CLABBY ANALYTICS

Research Report

IBM Operations Analytics - Log Analysis: Getting the Most out of Your Operational Big Data

Introduction

Operational data, such as log files and system metrics, provides important information about operations and system performance by collecting data from a variety of sources — including networks, web servers, transactions, databases, middleware and applications. The value of this information is well understood, but businesses grapple with how best to harness this type of Big Data. Searching, correlating and analyzing operational data can present a big challenge in that these files comprise huge volumes (some enterprises collect over a terabyte of log file information each day) of random structured and unstructured system information. Storing these files is extremely costly, so businesses regularly discard this data, even though it can offer meaningful insight into trends in IT operations.

A manual approach to log analysis requires an abundance of time and resources to sift through logs looking for clues in order to troubleshoot systems and application problems. Not only that, with different administrators analyzing siloed data, it can be difficult to view one piece of information in the context of another, leading to inaccurate problem diagnosis and longer resolution times. An automated solution with built-in correlation and analytics is much better equipped to zero in on problems proactively, identifying issues so they can be resolved before affecting users.

In fact, a study performed in 2014 by the IBM Institute for Business Value revealed that 40% of businesses surveyed are using big data and analytics to improve operational efficiency. Businesses that fail to take advantage of this big data resource and analytics capabilities will be at a competitive disadvantage.

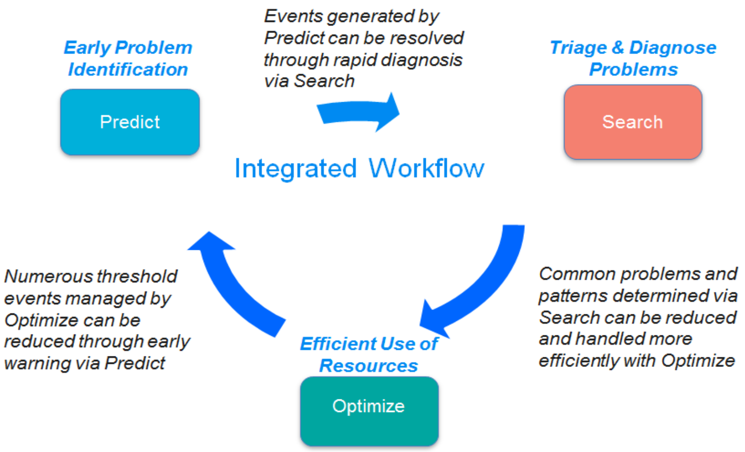
As a result, new, automated “analytics-based” log management tools have come to market. These programs have the ability to store, search and analyze a broad range of logs and other operational data including metrics, events, and machine data to identify the source of slow user response times, server/application performance issues, memory issues, high CPU network or disk usage, problems with database queries, and even potential security breaches. Stakeholders across the organization (including IT operations and support, application developers, and line-of-business managers) can take advantage of these tools to proactively manage infrastructure and applications through consolidated dashboards and reporting. With a range of pricing and delivery models–including SaaS and on-premise – organizations both large and small can reap the benefits of collecting and analyzing log data: faster problem resolution; improved accuracy and efficiency; and problem avoidance.

In this *Research Report*, *Clabby Analytics* looks at the IBM Operations Analytics Portfolio with a focus on IBM Operations Analytics – Log Analysis (Log Analysis). We describe how this solution works and how it can be used with other IBM Operations Analytics tools to learn systems behaviors and identify potential problems. We will highlight customers who are using the solution to lower management costs, improve service and make more informed business decisions. And finally, we will take a look at IBM's positioning relative to competitors such as Splunk , SumoLogic and ElasticSearch. We found that that there are distinct differences in pricing and delivery models; breadth and depth of capabilities; and in levels of integration with complementary solutions among these offerings.

IBM Operations Analytics Portfolio

IBM offers a unified, integrated suite of operational analytics tools that can rapidly search large volumes of structured and unstructured to predict, isolate and suggest a fix for problems before they become service impacting — as well as identify trends that can provide business insight for stakeholders across the organization (See ***Figure 1***, below).

