Cloud-based IT Log Analytics

Christian Beedgen
Kumar Saurabh
Agenda

Overview
Team
Market Size
Problem Statement
The Next Generation
Differentiators
Competition
Go To Market
Economics
Roadmap
Summary
Overview

*Cloud-based IT Log Analytics*

Service to manage and analyze IT logs

$2.5 Billion market size

Current products have high TCO, are services-heavy

*Easy to get started, lower TCO, superior intelligence*

Team of log management veterans, to be completed

Series A – customer-focused development process
Team

Christian Beedgen

**ArcSight** since 2001, Chief Architect, Director of Engineering
Lead ESM server developer
Built ESM server team, managing 20 people in server and UI teams
Named on 2 granted patents, 7 patent applications in process
Past experience at Amazon, Gigaton, Cleverlearn

Kumar Saurabh

Data Architect at **Mint.com**
Single handedly built Mint’s data analysis infrastructure
**ArcSight** 2001-2008, Director of Engineering, managing 12 people
Lead for Analytics and Solutions Team
Named on 2 granted patents, 2 patent applications in process
Market Size ~$2.5 Billion

- **Security Information Management**
  - ArcSight, EMC/RSA, Cisco, Splunk, Symantec, Q1 Labs, LogLogic

- **Event Correlation & Analysis**
  - Tivoli, BMC, CA, HP, Microsoft, Quest

Source: Gartner/Dataquest

**Key Drivers:** Compliance, Security, Operations
Key Drivers

Compliance is not optional

“What is the primary motivation for adopting or using security information management (SIM) within your enterprise?”

- Compliance and reporting: 32%
- Incident investigation: 21%
- Log management: 13%
- To demonstrate the effectiveness of our security program: 12%
- We neither use SIM nor have plans to adopt it in the next 12 months: 11%
- Event correlation: 7%
- Don’t know: 5%

Base: 1,335 North American and European enterprise and SMB security decision-makers who expressed interest in adopting SIM (percentages do not total 100 because of rounding)

Source: Enterprise And SMB Security Survey, North America And Europe, Q3 2008
Problem Statement

Today’s market leading products are:

Premise-based
Enterprise sales cycles, installation and upgrade hassles, expensive hardware, DBAs, sysadmins required

Not scalable
Not inherently clustered, scaling introduces tradeoffs and data fragmentation

Challenged with log parsing
Either simply don’t parse or require parsing at collection time, need constant software upgrades

Not context-aware
Identities, network assets, service dependencies are all critical for correlation and prioritization

Customers operate in silos
Insight gathered by one customer is hard to share; no cross-customer data mining

Not community-aware
Exchanging of solutions is a manual process, there’s no marketplace
The Next Generation

1. Cloud-based service
   Easy sale, quick delivery, ongoing upgrades, no care and feeding

2. Seamless scalability
   Built from scratch for big data, leverages large-scale processing

3. Machine-driven log parsing
   Extracting structure from raw logs is foundation for analytics

4. Context modeling
   Logs need to be analyzed in their real world environment

5. Global IT log intelligence
   Data mining leads to insight shareable across all customers

6. Built-in community
   Not everybody is an expert, and even experts exchange findings

Deliver superior log management for compliance, security and operations in a scalable, easy-to-adopt cloud-based service
Target Market

Medium Enterprises
Large Enterprise Departments

Large Enterprises

Use Cases

Compliance
- PCI, SOX, HIPAA, NERC
- Log Retention & Review
- User & Resource Access

Security
- Incident Response
- Data Protection
- Threat Intelligence

Operations
- Troubleshooting
- Business Continuity
- Service Levels

Use Cases for Medium Enterprises

Use Cases for Large Enterprises
High-level Solutions Architecture

Global IT Log Intelligence, Community

Compliance
- PCI, SOX, HIPAA, NERC
- Log Retention, Review
- User, Resource Access

Security
- Threat Analysis
- Incident Response
- Data Protection

Operations
- Troubleshooting
- Business Continuity
- Service Levels

Collect → Normalize → Correlate → Context → Business Impact

IT Logs
- Network
  - Router/Switch
  - Firewall/Proxy
  - IDS/IPS
- Systems
  - OS Logs
  - File Access
  - Virtualization
- Applications
  - Web Server
  - Database
  - Custom App

