

# When the Information is the Business

## Key Considerations for Analytics-as-a-Service Providers



Daniel Tkach

## TABLE OF CONTENTS

1	Overview
2	Introduction: The Growing Market for Data Aggregators
3	Analytics-as-a-Service: Four Phases
3	Phase 1: Extreme Data Management is the Norm
3	Phase 2: Power Users Pressured
4	Phase 3: Clients Demand Control
4	Phase 4: The Data Speaks for Itself
5	The Data Aggregation Process
6	Challenges in Data Aggregation
6	Sybase IQ: Addressing the Challenges In Data Aggregation
6	Cost and Performance
8	When the Information is the Business: Case Studies
8	Case 1: Experian Integrated Marketing
8	Case 2: Nielsen Media Research
9	Case 3: The Korea Health Insurance Review Agency
10	Case 4: The Alfred Wegener Institute
11	Conclusion
12	Bibliography
12	About the Author

## WHAT'S A DATA AGGREGATOR?

"Data aggregators" act as information brokers that collect data from numerous sources and offer access to or evaluation of that data to their customers. This spares decision makers from the complexity of attempting to gather and cleanse the data themselves, allowing them to focus on making better decisions that result in competitive advantage. The products and services provided by data aggregators go by many names: database marketing services, financial and credit information services, audience measurement services, market research, national statistical analysis, online shopping comparison, and many others.

## WHAT THE RESEARCH SHOWS

A survey of 70 information providers indicates that data aggregators evolved through a four-phase lifecycle of increasing sophistication and automation of the services they offer. This paper shows how these organizations can best meet their requirements at each stage

## OVERVIEW

Consider these scenarios:

*A global provider of marketing solutions to leading consumer brands creates information models that foster insight about consumer buying patterns and likely behavior. The information is used to identify sets of best-fit prospective customers that clients can target with specific marketing campaigns.*

*A government health agency in Asia analyzes medical treatment results, disease and injury factors, and aggregate trend analysis to produce comprehensive reports that supply information crucial to government decision-making and national healthcare planning.*

*A European institute dedicated to maritime research in the Arctic and Antarctic regions extracts information slices from billions of datasets as a basis for a long-term analysis comparing historical and modern-day measurements.*

What do these organizations have in common? All are tasked with providing information and analysis that will foster insights and improve the decision making of their customers. In today's business environment, critical decisions rely on accurate, complete and timely information presented to the decision makers — be they executives, managers or knowledge workers — in a comprehensible and actionable way.

To feed information to those decision makers, organizations have no shortage of data sources to choose from. Operational systems store transactional data in databases, documents are produced and placed in repositories, while CRM systems collect data about customer interactions and preferences. The growth of the Internet as a core channel for business operations has increased the immediate availability of data from vast and multiple sources, but has also created a data glut that has made it harder for those who rely on that data to make effective decisions.

This trend has led to the emergence of data aggregators; information services providers that capture, integrate, and analyze multiple vast stores of data to deliver the required information products and services to their clients, often in very short order. They support multiple industries and business processes, ranging from marketing services, telecommunications, financial services, and retail to public sector and healthcare organizations. Their business is the data they collect and the information they provide. Therefore, they are fully dependent on the technology and tools they choose to aggregate, manage, and analyze data. Their choices can make or break their business.

A recent survey of 70 providers of information and analytics services has revealed four progressive maturity phases for information providers:

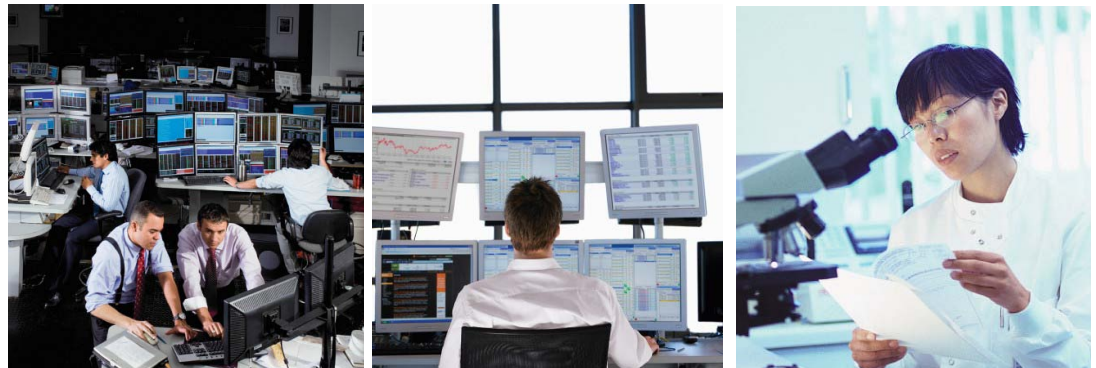
- Phase 1: Extreme Data Management is the Norm
- Phase 2: Power Users Pressured
- Phase 3: Clients Demand Control
- Phase 4: The Data Speaks for Itself

This paper explores how these organizations are addressing the challenges of evolving from one phase to the next. It presents how Sybase IQ is empowering the "business of information," by delivering high-performance analytics capability to multiple concurrent users, while lowering costs, and by providing simplicity and efficiency throughout each of these four phases.

