NEEDLE STACKS & BIG DATA: USING EVENT STREAM PROCESSING FOR RISK, SURVEILLANCE & SECURITY ANALYTICS IN CAPITAL MARKETS

JERRY BAULIER, DIRECTOR, EVENT STREAM PROCESSING
DAVID M. WALLACE, GLOBAL FINANCIAL SERVICES MARKETING MANAGER
NEEDLE STACKS & BIG DATA

AGENDA

• Big Data and the Three V’s
• Event Stream Processing: Continuous Data Analysis
• Surveillance for Risk Reduction
  • Roque trader surveillance
  • Market portfolio risk
  • Cyber security
• Wrap-up
### Data Intensity

*Terabytes per Revenue and Employee by Industry, 2011*

<table>
<thead>
<tr>
<th>Industry</th>
<th>TB/$M Revenue</th>
<th>TB/Employee</th>
</tr>
</thead>
<tbody>
<tr>
<td>Banking and FS</td>
<td>0.82</td>
<td>0.35</td>
</tr>
<tr>
<td>Media</td>
<td>0.76</td>
<td>0.32</td>
</tr>
<tr>
<td>Healthcare</td>
<td>0.65</td>
<td>0.05</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.46</td>
<td>0.26</td>
</tr>
<tr>
<td>Retail</td>
<td>0.23</td>
<td>0.08</td>
</tr>
</tbody>
</table>

NEEDLE STACKS | BIG DATA: GROWTH IN VOLUME AND VARIETY

Volume
Scale from terabytes to exabytes.

Variety
Access relevant information from structured and unstructured data.

Velocity
Collect, process, and deliver insight to support real-time business.

1 Terabyte = 1,024 Gigabytes
1 Petabyte = 1,024 Terabytes
1 Exabyte = 1,024 Petabytes
1 Zettabyte = 1,024 Exabytes

Source: CEB TowerGroup, Gain New Insight from Unstructured Data, 2013.
NEEDLE STACKS | BIG DATA: GROWTH IN VOLUME AND VELOCITY

OPRA Traffic Projections 2011 - 2015
Messages per Second

Source: Options Price Reporting Authority, Data Recipient Notices 2010-2013
NEEDLE STACKS

Find the needles in the haystacks

WHAT WE ASSUME WE WANT FROM BIG DATA ANALYTICS
I need to collect the needles while the data is in motion.
NEEDLE STACKS

WHAT WE WANT TO FIND IN THE NEEDLE STACK

Find me the golden needle...right now!
NEEDLE STACKS WHAT DO YOU CALL IT?

Real-Time Analytics  Continuous Intelligence
Complex Event Processing  Contextual Awareness
Event Stream Processing
Operational Intelligence  ESP  Situational Awareness
CEP  Event Driven Architecture  Real-Time Visualization
Data Stream Mining  Asynchronous Processing
COMPLEX EVENT PROCESSING

CEP usually refers to event processing that assumes an event cloud as input, and therefore can make no assumptions about the arrival order of events. *

- It's been around for a long time (mostly in form of proprietary solutions)
- It's only the name and availability of commercial frameworks that's relatively new.
- What is complex – the events or the processing?
- Two architectures: Rules-based, Continuous Query-based

* This is the definition provided by the Event Processing Technical Society
EVENT STREAM PROCESSING (ESP)

ESP is a subcategory of Complex Event Processing (CEP) focused on analyzing/processing ‘events in motion’ called Event Streams.*

* This is the definition provided by the Event Processing Technical Society
• Continuous queries on data in motion (with incrementally updated results)
• Very low (max) event processing latencies (i.e., \(\mu\)secs-\(\mu\)msecs)
• High volumes (>100k events/sec)
• Derived event windows with retention policies
• Continuously reduce event streams into actionable intelligence for alerts
• Predetermined data mining, decision making, alerting, position management, scoring, profiling, …
• Event out-of-order handling to ensure ordered source streams
EVENT STREAM PROCESSING ENGINE: EVENT-DRIVEN, FLOW-CENTRIC

EVENT STREAM SOURCES/PUBLISHERS

- Venues & Instruments
- Databases
- Applications
- Applications

EVENT STREAM PROCESSING ENGINE

- Publish / subscribe
- Inserts/updates/deletes
- Continuous queries
  - Aggregate (group by)
  - Correlate (join)
  - Compute
  - Filter
  - Procedural
  - User defined functions
  - Retention windows
  - Pattern matching
- Ad-hoc queries
- Command & control
- Security (Auth, Encrypt, AC)
- Persistence / recovery
- Fail-over
- Distributed services
- Model management

EVENT STREAM SOURCES/PUBLISHERS

- MOBILE BI / BAM ALERTS
- WORKFLOWS
EVENT STREAM PROCESSING CONTINUOUS QUERIES – DATA FLOW DIAGRAM
HYBRID DATA ANALYTICS

Data Sources
- Risk Data
- Market Data
- Reference Data
- Trades

Real-Time Analytics
- Event Stream Processing
- Pattern Matching
- Correlation
- Aggregation
- Thresholding
- Profiling

Advanced Predictive Analytics
- In-Memory Analytics Server
- Risk Analytics
- Fraud Analytics
- Behavioral Analytics
- ... Persistence Store (Hadoop)

