Research Report

IBM SmartCloud Analytics: Making the System Analyze Itself

Introduction

There is a major shift taking place in information systems management marketplace as new, automated "analytics-based" operational management programs come to market. These programs have the ability to look at information systems activities; learn system behavioral patterns; find abnormalities in collected systems data and predict future systems behavior based-upon the patterns and problems that are identified when analyzing operational data. Furthermore, these programs can analyze data exponentially faster than humans – leading to faster problem resolution, improved accuracy, as well as problem avoidance.

For years, systems vendors and management software makers have focused on simplifying user interfaces and on integrating management portfolios. These foci have improved information technology (IT) management productivity – and have helped lower data center management costs. But today these same vendors are turning their attention to the automated analysis of operational analytics data in order to further lighten IT manager/administrator workloads and to reduce human errors. These new operational analytics products are being used in conjunction with traditional graphical user interface/integrated management software products to automatically troubleshoot systems/software-related problems as well as to predict problems before they may occur.

The thing we like best about this new generation of automated analytics tools is that they make the system do the thinking. Until the arrival of operational analytics tools the burden of trouble-shooting systems/storage/network-related problems has fallen on already overworked information systems managers and administrators. These information technology (IT) professionals have had to develop metrics and set thresholds for system performance – and they have had to use event monitors and analyze log files in order to resolve IT issues. New generation operational analysis tools help identify problems more quickly than human operators – enabling new insights to be garnered from large, unstructured repositories; help isolate problems by bringing relevant structured and unstructured data into focus such that root cause analysis can commence more rapidly; and, these tools also help repair problems more quickly by giving operators the information they need without having to manually sift through volumes of log files and other operational data. With the time saved, IT managers and administrators are free to focus on higher-value management and tuning activities.

In this *Research Report*, *Clabby Analytics* discusses IBM's SmartCloud Analytics environment. We describe how it can be used with other IBM operational analytics tools (monitoring and log analysis) to learn systems behaviors and identify potential problems. We discuss how several IBM customers are using SmartCloud Analytics to lower management costs and improve service. And we also take a look at how SmartCloud Analytics is competitively positioned against competitors such as Splunk (with partner Prelert); Netuitive, and Hewlett-Packard (with IT Operations Analytics). Readers will find that there are several distinct differences in configuration (no more threshold setting); in establishing service models (none are required with SmartCloud Analytics); and in problem determination and learning behaviors. We conclude this report with a strong recommendation: *enterprises that wish to significantly reduce human labor costs related to IT management need to evaluate IBM's SmartCloud Analytics operational analytics offerings*.

The Big Change: Using Analytics to Quickly Resolve Problems and Predict Future Issues
Until recently, the approach used by IT managers and administrators to identify performance problems and to resolve IT issues has involved establishing metrics, checking event monitors and pouring through volumes of log files to identify anomalies. Metrics were established around elements such as availability, response time, latency, completion time, and instruction path length – these metrics have been used to analyze system performance. Event monitors have been used to track system behavior. And log files are used to scrutinize system activities – providing clues that can lead to problem resolution. The problem with each of these approaches is that they are too human labor intensive:

- *Systems tuning* – metrics need to be gathered from a number of sources in order to create a composite picture of system behavior before a system is tuned;
- *Event monitors* – have needed to be programmed in order to prioritize and track the events a data center manager may wish to monitor; and,
- *Log files* have presented the biggest challenge in that they are voluminous records (some enterprises collect over a terabyte of log file information each day) of random structured and unstructured system information. These files require a lot of human brainpower to sift through logs looking for clues in order to troubleshoot systems and application problems.

Over the past few years, however, machine-driven analytics has started to make its way into systems management marketplace. Numerous vendors now offer a variety of operational analytics products – including IBM, Hewlett-Packard (HP), Oracle, SAP, and dozens of smaller companies such as Netuitive and Splunk. Many of these vendors offer individual software products that tackle problems ranging from predictive outage avoidance to faster problem resolution; from improved visibility and insight into system resource relationships to optimized performance. IBM offers an entire, integrated suite of products that perform all of these functions (see Figure 1).