**INTRODUCTION: THE GROWING MARKET FOR DATA AGGREGATORS**

Data aggregation allows for better decision-making by presenting information in a context that makes it more meaningful. For instance, when buying a house, the price is an important piece of data but has no meaning by itself. The house record will probably include the house square footage, the address, the year it was built, and house features such as central air-cooling and heating, the number of stories, recent upgrades, and more. This information, however, is not enough to justify a decision to buy the house. The buyers will want to know how desirable is the neighborhood, the average price of a similar home in that location, how long the house has been on the market, how the nearby schools are rated, and how safe is the area before they make that decision.

This example illustrates why there is a growing market for data aggregators. Real estate agents are much more productive when they can obtain all of their information from a single source, rather than having to deal with many separate information sources. Their need for a single source of information has created the market for those who make a business of aggregating and supplying that data. The same principle applies across a host of industries and markets, as shown in Figure 1.



FINANCIAL SERVICES	MARKETING SERVICES	HEALTHCARE
<p>Data aggregators collect financial information from multiple sources, offering banks and other financial institutions opportunities to build stronger relationships with customers. These organizations maintain huge amounts of user behavior data — including transaction histories of accounts and credit cards and loan applications — which gives them the opportunity to offer new products and services as well as advice about investing and managing their money.</p>	<p>One of the largest providers of automobile marketing data compiles vehicle registration and sales data from more than 200 sources, including motor vehicle departments, insurance companies, automakers, and lending institutions. The company then repackages that data and sells it to dealers, manufacturers and marketing firms — anyone who wants detailed information about car-buying trends, such as the top-selling SUV for a particular ZIP code. The entire database comprises petabytes of data, and it continues to grow at a fast pace.</p>	<p>A project of the Defense Advanced Research Project Agency (DARPA) uses spatial and temporal data aggregation from multiple data streams such as counts of clinical diagnoses, sales of over-the-counter influenza remedies, and school absenteeism among a given age group to feed alerting algorithms that can warn of a public health threat. Basic organizational considerations for these systems include determining which records to count and how to group them in space and time.</p>

**Figure 1:** Data Aggregators help to fill the knowledge gap in a wide variety of industries and markets

Data aggregators quickly learn that decision support requires more than contextualizing related pieces of information. To be competitive, data aggregators need to provide users with the ability to analyze and obtain historical, current, and predictive views of business operations, all while facing the exponential growth of data involved in their analysis. Data aggregators are not only gathering and storing more operational data, but also have increased the granularity of the data captured and the number of data feeds integrated from market providers. High performance business analytics capabilities that can effectively handle this data glut and offer actionable business insights with fast turn-around time have become the key competitive differentiator for data aggregators.

#### **ANALYTICS-AS-A-SERVICE: FOUR PHASES**

In recent in-depth interviews conducted with several leading data aggregators who offer market segment-specific data and analytics services, Sybase confirmed that analytics has become an important strategic business driver. Observations from these leaders, some of whom are Sybase customers and some whom are not, as well as information gathered from 70 other Sybase customers from various industries who also offer analytics services, led Sybase to recognize four progressive maturity phases for the analytics services provided by data aggregators:

- Phase 1: Extreme Data Management is the Norm
- Phase 2: Power Users Pressured
- Phase 3: Clients Demand Control
- Phase 4: The Data Speaks for Itself

IN PHASE 1, INFORMATION PROVIDERS AGGREGATE DATA AND EQUIP THEIR CUSTOMERS WITH SUMMARIES AND REPORTS

##### **Phase 1: Extreme Data Management is the Norm**

In this phase, data aggregators build the foundation for future analytics services by acquiring data, enriching it and providing summarized data and basic reports. These reports may be standard (same for all clients) or semi-custom (similar for all clients but customized by data sources, segmentation or other attributes). The reports could be based on the data alone, or around information from campaigns or programs based on the data. Data delivery scheduling may vary, and the aggregators provide updates according to the clients' needs.

Data aggregators or information providers view capturing more data as the golden ticket to differentiate themselves from competitors, providing new services that can attract new customers in different segments. The data volume is ever increasing because data aggregators want to maintain historical data for longer periods of time ("ideally, we'd like to keep data on line forever"), and because they are constantly capturing different types of data from additional sources, often at a more granular or elemental level. All the surveyed companies in this phase were struggling with the challenges of managing increasing data volume, which is why this phase was dubbed "Extreme Data Management is the Norm".

IN PHASE 2, INFORMATION PROVIDERS MOVE BEYOND STANDARD AND CUSTOMIZED REPORTS TO PROVIDE TRUE ANALYTICS SERVICES.

