Expedia
Where you book matters.

1

Operational Intelligence on Big Data

Big Data Innovation Conference - 31ˢᵗ October & 1ˢᵗ November,13

Rohit Gossain , Analyst @ Expedia Inc.

✓ Linkedin @ rohitgossain
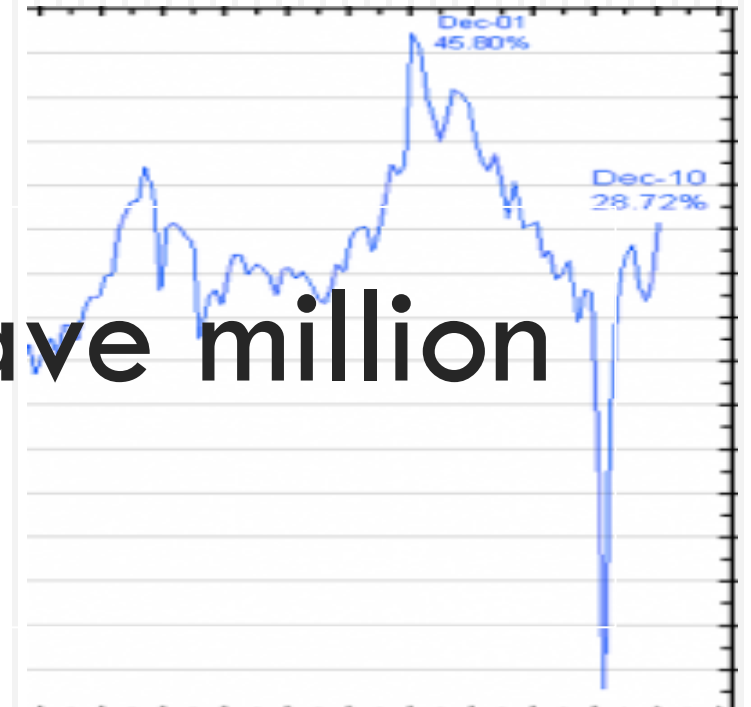
✓ rohitgossain88@gmail.com

Discussion Areas

- ☐ Opening Story
  - ◪ Operational Intelligence & Machine Data
- ☐ Difference with Business Intelligence
- ☐ OI Business Solutions
- ☐ Achieving Ah ha Moments
- ☐ Q&A

A story where we could save million dollar revenue
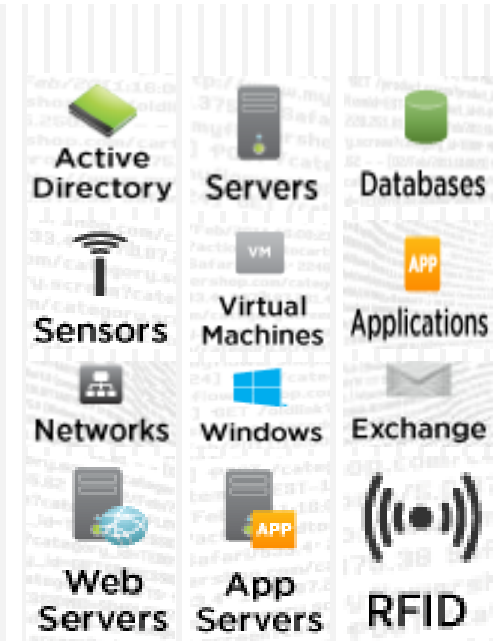
Solution – Operational intelligence

Typically used with **time-series unstructured or semi-structured data**. The data in operational intelligence systems enables you to see **what is happening now** and compare it to what has happened in the past.

What is Machine Data (Big Data) then??

☐ **Massive stream of unstructured data generated by machines,** in an array of unpredictable formats that are difficult to process and analyze.

- The volume, velocity and variety

- The most valuable insights from this data are often needed in *real time*.

