July 2013

MANAGEMENT BRIEF

Business Case for Enterprise Big Data Deployments

Comparing Costs, Benefits and Risks for Use of IBM InfoSphere Streams and Open Source Storm

International Technology Group

609 Pacific Avenue, Suite 102
Santa Cruz, California 95060-4406
Telephone: 831-427-9260
Email: Contact@ITGforInfo.com
Website: ITGforInfo.com

TABLE OF CONTENTS

List of Figures

EXECUTIVE SUMMARY

Real-time Economics

Few areas of IT have developed as rapidly as Big Data. Although definitions of this complex vary, the most visible area of activity has involved technologies that originated in Google, Yahoo!, Facebook, Twitter, LinkedIn and other social media companies.

One key area of evolution has been in real-time analytics. While Apache Hadoop remains the backbone of the Big Data world, it is designed for batch processing. New applications for continuous, real-time analysis of extremely large data volumes require a different approach – streaming data enables users to process extremely large, heterogeneous data volumes at unprecedented speeds.

Real-time applications are emerging among social media and ecommerce companies, and in a growing number of industry verticals. They deal not only with interactions through computers and mobile devices, but also with data generated by meters, sensors and other machine sources.

A number of streaming data analytics solutions have emerged. The leading players – which are the focus of this report – are IBM InfoSphere Streams, which has evolved since the early 2000s in a broad range of real-time analytics applications; and Storm, originally developed by Twitter for social media analytics.

Although there are some architectural commonalities between these, there are significant differences in other areas. The most obvious of these is that Storm is an open source offering, while Streams is managed, enhanced and supported by IBM.

Part of the appeal of Storm is that it may be downloaded free of charge. But overall economics are less clear-cut. More complex development and deployment processes, and higher costs for ongoing application maintenance and system administration may outweigh software savings.

This may be illustrated by comparisons of overall three-year costs in representative high-impact applications in four companies. Costs for use of Streams, as figure 1 shows, averaged 21 percent less than for Storm.

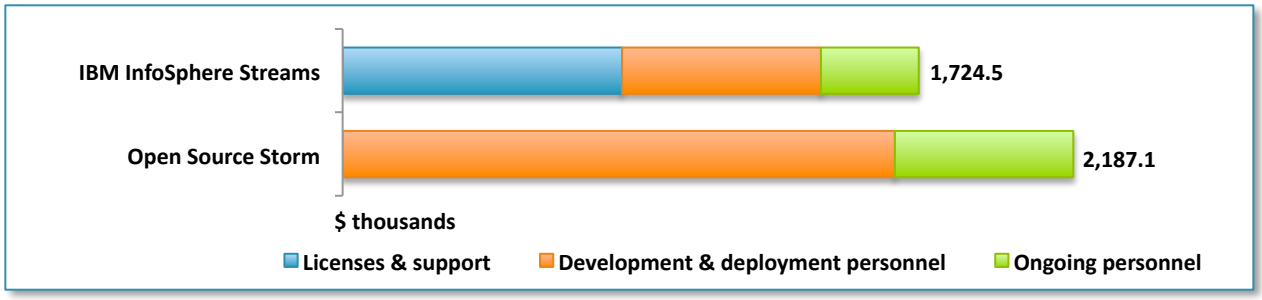

Figure 1: ***Three-year Costs for Use of IBM InfoSphere Streams and Open Source Storm – Averages for All Installations***

Personnel costs are for initial application development and deployment, as well as ongoing post-production operations over a three-year period.

Calculations include high-level specialists such as data scientists, along with project managers and developers for initial development and deployment; and developers and system administrators for post-production operations. Streams costs also include licenses and support.

Comparisons are for composite profiles of telecommunications, financial services, energy and utilities, and health care companies. Profiles were constructed based on information supplied by 16 organizations in the same industries, with similar business and application profiles, which employed Streams or Storm in comparable roles.

Further information on profiles, methodology and assumptions employed for calculations, along with cost breakdowns for individual companies may be found in the Detailed Data section of this report.

Differentiators

Overview

Streams and Storm occupy a distinct market space. Although they share many of the attributes of conventional complex event processing (CEP) systems, they are better geared than CEP systems to handle the data volumes that are characteristic of the Big Data world.

These solutions have, however, evolved in significantly different directions:

- ***Storm*** grew out of applications originally put in place by Twitter for real-time analysis of customer activity. It was originally developed by a start-up, BackType, which was acquired by Twitter in July 2011.