Figure 1 – The Use of Analytics in Systems Management

Source – IBM Corporation, May 2014

IBM's Operational Analytics Portfolio

IBM offers a unified suite of operational analytics tools that can rapidly search large volumes of structured and unstructured data (IBM is particularly strong in reading unstructured streamed data); that can help optimize IT infrastructure; and that can be used to predict problems before they occur. This portfolio consists of three elements:

1. *Predict: IBM SmartCloud Analytics - Predictive Insights* technology; and,
2. *Search: IBM SmartCloud Analytics -   Log Analysis* offering; and,
3. *Optimize*: IBM *SmartCloud Monitoring and IBM's SmartCloud Virtual Storage Center* (SCVSC – for more information on this offering, see this *Clabby Analytics* report).

The following subsections will describe each of these offerings in greater detail.

IBM's SmartCloud Analytics Predictive Insights

The origin of IBM's SmartCloud Analytics Predictive Insights environment can be linked with the development of IBM's Watson cognitive computing environment. Watson (the cognitive computing environment developed by IBM Labs to analyze very large volumes of structured and unstructured data) was the system used to show how a language-based cognitive computing environment could compete with human intelligence (Watson bested two Jeopardy game show champions in a system versus human knowledge testing event). IBM SmartCloud Analytics builds on this Watson analytics base by using customized algorithms to detect systems anomalies in performance data, and by predicting future system/application behavioral outcomes.

> *Clabby Analytics is a big believer in the Watson cognitive computing environment. The way we see it, Watson will – over time – become more and more aware (cognitive) of its surroundings and will be able to conduct a lot of analytics on its own using a wide range of advanced algorithms. Watson-based systems will become more and more able to take corrective actions by themselves – while also becoming an extremely valuable tool to systems managers and administrators by automatically finding issues and seeking guidance from humans on how to correct those issues. Watson will have the ability to learn from these engagements and will eventually be able to take on more and more tedious systems management tasks – freeing humans up to focus on adding value in application development, business process orchestration and in other high-value activities.*

As for its structure, IBM's SmartCloud Analytics Predictive Insights consists of two components:

1. An event viewer that consolidates detected anomalies.
2. Drill-down analysis view where you can review what has caused the event, the situation leading up to the event, and performance metrics correlated to this event

The way that this product works is that it "learns" system/application behavior patterns by comparing how a system behaves over a period of time (it analyzes time series data). By comparing "good" system behavior to "bad" system behavior, IBM SmartCloud Analytics is able to identify problems before they may occur – providing IT managers with a means to identify potential issues and fix those issues before they become an even bigger problem.

What makes IBM''s SmartCloud Analytics environment special is that it has been designed to exploit IBM''s InfoSphere Streams environment – an environment that can stream data to applications such that structured and unstructured data can be read in real time. Using streaming techniques, not all data that is streamed has to be stored and analyzed – instead the data streams

through a system and only the relevant information (such as anomalies) is captured and analyzed. This means that valuable system cycles are not used pouring through logs and other data sources to find anomalies because only the relative data is captured and acted upon. Further, it means that terabytes of storage can be saved (because, again, only the relative data is captured and stored).

> *IBM is particularly strong in comparison to its main competitors when it comes to data streaming technology. The ability to stream data and a rich suite of predictive analytics algorithms gives IBM a strong competitive edge in predictive analytics.*

Another element that makes IBM's SmartCloud Predictive Insights special is that it lightens deployment and management tasks by not requiring ongoing maintenance and or operator threshold setting. Further, setting up predictive, descriptive and decision service models is not required. In other words, the deployment and operation of IBM's SmartCloud Predictive Insights does not require a lot of operator intervention and ongoing maintenance.

IBM's SmartCloud Analytics – Log Analysis
IBM describes its SmartCloud Analytics - Log Analysis offering as an environment that "collects large volumes of obscure unstructured data and transforms it through analytics into actionable intelligence". This product has been designed to simplify anomaly detection using logs and metrics.

The product itself consists of three components:

1. A dashboard from which various activities are launched;
2. "Insight Packs" (Environment specific extensions that provide a collection of display tools that help managers and administrators visualize systems problems); and,
3. Intelligent support document integration using advanced text analytics techniques and expert advice portal integration.

The beauty of this product offering is that it can be used to analyze a wide breadth of data types including:

1. *Monitoring and metrics* – such as linking metrics in the context of log search results;
2. *Problem anomaly/detection* – such as detecting and alerting systems managers base on trends observed in the logs;
3. *Events* – such as searching event data and detecting seasonality trends
4. *Discovery and application performance management* – such as refining the scope of a search in logs and documents using topology and configuration context to help isolate performance problems;
5. *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.

