Operational Intelligence – Extracting the Business Value within Big Data

DJ Skillman
Director Technical Services, EMEA
What Does the Data Look Like?

Volume | Velocity | Variety | Variability

Machine-generated data is one of the fastest growing, most complex and most valuable segments of big data.

GPS, RFID, Hypervisor, Web Servers, Email, Messaging, Clickstreams, Mobile, Telephony, IVR, Databases, Sensors, Telematics, Storage, Servers, Security Devices, Desktops
Machine Data is a Gold Mine for Operational Intelligence

Customer adds products to ecommerce / website shopping cart

```
```

Illustrative Business / IT Insights

- Customer demand
- Shopping cart activity
- Website troubleshooting

Customer completes transaction through credit card purchase

```
2011-12-05 07:04:44 Id=00Q0000000d91EAJ City=New York
```

Illustrative Business / IT Insights

- Customer behavior
- Inventory updates
- Fraud heuristics analysis
- Real-time product sales tracking

Web server attempts to write to database

```
[1208/11 02:39:03:209 UTC] 000000c6 ConnectionError: A J2CA00561: The exception which was received is com.ibm.websphere.cm.ConnectionException: [IBM]
```

Illustrative Business / IT Insights

- Nature of application error
- Source of error
Harnessing Machine Data Requires a New Approach

Business Application Data
- Relational data, highly structured, based on inflexible schema
- Financial records, multidimensional data, math computation
- Monthly reporting, not for real-time events

Human-generated Data
- Generated by human-to-human interaction
- Includes email, IM, voice, video and text across
- Stored in centralized corporate servers, fileshares and desktops

Machine Data
- Time series unstructured data, no predefined schema
- Generated by all IT systems, highly diverse formats
- Massive volume; fast navigation and correlation paramount
Make machine data accessible, usable and valuable to everyone.
RDBMS/SQL – Early-structure Binding

SELECT customers.* FROM customers
WHERE customers.customer_id NOT IN(SELECT customer_id FROM orders WHERE year(orders.order_date) = 2004)

Structure
- Schema – created at design time
- Queries – understood at design time for maximum performance
- Homogeneous – must fit into tables or be converted to fit into tables
- Must exactly match constraints

Data
Late-structure Binding

**Structure**
- Schema-less
- Created at *search time*
- Queries/Searches can be *ad-hoc*
- Heterogeneous – can come from any text-based source
- Constantly changing
- No conversion required, no constraints

**Data**
- Heterogeneous – can come from any text-based source
- Constantly changing
- No conversion required, no constraints
Big Data is a Big Challenge

Ventana Research – Benchmarking Big Data

Top 4 Obstacles Using Hadoop

Integration: 63%
Real-time: 67%
Training: 74%
Staffing: 80%
Splunk Enterprise with Hadoop

GPS, RFID, Hypervisor, Web Servers, Email, Messaging, Clickstreams, Mobile, Telephony, Databases, Sensors, Telematics, Storage, Servers, Security devices, Desktops, CDRs, Applications

Ad hoc search  Add knowledge  Monitor and alert  Custom dashboards  Report and analyze

Splunk storage

Other Data Stores
Extracting Value from Big Data in the Real World
Proven Real-World Examples

<table>
<thead>
<tr>
<th>16TB/day</th>
<th>6TB/day</th>
<th>6TB/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>zynga</td>
<td>LinkedIn</td>
<td>Expedia</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2.5TB/day</th>
<th>1.5TB/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bank of America</td>
<td>intuit</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1TB/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>comcast</td>
</tr>
<tr>
<td>Viacom</td>
</tr>
<tr>
<td>metroPCS</td>
</tr>
<tr>
<td>salesforce.com</td>
</tr>
</tbody>
</table>

Diagram:
- App Mgmt
- IT Ops
- Security
- Compliance
- Web Analytics
- Business Analytics
Accelerate Games Releases with Big Data Insight

**Use:**
- Over 10 TB/day from scaled-out cloud and physical infrastructure
- Data indexed includes web server and application logs for games
- Splunk for operational visibility, troubleshooting and monitoring
- Users include: game operations, developers, and corporate IT

**Value Delivered:**
- Faster game releases with real-time visibility into production issues
- Reduced fault resolution time from hours to minutes
- Scale ops team to manage and monitor growing infrastructure

- Leading social gaming company globally
- 232 million monthly active users
- 60 million daily active users
Broad Use Drives Significant Value at Expedia