- Possible Types of Machine Data

  - Log files or event logs

  - Periodically sampled metrics such as processor usage and I/O statistics

  - Records of ad page impressions and video views

  - Diagnostic records from an organization's IT-enabled devices

  - Real-time location information from mobile devices

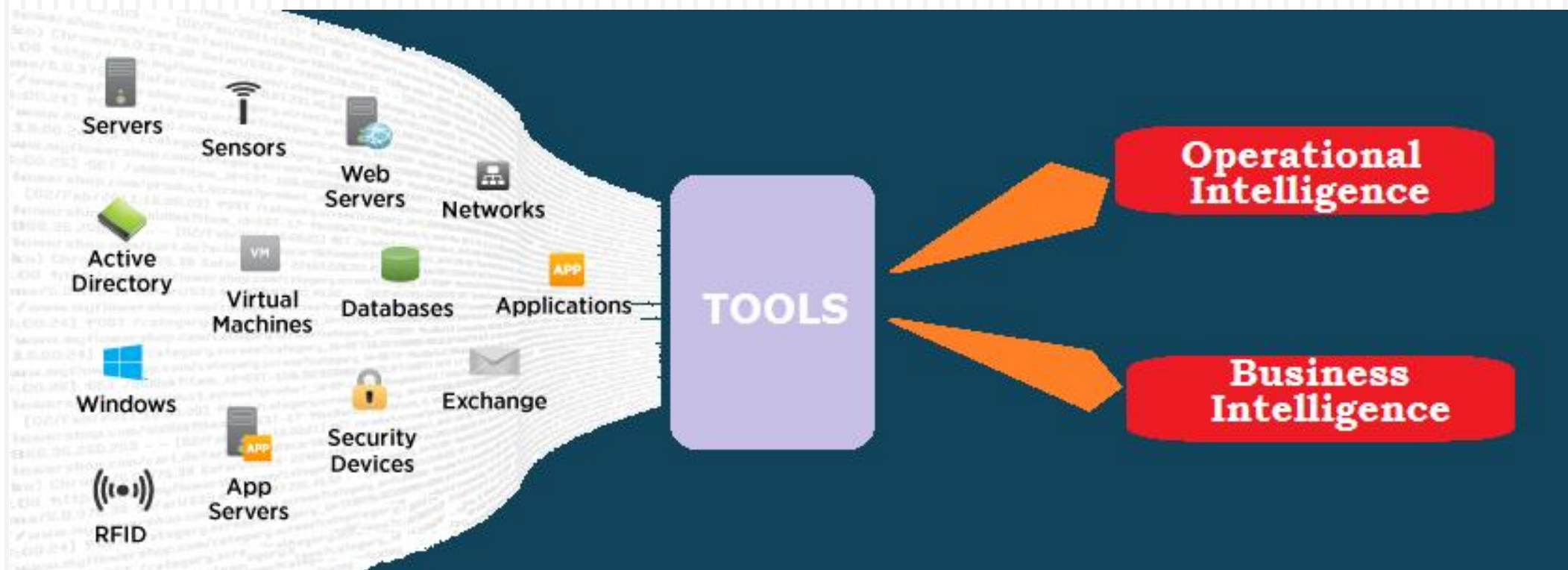Reference- splunk.com

We collect 4TB of machine data per day
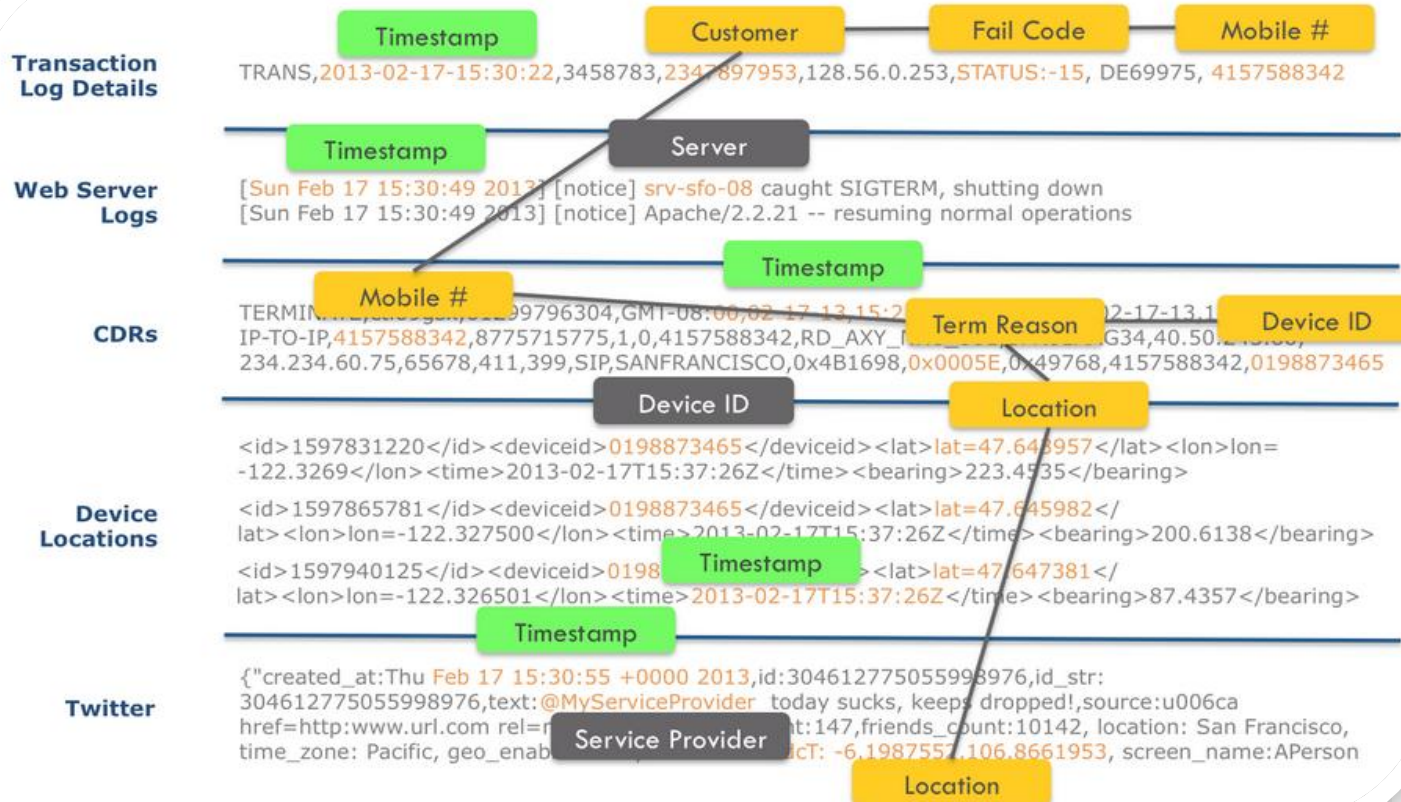
□ **Meaning in Machine Data (Insights)**
- ▪ Fail Code
- ▪ Timestamp
- ▪ Reason
- ▪ Location
- ▪ Customer ID
- ▪ Device ID

Reference - www.sqlstream.com

How can we utilize Machine Data?

Reference- splunk.com

Difference with BI

Basis	Operational Intelligence	Business Intelligence
Pattern Identification Approach	Real-time Dashboards (Proactive)	Reports (Reactive)
Source of Event Generation	CEP Queries (Activity Centric)	SQL Queries (Data Centric)
Value	Short term Planning (Next Best Action)	Long term Planning (Over the next days to weeks)