> *SmartCloud Analytics Log Analysis enables IT managers and administrators to more quickly isolate and repair problems by using analytics – rather than human brain power. Systems can sift through large volumes of structured and unstructured information and guide managers/administrators to problem resolution. As we evaluated other Log Analysis products we concluded that the two biggest differentiators for IBM's SmartCloud Log Analysis offering are the level of integration it has with other IBM operational analytics/system management offerings – and its ability to work with other trouble-shooting tools including*

metrics, configuration information, events, logs, traces and topology products in order to perform root cause analysis. Other differentiators include its pricing model (based on the average data consumption model, not on peak period use); its advanced text analytics facilities (only product in the industry that uses advanced text analytics to extract insights from unstructured sources); and its linkage with IBM's Big Data Platform.

IBM's SmartCloud Monitoring and SmartCloud Virtual Storage Center

Cloud management entails knowledge of systems, networks, and storage devices that make up a given cloud environment. To work together efficiently these resources need to be part of an interconnected cross-system cloud architecture that can be easily analyzed to trace the root cause of performance or other issues. IBM's SmartCloud management tools (that include IBM's Smart-Cloud Application Performance Management environment, SmartCloud Monitoring, SmartCloud Provisioning, and SmartCloud Control Desk) have been designed to work in concert to provide a unified view of application activity in the cloud, to allow for application troubleshooting and tuning – as well as to provide the tools needed to take corrective actions if necessary. SmartCloud Virtual Storage Center enables centralized management of heterogeneous storage while providing storage virtualization – so that storage resources can be pooled and provisioned and de-provisioned on-demand based on workload requirements.

IBM SmartCloud Monitoring

Cloud infrastructure offers many benefits to users, including flexibility, scalability, improved efficiency and cost reduction. Effective capacity planning is one of the keys to realizing these benefits and to exploiting the full potential of the cloud. Capacity planning tools help reduce hardware and software license costs as well as providing savings in energy costs and floor space requirements. Workload performance can be improved by optimizing resources and workload placement.

IBM SmartCloud Monitoring manages both physical and virtual cloud infrastructure and provides performance analytics capabilities and capacity planning tools to optimize workload performance and availability. Visibility through health dashboards monitors host performance and resource availability.   User configurable policy-based analytics enable intelligent workload placement to ensure that workloads are handled quickly and efficiently. IBM SPSS predictive analytics are included in IBM SmartCloud Monitoring for support of non-linear forecasting and data modeling that examines complex relationships in performance data for better predictions of trends, more accurate forecasting, and smarter decision making.

Smart Cloud Monitoring's "what-if" capacity analysis can model the impact of infrastructure changes, reducing risk and helping to improve availability. Energy and server costs are lowered by making optimal use of current capacity and planning the addition of incremental resources. IBM SmartCloud Monitoring protects enterprise investments in cloud technology by ensuring maximum efficiency in virtualization density while at the same time maximizing the performance of virtualized workloads.

IBM SmartCloud Virtual Storage Center

IBM's SmartCloud Virtual Storage Center includes a storage hypervisor, centralized management and application-aware snapshots for near instant back-up and recovery. The hypervisor isolates the underlying hardware, enabling services (such as I/O caching, thin provisioning, deduplication, com-pression, etc.) to be delivered and managed on a wide range of heterogeneous storage. This allows the pooling of storage from virtually any storage platform from any vendor (EMC, HP, Dell, Sun, Hitachi, NEC, IBM and others) and dramatically improves storage utilization. Services such as I/O

caching, thin provisioning, compression, automated tiering, application-integrated snapshot/mirroring, and mobility-driven disruption avoidance, are delivered consistently regardless of storage array choice, improving efficiency and easing management. Centralized management through an easy-to use GUI includes valuable features beyond storage hypervisor management. Analytics simplify management as well as automate configuration, storage tiering and data migration.

