

Acquire
- Structured Data
- Machine Data
- Location
- Click Stream

Store
- RDBMS
- EDW

Process
- BI / Analytics

Batch

Realtime

Acquire

Realtime barrier

Process in Memory
- IoT Analytics Applications

Deliver
- Visualizations
- Store
- Alerts
- Integrate

Reactive

Proactive
Minimize Time-to-Value – Perishable Value

Reduce the Latency to Capture, Analyze, and Ultimately Take Action to Increase Value

Based on concept developed by Richard Hackathorn, Bolder Technology
The Perfect Storm

**Perfect Storm**: An event where a rare combination of circumstances aggravate a situation drastically

- **Oil Rig Drill Sensor:**
  - Temperature up 10°C → continue drilling
  - Temperature up 10°C + Viscosity down → stop drilling

- **Hospital Bed:**
  - Blood O₂ below 93% → Patient went to the bathroom
  - Blood O₂ below 93% and Pulse @ 150BPM → send a doctor

You need realtime correlation of multiple data streams to handle the perfect storm
Actionable insights come from combining current events with context

\[
\text{Event} + \text{Context} = \text{Realtime Action}
\]

- Real-time Event Stream
  - e.g. Real-time Sensor Events

- Historical Context
  - Reference Data
  - e.g. Shopper profile, Store ID, Inventory, Profitability

- e.g. Present Next-Best-Action, update current price, modify sourcing
Because actionable insights come from combining current IoT events with context

- **Event:**
  - In the last 30 minutes, a store has sold $8,000.
  - In the last hour, 2 visits by shopper X in Store Zone 3 for 16 minutes.
  - A mobile subscriber drops 3 calls in 2 hours.

- **Context:**
  - This store typically sells $3,000 on Tuesdays in June.
  - Zone 3 has mobile phones. Shopper X due for device refresh.
  - A subscriber will drop 8 calls in a week before becoming a churn risk.

- **Realtime Action:**
  - Alert the store manager to require ID at checkout.
  - Offer promotion package for new device with 2 year contract renewal.
  - When a 611 call is made, alert the agent **NOT TO** offer a service discount.
IoT Platform Requirements

• Support Data in-Motion and Data at-Rest
  – Process events and groups of events (data windows) as Streams
  – Correlate multiple Streams in Realtime before disk storage
  – Leverage analyzed context from historic data sources
  – Store aggregate data, analyzed data, and raw payload on various storage frameworks

• Implement an Easy-to-Use Development environment
  – Allow users to quickly discover and analyze data
  – Convert analysis patterns into IoT Analytics Applications
  – Provide an easy-to-use development / deployment interface

• Address industrial and operational needs
  – Offer linear scalability
  – Run on commodity infrastructure / virtualized environments
  – Provide redundancy, failover, recovery
IoT Stream Analytics Ecosystem

Sources
- RDBMS
  - JDBC/SQL
  - Oracle CDC
  - MS SQL CDC
  - NonStop
  - GoldenGate
- Files
  - CSV/TSV
  - JSON
  - XML
  - Apache
  - Free-form
- Network
  - TCP/UDP
  - HTTP
  - SNMP/NetFlow
- Message Queues
  - JMS
  - Kafka
- Log Flows
  - Flume
  - Collectd
  - Windows Events
- BigData
  - HDFS

Applications
- Business-Level Logic
  - With Tungsten QL (extended SQL)

Real-time Dashboards
- Distributed Continuous Query Processor
- Distributed Results Cache

Delivery
- Message Queues
  - JMS
  - Kafka
- Alerting
  - Email
  - SMS
- Automated Workflows
- DB Persistence
  - JDBC/SQL
  - NoSQL
  - Vertica
- File Persistence
  - CSV/TSV
  - JSON
  - XML
- BigData
  - HDFS

External Context

WebAction

 PROPRIETARY & CONFIDENTIAL
WebAction Solution
How Does WebAction Work?

Assimilate
Structured and unstructured data

Process
Distributed, in-memory, as data is created

Deliver
Correlated, enriched, and filtered real-time big data records
Assimilate

- Data from transactional sources is acquired via redo or transaction logs
- Structured and non-Structured data
- No Production Impact
- No Application changes

