Using the IBM Big Data and Analytics Platform to Gain Operational Efficiency

IBM Redbooks Solution Guide

Organizations across many industries want to gain a full understanding of their customers, starting with what products they have, what product issues they are having, why they switch products, what they buy next, and what factors lead them to recommend a company to others. Capturing, integrating, and analyzing this type of customer information is strategic for virtually every organization.

IBM® Big Data and Analytics Platform helps businesses capture, integrate, and analyze customer information. One of the patterns this platform supports is the 360° view of a customer. This pattern can be implemented in many ways, but the scenario that is described here demonstrates how companies can collect machine and sensor data, integrate various sources of data, capture data real-time data, and analyze that data to gain customer insights. These insights provide a complete picture abut how to improve operational efficiency and the customer experience.

Figure 1. 360° view of a customer

Here are the software products that comprise this solution:

- IBM Netcool® Network Management
- IBM InfoSphere® BigInsights™ Enterprise Edition
- IBM InfoSphere Streams
- IBM Information Server
- IBM PureData™ for Analytics

- IBM SPSS® Modeler
- IBM Cognos® Enterprise

Did you know?

Most companies have client information spread throughout their organization. This information contains details about contracted services, complaints, billing and account balances, household information, and other campaign management programs. In the telecommunications industry, the client's usage data that is available in the network plant is key to monitoring client perception and sentiment of provided services. Each client usage or change in usage is reflected in network events. An attempt to call, an email check, a text message, access to a news site, a bank transaction, or even a dropped call are captured. Imagine what knowledge might be gained by seeing and correlating such information. What if the company knew that a service was down and quickly called their customer to warn them? Perhaps the company provides a quick apology and offer some compensation, such as a coupon or a month of service at no charge. These types of response might surprise and delight the client, improving their opinion of the company and retaining a valued customer.

Business value

Organizations today have incomplete and inaccurate customer information to make reliable and predictive business decisions. This challenge is found within every industry. The need is to have a dependable view of clients' interactions, how they use their company's products and services, and client sentiment toward these services. This type of insight helps businesses make better decisions, manage risk, and adapt new products. This merging of data and analysis also provides the following benefits:

- Improved customer experiences resulting in improved satisfaction and loyalty

- Coordinated communications across channels providing an integrated and unified customer experience

- Integrated intelligence from business areas, such as service assurance, retail store, online campaigns, and call centers, to intelligently maximize each customer touch point

- Improved operational efficiencies of call centers and marketing, and using self-service options

- Improved effectiveness of marketing and sales by cross-selling, upselling, bundling offerings, and offering cross product discounts

- Improved agility of sales and services systems

- Decreased customer churn

- Detect theft and fraud

- Forecast demand to construct price and promotion structures

- Reduce costs that are caused by fines (regulatory) that are received because of a services outage

- Define new pricing models for a specific customer's behavior and usage profiles

- Avoid major risks by performing what-if analyses

Using analytics and big data to understand each customer provides value to the business by decreasing the time it takes to respond to customer issues and improves the decision management process. It also provides a way to anticipate the business needs of the future.

Solution overview

IBM Big Data and Analytics Foundation describes a set of capabilities that help organizations collect and store information, and augment and analyze this data to achieve insight for effective business decisions. This platform encompasses an end to end set of capabilities that support organizations data and analytics initiatives. Here are the major capabilities of this foundation:

- Real-time processing of data
- Information integration and governance
- Operational systems to store data
- Exploration, landing, and archive capabilities for data
- Deep analytics and modeling
- Reporting and interactive analysis
- Decision management
- Predictive analytics and modeling
- Reporting analytics and content analytics
- Data discovery and exploration
- Systems-Security-Storage-Cloud