Reference-Research Reports
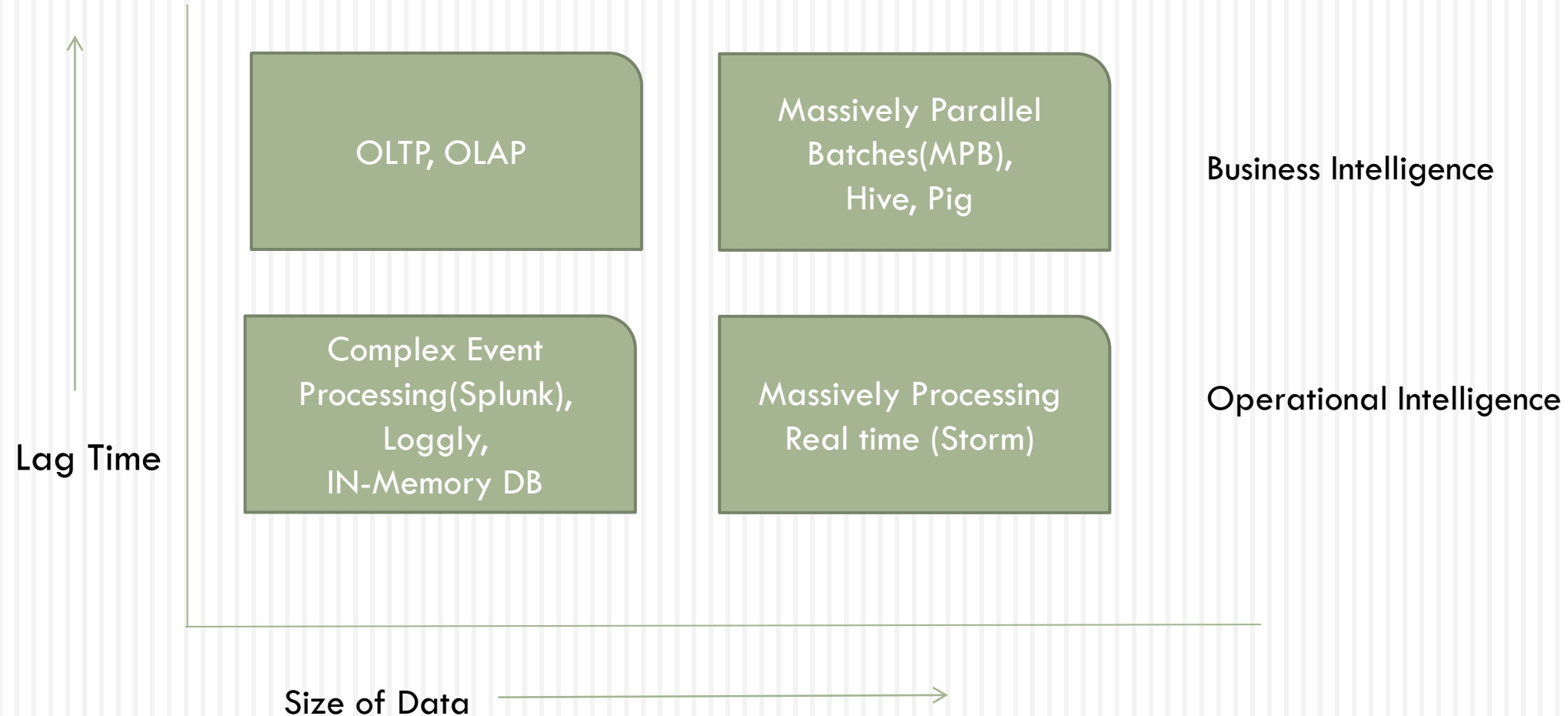
Is Hadoop Suitable for Operational Intelligence??

☐ Many believe that Hadoop is a platform for real-time, low latency analytics. This is not the case. It is a multi-purpose data storage engine but **not a real-time.**

▫ The parallelism of Hadoop is **great for processing the data once stored,** but has high throughout latency (mainly for **Batch Processing**)

