PREPARING FOR A UNIFIED IMC ARCHITECTURE BY 2020

STEVE WILKES
CO-FOUNDER & CTO OF STRIIM
EVERYTHING IS CONVERGING TOWARDS IN-MEMORY COMPUTING

"CONVERGENCE
the merging of distinct technologies, industries, or devices into a unified whole"
ENTERPRISE, CLOUD AND IOT ARE NOT ISLANDS

Enterprise

Cloud

IoT
THEY ARE PART OF A CONNECTED ECO-SYSTEM

Enterpris

Cloud

IoT
PART OF A DIGITAL TRANSFORMATION THAT INCLUDES AI

Enterprise

Cloud

IoT

Predictive maintenance/part management

AI driven adaptive AML

Cybersecurity Analytics

NLP Call Center/Sentiment Analysis Retail Banking

Automating customer engagement through smart bots

CGM predictive monitoring

Machine Learning
EVERY INDUSTRY IS UNIFIED BY DIGITAL TRANSFORMATION

Financial Services

Healthcare

Manufacturing

Retail

Communication

Transportation/Logistics

Insurance

Public Sector
DATA GENERATION RATES ARE GROWING EXPONENTIALLY

Today we generate around 16ZB data annually. By 2025, this will leap to 160ZB. By 2025, 25% of all data will be real-time. Only a small percentage of this data can be stored. About 5% of this is real-time data. 95% of real-time data will be generated by IoT.

*Data Age 2025: The Evolution of Data to Life-Critical. An IDC White Paper, Sponsored by Seagate
IF YOU CAN’T STORE ALL DATA – WHAT CAN YOU DO?

PROCESS AND ANALYZE DATA IN-MEMORY IN A STREAMING FASHION
How do you avoid losing or ignoring valuable data, while still storing only the minimum?

How do you correlate all events for immediate insights and proactive responses?

How do you act promptly to better serve customers, protect reputation, and beat competitors?
REAL-TIME USE CASES CONVERGE ACROSS MANY INDUSTRIES

**Financial Services**
- Anti-money laundering
- Fraud prevention
- Management
- Customer service

**Healthcare**
- Proactive illness detection
- Allocation optimization
- Compliance verification

**Manufacturing**
- Quality management
- Predictive maintenance
- Equipment monitoring

**Retail**
- Fraud and theft detection
- Real-time offers
- Targeted marketing
- Dynamic pricing

**Communications**
- Network health monitoring,
- Predict network failures
- Proactive service outreach
- Location-based advertising

**Transportation/Logistics**
- Connected car
- Predictive maintenance
- Asset tracking
- Route optimization

**Public Sector**
- Crime detection and prevention
- Cyber security
- Traffic management
- Connected City

**Insurance**
- Claim fraud detection
- Agent fraud detection
- Risk-based policy pricing
- Usage-based insurance

**IT**
- Cyber security
- Replication validation
- API usage monitoring
- SLA monitoring
ALL DATA ARRIVES IN STREAMS NOT BATCHES

streaming

... stream processing has emerged as a major infrastructure requirement

database

logs

events

human s

machines

devices
STREAM PROCESSING REQUIRES A COMPLETE IMC PLATFORM

IN-MEMORY COMPUTING PLATFORM

VALUE EXTRACTED IMMEDIATELY

CONTINUOUS DATA COLLECTION
REAL-TIME STREAM PROCESSING
REAL-TIME STREAMING ANALYTICS & ALERTING
CONTINUOUS INFORMATION STORAGE

CONTEXT ADDED WHILE PROCESSING

STREAM PROCESSING REQUIRES A COMPLETE IMC PLATFORM

IN-MEMORY COMPUTING PLATFORM

VALUE EXTRACTED IMMEDIATELY

CONTINUOUS DATA COLLECTION
REAL-TIME STREAM PROCESSING
REAL-TIME STREAMING ANALYTICS & ALERTING
CONTINUOUS INFORMATION STORAGE

CONTEXT ADDED WHILE PROCESSING
GARTNER TAXONOMY OF IN-MEMORY COMPUTING TECHNOLOGIES

**Memory-Intensive Computing Platform**
- (DRAM, Flash, SSD, Multicore, InfiniBand, Clusters, Grid, Cloud)

**In-Memory Data Management Platforms**
- In-Memory DBMSs
- In-Memory Data Grids

**In-Memory Application Platforms**
- In-Memory Analytics and Visual Data Discovery
- Stream Processing Platforms
- Other Application Platforms

**High-Performance Message Infrastructure**

**Memory-Intensive Computing Platform**

Source: Gartner (January 2017)
IN-MEMORY COMPUTING USED FOR HTAP & HIP

**HTAP**
Hybrid Transactional Analytics Processing

**HIP**
Hybrid Integration Platform
ALSO FOR STREAMING INTEGRATION AND ANALYTICS

In-Memory Application Platforms
- In-Memory Analytics and Visual Data Discovery
- Stream Processing Platforms
- Other Application Platforms

In-Memory Data Management Platforms
- In-Memory DBMSs
- In-Memory Data Grids
- High-Performance Message Infrastructure
UNIFIED IMC ARCHITECTURE FOR STREAMING ANALYTICS

Stream Processing & Analytics

Distributed Results Storage

Distributed In-Memory Data Grid

Distributed High Speed Message Infrastructure

Dashboards & Visualization
For In-Memory Analytics / Visual Discovery

CLUSTER

Sources

Targets

Data Collection

Stream Processing & Analytics

Data Delivery

In-Memory Computing SUMMIT 2017

striim
HOW DO YOU GET THERE?