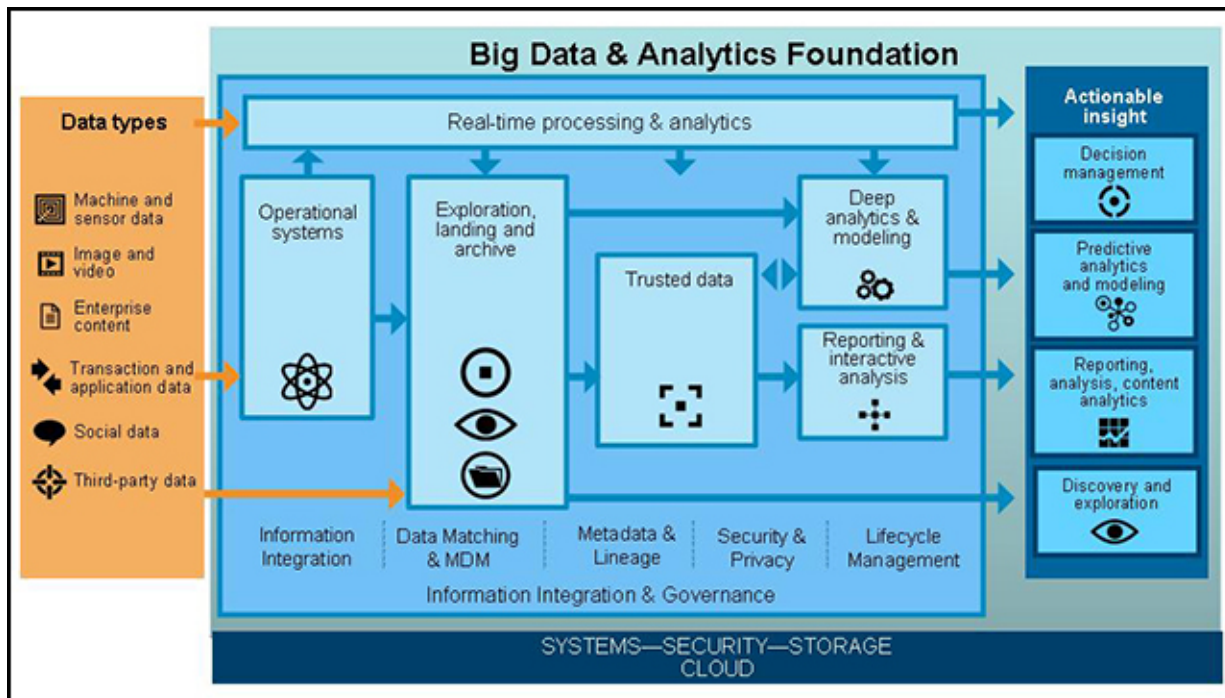Figure 2 shows each of these capabilities in this foundation.

Figure 2. IBM Big Data and Analytics capabilities

Here are the descriptions of the capabilities that are shown in Figure 2:

- *Real-time processing and analytics* allow your organization to analyze data in motion, that is, you can capture, analyze, and correlate information as it arrives from thousands of sources.

- *Operational systems* provide the traditional data management systems and acceleration capabilities to support transaction and analytic workloads.

- *Exploration, landing, and archive* bring together the capabilities of Hadoop and relational databases, enhanced storage and data access, advanced workload management, multi-tenancy, and an enterprise-ready security framework. Trusted data, deep analytics and modeling, reporting, and interactive analysis provide a rich set of analytics capabilities that find insights in new and emerging data and content. All of these capabilities comprise a logical Data Warehouse within the organization.

- *Decision management* uses the available data to make real-time decisions by combining predictive analytics, business rules, and optimization.

- *Predictive analytics and modeling* use the data to find patterns and see what is likely to happen or predict what might happen.

- *Reporting analysis and content analytics* uses the available data to understand what has happened.

- *Discovery and exploration* provide the capability to discover, understand, search, and navigate federated sources of big data.

- *Information Integration and Governance* facilitate the integration of data across an organization to ultimately address the data quality issues that are expected when augmenting enterprise databases with comprehensive sources of unstructured data. Information becomes more trustworthy as governance policy drives compliance feedback to improve relevance, consistency, and security.

- *Systems-Security-Storage-Cloud* provides the infrastructure to support big data workloads running both on the traditional physical hardware and in a cloud. The storage supports where the data is for both unstructured and structure data. The capabilities provide support for end to end security functions.

Not every solution starts out using all of these components, which allows organizations to implement big data in a phased approach and expand the analytic insights as their data foundation matures. The next section describes one example of how to implement a 360° view of a customer.