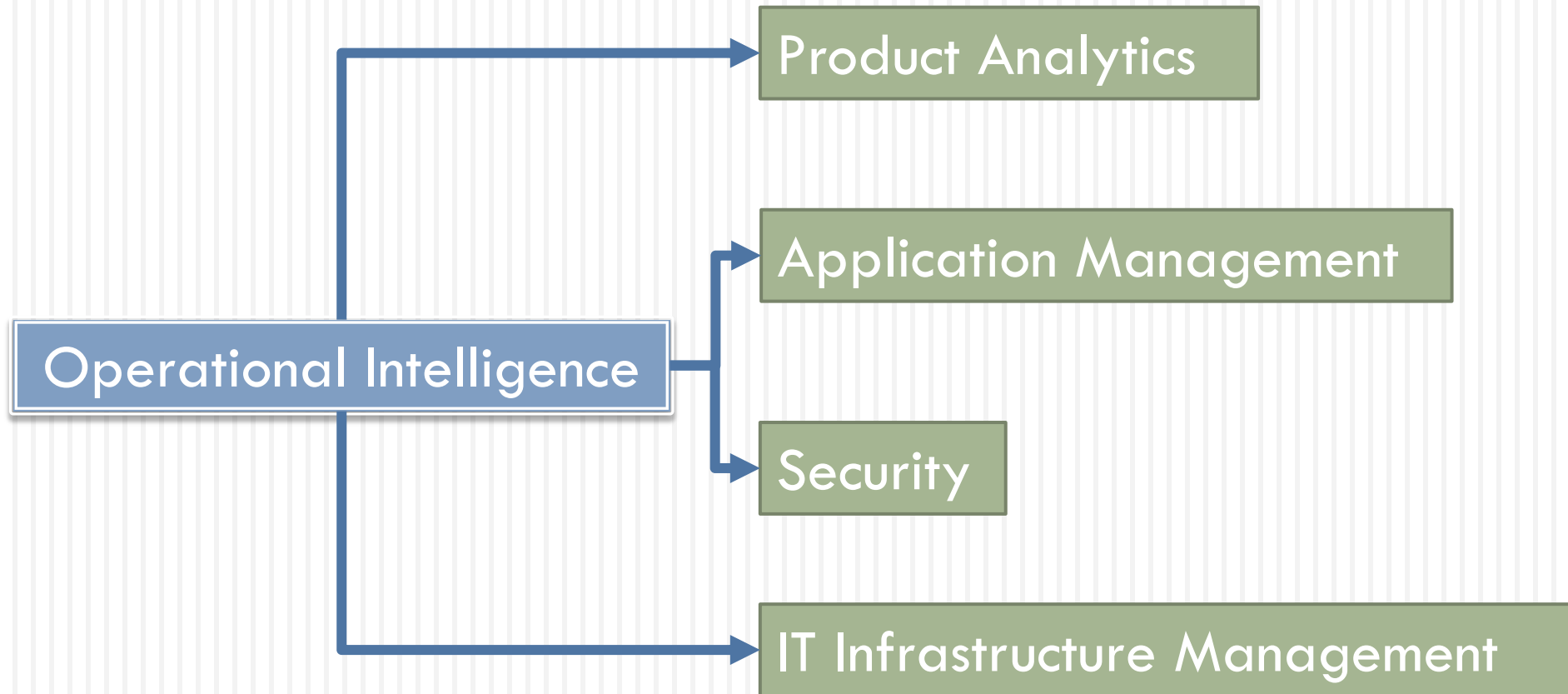# Market Watch

Lag Time

| OLTP, OLAP | Massively Parallel Batches(MPB), Hive, Pig | Business Intelligence |
| Complex Event Processing(Splunk), Loggly, IN-Memory DB | Massively Processing Real time (Storm) | Operational Intelligence |

Size of Data

OI - Business Solutions

Operational Intelligence

- Product Analytics
- Application Management
- Security
- IT Infrastructure Management