##### **Phase 2: Power Users Pressured**

All the leading data aggregators and most of the 70 other organizations in the Sybase study reported that they had begun to offer some level of analytics services directly to clients. In this phase, these services tend to be highly consultative and quantitative, and are performed by analytics experts such as statisticians. The name of this phase, "Power Users Pressured," reflects the fact that power users, often within the data aggregator's own organization, control the design of questions and the access to answers. The experts build models to describe segmentation, predictions, and scoring and provide reports to the clients based on these analytic models. These "power analytics" services offer substantial business value, providing higher revenue, greater client "stickiness", and competitive differentiation in the marketplace.

PHASE 1	PHASE 2	PHASE 3	PHASE 4
<p><b>The Korean Health Insurance Review Agency</b>, a body that examines and evaluates the medical expenses of 45 million people, needed to consolidate data that was dispersed and managed separately. Using Sybase IQ, the agency developed an electronic data warehouse, significantly reducing query times, and enabling immediate access to years of data while lowering raw data storage requirements. This solution allowed the agency to reduce the time for analysis from four days to 30 seconds, while maintaining five years of data — up from five months — and reducing the volume of raw data by up to 80 percent .</p>	<p><b>Experian Integrated Marketing (EIM)</b>, a global information services company, needed to rapidly design, build and manage large-scale, high performance data solutions to deliver insight and analytics to its clients. In looking for a provider, EIM sought a solution that offered flexible capabilities, fast performance, and the ability to scale to meet increased future demand. EIM also wanted a solution that would allow it to support both analytical and operational data requirements from a single data source; they chose Sybase IQ.</p>	<p><b>Nielsen Media Research</b>, provides measurements of television audiences and related services. With hundreds of broadcast channels running 24/7, NMR collects hundreds of terabytes of data on audience demographics and viewing habits residing in the company's data center. This data is critical to broadcasters and advertisers, who rely on it to make and fine-tune programming decisions and advertising placements. Using Sybase IQ, Nielsen developed an online audience data warehouse, enabling its clients to access audience data online. Clients were able to use the Nielsen data to determine for themselves where to place their ads to get the product message to the right audience at the right price.</p>	<p>Phase 4 is an emerging phase that information providers are experimenting with in partnership with innovative clients. System-to-system analytics is also used within single organizations such as <b>Material Requirements Planning (MRP)</b> systems in manufacturing, or trading systems in capital markets. The use of B2B analytics between trading partners in e-commerce is also emerging.</p>

**Figure 2:** Success through the Four Phases

IN PHASE 3, THE INFORMATION PROVIDER'S CUSTOMERS BEGIN TO TAKE A MORE ACTIVE ROLE IN DEFINING AND ACCESSING THE DATA ANALYSIS THEY NEED.

**Phase 3: Clients Demand Control**

In this phase, the information providers create a web-based portal for clients to do some analysis of their own. This type of environment has been called “self-service analytics” or “on-demand analytics,” and the data aggregators and their clients have greater parity: both are experts, but in different ways. The data aggregators remain the experts in the data and programs, while the clients are the experts in the application of the data and programs.

The level of analytics sophistication can vary: from data aggregator-built models with semi-customizable reports that clients can modify by requesting different segments and parameters, to services that provide clients with full advanced analytics, ad hoc analysis and reporting capabilities.

Self-service analytics is not a full substitute for power analytics, but opens up new service areas for data aggregators to sell new products to credit marketers, affinity partners, customer service representatives and other users with knowledge of analytics. In addition, self-service analytics may drive the need for more power analytics as clients reach the edge of their own capabilities and need more from the data aggregators. It also broadens the market by providing lower-cost services for novice users, especially in the mid-market.

IN PHASE 4, THE INFORMATION PROVIDERS INTEGRATE KEY ANALYTICS DIRECTLY INTO THEIR CLIENT'S BUSINESS PROCESSES AND SYSTEMS

**Phase 4: The Data Speaks for Itself**

During this phase, data aggregators integrate analytic systems directly into their client's business processes and automate decisions and actions. Similar to Phase 2 and Phase 3, Phase 4 augments rather than replaces the analytic services that came before. If Phase 2 is about people delivering analytics to people and Phase 3 is about systems delivering analytics to people, then Phase 4 can be described as about systems delivering analytics to other systems.

The “Data Speaks for Itself” phase represents a whole new level of business development for data aggregators. Clients who seek out this type of service will do so in order to lower the cost of conducting business and provide unique service offerings to their customers.