Solution architecture

Delivering a 360° view of a customer involves understanding all the customers' interactions with an organization across their various channels, products, and services. Companies capture customer interactions through machine and sensors, or within existing applications that are used by a call center, or by capturing social media and geospatial data. This data is not integrated, but stored in silo applications. Business insights are discovered when this data is integrated, analyzed, and correlated.

Consider managing information streaming from the web or capturing personal location data to offer special services to customers. IBM Big Data and Analytics Platform provides capabilities that support collecting large amounts of data in motion. This data can be pre-processed to determine which data is the most valuable and store only that data. This improves operational efficiency by augmenting the data warehouse environments.

Businesses provide company promotions by using location-based services and correlating that data with loyalty programs to trigger coupons and sales in nearby retail stores or restaurants. They use data in motion, real-time analytics, and a Hadoop cluster for data storage.

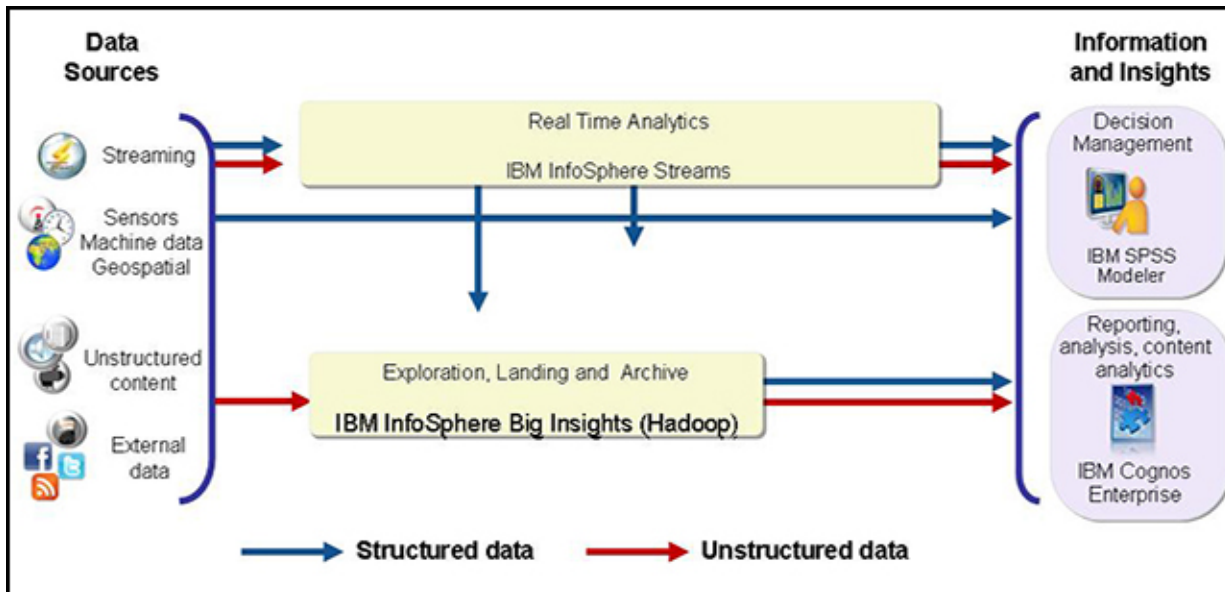Figure 3 shows the components of this type of solution.

Figure 3. Real-time analytics and Hadoop capabilities

You can expand the above solution further by adding the capability to integrate this data with existing application systems. For example, in the Telecommunications Industry, devices and sensors are monitored that capture network activity, status, and failures in one system. Customer complaints are documented by the call center in another application. Correlating these disparate sets of information provides information about how the business is running and insight on actions to take to improve.

IBM Big Data and Analytics Platform provides tools to integrate, cleanse, and transform this data, storing only the valuable data in the warehouse environment. IBM offers a data warehouse appliance for those businesses that want to integrate quickly the database, server, storage, and analytic capabilities in to a single easy to manage system.

After the data is captured, the information can be viewed through a dashboard of reports that are delivered in a browser or on a mobile device. Tools are available to model and analyze the data. Simulation and predictive techniques help companies gain insight in to a business process and assets. This insight leads to quicker and more accurate decision making.