Use:
- Monitoring 100% of their infrastructure
- Over 11,000 servers sending data / over 6.5TB per day
- Use cases: application monitoring, infrastructure management, and web analytics

Value Delivered:
- Over 25X ROI in just 9 months
- Millions in cost reduction by removing unused servers
- 75% reduction in MTTR and improved up time
- Focus on innovation vs. operations; 4000 users of Splunk

- Largest online travel company in the world
- Over 100 local Expedia and Hotel.com sites
- Revenues over $3.0 billion
Expedia Big Data Facts

12,000+ Servers
27,000+ Hosts
900+ Source Types
227,000 Sources

22 Indexers, 9 Search heads
> 6TB per day indexed

20+ Different Solutions for RCA
All Migrated to Splunk in 3 Months
Expedia: Integrate External Data

Extend search with lookups to external data sources.

Correlate across multiple data sources and data sets using indexes and keys
Visibility to Drive Innovation

Splunk Use:
- Single ‘pane of glass’ across their enterprise cloud computing environment
- Role-specific dashboards for developers, operations staff and product managers

Value Delivered:
- Refine product features for better use
- Product innovation by building new capabilities
- Indexing more than 1TB per day across multiple datacenters from the entire cloud stack - including application servers, web servers and email servers
Search Analytics to Improve Content Delivery

**Search (Machine Data)**

- Device (MAC Address)
- Time of Search
- Search Results (IDA #s)

**Billing (Structured Data)**

- Device (MAC Address)
- Content Purchased (IDA #)
- Amount of Purchase ($)
- Time of Search

**Correlation Criteria**
- Is MAC address same?
- Was content purchased in the search results?
- Was content purchased within 1 min of search?

**Business Value**
- Revenues driven by Search
- Improving local content mix to drive revenues
- Delivering better search results
- Tailor content promotion

© 2012, Splunk Inc.
MUVE Service – Understanding the Customer

User authenticates & receives an IP address (authentication log has phone # / IP address)

Activity Tracked by IP
- Searches
- Downloads
- Browsing

Authentication Logs

Search Logs
Download Logs
Application Logs

Correlate Phone # (MDN) and IP Address

Key to understanding individual customer behavior – insights for support, IT and business
Content Analytics to Improve Experience

- **Real-time content analytics**
  - Top 10 customer searches
  - Top song downloads
  - Top artist downloads
  - Top played songs
  - Top artists searched for

- **Business value**
  - Increasing banner hit rates
  - Improving content mix
  - Understanding customer transactions
IT & Business Insight Dashboards

**Metrics Tracked:**
- Uploads (success by browser)
- Breakdown of upload types
- Capacity modeling metrics

**Benefits:**
- Results of marketing activity related to feature release
- Insight into trending analysis
- Capacity planning and scaling
Customer Interactions with Mobile Apps

Benefits:

- Visibility into mobile app usage and trends
- Helps understand investment from a development standpoint
- Highlight areas of improvement / issues that customers have
- Better customer campaigns for new features
Analytics Deliver Competitive Advantage

“[Having access to this data] delivered a competitive advantage for us—across product planning, marketing and support.”

Lucia Cheng
Director
Taiwan Mobile

- Marketing visibility into devices and apps to improve service offerings
- Visibility into conversion rates helps Marketers optimize AdWord spend
- 1500 customer service reps solve problems quickly and have increased customer satisfaction scores
- Splunk is the mandated operational & real-time analytics platform
Tesco – Understanding Customer Behavior

Splunk Use:
- Across large scale production environment
- 400 servers in 10 environments / 50 different log types
- Monitoring, transaction tracking, & web analytics

Value Delivered:
- Faster mean time to resolution and better monitoring
- Improved customer satisfaction and experience
- Better understanding of customer interactions
- Business visibility and performance awareness

“Engineering team have been impressed with the information that Splunk can provide. They are now developing future applications with Splunk in mind”
Summary

- Big data is a big challenge, but also a big opportunity
- Machine-generated data is one of the fastest growing, most valuable segment of big data
- Turn big data into real-time operational intelligence
- This data exists in your environment today, make use of it.
Thank you.