Figure 1 – Integrated Value of Operations Analytics



Early Problem Identification
Predict
Events generated by Predict can be resolved through rapid diagnosis via Search
Triage & Diagnose Problems
Search
Integrated Workflow
Numerous threshold events managed by Optimize can be reduced through early warning via Predict
Common problems and patterns determined via Search can be reduced and handled more efficiently with Optimize
Efficient Use of Resources
Optimize

Source: 1BM 2015

This portfolio includes:

- *Predict: IBM Operations Analytics - Predictive Insights* (now available on-premise and as-a-service) leverages IBM's strength in analytics using Watson Research and customized algorithms, cognitive intelligence (e.g. behavioral learning) techniques and automated thresholding to detect anomalies in performance data, thereby alerting to the emergence of service impacting issues before traditional threshold based tools could have detected the problem (additional information here);
- *Search: IBM Operations Analytics - Log Analysis* offering (discussed in detail in remainder of this document); and
- *Optimize: IBM Netcool Operations Insight* combines infrastructure and operations management into a single unified view of business applications, virtualized servers, network devices and protocols, internet protocols, and security and storage devices for more efficient resource usage. Real-time alarm and alert analytics, combined with broader historic data analytics enable a 98% reduction in critical events according to IBM.

  Log Analysis Service Desk Extension helps customers who have a help desk solution cut down their Mean Time to Resolution (MTTR) for incidents and service requests, improve first-time response service levels, and user satisfaction. It provides analytics on help desk tickets, such as understanding incident or service request hot spots and trending issues at-a- glance to support more informed personnel assignments and speedy problem resolution. The extension supports a range of help-desk tools including IBM Control Desk, BMC Remedy, and ServiceNow.

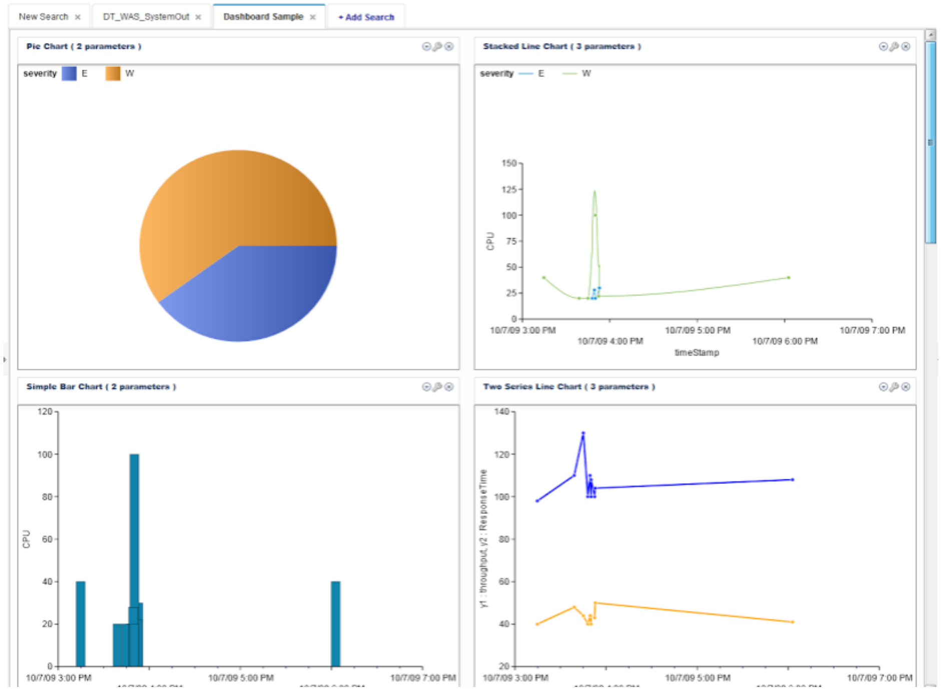
IBM Operations Analytics-Log Analysis – A Closer Look

Log Analysis is a global solution (support for 10 languages) that provides the insight required to rapidly diagnose and resolve application, infrastructure, and networking issues using all your operational data including log files and system metrics. IBM's proven search technology, which is employed across the IBM Analytics portfolio, combined with the depth and breadth of Watson-based advanced text analytics provides a platform that collects, consolidates, correlates and analyzes operational data in context, historically and in real-time. This solution analyzes a wide range of data types including:

1. *Monitoring and metrics* – such as linking metrics in the context of log search results and correlating data from IBM and third-party monitoring tools;
2. *Log files*– middleware, database, transactions, systems, applications, zLinux and zOS;
3. *Events* – such as searching event data and detecting seasonality trends
4. *Service desk solutions* – such as searching and analyzing service tickets, or searching event logs and documents using a ticket context to look for specific types of events.