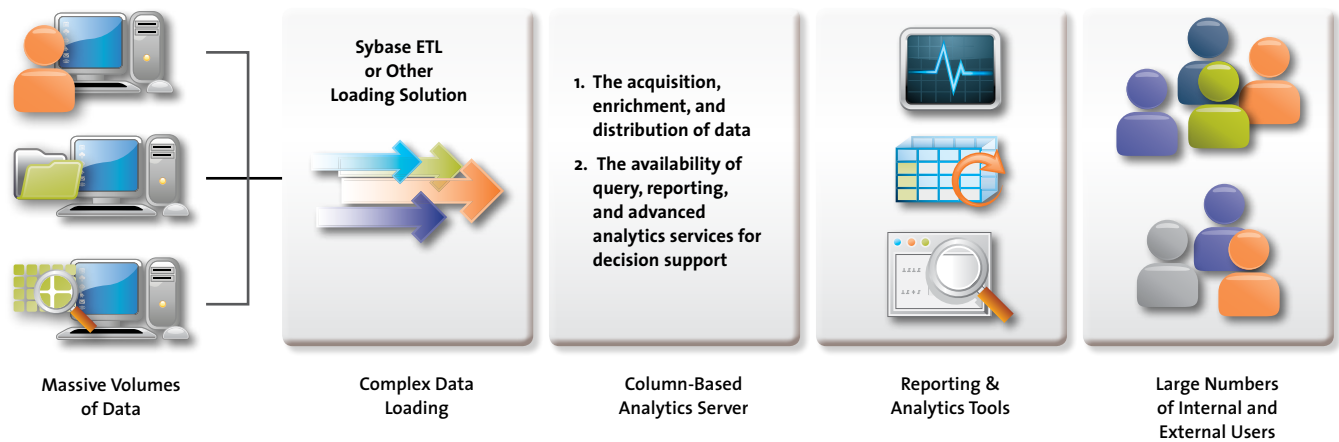
### THE DATA AGGREGATION PROCESS

Data aggregators need to optimize two main functions:

- the acquisition, enrichment, and distribution of data
- the availability of query, reporting, and advanced analytics services for decision support

Optimization of these functions is a key competency because the aggregation is often performed under high-pressure requirements — hundreds or thousands of concurrent users that analyze very high volumes of fast changing data using complex and ad hoc queries.

Data aggregators are constantly gathering large amounts of data from heterogeneous sources such as diverse repositories and databases, operational applications, news feeds and the Internet. Depending on the services provided, the data could be stored raw; alternatively, it may need to undergo a set of transformations: converting, cleansing, and merging with other data, as shown in Fig.3. The resulting stream is then loaded into a data repository where it will either create new or update existing data.



**Figure 3:** From raw data to actionable user information with Sybase IQ

Advanced analytics functions, such as data mining, require that large amounts of data be available in order to discover previously unknown patterns and correlations, and to predict future outcomes. The more data that is accessible for analysis, the more accurate predictive models are likely to be. Performing analysis of trends that is sufficient to make meaningful predictions may require many years worth of historical records, and many data aggregators are now maintaining data warehouses that incorporate hundreds of terabytes of information.

## **CHALLENGES IN DATA AGGREGATION**

Two conflicting challenges that data aggregators face are response time and cost of operations. In order to conform to the response time service level agreements, they often add more resources to their data centers. The resulting complexity of the system now requires additional administration resources for hardware and software installation, as well as maintenance, configuration, and deployment of ongoing administration and performance monitoring and tuning. In many data centers, there is a trend toward using multiple inexpensive processors in lieu of large mainframes to reduce hardware costs. Unfortunately, this also results in a proliferation of software components, higher administrative and operational complexity, and ultimately higher management costs. Various studies estimate that roughly 80% of the Total Cost of Ownership (TCO) in large data centers relates to software operation and management, and this is likely to increase further. These issues are driving the need for a faster, more efficient way for data aggregators to manage and store their ever growing business data, as analytic databases are now so large that the hardware required to use relational systems can be too expensive, even at current prices.

Providing comprehensive analytic services to clients presents a challenge of its own, because maintaining a large, highly skilled analytics staff is expensive. Therefore, user-friendly, end-user analytics tools need to be integrated into the data aggregators' frameworks. However, as clients start using analytics heavily for their decision support processes, they become more demanding: they want to analyze more data, in more detail, and with more customization. Many data aggregators are finding that the Power Users business model is not scalable enough to meet growth requirements — especially when expanding into the mid-market. But self-service analytics also pushes system scalability because users are numerous and concurrent. Security can also be a challenge, especially if corporate data is exposed to the outside world.

## **SYBASE IQ: ADDRESSING THE CHALLENGES IN DATA AGGREGATION**

The choice of database architecture is a critical factor in how well data aggregators are able to address these challenges. Traditional row-based databases are not well suited for large scale analytical processing because they are designed to retrieve all elements from a number of records, while analytical queries typically look at all the records but need only a few elements of each. Column-oriented databases allow all instances of a single data element, such as car license numbers, to be stored together so they can be accessed as a unit. Column-oriented databases are therefore particularly efficient at analytical queries, where the system reads a small number of data elements but needs to work with all instances of these elements.

### **Cost and Performance**

Sybase IQ, the industry's leading column-oriented database server, excels in solving the two main problems data aggregators face: the cost associated with managing large datasets, and the performance of analytics.