Figure 4 shows an architectural overview of the components of the next phase of the solution.
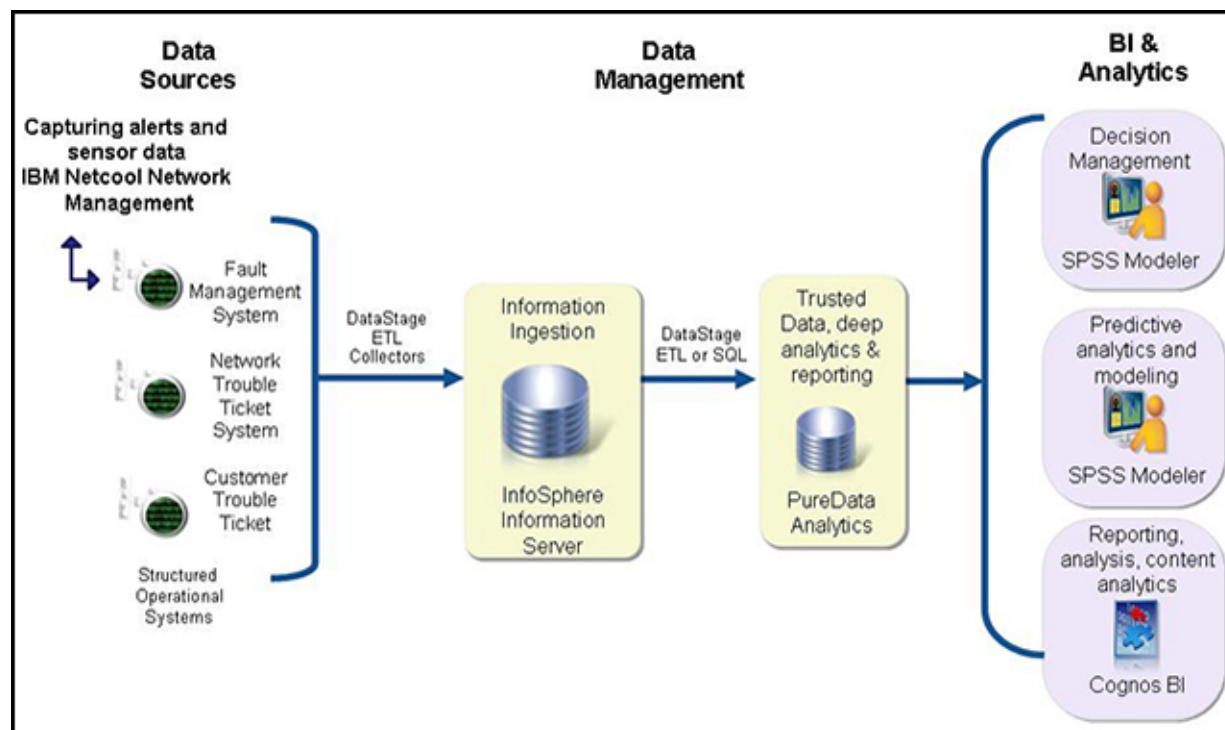

Figure 4. 360° view of a customer by integrating various sources of customer data

Here are descriptions of these products:

- *IBM InfoSphere Streams* provides continuous and fast analysis of massive volumes of information-in-motion to help improve business insights and decision making. It supports high-volume, structured and unstructured streaming data sources, such as images, audio, voice, VoIP, video, TV, financial news, radio, police scanners, web traffic, email, chat, GPS data, financial transaction data, satellite data, sensors, and badge swipes. It also provides an execution platform and services for user-developed applications that ingest, filter, analyze, and correlate potentially massive volumes of continuous data streams. It supports the composition of new applications in the form of stream processing graphs that can be created dynamically, mapped to various hardware configurations, and adapted as requests come and go, and relative priorities shift.

- *IBM InfoSphere BigInsights Enterprise Edition* manages and analyzes Internet-scale volumes of structured and unstructured data. Built on the open source Apache Hadoop software framework, it enhances this technology by adding administrative, workflow, provisioning, and security features, along with sophisticated analytical capabilities from IBM Research.

- *IBM Netcool Network Management* helps an organization visualize and understand the layout of complex networks and the impact of events upon them. Its discovery, visualization, and event-correlation/root-cause analysis helps Network Operations Centers' operators to work more efficiently by focusing time and attention on root-cause events and identifying symptom events that can be filtered into a separate view. It helps deliver cost-effective event, network, configuration, and compliance management, enabling service assurance of a dynamic IT and network infrastructure in a single offering.