  Storm became an open source offering in September 2011. It may currently be downloaded from the GitHub website free of charge under an Eclipse Public License. GitHub provides online hosting services for a variety of open source projects.

  Specifications and enhancements are managed by a group of contributors operating through GitHub, which currently lists 38 individual contributors and 45 active user companies. In practice, however, a handful of developers account for a disproportionate amount of Storm code, and levels of commitment by user companies vary.

  Among the latter, the largest groups are social media and ecommerce businesses (13) and suppliers of advertising, marketing and other services that involve tracking and responding to online activity (21). Only four can be characterized as conventional businesses.

- ***Streams*** originated in the early 2000s as a major research project undertaken by IBM with the U.S. federal government. Initially named System S, it was later deployed by numerous universities, research institutions and defense and intelligence users.

  A commercial version, InfoSphere Streams, was introduced in 2009. According to IBM, it represents more than 400 person-years of development effort.

  Streams evolution is reflected in different market demographics. The 76 Streams users identified during research for this report included academic and research (18) and government (12) organizations.

  The largest group, however, was comprised of conventional businesses in telecommunications, financial services, energy and utilities, health care, manufacturing, transportation and other industries (38). Social media and IT services companies accounted for a further eight users.

  Deployed applications covered a broader range than for Storm. Figure 2 shows examples.

Telecommunications & network services	Health care
CDR processing	Patient monitoring
Campaign management	Medical research
Churn prediction/prevention	**Manufacturing & logistics**
Performance monitoring & optimization	Operational monitoring
Problem detection & management	Civilian government
Financial services	Traffic management
Trading applications	Environmental monitoring & research
Trend analysis	**Defense & intelligence**
Fraud detection/prevention	Surveillance applications
Regulatory compliance	Command & control
Energy & utilities	Classified
Meter & sensor data analysis	**Scientific & engineering**
Grid monitoring & management	Radio astronomy
Exploration & production applications	Physics research
Social media	Satellite signal processing
Pattern & trend analysis	Weather monitoring & prediction
User identification & profiling	**Cross industry**
Sentiment, buzz, intent & ownership tracking	Cybersecurity
Campaign management	

Figure 2: ***Deployed IBM InfoSphere Streams Applications***

Although Storm remains popular among social media and ecommerce companies, Streams has gained a great deal more traction in conventional business verticals.

IBM has, however, significantly expanded targeting of social media and ecommerce companies during 2012 and 2013. Streams has been increasingly deployed for event tracking for marketing, brand management, customer service and related applications in conventional business verticals.

Development and Delivery

Differences in development focus and delivery models should be highlighted. The Storm community is focused primarily on social media businesses. It tends to assume that its developers are highly skilled, and that they are capable of working with "raw" code and configuring and modifying software components on a case-by-case basis as needs change.

Manual coding is, moreover, the norm. Experiences with manual techniques have shown that these tend to result in low developer productivity and high error rates. This is particularly the case if less skilled and/or experienced personnel are employed.

Documentation processes typically suffer, and it becomes more difficult to re-use code in subsequent application projects. Over time, organizations may accumulate masses of "spaghetti code" that are expensive to maintain and enhance over time.

Challenges are compounded when interoperability with existing applications and databases is required. Implementing real-time analytics is a significantly more difficult task if new functionality must be integrated into broader infrastructures. Storm capabilities in such areas as system management, availability, recovery and security also tend to be problematic.

IBM has placed a stronger focus on ease of development and deployment, and on such issues as interoperability, manageability and resiliency.

Streams has been expanded in three main areas:

1. ***Application development and deployment*** tools include the Eclipse-based InfoSphere Streams Studio, which provides a common, high-productivity development and administration interface.

   Modular ***analytics toolkits*** address functions that range from standard network and database interfaces to higher-level capabilities such as statistical analysis, complex event processing, data mining and geospatial and temporal data analysis.

   Accelerators and ***integration toolkits*** also contribute to a portfolio that currently includes the tools illustrated in figure 3.