Context
- Active Directory
- Vulnerability Scans
- Custom Source
High-level Platform Architecture

1. Collector
   - Customer C Apps in Cloud

2. Data Processing
3. Log Parsing
4. Context
5. Full Text Search
6. Community

- Global IT Log Intelligence
- Collector
  - SaaS Access Logs

Collector
- Customer A
- Customer B
- Customer C
Log Management Architecture Today

- Analytics Workflow
- Compressed log store
- Full-text search
- Collection
- Log parsing

Components:
- Correlation Engine
- Database
- Log Store
- Collector
- Raw Data

Flow:
- Raw Data to Collector
- Collector to Log Store
- Log Store to Correlation Engine
- Correlation Engine to Analytics Workflow
Log Management Architecture Today

Log parsing challenges

Parser creation and maintenance is manual and expensive

Limited vendor support for custom log sources like applications

Data is often not parsed, or parsed incorrectly; custom applications ignored
Next Generation Architecture

*Machine-driven Log Parsing*

1. Collector
2. Collector
3. Parsing
4. Raw Data

Efficient parser development driven by structure inference

Customer can add knowledge about custom sources like applications

**Better parser support enables more log sources, which enables superior analytics**
Database hardware is expensive, performance is still a major problem.

Most data in the Log Store is not available for analytics.

*Performance or intelligence – choose one*
Next Generation Architecture

*Seamless scalability*

Distributed store is built for scale, no excuses

Elastic compute layer can tackle any analytics efficiently

Better intelligence and faster response times
Log Management Architecture Today

Customers operate in silos

Advanced analytics cannot run due to database limits

Customers don’t always have analysts on staff, vendors don’t have the data

*Incidents that could have been detected with better analytics are missed*
Next Generation Architecture

- Makes it feasible to analyze all data from every customer.
- Enables deep analytics – root cause detection, network graph analysis, anomaly detection.
- Extracts and delivers actionable insights.

Diagram:
- Collector
- Collector
- Raw Data
- Distributed Store
- Elastic Compute
- Parsing
- Alerting
- Reporting
- Community
- Global Log Intelligence
<table>
<thead>
<tr>
<th>Differentiator</th>
<th>Customer Benefit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud-based service</td>
<td>Quick initial service delivery</td>
</tr>
<tr>
<td>No enterprise software release cycles</td>
<td>Seamless ongoing upgrades</td>
</tr>
<tr>
<td></td>
<td>No deployment or sizing hassles</td>
</tr>
<tr>
<td>Seamless, transparent scalability</td>
<td>Instant slice and dice analytics</td>
</tr>
<tr>
<td>Headroom to deal with all the data</td>
<td>No data fragmentation</td>
</tr>
<tr>
<td></td>
<td>100 EPS today, 10,000 EPS tomorrow</td>
</tr>
<tr>
<td>Intelligently evolving log parsing</td>
<td>Custom app logs useful right away</td>
</tr>
<tr>
<td>Automated structure inference</td>
<td>Structured data beats unstructured data</td>
</tr>
<tr>
<td></td>
<td>Once seen, available everywhere</td>
</tr>
<tr>
<td>Context modeling</td>
<td>Real world environment context for logs</td>
</tr>
<tr>
<td>Identities, network elements, services</td>
<td>Business impact correlation</td>
</tr>
<tr>
<td></td>
<td>Enables risk modeling</td>
</tr>
<tr>
<td>Global IT log intelligence</td>
<td>Solution recommendation system</td>
</tr>
<tr>
<td>Pattern discovery</td>
<td>Zero day discovery of emerging threats</td>
</tr>
<tr>
<td></td>
<td>Defense for all customers</td>
</tr>
<tr>
<td>Built-in community</td>
<td>Expert exchange</td>
</tr>
<tr>
<td>Frictionless sharing</td>
<td>Sharing of analytics content</td>
</tr>
<tr>
<td></td>
<td>Service ➔ Platform</td>
</tr>
</tbody>
</table>
Competition