- *IBM Information Server* allows organizations to integrate disparate data and deliver trusted information in line and in context to specific people, applications, and processes. It helps business and IT personnel to understand the meaning, structure, and content of any type of information across any source. It provides breakthrough productivity and performance for cleansing, transforming, and delivering this information consistently and securely throughout the enterprise.

- *IBM PureData for Analytics* is a purpose-built, standards-based data warehouse and analytic appliance that architecturally integrates database, server, storage, and advanced analytic capabilities into a single, easy-to-manage system. It is designed specifically for running complex analytics on large data volumes at high speeds and delivers the performance, scalability, intelligence, and simplicity that organizations need to dive deep into their data.

- *IBM SPSS Modeler* is a comprehensive predictive analytics platform that is designed to bring predictive intelligence to everyday business problems, enabling front-line employees or systems to make more effective decisions and improve outcomes. It provides a range of advanced analytics, including text analytics, entity analytics, social network analysis, automated modeling, and data preparation, in addition to decision management and optimization.

- *IBM Cognos Enterprise* delivers self-service analytics with cost-effective scale for your Business Intelligence and Performance Management initiatives. Users can freely explore information, analyze key facts, and quickly collaborate to gain alignment with key stakeholders and act on insights. Cognos Enterprise supports delivering consistent information and analytics on the web, mobile, or desktop, and can be embedded in other applications.

Usage scenarios

Let us take an example of a telecommunications company. This company wanted to improve their operational efficiency, reduce customer churn, and improve client satisfaction. They were not able to identify the root-cause of network trouble tickets, accurately measure the impact of network unavailability events, or understand the inventory of their logical and physical network. There was a lack of traceability on customer trouble tickets and service-level agreements with contractors, no correlation of alarms, and they were not capturing historic data for predictive trend analytics. They decided to start a phased approach that starts with capturing the network alerts and correlating the trouble tickets.

The organization has various sources of data. Trouble tickets were reported by their customers. Thousands of network alerts were happening per second. Customer information was stored around the company in various applications. There was much data, but no context.

They wanted to implement an automated process to capture equipment events and sensor alerts automatically. They used IBM Netcool Network Management for the collection of the machine and sensor alarms, creating a fault management system. They were then able to use alarm data to gain a deeper understanding of the network issues. By automating the collection of these alerts from the network elements, they could understand and act quickly knowing the root cause of a fault.

They also wanted to correlate the customer complaints and the network events. They already had a third-party application that captured the customer trouble tickets that the call center received. They also had an application that captures the network management of trouble tickets. They used IBM Information Server to integrate these disparate sets of transaction and application data. Using the collector capability, the data was cleansed, integrated, and transferred to a data repository.

The data and analytics capability needed to be running in a matter of days. Therefore, IBM PureData for Analytics appliance was used for the data warehouse and analytics requirements. The data was inspected and cleansed for quality and then loaded into the repository.

The IBM Cognos Enterprise product suite was used to align data with analytic categories, including dimensions for time, products, customer and location. Insight that is related to various metrics that are produced with SPSS Modeler were highlighted with analytical reports, trend analysis, and what-if scenarios. Analytic results, along with related business details, were shared with desktop, mobile, and web stakeholders using IBM Cognos Enterprise Reporting.

The IBM SPSS Modeler product allowed the business users to analyze and model the data to identify what actions were needed, what might happen through simulation, forecast the trends, and predict what might happen next. Here are some of the business decisions that this data helped support:

- Determine if a network site is expanded based on potential revenue loss because of congestion
- Understand a customer's propensity to churn based on poor broadband experience
- Identify high-spend customers that are likely to churn because of poor user experience
- Generate revenue and defer network investment costs
- Target business users with appropriate discounts
- Push information services appropriately as added value services
- Identify and analyze sequence of events that anticipate an outage of service to avoid service interruption

The company then wanted to capture social media data to gain more insight in customer sentiment. They also wanted to collect the geospatial data to offer real-time product promotions to their customers. IBM InfoSphere Streams was used to capture data in motion. IBM InfoSphere BigInsights Enterprise Edition provides a Hadoop cluster to store and analyze the unstructured data, augmenting the existing data warehouse sources. This data, combined with the capabilities of advanced analytics, uncovered patterns and data correlations that were not understood before.