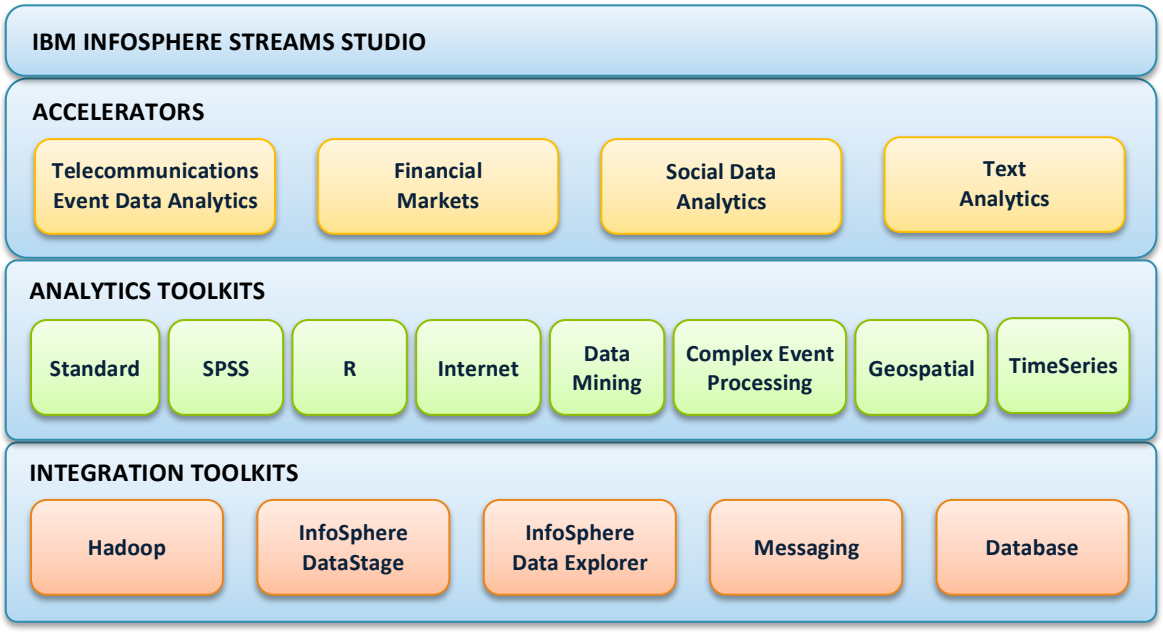

Figure 3: ***IBM InfoSphere Streams Development Tools***

Accelerators include prebuilt templates, components and interfaces designed to facilitate development of industry- and application-specific solutions. Accelerators were developed based on customer experiences, and have materially improved "time to value" for development and deployment of Streams-based applications.

Current accelerators address applications for financial markets (recognition of patterns in high-volume trading), telecommunications (call detail record and event analysis), social media analysis and text analysis (extraction of information from text).

The scope of accelerator-based solutions extends to IBM and third-party applications. The Telecommunications Event Data Analytics Accelerator, for example, may be integrated with IBM (formerly Unica) and third-party campaign management solutions. Comparable integration is available for other industries and areas of functionality.

Integration toolkits address interoperability with Hadoop, including the IBM InfoSphere BigInsights Hadoop distribution; InfoSphere DataStage extract, transformation and load (ETL) tooling; WebSphere MQ and JMS messaging systems; and a variety of databases including IBM DB2, Informix, Netezza and solidDB, along with Oracle, Microsoft SQL Server, MySQL and others.

2. ***Infrastructure*** features include extensive management services for resiliency functions. These include monitoring, scheduling and job management, authentication and authorization, and failover and recovery. An embedded recovery database enables retention and reinstatement of service data in the event of an outage.

   Streams infrastructure may be reinforced by use of IBM ***General Parallel File System (GPFS)***. This is a parallel file system that has been deployed for more than a decade for scientific and technical, as well as commercial high performance computing. GPFS is a mature solution that has been extensively enhanced by IBM for use in a wide range of applications.

   In addition to offering high levels of scalability – installations with more than 1,000 nodes are common, and some exceed 5,000 nodes – GPFS is a recognized industry leader in cluster availability, system and data management, failover and recovery, and security.

3. ***Enterprise integration*** has been a major IBM focus for Streams enhancement since the late 2000s. Interoperability has been developed with open source components defined by Apache Foundation and other communities, and with numerous IBM and third-party solutions employed in conventional enterprise IT environments.

   The IBM Big Data Platform, for example, allows Streams integration with InfoSphere BigInsights; the IBM PureData System family of analytics appliances; InfoSphere Information Server, one of the industry's broadest and most sophisticated data integration toolsets; and others.

   Streams interoperability is also provided with the industry's major relational databases, including IBM DB2 and Informix, Oracle, Microsoft SQL Server, Teradata and MySQL.