<table>
<thead>
<tr>
<th>TYPE</th>
<th>EXAMPLE</th>
<th>COMPLEXITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common File Format</td>
<td>CSV, JSON, XML</td>
<td>SIMPLE</td>
</tr>
<tr>
<td>Social Feeds</td>
<td>Facebook, Twitter</td>
<td>SIMPLE TO MEDIUM</td>
</tr>
<tr>
<td>System/ IT Data</td>
<td>Syslogs, weblogs, event logs</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Device Data</td>
<td>SmartMeter, Medical Device, RFID, Netflow, iBeacon, CDR</td>
<td>MEDIUM</td>
</tr>
<tr>
<td>Industry Data</td>
<td>SWIFT, HL7, FIX, ASN</td>
<td>HIGH</td>
</tr>
<tr>
<td>Real-Time Transaction Data</td>
<td>Oracle, DB2, SQLServer, MySQL, HP NonStop</td>
<td>VERY HIGH</td>
</tr>
</tbody>
</table>
Process

- Enrich live Big Data with historical data sources
- Process Big Data faster using partitioned streams, caches, and additional nodes
- Execute SQL-like queries of in-memory Big Data
- Alert in real-time based on predictive analytic model results
Deliver

- Continuous Big Data Records
- Realtime Drag & Drop Dashboards
- Predictive Alerts
- Business Trends
- Data Patterns
- Outliers

Assimilate
Structured and unstructured data

Process
Distributed, in-memory, as data is created

Deliver
Correlated, enriched, and filtered real-time big data records
WebAction Platform Architecture

Stream Analytics Applications

High Speed Data Acquisition

Node 1
Node 2
Node 3
Node n

Event Windows
Context Cache

Distributed Query Processor

Distributed Results Cache
External Targets & Alerts

Results Persistence

Command Line
Visual Designer

CREATE APPLICATION MultiLogApp;
CREATE FLOW MonitorLogs;
CREATE SOURCE AccessLogSource USING ...
CREATE TYPE AccessLogEntry ...
CREATE STREAM AccessStream OF ...
CREATE CQ ParseAccessLog ...
W >

Drag & Drop Stream Dashboards

Enterprise Apps & Workflows
RDBMS
Enterprise Data Warehouse
Big Data Infrastructure

PROPRIETARY & CONFIDENTIAL
# Core WebAction Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source</td>
<td>Access external data and provides realtime continuous events into streams</td>
</tr>
<tr>
<td>Stream</td>
<td>Carries data between components and nodes</td>
</tr>
<tr>
<td>Window</td>
<td>Provides moving snapshot/collection of events for aggregates and models</td>
</tr>
<tr>
<td>Cache</td>
<td>External contextual data made available using distributed in-memory grid</td>
</tr>
<tr>
<td>CQ (Continuous Query)</td>
<td>A Continuous Query emits big data records after processing realtime streaming events (can process data from streams, windows, caches, event tables, and stores)</td>
</tr>
<tr>
<td>WAction Store (big data records)</td>
<td>Resulting big data records from processing (aggregates, correlates, anomalies, predictions) - can be in-memory only or persisted to Elasticsearch / database</td>
</tr>
<tr>
<td>Target</td>
<td>Outputs realtime big data records to external systems</td>
</tr>
<tr>
<td>Application</td>
<td>A combination of the above components performing business logic</td>
</tr>
<tr>
<td>Dashboard</td>
<td>A drag and drop realtime view into stores, caches, and streams</td>
</tr>
</tbody>
</table>
How to build Applications
Application Developer UI

- Create Applications
- Add & Navigate through Flows
- Design Data Model
- Drag & Drop Components
- Configure Components
- Deploy Applications
- Start / Stop Applications
- View Alerts
- View Event Flow Rate
Building a Stream Analytics Application

- Add a new application
- Add a source (CDC, structured, semi-structured, etc.)
- Configure Source
Building a Stream Analytics Application

- Add a typed stream
- Add a CQ to transform data types
- Add a Cache and CQ for context enrichment
Streaming Visualizations
Realtime Dashboard Designer UI

- Design dashboard
- Create multiple pages / drilldowns
- Define data through queries
Building a Realtime Dashboard

- Add a new Dashboard
- Add a Visualization
- Configure Query and Visualization
Realtime Dashboard

- View Data Visualized
- Filter data in a page
- Drilldown to related and detail pages
Under the Hood

- Realtime log / database CDC reading in addition to push sources like TCP/JMS
- Bytecode generation for data types and query processing
- Scaling across multiple nodes with flexible deployment
- Auto failover of application components from one node to another
- Nodes can be added and removed while applications are running
- Recovery ensures no events are missed or processed twice
- Recovery takes window contents into account
- Role based security at the application through component level
- Integrated realtime dashboard visualizations using server push
Continue the Conversation...

WebAction Headquarters
info@webaction.com
+1 (650) 241-0680

LinkedIn.com/company/WebAction-Inc
facebook.com/WebActionSoftware
Twitter.com/WebActionInc