Reference- splunk.com
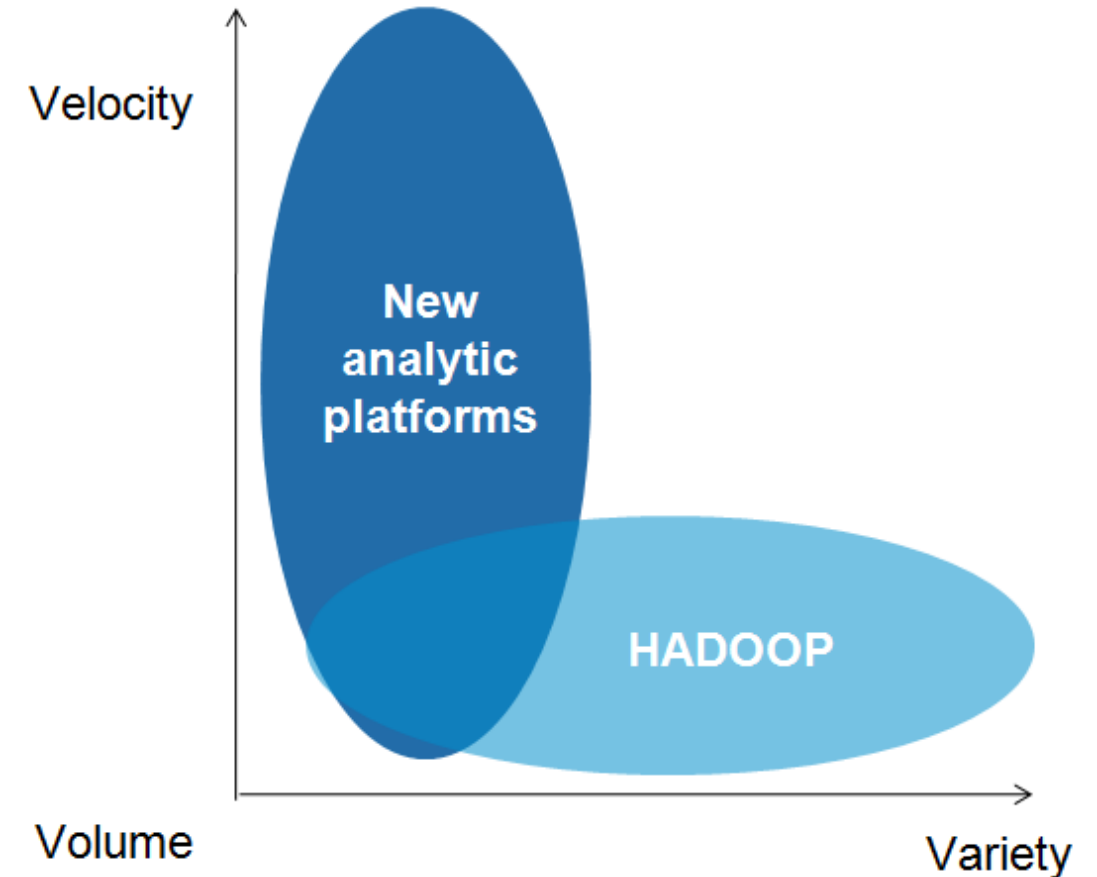
Product Analytics – Effecting Top Line Revenue
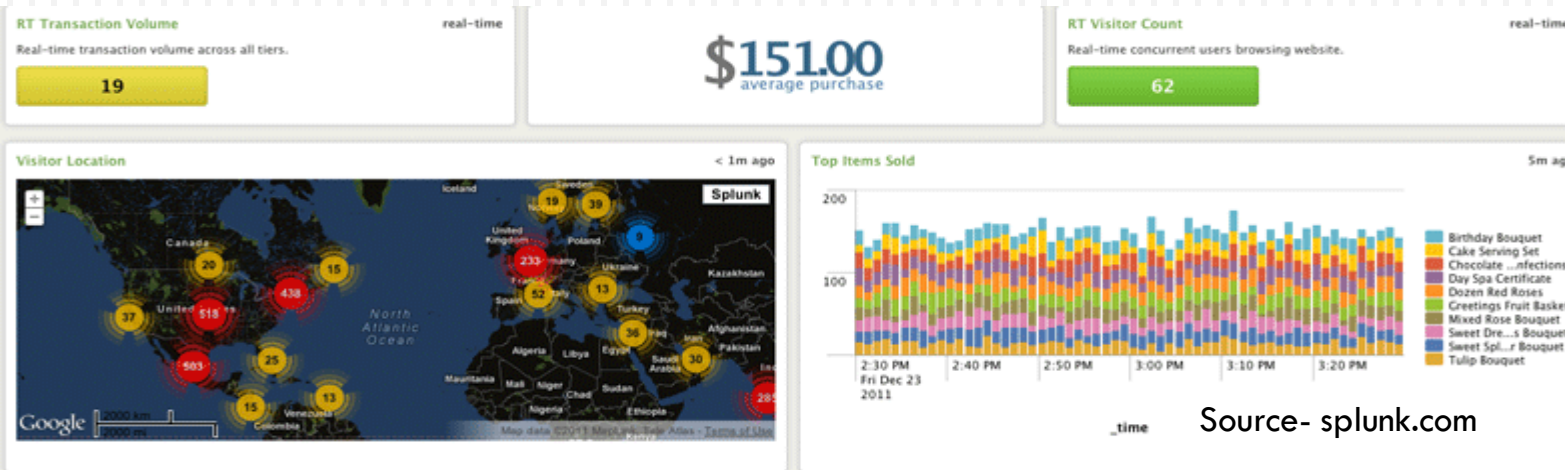
Challenges

- Lack of Visibility in Business Behavior
- A/B Test Results in nearly a month.

OI Business Impact

- Visibility into critical business metrics (KPI- Units Sold, Revenue, Visitors, Conversion rate & Opportunity).
- A/B Test Results in 2 hours effecting Top-line revenue.

Source- splunk.com

Key Takeaway –
Understand Real-time Business Behavior.

Application Management – Effecting Customers

Challenges

☐ Limited visibility across their entire application stack.

☐ Needed to be able to quickly respond to potential customer issues such as login, Flight search or payment errors.