Comparing Predictive Analytics Offerings
Competitive offerings that we believe compete most directly with IBM's SmartCloud Predictive Insights offering include:

1. Splunk with Prelert
2. Hewlett-Packard with IT Operational Analytics; and,
3. Netuitive.

Splunk/Prelert
Splunk is an easy-to-use, easy-to-install software platform that collects and indexes any IT streaming machine and historical data from virtually any source including event logs, web server logs, live application logs, network feeds, archive files, mobile devices, sensors etc. in real time. The latest version also offers data correlation, visualization, reporting and analytics capabilities. The Predictive Analytics dashboard (from PRELERT) searches for anomalies in Splunk's data, to provide statistical information about the results, and based on preset thresholds, identify anomalies that could indicate potential problems.

Using the Splunk framework and GUI, Prelert Anomaly Detective's self-learning techniques automatically establish models of normal behaviors it observes in huge volumes of data (aggregated through Splunk). Pattern detection algorithms sort both historical and real-time data into discrete activities and then statistical analysis is used calculate a normal range of values of a given key performance indicator (KPI). Values outside the range are identified as anomalies and correlated results illustrate the severity or impact of the out-of-range data that could pose performance problems/security threats. By identifying causal chains of activity, Prelert can also identify root cause.

Strengths:
- Splunk's broad range of supported machine data;
- Self –learning Prelert Analytics; and,
- Splunk is considered by some customers as a "best-of-breed" solution.

Weaknesses:
- Two products from two different companies lengthens sales and implementation process
- Cannot correlate multiple types of metrics and predict anomalous behavior
- Integration required may mean a lag in access to new features
- Prelert's size (20 people) may make them an acquisition target
- Point solution not integrated with a broader management portfolio
- Prelert OEM relationship with CA Technologies (Prelert technology is the basis for CA's Application Behavior Analytics)
- No support for 3rd party monitoring tools

HP Operations Analytics
HP Operations Analytics is a new capability of HP Business Service Management, which delivers advance analytic intelligence for structured and unstructured data. This intelligence includes: machine data, logs, events, topologies and performance statistics. HP Operations Analytics collects and analyzes operational data from a variety of HP products, including HP ArcSight Logger and the HP Business Service Management portfolio, as well as third-party data sources, such as Splunk and includes the following capabilities:

- *IT Search* - IT Search enables IT operations to use a simple query language to search across all collected IT operations data. This language facilitates and accelerates search across the big data stored datasets for logs, metrics, events, and topology and displays contextually relevant information through an analytics dashboard.

- *Visual Analytics*- Allow users to easily examine anomalies in specified time periods or by location or department. Data can be displayed in role-based views:
  - Operations manager view for analyzing event data, logs, and metrics;
  - Application, system, and network management team view for application/infra-structure problem isolation using guided analytics and the IT search capability for ad hoc use cases for problem isolation; and,
  - Data analysts and specialists view for the creation of ad hoc queries against metrics, events, topology, and logs.

- *Guided Trouble Shooting* - guided troubleshooting makes it easy to review past service performance to identify and isolate the exact time and origin of failures. In the event of a performance issue, a contextual step-by-step process can be initiated. First, relevant search topics will be suggested. Based on the selected topic, a list of potential issues will be prioritized and after identifying the issue, a fix will be recommended.

HP Service Health Analyzer is a predictive analytics tool that correlates data from HP Business Service Manager (BSM) to look at various performance metrics. This correlation, used with advanced analytics and algorithms, can forecast problems and prioritize issues based on business impact. It is self-learning and can analyze historical data to automatically create real thresholds—then combine anomaly data to predict potential problems. HP SHA is not currently integrated with HP Operations Analytics.

Strengths:
- Role-based visualization and expert advice in context.

Weaknesses:
- Self-learning only available in HP SHA
- Thresholds/service models must be set manually
- Weak integration across management product line
- Limited 3rd party monitor support
- Both HP SHA and HP Operations Analytics require HP BSM as a prerequisite. Correlation is established using BSM topology mapping
- Predictive accuracy depends on how up-to-date the BSM models are

Netuitive
Netuitive is the pioneer of predictive operational analytics. *Clabby Analytics* first reviewed the company in July 2010 and published an update in February 2012 available here. The Netuitive strategy has always been to collect data from businesses' existing monitoring tools including server,

storage, network, applications from CA Technologies, IBM,BMC and others, and then consolidate, correlate and analyze that data. Netuitive is also integrated with Splunk Enterprise to provide a rich source of machine data.