   In comparison, Storm users typically cite requirements for integration with Apache Hadoop and open source tools such as Kafka, RabbitMQ and ZooKeeper (which forms part of the Storm stack); and NoSQL databases such as Cassandra, DynamoDB, MongoDB and Redis. Among social media companies, entire application portfolios and IT infrastructures have often been customized "from the ground up" using open source software.

   It can be expected that, as adoption of real-time analytics spreads beyond this group, easier integration of Streams with conventional IT environments will represent an increasingly significant differentiator.

These differences have a number of implications for users. One is that, as this report shows, the economics of application development, deployment, maintenance and administration over time may be more favorable to Streams. A second is that the degree of risk to which organizations are exposed may be significantly different.

Risk Exposure

Commitment to Storm for business-critical deployments raises a number of risk issues. These include:

- ***Focus***. Storm is managed and enhanced primarily to meet the needs of social media companies. While few would dispute the technical competence of the Storm community, its members have tended to focus on the interests of this group. Other types of business and application have received less attention. There is no obvious reason to expect that this will change in the future.

- ***Continuity***. Although some major companies use Storm, it lacks the broad base of contributors and vendor support enjoyed by Hadoop. Its design and evolution are dependent upon a small group of individuals whose long-term commitment cannot be taken for granted.

There have been signs that both the rate of Storm enhancement and the speed of response to user queries have slowed during 2013. Although it is unclear whether this is a long-term trend, it highlights the risks to which Storm users are potentially exposed.

- ***Stability***. Even if the Storm community remains vibrant, experience with open source technologies has shown that these may evolve in an unpredictable manner. This may pose significant challenges for users. It may become more difficult to plan technology strategies, and to predict project schedules and costs.

  Organizations may, moreover, standardize upon individual components only to find that these receive declining attention over time. The pace of technology change among social media companies is a great deal faster than users in most other industries are accustomed to.

  Storm is typically used with other open source components whose evolution is determined by different communities. Enhancement paths tend to move at varying speeds, and may reflect diverse priorities. New version release dates are at best loosely coordinated. Developers may be exposed to a continuous stream of changes.

- ***Resiliency***. Although Storm includes mechanisms designed to maintain availability in the event of a server failure – guaranteed message delivery is most commonly cited by users – it falls short of the capabilities normally considered mandatory for high availability operations.

  For example, while Storm's structure of master and worker nodes is often described as "fault tolerant," in practice this means that the system can detect worker node failures and transfer workloads and reroute messages to others. Master nodes, however, represent, according Storm documentation, "a sort of SPOF (single point of failure)."

  A further vulnerability is that key Storm mechanisms are still comparatively immature and – as is the case for any software environment characterized by manually developed code and hand-configured components – error-prone. It is a great deal easier to maintain high availability for stable rather than unstable environments.

  Some Storm users have realized high levels of availability. Major investments have, however, typically been required to harden software, ensure redundancy and provide in-depth staff and procedures for operational monitoring and contingency response. Organizations with fewer resources and less developed skill bases may find this approach less attractive.

- ***Support***. Storm is available only with community support – i.e., users rely upon online peer forums for enhancements, technical advice and problem resolution.

  This may prove appropriate for commonly encountered issues, although it is dependent on the willingness of others to share time and experience. But it is a great deal less reliable in dealing with organization-specific issues, and as Storm environments become more diverse, it will be increasingly difficult to draw upon adequate bases of skills and experience.

  The bottom-line implications may be substantial. Delays in resolving problems may undermine developer productivity, and may result in application errors, performance bottlenecks, outages, data loss and other negative effects.

This situation is reminiscent of earlier waves of open source activity, including the complexes surrounding the Internet and more recently Apache Hadoop. In these cases, the demands for more robust functionality, greater stability and in-depth support led to the appearance of enhanced vendor offerings.

In the Hadoop world, vendor-managed and supported distributions are offered by Amazon, Cloudera, EMC, Hortonworks, IBM, Intel, MapR and others. There has to date been no comparable trend for Storm.

In comparison, the risk that IBM will fail to maintain its commitment to Streams is small. The company has a longstanding business among scientific, technical and government users worldwide. Streams is also a significant factor in such markets as analytics, social media, high performance computing and advanced public sector applications that represent top strategic priorities for IBM.

Conclusions

The choice between Streams and Storm is currently between an open source environment that requires exceptionally high skill levels to implement effectively, offers lower long-term development and administration productivity and represents higher risk exposure; and a more mature, stable and better-supported solution set subject to license and support fees.