Sybase IQ offers powerful compression, resulting in significant cost reduction due to reduced storage requirements. Sybase IQ creates a two- to five-fold reduction of raw data size, depending on the structure of the data. These results exceed those obtained with a traditional database because storing data in a column-oriented fashion greatly increases the similarity of adjacent records on disk, and consecutive entries in a column are often quite similar to each other. Column-oriented compression schemes also improve performance by allowing database operators to operate directly on compressed data. In addition to its column-oriented storage, Sybase IQ delivers a number of specialized indexes in order to further accelerate ad hoc query performance.

“SYBASE IQ IS A COLUMNAR DATABASE, WHICH GIVES IT ITS STRENGTH AT QUICK LOADING AND QUICK RETRIEVAL OF DATA. IT PLAYS AGAINST THE APPLIANCES IN TERMS OF LOAD SPEED, QUERY SPEED, AND PRICE PERFORMANCE. HOWEVER, IT IS ALSO A FULL ANALYTIC DATABASE WITH MIXED WORKLOAD ABILITY, MEANING IT CAN SCALE UP BOTH THE AMOUNT OF DATA CONTAINED IN THE PRODUCT AND THE NUMBER OF USERS AND RANGE OF QUERIES THAT IT CAN HANDLE. FLEXIBILITY TO HANDLE A RANGE OF QUERIES HAS ALWAYS BEEN THE KEY CHALLENGE FOR ANY DATA WAREHOUSING OR BUSINESS INTELLIGENCE PROJECT, AND WILL CONTINUE TO BE A DRIVER OF SUCCESS FOR SYBASE.”

— Alys Woodward  
IDC

Sybase IQ’s column-oriented architecture delivers the high performance data aggregators need. The database was designed from the ground up for complex analytics on massive amounts of data. Because analytical queries typically read a few elements in all records, the column-oriented structure lets the database engine load only the information it needs. Only the columns required to answer a query need to be read — in effect, every possible ad hoc query has performance akin to a covering index. This reduces the amount of disk access and speeds the result. The queries focus on specific columns, and only those columns requested need to be loaded across the memory hierarchy. Additionally, each load is more efficient due to the sequential layout of each column. Furthermore, enhanced Parallel Processing provides Sybase IQ with very high performance data loading by fully utilizing modern day hardware technologies, thus giving customers a highly economical solution.

Sybase IQ provides both multithreading and 24/7 high availability features (including partnerships with relevant storage vendors for high availability and disaster recovery). In particular, separate read and write nodes allow for procedures to be executed in parallel, without affecting one another. In the latest release of the Sybase IQ analytic server, significant additional parallelization capabilities speed up both query and load performance. Separate read nodes are particularly useful for data aggregators that offer multi-client analytics services, as a node can be assigned to an individual account for later chargeback.

As of release 15.1, Sybase IQ also delivers in-database analytics functionality to run predictive analytics business logic directly in the database, removing the need to move data out of the database and significantly reducing time to results. It is the only high performance column-based analytics server that supports the execution of hundreds of statistical and data mining functions completely in-database while supporting hundreds of concurrent users at high performance levels.