Data is collected and annotated and indexed to facilitate quick, meaningful searches. Embedded correlation and analytics identify patterns, errors and trends — consolidating all collected data types in order to diagnose and solve problems. Data is visualized through a single pane of glass dashboard that automatically refreshes as information changes. A range of charts and graphs filter and display information that can identify times of day when response time is slowest or even which customers are impacted by slow response time, for example. Real-time alerting communicates this information quickly to administrators who can drill down to identify root cause. A unified view across infrastructure provides more accurate, more comprehensive anomaly detection and problem resolution. See ***Figure 2***, below.

Figure 2 - Log Analysis Dashboard



Source: IBM 2015

Other features include:

- *Insight Packs-* These add-ons extend the capabilities of Log Analysis to provide easier and deeper analytical coverage to additional applications, infrastructure elements, and storage used in IT operations. The Insight Pack provides a set of related queries, routines, and programs for ingesting, transforming, visualizing, and analyzing content from a data source (application, operating system, technology, or solution). They offer content specific visualization and insights with pre-configured custom dashboards, tailored content ingestion programs, quick searches and embedded analytics for handling log content. Some of these Insight Packs are provided with the base product, some are available by direct download and others by purchase of a standard or premium additional license. Examples include Microsoft Exchange Server, Siebel/Oracle database, IBM DS8000, IBM Tivoli Storage Manager, Websphere Application Server (WAS), DB2, and Cisco UCS Manager. In addition, product documentation provides tools and techniques that enable customers to create their own Insight Packs.
- *Expert Advice –*Log Analysis searches identify the source of a potential problem and determine root cause. From the same management screen, a link to expert advice (scored entries that match problem and relevance), IBM support or an integrated trouble ticket dashboard, is provided. A direct link to a community forum is also available as a way to pose specific questions to product experts.
- *Searchability -* After files are collected, parsed, annotated and indexed — text analytics and search features enable natural language, free-form and saved searches, searches along single or multiple log sources, and the use of search parameters, such as time or log source.
- *Integration with other IBM solutions*
    - Out-of-the-box integrations to IBM Application Performance Management (APM), IBM Tivoli Monitoring (ITM) or 3rd-party monitoring solutions enable monitoring data to be correlated and analyzed as another source of operational data, to proactively identify performance issues.
    - Log Analysis is included in IBM's Operations Management solution, Netcool Operations Insights, to provide event analytics. This can be extended to include additional data sources for rapid diagnostics.
    - IBM Control Desk integration improves MTTR and, as a result, customer satisfaction using analysis and categorization of ticket data to determine hot spots and trends in your service desk.
    - Integration with IBM Big Insights Hadoop platform enables Log Analysis users to economically search historical data to analyze long term trends and patterns. IBM Big Insights uses compression algorithms to provide a more efficient method for long term data storage, storing months or years of data from annotated log files (up to hundreds of terabytes) on the Apache Hadoop platform. Data can be easily searched to meet compliance requirements or for historical analysis. IBM's Log Analysis product uses IBM Big Insights for operational data storage, saving licensing costs for customers wanting long term access to operational big data.

Customer Use Cases

Barclay's Bank

Barclays Bank was having issues with ad-hoc, siloed processes for analyzing customer's channel (ATM, mobile, online etc.) usage reports. Some channels were using excel spreadsheets, while others were using Cognos to collect and analyze data, making it difficult to see the "big picture". A lack of a central hub for collecting and searching data was affecting accuracy and timeliness of analysis. Using Log Analysis, reports from each channel are consolidated and correlated for centralized analysis, revealing information about application issues and customer behavioral patterns. With Log Analysis, Barclays is getting results 60% faster for better insight that allows them to make changes on-the-fly in individual channels in order to optimize availability and channel usage.

China Merchants Bank

China Merchants Bank (CMB) was suffering service outages and transaction downtime, costing the bank money and damaging their reputation. They were looking for a scalable platform that would help manage events from these transactions to prevent outages and downtime, as well as reduce mean-time-to-repair (MTTR). Log Analysis was able to provide CMB with a system that provided: log collection; storage and retrieval; trade volume monitoring; and application logs, metrics and reports in an easy-to-use dashboard. Integration with IBM Operations Analytics – Predictive Insights provided self-learning behavioral analytics and correlation that could automatically track numerous Key Performance Indicators (KPIs) to prevent outages before they occurred.