OPEN SOURCE

HYBRID
“OPEN CORE”

PROPRIETARY
BUILDING THIS FROM OPEN SOURCE

- Development
  - Distributed High Speed Message Infrastructure
  - Distributed In-Memory Data Grid
  - Distributed Results Storage

- Processing & Analytics

- Dashboards & Visualization
  - API Connectivity / Abstraction Layer / Web Server

- Sources
  - Streaming
  - Data Collection
  - Distributed In-Memory Data Grid

- Targets
  - Glue-Code
  - Clustering
  - Scalability
  - Reliability
  - Security
  - Management
OPEN SOURCE DEVELOPMENT PROCESS

Build From Open Source

- Design
- For Each Component
  - Identify
  - Evaluate
- Maintain
  - Deprecated
  - Upgraded
- Integrate
- Test
- Vendor or Community Support
- Build Applications
- Test
- Deploy
ADVANTAGES OF HYBRID “OPEN CORE” PLATFORMS

HYBRID “OPEN CORE”
Commodity meets enterprise grade. Combines rapid innovation & economies of commodity software of open source with security, unique IP, and last mile integration of proprietary

OPEN SOURCE
Commodity Technology
Extensible Technology
Critical Mass Technology

PROPRIETARY
Business Logic Intensive
Unique Integration
Niche Technologies
HYBRID “OPEN CORE” DEVELOPMENT PROCESS

- Install
- Hybrid
- Build
- Applications
- Test
- Deploy
- Hybrid
- Support

- Build From Open Source
  - Design
  - For Each Component
    - Identify
    - Evaluate
    - Integrate
  - Maintain
    - Deprecated
    - Upgraded
    - Test
IMC NEEDS TO BE ENTERPRISE GRADE FOR MISSION CRITICAL APPS

Scalability
Security
Enterprise Grade
Reliability
Integration
"Scalability is a characteristic of a system that describes its capability to cope and perform under an increased or expanding workload."

Scalability in IMC:
- Ingestion volume
- Processing
"Reliability is the ability of a system to consistently perform its intended or required function, on demand without degradation or failure."

Reliability in IMC:
- Ingestion
- Processing
"Security is the mechanism by which a system is protected from data corruption, destruction, loss, interception, or unauthorized access."

Security in IMC:
- Authentication
- Authorization

In-Memory Computing Summit 2017
striim
"Integration is the bringing together of component subsystems into one system and ensuring that the subsystems function together."

Integration in IMC:
• Ingestion
• Enrichment
EXAMPLE – STRIIM’S HYBRID ARCHITECTURE

Server
Drag and Drop UI + Command Line Interface

Distributed High Speed Message Infrastructure
Distributed In-Memory Data Grid for Metadata / Control
Distributed In-Memory Data Grid for Context Data
Distributed Persistent High Speed Message Infrastructure
Distributed Results Storage

Real-Time Streaming Dashboards to Surface In-Memory Analytics
TQL / JDBC / ODBC / REST / WS APIs
SQL-Based Processing and Analytics
Continuous Data Collection
Continuous Data Delivery

Sources
Kafka
HDFS

Targets
Kafka
HDFS

Scalability, Distribution, Clustering & Failover
Reliability, Recovery & EIP
Management & Monitoring
Role-Based Security & Encryption

Sources
Flume
HBase

Targets
STRIIM CLUSTER

Elastic

JCache
Hazelcast
Kafka
JMQ + Kryo
Hive

In-Memory Computing
Striim

HBase
Hive
EXAMPLE PROPRIETARY IP IN THE HYBRID MODEL

Non-Intrusive Log-Based Change Data Capture
- Change Data Capture not available as Open Source
- Captures DML / DDL as Change Stream

Distributed Stream Processing
Cache Integration and CEP
- Patented technology ensures scalability
- Pre-integrated distributed cache avoids adding latency

Fault-Tolerant Exactly-Once Processing
- Fully rebuilds transaction state for rollback and replay
- Supports jumping and sliding time-windows

Flow Designer and Dashboard Builder
- Enables ease of use and productivity
- No integration required for built-in visualization

End-to-End Security
- Encryption of all data over the wire
- Single role-based security policy across all components
EXAMPLE USE CASES

- Hybrid-Cloud Integration
- Real-Time Streaming Integration
- Cyber Security
- Production Quality
- Health Care Device Monitoring
- Location Tracking
HYBRID-CLOUD INTEGRATION