DJ Skillman
dj@splunk.com
Splunk Collects and Indexes Any Machine Data

Customer Facing Data
- Click-stream data
- Shopping cart data
- Online transaction data

Outside the Datacenter
- Manufacturing, logistics...
- CDRs & IPDRs
- Power consumption
- RFID data
- GPS data

Windows
- Registry
- Event logs
- File system
- sysinternals

Linux/Unix
- Configurations
- syslog
- File system
- ps, iostat, top

Virtualization & Cloud
- Hypervisor
- Guest OS, Apps
- Cloud

Applications
- Web logs
- Log4J, JMS, JMX
- .NET events
- Code and scripts

Databases
- Configurations
- Audit/query logs
- Tables
- Schemas

Networking
- Configurations
- syslog
- SNMP
- netflow

Splunk Collects and Indexes Any Machine Data

Copyright © 2012, Splunk Inc.
Splunk Collects and Indexes Any Machine Data

- Click-stream data
- Shopping cart data
- Online transaction data

**Customer Facing Data**

**Windows**
- Registry
- Event logs
- File system
- sysinternals

**Linux/Unix**
- Configurations
- syslog
- File system
- ps, iostat, top

**Virtualization & Cloud**
- Hypervisor
- Guest OS, Apps

**Applications**
- Web logs
- Log4J, JMS, JMX
- Code and scripts

**Databases**
- Configurations
- Audit/query logs
- Tables
- Schemas

**Networking**
- Configurations
- Log
- SNMP
- Netflow

**Outside the Datacenter**
- Manufacturing, logistics...
- CDRs & IPDRs
- Power consumption
- RFID
- GPS
- Data

- Manufacturing, logistics...
- CDRs & IPDRs
- Power consumption
- RFID
- GPS
- Data

- No upfront schema
- No custom connectors
- No RDBMS
- No need to filter/forward

Any amount, any location, any source.
Create and Share Dashboards in Minutes

Deliver new levels of visibility and insight for IT and the business from operational data
250+ Splunk Apps Support Customer Success

<table>
<thead>
<tr>
<th>Community</th>
<th>Technology Partners</th>
<th>Developers</th>
<th>Splunk-built</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather</td>
<td>Ruby on Rails</td>
<td>Twigy</td>
<td>Security</td>
</tr>
<tr>
<td>Big Fix</td>
<td>Google Maps</td>
<td>Windows</td>
<td>Java</td>
</tr>
<tr>
<td>Sendmail</td>
<td>Whois lookup</td>
<td>Nagios</td>
<td>Javamail</td>
</tr>
<tr>
<td>PDF</td>
<td>WHOIS</td>
<td>*nix</td>
<td>Blue Coat ProxySG</td>
</tr>
<tr>
<td>Report Server</td>
<td>PCI Compliance</td>
<td>Linux and Linux</td>
<td>Solaris Deepide</td>
</tr>
<tr>
<td>FS</td>
<td>Puppet Conf. Mgt</td>
<td>Sourcefire</td>
<td>IMAP</td>
</tr>
<tr>
<td>Radio Stages</td>
<td>Python Mail</td>
<td>Splunk Monitoring</td>
<td>YouTube</td>
</tr>
<tr>
<td>WebSphere</td>
<td>NetFlow</td>
<td>SNORT</td>
<td>Encrypt/Decrypt</td>
</tr>
<tr>
<td>XenDesktop NetScaler</td>
<td>Audible Alerts</td>
<td>FireEye Malware</td>
<td>Enterprise Security</td>
</tr>
<tr>
<td>Multicast</td>
<td>Stock Quote</td>
<td>POST/GET Requests</td>
<td>AS/400 - Series</td>
</tr>
<tr>
<td>MS Exchange</td>
<td>FISMA Monitoring</td>
<td>Citrix NetScaler</td>
<td>Transaction Profiling</td>
</tr>
</tbody>
</table>

Developer Framework

250+ Apps and growing
Massive Linear Scalability to Tens of TBs/Day

Send data from 1000s of servers using combination of Splunk Forwarders, syslog, WMI, message queues, or other remote protocols

Auto load-balanced forwarding to as many Splunk Indexers as you need to index terabytes/day

Offload search load to Splunk Search Heads
Splunk is a Big Data Solution

**Product-based solution**
- Easy to download and deploy
- Pre-integrated, end-to-end functionality
- All the enterprise-grade features you expect!

**Powerful all-in-one solution**
- Collects data from tens of thousands of sources
- Advanced real-time and historical analysis of data
- Fast, custom visualizations for IT and business users

**Performance at scale**
- Proven at multi-terabyte scale per day
- Upwards of PB under management
- 3,300+ customers
Broad Adoption Across 3,700+ Customers

Financial Services & Insurance

Retail

Technology

Cloud and Online Services

Government

Healthcare

Manufacturing

Media & Entertainment

Energy and Utilities

Education

Telecommunications

Travel and Leisure

Copyright © 2012, Splunk Inc.