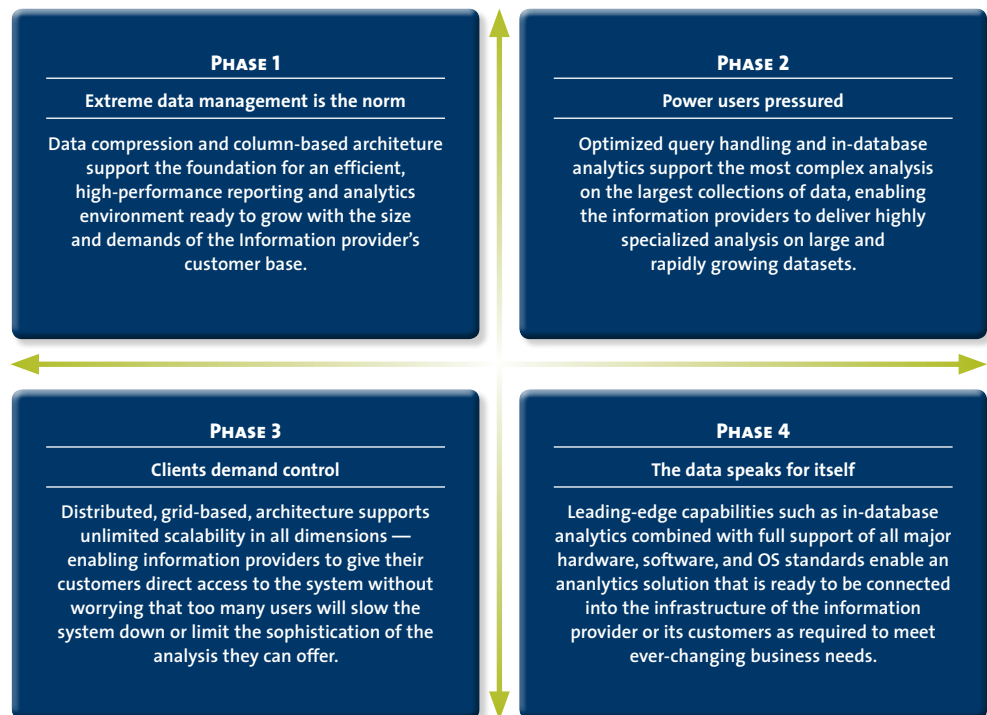


Figure 4: Sybase IQ provides critical capabilities at each phase

“WE CAN DEPLOY MULTIPLE CAPABILITIES FOR A SINGLE CLIENT ON THE SAME SYSTEM AND CONFIGURE EACH ONE FOR SPECIFIC RESPONSIVENESS. THIS WAS A BIG BENEFIT, BUT WE WERE ULTIMATELY CONVINCED BY SYBASE IQ’S PERFORMANCE AND ABILITY TO SCALE WITHOUT PERFORMANCE DEGRADATION.”

— Tony Mooney  
Experian Integrated Marketing

## WHEN THE INFORMATION IS THE BUSINESS: CASE STUDIES

Data aggregation extends the scope of decision support for a given client by aggregating multiple sources of data that broaden the understanding of the context of the problem and lead to better, and often unexpected and fruitful insights. The value of these insights increases significantly if the related information is delivered quickly and accurately to all the users that need it. The following are real case studies that describe the problems some data aggregators faced, the solutions they provided, and why Sybase IQ was the data aggregators’ data and analytics platform of choice.

### Case 1: Experian Integrated Marketing

Experian Integrated Marketing (EIM) delivers best-in-class, intelligence-driven personal marketing solutions to leading consumer brands in more than 65 countries, keeping clients a step ahead of the competition. EIM enhances their clients’ raw customer data and provides segmentation tools and predictive models that foster insight about consumer buying patterns and likely behavior, identifying sets of ‘best fit’ prospective customers that the client will target with specific marketing campaigns.

*“We consider ourselves a service business that offers marketing strategy solutions,”* says Tony Mooney, Managing Director for Experian Integrated Marketing. *“For example, one of our services takes call centre info, wraps in campaign management and technology data, and then offers value and propensity modeling.”*

The analytics requested by EIM clients requires a data processing system that can handle very high volumes. *“Our system has relatively high I/O and memory-intensive functions,”* Mooney explains. *“We process 600 million records per day, and as the #1 marketing data hosting service in the UK, we need to house and query very large volumes at fast speeds.”* In looking for a provider, EIM sought a solution that offered flexible capabilities, fast performance, and the ability to scale to meet increased future demand. It also wanted a solution that would allow it to support both analytical and operational data requirements from a single data source; they chose Sybase IQ.

*“We can deploy multiple capabilities for a single client on the same system and configure each one for specific responsiveness,”* Mooney says. *“This was a big benefit, but we were ultimately convinced by Sybase IQ’s performance and ability to scale without performance degradation.”*

### Case 2: Nielsen Media Research

Nielsen Media Research provides television audience measurement and related services. They cover hundreds of broadcast channels that run 24/7, and their data center houses hundreds of terabytes of data on audience demographics and viewing habits.

Sybase IQ enabled Nielsen to address a number of formidable challenges inherent in building and maintaining a very large data warehouse. Nielsen loads almost 300 million rows a day to the warehouse, and this needs to be done in a very short time frame — an hour or two at the most. Nielsen found Sybase IQ’s load speed to be eight times faster than the solution it previously used.

*“I would venture to say Sybase IQ is the fastest in the industry,”* says Craig Silver, senior database architect of the data technology group at Nielsen. *“We load almost 10 billion rows a month ... With Sybase IQ, it just flies.”*

Sybase IQ’s data compression functionality lowers Nielsen Media Research’s cost of operations by reducing the company’s storage requirements by about 70 percent. *“When you’re dealing with less than a terabyte of data,”* notes Kim Ross, CIO, *“the cost of storage is not that significant a factor. When you’re dealing with tens of terabytes, as we are, however, it’s a very significant factor — you’re talking about millions of dollars. To be able to save that kind of money is very important to us.”*

“OUR CLIENTS WANT MORE POWERFUL ANALYTICAL APPLICATIONS AND EVER MORE DETAILED AMOUNTS OF INFORMATION; MEETING OUR CLIENTS’ NEEDS MEANS BUILDING MORE POWERFUL AND MORE USEFUL APPLICATIONS.”

— Kim Ross  
Nielsen Media Research

Most important to Nielsen Media Research’s clients, perhaps, is Sybase IQ’s ability to deliver answers to even the most complex queries 10 to 100 times faster than traditional data warehouse technologies. This enables them to sort through enormous volumes of data (the warehouse currently contains 20 terabytes of data and is growing at the rate of two to three gigabytes a day) to get the intelligence they need to make the most effective and profitable business decisions.