Approach
- Use Initial Load + CDC to Move Data
- True Real-Time Integration
  - CDC pushes new data real-time
  - Process as necessary
  - Monitor and alert on issues

Benefits
- Streaming not Batch
- Cloud Always Up to Date
- Not Limited to Single Target
REAL-TIME STREAMING INTEGRATION

Approach
- Collect, Prepare and Enrich Streaming Data for Delivery to Multiple Targets

Simple SQL-Based Processing
- Filter, Transform, Aggregate & Enrich Streaming Data
- Many Targets in one flow

Benefits
- Easy to Collect Real-Time Data
- SQL Enables Non-Developers
- Simple Deployment / Monitoring
CYBER SECURITY

**Approach**
- Collect and Correlate Data From Network, VPN, Firewall, Devices, Motion Sensors, etc.
- Identifies Multi-Phase Attacks
  - Port Scans + External Access
  - Operationalize AI
  - Unusual User & Machine Behavior

**Benefits**
- Instant Insights
- Proactive vs Reactive
- Not Limited to Single Solution
PRODUCTION QUALITY

- **Approach**
  - Collect and Analyze Device Data and Predict with Machine Learning

- **Real-Time Monitoring**
  - Sensor and Device Activity
  - On-Going Quality Expectations
  - Alert on Predicted Issues

- **Benefits**
  - Flexible Extensible Architecture
  - Scales With Your Business
  - Real-Time Insights & Fast Actions
HEALTH CARE DEVICE MONITORING

- Approach

- Real-Time Patient Monitoring
  - Multiple Medical Measurements
  - Use ML on Anonymous Data
  - Look for Anomalies / Issues

- Benefits
  - Doctors Have Real-Time Insights
  - React Immediately
  - Large Scale Data for Trends
LOCATION TRACKING

Approach
- Collect and Analyze Location Data Enriched With Contextual Information And Zones

Real-Time Tracking
- 1000s Real-Time Locations
- Multiple Active Zones
- Identify Entry / Exit / Wait / etc.

Benefits
- Spot Unusual Activity
- Integrate With Existing Context
- Real-Time Insights & Alerts
KEY TAKEAWAYS

IMC COMPONENTS DEPEND ON USE-CASE BUT WILL CONVERGE
• HTAP / HIP / SI & SA Each Use Subsets of all IMC Components
• Future IMC Platforms will merge to provide complete IMC capabilities

IT’S THE RIGHT TIME FOR STREAMING FIRST
• Customer demand now requires immediate insight and action for operational excellence
• Growing data volumes require pre-processing in-flight before storing data

YOU NEED A FULL END-TO-END PLATFORM
• Streaming data architecture addresses both concerns
• Add an enterprise-grade streaming data platform to existing infrastructure with small use cases and expand gradually - no need to rip and replace batch solutions
• Hybrid “Open Core” Platforms provide speed to solution while embracing open source
ABOUT STRIIM
STRIIM IS A COMPLETE END-TO-END PLATFORM

**Continuous Data Collection**
- DBs (thru CDC), files, HDFS, system logs, message queues, sensors

**Stream Processing**
- Real-Time Filtering, Transformation, Aggregation, Enrichment

**Streaming Analytics**
- Correlation, CEP, Statistical, ML, Alerts and Visualization, Trigger External Systems

**Continuous Results Delivery**
- Enterprise & Cloud DBs, files, Big Data, Blob Storage, Kafka, etc.

**Flexible Architecture With Deployment**
- On-Premise / At The Edge / In The Cloud

**Integration With Existing Enterprise Software**

**Enterprise Grade Streaming First Architecture**
- Clustered, Distributed, Scalable, Reliable and Secure

**Streaming Integration & Analytics Platform**
- Supporting Enterprise, Cloud and IoT
INTEGRATION AND ANALYTICS THROUGH DATA FLOWS
VISUALIZATION THROUGH STREAMING DASHBOARDS
Striim is unique in the market by providing all 4 of the following in a single platform.

- **Single Platform for Collection, Processing, Analysis, Delivery and Visualization of Streaming Data**
- **Supports wide variety of data sources, targets, and data types**
- **Converged In-Memory Platform**
- **Consistent end-to-end UI**

- **Log-based Change Data Capture**
- **Deep integration with Kafka**
- **Integrates with other technologies easily to collect data and distribute**
- **Top 3 Cloud Platforms**
- **Top 3 Big Data Platforms**
- **Major Enterprise Databases**
- **Multiple Open Source Solutions**

- **End-to-End**
  - Low configuration installation
  - Fast to build and deploy apps in days
  - Easy to iterate using SQL-like language
  - Continuous ingestion and processing
  - Multi-stream correlation
  - Time series/windowing

- **Easy to Use**
  - Secure with built-in authentication, protection and encryption
  - High performance and highly scalable with distributed architecture
  - Reliable with fault-tolerant architecture and “exactly once” processing

- **Easy to Integrate**
  - Supports wide variety of data sources, targets, and data types

- **Enterprise Grade**