Integration

Here are other IBM products that integrate with this solution:

- *IBM Campaign* helps marketers plan, design, execute, measure, and analyze multi-wave, cross-channel, and highly personalized marketing campaigns. Its recognized best-of-breed functionality and scalability play a key role in defining and executing interactive dialogs with customers and prospects.

- *IBM Enterprise Marketing Management (EMM) Cross-channel Marketing Optimization* enables marketers to engage each customer and prospect in a cross-channel dialog that builds upon past and current behavior to deepen customer relationships, increase customer lifetime value, and strengthen customers' return on marketing investment.

- *IBM Price Promotion, and Product Mix Optimization (formerly IBM DemandTec® solutions)* offers businesses the ability to model and analyze price sensitivity, brand preference, promotional response, and other consumer behavior online and in-store. Use predictive simulation and optimization to make better price, promotion, and assortment decisions that increase revenue and profitability.

- *IBM Customer Experience Management (formerly IBM Tealeaf® solutions)* provides unprecedented visibility into the online and mobile customer experience. Businesses can see the way their mobile or desktop websites work through the eyes of each individual customer. Collect, analyze, and report on customer and digital channel behavior in real time to gain insight in to customer interactions with your digital channels and understand why they are, or are not, successful. Use this data to optimize your web and mobile channels by removing friction that causes customer struggle and impacts business objectives.

Supported platforms

Solutions are available on the traditional platforms such as IBM Power Systems™, IBM System z®, and hardware appliances. This solution can also be delivered on the cloud.

Ordering information

Ordering information is shown in Table 1.

Table 1. Ordering part numbers and feature codes

Program name	PID number	Charge unit description	Announcement letter
IBM Netcool Network Management	5724-W11 5724-S44 5724-S45	Resource Value Unit (RVU)	http://bit.ly/1lkUKkH
IBM InfoSphere BigInsights Enterprise Edition	5725-C09	Resource Value Unit (RVU)	http://bit.ly/1lkVbLy
IBM InfoSphere Streams	5724-Y95	Resource Value Unit (RVU)	http://bit.ly/1mvA5tS
IBM Information Server	5724-Q36 5725-G05 5725-C81 5725-C80	Process Value Unit (PVU)	http://bit.ly/1i9UfHd
IBM PureData for Analytics	5725-I26 5725-I27 5725-I28 5725-I29 5725-I30 5725-I31 5725-I32	Appliance install	http://bit.ly/1hvCYua
IBM SPSS Modeler	5725-A60 5725-A54 5725-A58 5725-A56	Process Value Unit (PVU)	http://bit.ly/1ko7LbB
IBM Cognos Enterprise	5724-W12	Process Value Unit (PVU)	http://bit.ly/1hvDzvV

Related information

For more information, see the following documents:

- *Big Data Network Storage Solution for Hadoop*, REDP-5010
  http://www.redbooks.ibm.com/abstracts/redp5010.html

- *Smarter Analytics: Information Architecture for a New Era of Computing*, REDP-5012
  http://www.redbooks.ibm.com/abstracts/redp5012.html

- *Smarter Analytics: Making Better Decisions Faster with IBM Business Analytics and Optimization Solution*, REDP-4886
  http://www.redbooks.ibm.com/abstracts/redp4886.html

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service. IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law : INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you. This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

This document was created or updated on April 1, 2014.

Send us your comments in one of the following ways:
- Use the online **Contact us** review form found at:
  ibm.com/redbooks
- Send your comments in an e-mail to:
  redbook@us.ibm.com
- Mail your comments to:
  IBM Corporation, International Technical Support Organization
  Dept. HYTD Mail Station P099
  2455 South Road
  Poughkeepsie, NY 12601-5400 U.S.A.

This document is available online at http://www.ibm.com/redbooks/abstracts/tips1170.html .

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the web at http://www.ibm.com/legal/copytrade.shtml.

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

BigInsights™
Cognos®
DemandTec®
IBM PureData™
IBM®
InfoSphere®
Netcool®
Power Systems™
PureData™
Redbooks®
Redbooks (logo)®
SPSS®
System z®
Tealeaf®

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.