Comparing Log Search/Analysis Offerings

Competitive offerings that we believe compete most directly with IBM's Operations Analytics-Log Analysis include:

1. Splunk
2. SumoLogic
3. ElasticSearch

Splunk

Splunk is an easy-to-use, easy-to-install operational intelligence software platform that collects, indexes and searches any IT streaming machine and historical data from virtually any source in physical, virtual and cloud environments, including event logs, web server logs, live application logs, network feeds, archive files, mobile devices, sensors etc. in real time. Splunk also includes single-pane-of-glass visualization, reporting, ad-hoc query and charting capabilities for both real-time and historical data. It enables users to troubleshoot and monitor infrastructure for potential performance problems and supports a wide range of log management use cases including log consolidation and retention, security, IT operations troubleshooting, application troubleshooting and compliance reporting.

The solution is available as a software download or cloud-based service: Splunk Enterprise (on-premise); Splunk Cloud (software as-a-service); Splunk Light (up to 5 users – single server). Splunk has recently announced additional products: Splunk Enterprise Security, and Splunk User Behavior Analytics, aimed at detecting security breaches by identifying cyberattacks and insider threats.

Strengths:

- Splunk's broad range of searchable supported machine data
- Searches historical and real-time data
- Licensing options: Splunk Enterprise; Splunk Cloud; Splunk Light
- Dashboard/Reporting/Charting features
- Many available plug-in Splunk apps

Weaknesses:

- Cannot correlate multiple types of metrics to predict anomalous behavior
- Additional solutions in portfolio primarily focused on security
- High-scale environment requires the installation and configuration of a dedicated cluster.
- Price
  - Splunk Enterprise pricing based on peak usage rather than average usage in a 30 day timeframe. So, for example, if average usage is 30 Gb/day, but peak periods require 100 Gb/Day, for most days you will pay for more than is being used.
  - Add-on functionality priced separately Hadoop connection, anomaly detection, Splunk Enterprise Security, Splunk User Behavior Analytics), adds to base price.

    For example, Hunk (Hunk Splunk Analytics for Hadoop and NoSQL Data Stores) pricing is based on the number of TaskTrackers (Compute Nodes) in Hadoop clusters. Pricing for a one-year term license of Hunk starts at $3,000 per Hadoop TaskTracker/Compute Node with a minimum of ten TaskTrackers/Compute Nodes ***which results in an additional cost of $30K***.

Sumo Logic

Sumo Logic is a SaaS-only log analysis tool with full-stack visibility across applications and infrastructure. Sumo Logic collects a wide range of log data from a wide range of sources (including Java, .NET, Hadoop, Apache, JBoss, Oracle WebLogic, WebSphere, IBM DB2, Microsoft SQL Server, Linux, Windows, VMware, Amazon, Salesforce, Cisco etc.) and uses patented analytics to identify patterns and anomalies. In a February 2015 interview in Forbes, Splunk CEO Ramin Sayar stated that SumoLogic is pursuing three markets: (1) DevOps (2) Security and (3) IT infrastructure and operations testing services.

Lightweight collectors receive log data from one or more sources, compress it, encrypt it, and send it to the Sumo Cloud, in real time. A single collector can collect up to 15,000 events per second and provides fault tolerance during network or service outages. Collector and search API's enable administrators to develop and integrate other data sources.

Browser-based search capabilities for collected log data enable forensic analysis, troubleshooting, and system health checks. Real-time alerting allows users to establish baselines for key metrics which, when exceeded, notify administrators – helpful for testing software prior to roll-out or to detect security breaches A dashboard interface can be used for queries or to create charts. Other features include auto-scheduled searches, data retention periods, and user profiles.

Pricing: Free lite version; $90/month (Professional) for each GB/day needed $150/month (Enterprise) for each GB/day needed .