If the future evolution of Storm follows the same pattern as earlier waves of open source technologies, however, it will tend to converge with the Streams model. It can be expected that more productive and robust vendor-supported "distributions" will be offered through license and/or fee-for-service arrangements.

For enterprise users, however, it can be expected that the appeal of Streams will remain differentiated. Deployment trends among this group will tend to leverage broader IBM strengths.

These include established competencies in ***software engineering*** (Streams components are not only pre-integrated, but also extensively tested for optimum performance and functional transparency), ***customization*** (the ability of IBM services organizations to deliver industry- and organization-specific solutions has already proved to be popular) and ***customer support***.

IBM has also, as in other areas of its software business, moved aggressively to recruit and support ***business partners***. These currently include more than 50 independent software vendor (ISV) and services firms worldwide, including suppliers of a wide range of complementary tools and industry-specific solutions. The number is expanding rapidly.

The Storm open source community will no doubt remain active. But for mainstream businesses, the future evolution of the Storm environment will be toward capabilities that exist today for Streams. For organizations that are candidates to use real-time analytics for Big Data, it may make sense to use IBM InfoSphere Streams sooner rather than later.

SOLUTION SET

Overview

Streams Evolution

Streams has evolved rapidly since version 1.0 was introduced in May 2009, as IBM has progressively enhanced all major areas of functionality. Versions 2.0 and 3.0 appeared in April 2011 and November 2012 respectively, and Version 3.1 in May 2013.

According to IBM, the company plans to continued expansion of its portfolio of development tools and accelerators. This currently includes the components shown in figure 4.

New industry-specific tools, including accelerators designed for energy and utilities, and health care applications, are expected in the future. The company also plans to expand interoperability with its other Big Data offerings, which are described below, and with non-IBM applications and databases.

Further enhancements are expected in the Streams Processing Language (SPL), and in such areas as usability, manageability, availability, recovery and security across the Streams runtime system.

Deployment Options

Streams is a Linux-based system that may be deployed on recent versions of Red Hat Enterprise (RHEL), SUSE Linux Enterprise Server (SLES) and CentOS on x86 servers; and on RHEL on IBM Power Systems. VMware ESXi, Kernel Virtual Machine (KVM) and – on Power Systems – IBM PowerVM hypervisors are supported.

Although Streams may run on other x86 hardware, IBM offers a set of prepackaged configurations built around its System x3550 M4 and x3690 X5 dual-socket servers equipped with Intel E5 and E7 processors respectively. The x3690 X5 implements IBM X-Architecture, a customized Intel-based design that offers higher levels of memory and I/O performance than standard x86 servers.

System x3550 M4 and x3690 servers may be included in customized IBM Intelligent Cluster offerings. These combine preintegrated and pretested servers, switches, adapters and other components in rack-mounted configurations with IBM installation and support services.

Big Data Platform

The IBM Big Data Platform is the company's name for a broader set of offerings that enable Big Data solutions to be integrated into broader enterprise application suites and infrastructures. In addition to Streams, these include the following:

- ***InfoSphere BigInsights*** combines a full suite of standard open source Hadoop components with IBM enhancements in development and interoperability tooling, system management, availability and recovery, security and other areas comparable to those for Streams.

  Certain components, including Social Data Analytics Accelerator and text analytics, are shared by Streams and BigInsights, and IBM has placed a strong emphasis on customized integration of and support for the two solution sets. Many customers have deployed both. BigInsights also supports use of GPFS.