*“Our clients want more powerful analytical applications and ever more detailed amounts of information; meeting our clients’ needs means building more powerful and more useful applications,” says Ross. “It also means building and managing larger, more complex databases and serving that information to clients in more convenient ways — whether that means through the wireless Internet environment where they’re using wireless laptops or enabling them to access it on their PDAs. Thanks to Sybase IQ, we’re confident we’ll be able to continue to deliver value to our clients, while at the same time producing a significant, bottom line ROI for Nielsen Media Research.”*

### Case 3: The Korea Health Insurance Review Agency

Korea Health Insurance Review Agency (HIRA) stores medical care details — such as medical treatment results and expenses to be paid — for Korea’s 45 million people in accordance with Korea’s National Health Insurance Act. To manage this large flow of data and produce meaningful analysis, the agency needed to consolidate data for all of Korea, currently dispersed and managed separately across seven branch institutions. Furthermore, it needed to provide immediate access to years of data that the individual branches kept available for only five to six months.

HIRA considered using its legacy system provider, Oracle. However, the raw data storage volume was burdensome, so HIRA conducted examinations using simplified data and documents. Using Sybase IQ’s unique compression and indexing technologies, HIRA reduced the volume of raw data by 60 to 80 percent. This comparison strongly favored Sybase IQ, which drastically cut storage, administration and maintenance costs.

General Manager Han Beom-Su explained one reason for HIRA’s selection of Sybase IQ as their data warehouse. *“According to published benchmark results, performance of NCR Teradata was also excellent,” he said. “But the weak point with NCR Teradata was that it was hardware-dependent. That would cause a great deal of difficulty in linking with other hardware and expanding its capacity.”*

With Sybase IQ, the volume of raw data was decreased by 80 percent, and queries were reduced from four days to 30 seconds. Korea HIRA has also consolidated a national record of treatment details, and Sybase IQ provides immediate access to five years’ worth of data.

*“It used to take three to four days to analyze non-standardized data for one month’s activities,” said Lee Ji-Seung, deputy general manager of information communication division, Korea Health Insurance Review Agency. “Now it takes a minimum of 30 seconds to a maximum of three hours.”*

HIRA staff — formerly mired in manually intensive data collection, statistical analysis and distribution — can now focus on more value-driven efforts. The Sybase IQ data warehouse delivers electronic data processing and analysis of medical treatment results, disease and injury factors, and aggregate trend analysis, enabling data teams to focus on more comprehensive analysis and data utilization. HIRA now produces better reports and supplies information crucial to government decision-making and health care planning for all of Korea.

“IT USED TO TAKE THREE  
TO FOUR DAYS TO ANALYZE  
NON-STANDARDIZED DATA  
FOR ONE MONTH’S ACTIVITIES  
NOW IT TAKES A MINIMUM OF  
30 SECONDS TO A MAXIMUM  
OF THREE HOURS.”

— Lee Ji-Seung

Korea Health Insurance Review

#### Case 4: The Alfred Wegener Institute

The Alfred Wegener Institute for Polar and Maritime Research (AWI) in Bremerhaven, Germany conducts research in the Arctic and Antarctic regions, as well as marine areas in temperate regions. “Information slices” are extracted from roughly 1.6 billion datasets in the information system, and the resulting data is formatted by a special graphics application, allowing scientists to study local concentration readings on a multi-colored world map. The PANGAEA data system developed by the institute stores the information and offers a list of 46,000 parameters that can be selected for ad hoc queries.

Measurements of ozone, particles, greenhouse gases, trace gases, aerosols, turbidity and other parameters are taken at regular intervals (minutes and hours). Present-time atmospheric measurements are complemented by ice core drilling samples to trace back the evolution of the earth’s atmosphere during the last one hundred thousand years. All of the measurements gathered are archived in PANGAEA as a basis for a long-term series analysis comparing historical and modern-day values.

Over the last few years, the contents of PANGAEA have increased exponentially. Since 1996, the data volume has doubled each year. By November 2006, 136 research projects have entered 470,000 datasets that included 1.7 billion data points, amounting to a volume of 1.2 terabytes. And the data keeps growing: several observatories are feeding in new values every minute. The stream of data is expected to grow by one or two orders of magnitude once the sensor networks currently under construction begin to deliver additional oceanographic, meteorological and geophysical data.

Sybase IQ, the database on which PANGAEA is built, stores the validated and published data. The ability to provide a multitude of attributes and measured values is due to Sybase IQ’s high compression ratios. This allows raw data to be compressed to between one-sixth and one-ninth of its original volume.

The PANGAEA system also leverages Sybase IQ’s inherent indexing architecture. A traditional relational database requires multiple indices to be created before qualified searches can be performed. This inflates the data volume, which limits the indices and compromises query flexibility. In contrast, Sybase IQ organizes data in columns rather than rows, allowing any field to serve as an index. This enables very flexible searches: the database always offers features that support access to whatever search criteria the user selects. Since queries only read the selected value in the respective column instead of the entire row, they are processed very rapidly even in very large datasets.