OI Business Impact

☐ Helps in Root Cause Analysis and enables them to triage errors and focus on trends and solutions -- cut MTTI/MTTR by 95%

☐ Troubleshooting tools and applications across complex environments and multiple
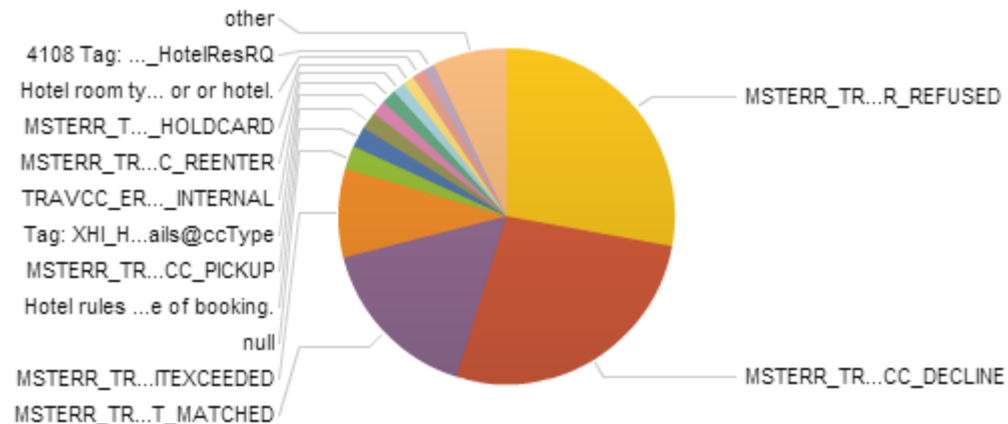
GDS Errors - Last 12 Hours
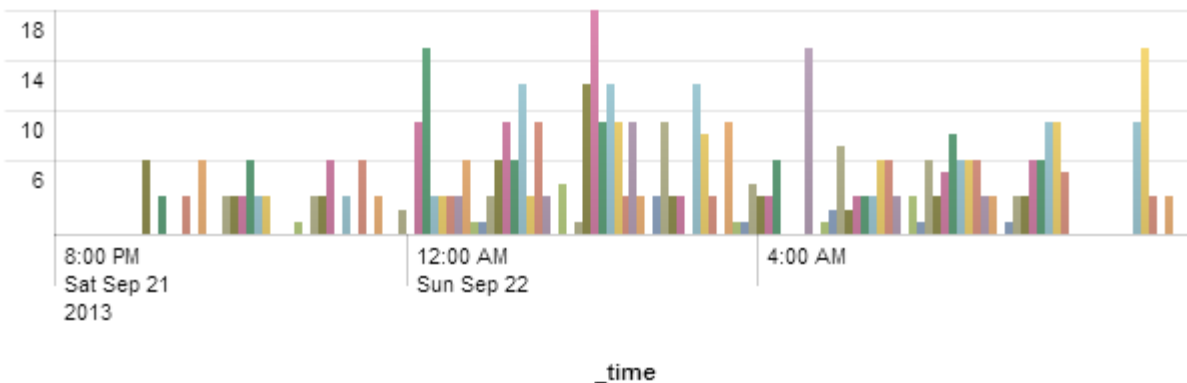
Key Takeaway –

Improve Customer Satisfaction &

Improve Opportunity Revenue

Security – Data Lost and Financial Impact

Challenges

☐ Rich database of information was under constant threat from attackers "scraping" data and abusing the payment part. We needed a faster, more efficient way to detect and mitigate fraud activities.

OI Business Impact

☐ Catching unknown threats and new zero-day attacks

☐ IP address and a timeframe, they analyzed and understood the pattern, then coordinated remediation to block the IP addresses.

Key Takeaway –
Save Intellectual Property and Data

# IT Infrastructure Management – Resource Optimization

Challenges

- IT organization was inhibited without a centralized view of its infrastructure

- Often missed items such as Apache 500 errors (An impact on the customer experience)

IIS Status Codes   6m ag

18,000	■ 400-Bad Request
14,000	■ 401-Unauthorized
10,000	■ 403-Forbidden
6,000	■ 404-Not Found
	■ 406-Not Acceptable
5:00 AM	■ 408-Request Timeout
Sun Sep 22	■ 413-Request Entity Too Large
2013	■ 500-Internal Server Error
	■ 502-Bad Gateway
_time	■ 503-Service Unavailable