STREAMS STUDIO	
Base solution	Eclipse-based integrated development environment (IDE) supports rapid development includes editors, wizards, application structure graphs & runtime monitoring. ***Streams Explorer*** enables simplified installation & administration of all components; drag & drop graphical editor automatically synchronizes graphical & SPL source code views; ***live (instance) graph*** provides visual monitoring of & metrics for application health & allows rapid identification of issues using customizable views; ***data visualization*** offers predefined & customizable charts.
ACCELERATORS	
Telecommunications Event Data Analytics	Provides real-time analytics support for call detail records (CDR) processing for churn reduction, revenue assurance, fraud detection & other applications. Integrates with IBM (ex-Unica) Campaign & third-party applications.
Financial Markets	Provides real-time analytics functions for automated real-time equities & options trading. Supports Financial Information Exchange (FIX) & other industry standards, along with WebSphere Front Office (WFO) for Financial Markets & MQ Low Latency Messaging.
Social Data Analytics	Extracts data from tweets, blogs & other sources, identifies user characteristics, develops profiles & associates these with sentiment, buzz, intent & ownership for products, brands & companies. Enables customization for specific use cases (e.g., lead generation, brand management, micro-segmentation, campaign effectiveness measurement), & industries (e.g., retail, financial, media & entertainment). Also part of BigInsights.
Text Analytics	Enables extraction of structured information from text using Annotation Query Language (AQL). Also part of BigInsights. Based on IBM System T technology in Watson software.
ANALYTICS TOOLKITS	
Standard	Base Streams Processing Language (SPL) includes common operators & functions, including relational, statistics & XML data support.
SPSS	Exploits SPSS Modeler to develop & build predictive models, deploys & refreshes these.
R	Enables integration with R statistical analysis & graphics environment.
Internet	Enables integration of HTTP/FTP & RSS XML data.
Data mining	Enables scoring of real-time data for data mining using Predictive Model Markup Language (PMML) models.
Complex Event Processing	Enables pattern detection & processing of complex events in data streams.
Geospatial	Enables analysis & processing of geospatial data for location-based services, geospatial data types (e.g., point & polygon) & geospatial functions (e.g., distance between locations).
TimeSeries	Enables generation (synthesis or extraction), preprocessing (preparation & conditioning), analysis (statistics, correlations, decomposition & transformation) & modeling (prediction, regression & tracking) of data streams over time.
INTEGRATION TOOLKITS	
Hadoop	Enables interoperability with Hadoop systems, including IBM BigInsights & other enhanced Hadoop distributions.
IBM InfoSphere DataStage	Enables exchange of Streams data with DataStage extract, transformation & load (ETL) a.k.a. data integration toolset.
IBM InfoSphere Data Explorer	Pushes Streams data to Data Explorer federated navigation & discovery toolset.
Messaging	Implements IBM WebSphere MQ classes for Java Message Service (WebSphere MQ, JMS), allowing JMS applications to access WebSphere MQ systems. Enables exchange of Streams data with WebSphere MQ topics or queues.
Database	Enables interaction with IBM DB2, Informix, Netezza & SolidDB, Oracle, SQL Server, MySQL & other databases.

Figure 4: ***IBM InfoSphere Streams 3.1 Development Tools***

- ***InfoSphere Data Explorer*** is a high-volume federated navigation and discovery tool that may be employed across a broad range of structured and unstructured data types. It is based on Vivisimo Velocity Platform, an enterprise search and retrieval engine incorporating innovative federation, clustering and data deduplication technologies. IBM acquired Vivisimo in April 2012.

  There are currently more than 200 InfoSphere Data Explorer and Vivisimo customers in industries such as defense and intelligence, civilian government, life sciences, manufacturing, financial services and other industries.

- ***InfoSphere Warehouse*** is an IBM DB2-based data warehouse framework that incorporates features for partitioning, data compression, online analytical processing (OLAP), data mining, access control, high availability clustering and other functions. It also includes an extensive suite of management, optimization and development tools.

  A key feature of InfoSphere Warehouse is that it supports Continuous Ingest capability in DB2 10. This employs IBM parallel loading technology for extremely fast, low-overhead data transfers. This capability, which offers a higher-performance alternative to conventional batch and "trickle feed" techniques, is highly synergistic with use of Streams.

- ***InfoSphere Information Server*** is an enterprise-class data integration solution set. It provides capabilities for information discovery and analysis, ETL (the IBM InfoSphere DataStage product is included), data quality management, business alignment, integration with service oriented architecture (SOA) environments and other functions.

  InfoSphere Information Server is supported by IBM for use with Apache Hadoop, Streams and a wide range of other open source and IBM proprietary solutions.

Other IBM offerings also form part of the Big Data Platform, which is illustrated in figure 5.