Strengths:

- SaaS license model lowers barriers to entry
- Broad range of search and reporting capabilities
- On-demand scalability
- Enterprise class features: real-time alerts, role-based views, automated searches, and charting

Weaknesses:

- Pricing can become high at scale
- Hosted version relies on AWS
- May not support obscure, less mainstream tools and data sources
- No integration with broader product set
- Movement of data to and from cloud may be cumbersome

Elasticsearch

An element of the ELK stack (Elasticsearch , Logstash, Kibana) driven by the open source vendor, Elastic, Elasticsearch is the log search/analytics engine component of the combined solution, with the ability to perform real-time data extractions and data analytics from structured or unstructured data sources. Logstash allows parsing of data and converges on a common format to prepare the data for further analysis. Kibana is a simple visualization tool—a log data dashboard that includes point-and-click pie charts, bar graphs, trendlines, maps and scatter plots that can be used to illustrate and identify trends and patterns in collected data. These three products have been designed to work together, but each is its own project and can be used separately or with other products.

ElasticSearch is easily scaled-out with nodes added to the cluster automatically and transparently. Node failures are automatically detected and removed. Elasticsearch has full-text search (using Apache Lucene open-source Java full-text search library) with features such as multi-language support, an extensive query language for structured and unstructured data, geolocation support, context-sensitive suggestions, and autocompletion.

Strengths:

- Free open–source solution
- Works with other Open Source solutions: native Hadoop integration, APIs, ELK
- Quick and easy horizontal scaling

Weaknesses:

- No support for scheduled searches
- No real-time alerting , triggers or thresholds
- Basic charting only
- “Free” can be become expensive quickly when skilled engineers are required to integrate with other solutions and maintain code
- No built-in use cases

Figure 3 (next page) provides a side-by-side comparison of these offerings:

Figure 3: A Side-by-Side Comparison of Leading Log Analysis Environments

	IBM Log Analysis	***Splunk***	***Sumo Logic***	***Elasticsearch***
Real-time alerts	Yes	Yes	Yes	No
Searchable/data/features	Searches metrics, alerts, configuration information, events, logs (including mainframe), traces and documentation. Broad range of search types and features	Searches machine and historical data. Reporting, charting queries for real-time and historical data	Searches broad range of log data but may not support less popular tools	Requires integration with ELK solution
Use Cases	Extensive, in Insight Packs	Yes	Some	No
Embedded Analytics/Correlation	Some. Note significant analytics and correlation capabilities extended with IBM Predictive Insights	Some analytics but limited correlation capabilities	Some	No
In context expert advice	Yes- links to multiple ranked potential solutions, IBM support or user forum	No	No	No
Additional Capabilities/Optimization/Integration for IT Operations Analytics	Broad support across multiple technologies (Netcool Operations Insight, IBM APM and ITM, IBM Predictive Insights, IBM Control Desk, Hadoop)	Other products in the portfolio largely data security related rather than focusing on operational data	Limited	Integrated with other Open Source solutions but each is maintained as a separate project
Pricing/Delivery Model	On-premise with many additional features free-of-charge. Aggressively priced below competition.	Both on-premises and SaaS. Enterprise pricing high and many add-on features are priced separately	SaaS only	Free Open Source

Summary Observations

In evaluating Log Analysis products, it became clear that the biggest differentiators for IBM Operations Analytics - Log Analysis are the level of integration it has with other IBM solutions; its breadth and scope of searchable data (even mainframe logs); and embedded analytics, all working together in order to quickly perform root cause analysis and to prevent issues. This offering is also unique in its ability to provide real-time access to expert advice through multiple channels, saving administrators the time and effort required to diagnose a common problem (or maybe not so common). Other differentiators include its pricing model (based on the average data consumed, not on peak period) and Insight Packs which enable users to quickly and easily get specific, pre-configured custom dashboards, reports and queries for a particular technology, improving administrator efficiency and lowering costs.

There are many players in the operational analytics market, offering a range of solutions for searching, correlating and analyzing operational Big Data. When evaluating these products, look carefully at: (1) pricing models (2) range of searchable data and features (3) integration with other complementary solutions (4) use cases and (5) levels of support. In our opinion, IBM's Operations Analytics solutions offer an entire, integrated suite of products that ticks all the boxes, while still providing a cost-effective solution for a range of customers, both large and small.

Clabby Analytics
http://www.clabbyanalytics.com
Telephone: 001 (207) 846-6662

Clabby Analytics is an independent technology research and analysis organization. Unlike many other research firms, we advocate certain positions — and encourage our readers to find counter opinions —then balance both points-of-view in order to decide on a course of action. Other research and analysis conducted by Clabby Analytics can be found at: www.ClabbyAnalytics.com.