Visualization, Alerts, Decision Management
- Visualization
- Decision Management
- Alert Management

Macro Orchestration (Event Stream Processing)
NEEDLE STACKS | EVENT STREAM PROCESSING: RISK SURVEILLANCE METHODS

- Position monitoring
- Thresholding
- Profiling
  - individual signatures, group signatures, fraud signatures, …
- Pattern matching
  Event A followed by (time) Event B followed by (time) not Event C
- Neural networks
- Cognitive learning
TRADING SURVEILLANCE / FRAUD MANAGEMENT PROCESS FLOW

Operational Data Sources
- Transactions
- Entities
- Internal Data
- Market Feeds

Continuous Surveillance with Event Stream Processing

Fraud Data Staging

Intelligent Fraud Repository

Advanced Predictive Analytics and Alert Generation Process
- Business Rules
- Alert Administration
- Analytics
  - Anomaly Detection
  - Predictive Modeling

Fraud Network Analysis
- Network Rules
- Network Analytics

Alert Management & Reporting

Learn and Improve Cycle

Case Management Compliance Investigators

Business Rules Update Process

Directed Alerts to ESP

Directed Alerts from ESP
**ROGUE TRADER SURVEILLANCE USE CASE**

**POST-TRADE ORDER PRACTICE & COMPLIANCE ALERTING**

**Event Stream Sources/Publishers**
- Trades Market Feeds
  - Trades
  - Brokers of Interest
  - Restricted Securities
  - Venues Trades
- Brokers
- Restricted Securities
- Venue Trade Windows

**Event Stream Processing Server**
- Trades
- Brokers of Interest
- Restricted Securities
- Venues Trades
- Trades of Large Size (filter)
- Trades of Interest (join)
- Restricted Sales (join)
- Marking Open/Close Patterns (procedural)
- Front-running Patterns (procedural)
- Broker Alerts (aggregate)

**Broker Alerts**
- Broker aggregates for each alert type and total.

**Event Descriptions**
- **Front-running:** broker buys securities for his own account before buying the same securities for his customer, then sells when the price rises; or broker sells securities out of his personal accounts prior to selling the same securities for his clients.
- **Restricted Sale:** sales of securities that have ownership restrictions.
- **Marking the Open:** attempting to influence the opening price of a security by making trades at the opening of normal trading hours.
- **Marking the Close:** attempting to influence the closing price of a security by executing purchases at the close of normal trading hours.

**Case Management GUI**

**Copyright © 2012, SAS Institute Inc. All rights reserved.**
RISK DATA AGGREGATION USE CASE
CYBER SURVEILLANCE USING HYBRID ANALYTICS

Data Sources
- Syslogs
- SIEM (Security Info & Event Mgmt)
- Reference Data
- Network Monitors

Real-Time Analytics
- Event Stream Processing
  - Pattern matching
  - Trending
  - Individual signatures
  - Known threat signatures
  - Consolidated views
  - Causality

Advanced Predictive Analytics
- In-Memory Analytics Server
  - Predictive analytics
    - Decision trees
    - Neural networks
    - Gradient boosting
  - Case management
  - Alert management
  - Rule effectivity
  - Prevention
  - Causality

Visualization, Alerts, Decision Mgt.
- Visualization
- Case Management GRC

Macro Orchestration
- (Event Stream Processing)
ADVANTAGES OF EVENT STREAM PROCESSING

- Much lower latencies for action, hence new opportunities
- Big data analytics – ability to deal with the growing huge data volumes & reduce to knowledge
- Move back-end analytics to continuous analytics with a working resultant set
  - Reduced storage requirements
  - Reduced computational requirements
  - Focused on action and new patterns of interest
- Hybrid analytics = closed loop analytics
  - ESP front-ends advanced predictive analytics
- Knowledge based analytics can be applied to many problems
NEEDLE STACKS

FINDING THE RIGHT NEEDLE IN THE STACK

Continuous Event Stream Processing + Advanced Predictive Analytics Delivers the Gold!
THANK YOU!
QUESTIONS? CONTACT:
DAVID.M.WALLACE@SAS.COM
ROBUST ARCHITECTURE

REQUIRES ENTERPRISE ARCHITECTURE APPROACH

DATA SOURCES

DATA SERVICES

ENTERPRISE ANALYTICAL FRAMEWORK

EDW
GRID COMPUTING
IN DATABASE
IN MEMORY
ADW

FOUNDATIONAL ENTERPRISE & ANALYTICAL DATA WAREHOUSE

BUILT FOR PURPOSE ANALYTICAL DATA STORES

ANALYTICS SERVICES

ANALYTICAL INSIGHTS
OPERATIONAL DECISIONS
MARKET RISK MANAGEMENT USE CASE

Data Sources
- Risk Data
- Market Data
- Reference Data
- Trades

Real-Time Risk Analysis
- Data Integration
- Event Stream Processing
- Positions & Limits Monitoring

Analytics
- In-Memory Analytics Server
- In-Memory Risk Analytics
- Valuation Aggregation
- Stress Testing

Data Visualization
- Visualization
- Other Reporting

Macro Orchestration
(Event Stream Processing)