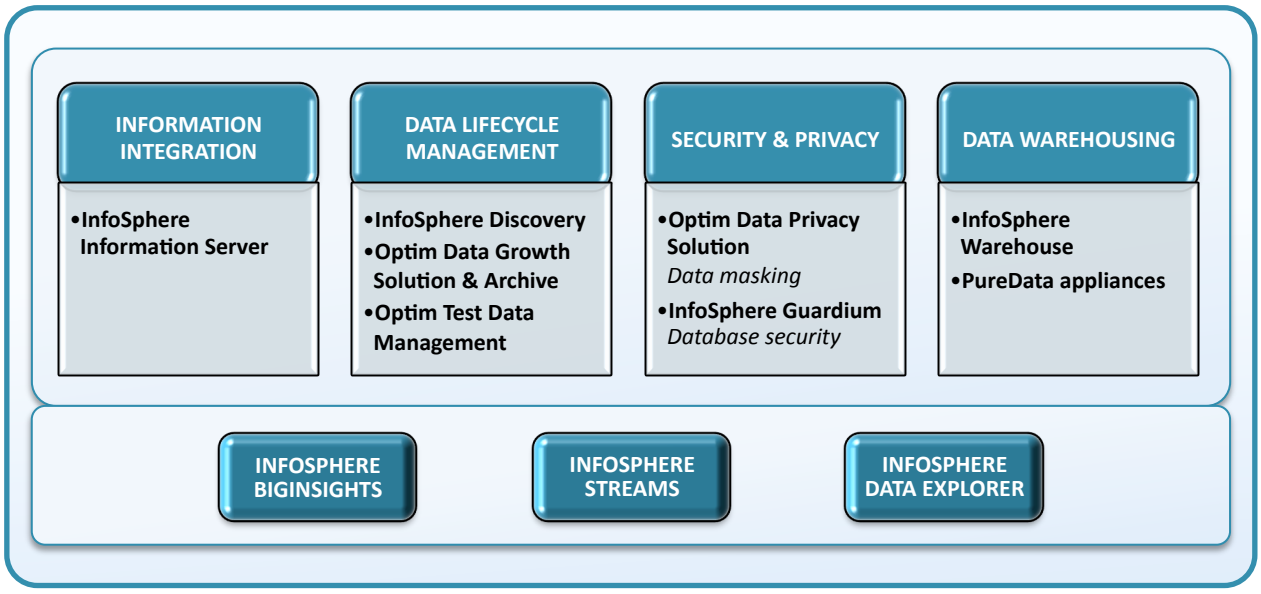

Figure 5: ***IBM Big Data Platform***

Offerings include Optim data management and Guardium database security solutions; and IBM appliances including PureData System for Analytics (formerly Netezza), PureData System for Operational Analytics (formerly Smart Analytics System) and PureData System for Hadoop

DETAILED DATA

Composite Profiles

The calculations presented in this report are based upon the four composite profiles shown in figure 6.

TELECOMMUNICATIONS COMPANY	FINANCIAL SERVICES COMPANY	ENERGY & UTILITIES COMPANY	HEALTH CARE COMPANY
BUSINESS PROFILE			
Mobile telephone & Internet access, related services 35 million subscribers	Brokerage arm of diversified retail bank 5 million accounts	Diversified electric & natural gas utility 6 million customers	Health care services company + major medical center
APPLICATION			
Process 400 million CDRs/day for mediation, revenue assurance, fraud detection & campaign management applications	Process 250 million trades/day for trend identification, fraud & anomalous activity (e.g., money laundering) detection	Process 200 million intelligent meter readings/day for customer insights, anomaly detection & demand-side management programs	Process 50+ million instrument readings/day for patient monitoring alerts & offline research; trial predictive analysis project
INTERFACES			
Network infrastructure & management systems, campaign management system, ETL, data warehouse	Proprietary systems, data warehouse, decision support & visualization tools	Advanced Metering System, CIS, back-end business systems, data mart	Monitoring systems, custom data acquisition software, analytical tools & databases
IBM INFOSPHERE STREAMS CONFIGURATIONS			
4 nodes x 2 Intel E5 8-core processors	4 nodes x 2 Intel E5 6-core processors	4 nodes x 2 Intel E5 4-core processors	4 nodes x 2 Intel E5 4-core processors
IBM INFOSPHERE STREAMS FTEs			
Development & deployment (7 months): 6.15 Post-production: 0.75	Development & deployment (6 months): 6.0 Post-production: 0.65	Development & deployment (6 months): 4.95 Post-production: 0.55	Development & deployment (4 months): 3.75 Post-production: 0.35
OPEN SOURCE STORM FTEs			
Development & deployment (12 months): 9.75 Post-production: 1.75	Development & deployment (12 months): 9.05 Post-production: 1.5	Development & deployment (10 months): 7.25 Post-production: 1.15	Development & deployment (6 months): 6.25 Post-production: 0.55

Figure 6: *Composite Profiles*

Profiles were constructed using information supplied by seven companies using Storm, and nine using Streams. Companies were based in North America and Asia.

Companies supplied information on applications; development and deployment times for these; and numbers of full time equivalent (FTE) personnel for application development and deployment, and (separately) post-production operations.