OI Business Impact

- End-to-end visibility across their applications infrastructure to optimize capacity planning

- Enhanced the service levels and high availability of cloud-based service, accelerating operations response time from 24 hours to minutes.
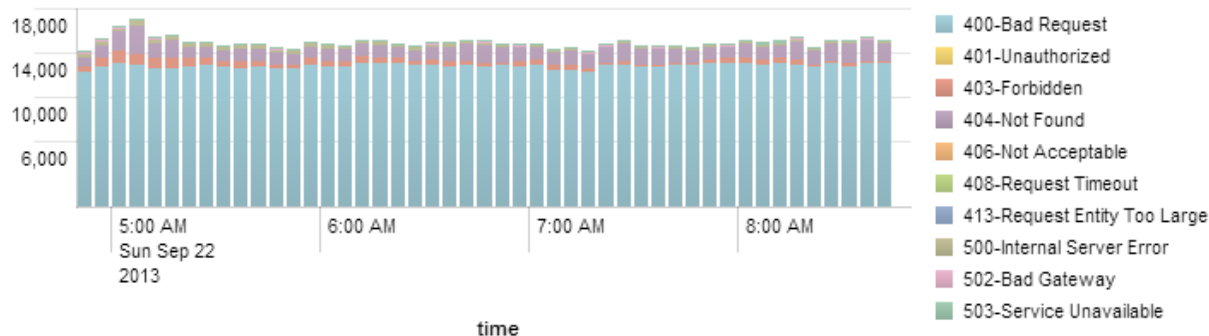
Key Takeaway –
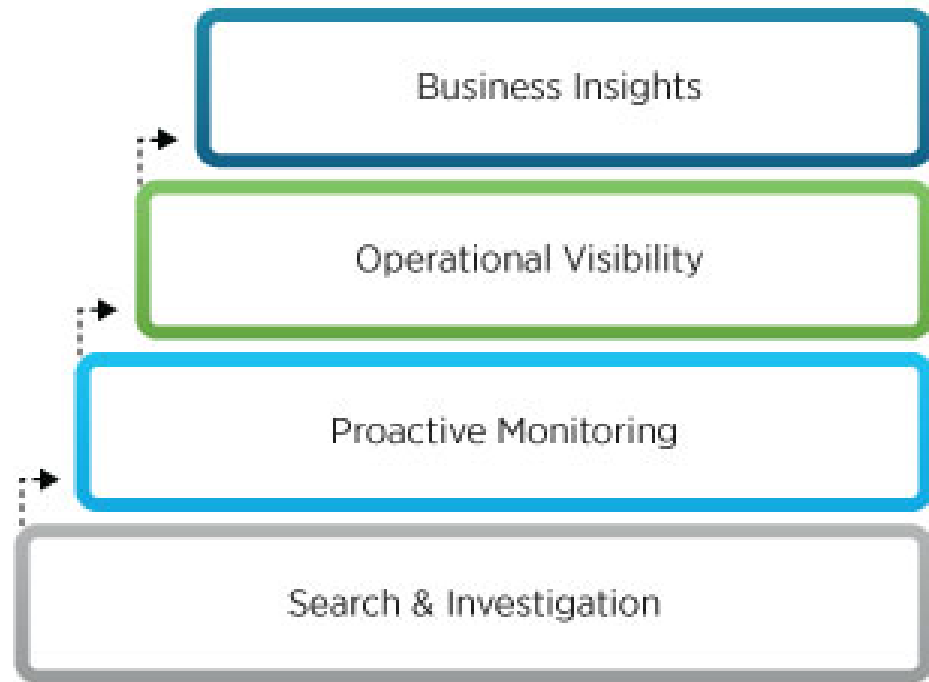*Optimize Infrastructure Capacity and Save Money!*

Final Wording - Reach to Ah ha Moment :D

Business Insights

Operational Visibility

Proactive Monitoring

Search & Investigation

Source- splunk.com

- ☐ Root Cause Analysis - Find and fix problems and investigate incidents across all your IT systems
- ☐ Keeps watch of specific patterns, trends in your machine-generated data. Trigger notifications in real time using alerts.
- ☐ Devise thresholds, SLAs, KPI's and models around it for Operational visibility.
- ☐ Make better business decisions based on real time business dashboards and act on situations

Q&A

Thank you!
Rohit Gossain