<table>
<thead>
<tr>
<th>ArcSight</th>
<th>loglogic</th>
<th>splunk</th>
<th>Alertlogic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enterprise software &amp; services, appliances</td>
<td>Appliance-based</td>
<td>Downloadable software</td>
<td>SaaS, Appliance</td>
</tr>
<tr>
<td>RDBMS limitations, data fragmentation</td>
<td>RDBMS limitations, data fragmentation</td>
<td>Full-text index, cost of repeated parsing</td>
<td>Storage backend tradeoffs unknown</td>
</tr>
<tr>
<td>One schema, parse at collection, limited SDK</td>
<td>Schema per device, limited device support</td>
<td>Parsing on access via regular expressions</td>
<td>Schema per device</td>
</tr>
<tr>
<td>Limited context model, only network</td>
<td>No context model</td>
<td>No context model</td>
<td>Limited context model, only network</td>
</tr>
<tr>
<td>No intelligence shared among customers</td>
<td>No intelligence shared among customers</td>
<td>No intelligence shared among customers</td>
<td>No shared intelligence, some SOC service</td>
</tr>
<tr>
<td>Informal community</td>
<td>Informal community</td>
<td>Informal community</td>
<td>Informal Community</td>
</tr>
</tbody>
</table>

Cloud-based Log Analytics

- Cloud-based service
- Seamless, transparent scalability
- Intelligently evolving log parsing
- User, network, service model, extensible
- Global IT log intelligence
- Community baked into the service
Go To Market

Self-serve, easy to try, buy and use
Instant gratification
Free trials, freemium model also possible
Value-before-commit
Tiered pricing
Pay for what you use
Web sales, telesales, focused direct touch, channel
Lower cost of sales
Leveraged partnerships
PaaS add-on sale
Economics

Deliver service at high gross margin

Network, Storage, CPU

Inbound network traffic dominates outbound traffic
Storage needs to consider monthly charge due to retention
CPU can be optimized because of elasticity

<table>
<thead>
<tr>
<th>Tier</th>
<th>Events/Sec</th>
<th>GB/Day</th>
<th>COGS $/Month</th>
<th>MRR*</th>
<th>ACV*</th>
<th>AlertLogic ACV</th>
<th>ArcSight Deal Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial</td>
<td>5</td>
<td>0.5</td>
<td>$5</td>
<td></td>
<td></td>
<td></td>
<td>$2,148</td>
</tr>
<tr>
<td>Silver</td>
<td>120</td>
<td>10</td>
<td>$101</td>
<td>$506</td>
<td>$6,072</td>
<td>$36,000</td>
<td>$100-500k</td>
</tr>
<tr>
<td>Gold</td>
<td>1,200</td>
<td>100</td>
<td>$1,012</td>
<td>$5,063</td>
<td>$60,756</td>
<td>$153,000</td>
<td>$0.5-2M</td>
</tr>
<tr>
<td>Platinum</td>
<td>3,000</td>
<td>250</td>
<td>$2,531</td>
<td>$12,656</td>
<td>$151,872</td>
<td>$324,000</td>
<td>$2-6M</td>
</tr>
<tr>
<td>Diamond</td>
<td>12,000</td>
<td>1000</td>
<td>$10,125</td>
<td>$50,625</td>
<td>$607,500</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Priced at 80% gross margin
## Roadmap

### Release 1
- **Intelligence**: Anomaly Detection
- **Operations**: Troubleshooting
- **Security**: Threat Analysis
- **Compliance**: PCI Pack
- **Platform**: Collection, Search, Reporting, Assets
- **9 – 12 Months**

### Release 2
- **Intelligence**: Pattern Mining
- **Operations**: Business Continuity
- **Security**: Incident Response
- **Compliance**: SOX, HIPAA, NERC
- **Platform**: Workflow, Dashboards, Trending, Identities

### Release 3
- **Intelligence**: Predictive Analytics
- **Operations**: Service Levels
- **Security**: Data Protection
- **Compliance**: Fraud, Risk
- **Platform**: Context Modeling, Community

---
Summary

Cloud-based IT Log Analytics

Large opportunity in growing enterprise market

Team of veteran log management experts

Game-changing *functionality* and ease of adoption