Netuitive uses a mathematical correlation and statistical engine to "self-learn" the IT infrastructure, collecting and correlating the data and establishing behavior profiles, classifying what is "normal" and what is "abnormal" in any given situation, enabling end-to-end proactive management. The more data that's collected, the "smarter" Netuitive becomes. Based on this analysis, Netuitive can identify anomalies and predict potential performance issues days in advance, speeding root cause analysis, problem resolution and reducing false alerts. Netuitive can take this breadth and depth of individual performance metrics, apply the correlation algorithms, and provide a composite health score. In this way, all these indicators of performance can be merged to answer the question, "Is my IT environment healthy?" Netuitive includes adaptive behavior profiles that show normal behavior profiles for hour of the day, day of the week, and cyclical events, correlating all key metrics together in context. Netuitive also can be tailored to weigh certain performance attributes more than others based on business requirements.

Other features include:

- Performance Assistant automates diagnostics and root cause isolation when performance problems occur and can drill down to each metric;
- Correlation Assistant keeps track of relationships between business service components and performance metrics to help with root-cause analysis as well as determine how a potential performance issue may impact users or applications;
- Smart Search capability can query millions of data points to identify systems by element attribute or performance levels; and,
- Best practice templates use expert knowledge and statistical techniques to help determine which metrics are the best indicators of performance for certain infrastructure components.

Strengths:
- Third-party monitor support – works with what customer already has;
- Support for wide range of data sources;
- Long history in self-learning analytics; and,
- Correlation of multiple metrics to determine business impact.

Weaknesses
- Point solution for operational analytics; no integration with broader portfolio
- Requires a business service model to be built and maintained to establish on-going correlation between metrics
- Time/resources required to build and maintain service model; not an "out-of-the box" solution

For a side-by-side comparison of all of these product offerings as contrasted with IBM's SmartCloud Predictive Analytics, see Figure 2, next page.

Figure 2: A Side-by-Side Comparison of Leading Predictive Analytics Environments

	IBM Predictive Insights	HP Operations Analytics	Splunk/Prelert	Netuitive
Problem Determination	Yes	Yes	Yes	Yes
Root cause Analysis	Yes through SmartCloud Analytics Log Analysis	Yes	Some	Yes
Set thresholds /service models	Set Automatically	Must be set by administrator	Set Automatically through Prelert	Service models and configuration is required.
3rd Party Monitor Support	Yes through embedded mediation tools and optional IBM Data Stage ETL tool.	No	No	Yes through adapters and APIs
Search diagnostics	Yes with integration to SmartCloud Analytics – Log Analysis	Yes with IT Search	No	Yes with Smart Search
Self-learning	Yes	No	Yes	Yes
In context expert tips	Yes with Expert Advice application	Yes with Guided Troubleshooting	No	No
Additional Optimization	Broad support across multiple technologies (Netcool Operations Insight, Application Performance Management)	Not well integrated with other offerings (HP Service Health Analyzer for example)	Point solution	Point solution
Other	Very comprehensive user documentation - broad and deep			

Source: Clabby Analytics, April, 2014

Summary Observations

IT managers and administrators now have at their disposal new-generation analytics-based tools that can ensure the availability of applications and services; can find and correct problems faster than human beings; that can help determine actions required to resolve issues; and that can enhance manager/administrator visibility into systems resource relationships. Further, these tools have the ability to self-tune for performance; automatically monitor, control and fix system errors – and they are able to predict system failures at the hardware as well as software levels before those failures occur and take remediating action. Workload and infrastructure performance and availability in the cloud is optimized using predictive capacity planning tools, and storage management features such as virtualization, tiering, snapshots and compression techniques.

Operational   analytics tools offer great productivity improvements over labor intensive event monitoring/log search approaches – and can very significantly help drive down management costs while improving service levels (due to faster troubleshooting and advanced predictive maintenance). Not only that, predictive analytics can support capacity planning, as well as workload and infrastructure optimization, maximizing resource usage and driving down hardware, software, energy and floor space costs. Enterprises evaluating operational analytics for proactive predictive management as a means to achieve cost, availability and performance benefits should closely examine IBM's comprehensive set of [SmartCloud Analytics offerings](#)*. IBM's broad support of a range of analytics tools as well as deep expertise in analytics sets them apart from other vendors in this space.*

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.