Because job descriptions and titles often varied between companies, equivalent numbers of FTEs for equivalent specializations were in some cases estimated by the International Technology Group.

Cost Calculations

Personnel costs were calculated for numbers of FTEs for applicable periods. For use of Storm at the telecommunications company, for example, costs were calculated for development and deployment FTEs for 12 months, and for post-production FTEs for 36 – 12 = 24 months.

Calculations were based on the annual salary assumptions shown in figure 7. Salaries were increased by 55.48 percent to allow for benefits, bonuses and other per capita costs. The same assumptions were employed for use of Storm and Streams.

TITLE	SALARY
Subject matter expert [1]	
- Telecom company	$185K
- Financial services company	$200K
- Energy & utilities company	$175K
- Health care company	$160K
Project manager [1]	$165K
Developer [1] [2]	$147K
System administrator [2]	$104K

[1] Development & deployment
[2] Post-production operations

Figure 7: *FTE Salary Assumptions*

The term subject matter expert (SME) is employed here for individuals with skill sets that would normally be considered those of a "data scientist," although only two companies cited this term.

License and support costs for use of Streams were calculated based on discounted IBM list prices.

Cost Breakdowns

Breakdowns for individual profiles are shown in figure 8.

	COMPANY TYPE			
	Telecom	Financial Services	Energy & Utilities	Health Care
IBM INFOSPHERE STREAMS				
Licenses & support	**1,202.43**	**901.82**	**827.90**	**413.95**
Personnel				
Development & deployment	850.51	713.26	582.86	234.26
Ongoing operations	373.86	329.62	280.84	186.58
Total	**1,224.37**	**1,042.88**	**863.69**	**420.83**
TOTAL ($ thousands)	**2,426.80**	**1,944.71**	**1,691.60**	**834.78**
OPEN SOURCE STORM				
Licenses & support	**0**	**0**	**0**	**0**
Personnel				
Development & deployment	2,308.49	2,155.34	1,419.73	728.81
Ongoing operations	733.09	618.81	511.54	272.48
Total	**3,041.58**	**2,774.15**	**1,931.27**	**1,001.29**
TOTAL ($ thousands)	**3,041.58**	**2,774.15**	**1,931.27**	**1,001.29**

Figure 8: *Three-year Cost Breakdowns*

ABOUT THE INTERNATIONAL TECHNOLOGY GROUP

ITG sharpens your awareness of what's happening and your competitive edge . . . this could affect your future growth and profit prospects

International Technology Group (ITG), established in 1983, is an independent research and management consulting firm specializing in information technology (IT) investment strategy, cost/benefit metrics, infrastructure studies, deployment tactics, business alignment and financial analysis.

ITG was an early innovator and pioneer in developing total cost of ownership (TCO) and return on investment (ROI) processes and methodologies. In 2004, the firm received a Decade of Education Award from the Information Technology Financial Management Association (ITFMA), the leading professional association dedicated to education and advancement of financial management practices in end-user IT organizations.

The firm has undertaken more than 120 major consulting projects, released more than 250 management reports and white papers and more than 1,800 briefings and presentations to individual clients, user groups, industry conferences and seminars throughout the world.

Client services are designed to provide factual data and reliable documentation to assist in the decision-making process. Information provided establishes the basis for developing tactical and strategic plans. Important developments are analyzed and practical guidance is offered on the most effective ways to respond to changes that may impact complex IT deployment agendas.

A broad range of services is offered, furnishing clients with the information necessary to complement their internal capabilities and resources. Customized client programs involve various combinations of the following deliverables:

Status Reports	In-depth studies of important issues
Management Briefs	Detailed analysis of significant developments
Management Briefings	Periodic interactive meetings with management
Executive Presentations	Scheduled strategic presentations for decision-makers
Email Communications	Timely replies to informational requests
Telephone Consultation	Immediate response to informational needs

Clients include a cross section of IT end users in the private and public sectors representing multinational corporations, industrial companies, financial institutions, service organizations, educational institutions, federal and state government agencies as well as IT system suppliers, software vendors and service firms. Federal government clients have included agencies within the Department of Defense (e.g., DISA), Department of Transportation (e.g., FAA) and Department of Treasury (e.g., US Mint).

International Technology Group

609 Pacific Avenue, Suite 102
Santa Cruz, California 95060-4406
Telephone: 831-427-9260
Email: Contact@ITGforInfo.com
Website: ITGforInfo.com