Bremen University is developing a universal access tool that will allow user-defined subsets to be extracted from the pool of data. Sybase IQ supports the implementation of this tool through its efficient indexing of time information. The tool will allow scientists to build and view custom subsets of time series. For instance, in the exceptionally clean air of Antarctica, the atmospheric chemistry observatory can track long-term developments without being compromised by local, human influences.

## **CONCLUSION**

Data aggregators, also known as information providers are the driving force behind the emergence of Analytics-as-a-Service. They face the challenge of capturing, integrating, and analyzing multiple vast stores of data to deliver information-related products and services to their clients. For data aggregators, information is the core of their business — their services are only as good as the speed, quality, and completeness of their reporting and analytical capabilities. Sybase IQ's column-oriented architecture and innovative features enable data aggregators to serve large numbers of users' ad hoc and complex queries with a very fast response time. Sybase IQ's data compression algorithms support the largest databases in an efficient and highly cost-effective way.

Recent additions to Sybase IQ's capabilities include in-database business analytics plus an enhanced grid architecture that enables the database to scale to thousands of concurrent users. Information providers can grow their decision support services by running real-time predictive analytics directly in the Sybase IQ database server, increasing speed and accuracy while utilizing years' worth of organizational data.

Sybase IQ distinguishes itself from traditional OLTP databases with its column-based architecture designed for analytics performance and its patented data compression, which ensures a lower total cost of ownership than traditional solutions can provide. Sybase IQ distinguishes itself from new players in the column-based analytics server space by the completeness and maturity of the offering, as demonstrated by features such as unique and proven query optimization technology. Sybase IQ distinguishes itself from data warehouse appliances and other hardware-based offerings with its flexible scalability and open support of any hardware, storage, or operating system.

Such advantages help to explain why Sybase IQ is a leader in the specialty analytics server market, and has in fact been the leading column-based analytics server for more than 10 years. As the case studies above indicate, Sybase IQ has repeatedly distinguished itself as the right platform for the widest variety of data aggregation challenges. As shown in Figure 4, and as demonstrated in some of the most complex data aggregation and analytics environments in the world, Sybase IQ is uniquely equipped to meet the challenges faced by those for whom the information is the business.

## **BIBLIOGRAPHY**

Roger MacNicol, Blaine French, *Sybase IQ Multiplex — Designed For Analytic*, Proceedings of the 30th VLDB Conference, Toronto, Canada, 2004

Philip Howard, *Sybase IQ 15 In Detail*, Bloor Research, 2009

Alys Woodward, *Sybase Releases IQ 15 to Capitalize on 2008 Success in Europe*, IDC Link, 2009

*Specialty analytics servers: Enabling decision-making @ the speed of business*, peer2peer, 2008

Kyriaki Venetis, *Realty Data forms alliance and rolls out database*, National Mortgage News, 2002

Todd Spangler, *Polk Revs Its Data Engine*, Baseline, 2006

David Loshin, *Gaining the Performance Edge Using a Column-Oriented Database Management System*,

Sybase white paper, 2009

David M. Raab, *How to Judge a Columnar Database, Revisited*, DM Review, 2008

Adam Cherubini, *Online Aggregators Can Help Open Doors*, National Underwriter Property & Casualty, 2008

Chris Kentouris, *Where Low (Latency) Is High (Value)*, Securities Industry News, 2007

George Plosker, *Making Money as an Aggregator*, Online, 2004

Hiroshi Fujii, Taeko Okano, Stuart Madnick, Michael Siegel, *E-Aggregation: The Present and Future of Online Financial Services in Asia-Pacific*, MIT, 2003

## **ABOUT THE AUTHOR**

Daniel Tkach is CEO of PartnersMarket Consulting, Inc. [www.partnersmarket.com](http://www.partnersmarket.com). Previously he was VP of Marketing with Sinapsis USA and Senior Director of Outbound Marketing with Tasman Networks and AI Tech Software. He held senior marketing positions with Oracle and IBM, was Principal and Practice Leader with IBM Global Services and served as Director of Technology of the IBM Institute for Knowledge Management. Daniel Tkach has earned MBA, Ms Cs and BSEE degrees. He has authored numerous white papers and journal articles and two books on Application Development with Object Technology. He serves as Vice Chair of the TechExecs Silicon Valley Chapter and is a Member of the Industry Advisory Board of the Department of Computer Science, California State University, Chico.



SYBASE, INC.  
WORLDWIDE HEADQUARTERS  
ONE SYBASE DRIVE  
DUBLIN, CA 94568-7902  
U.S.A.  
1 800 8 SYBASE

[www.sybase.com](http://www.sybase.com)

Copyright © 2010 Sybase, Inc. All rights reserved. Unpublished rights reserved under U.S. copyright laws. Sybase, the Sybase logo and Sybase IQ are trademarks of Sybase, Inc. or its subsidiaries. All other trademarks are the property of their respective owners. ® indicates registration in the United States. Specifications are subject to change without notice.  
04/10 L03271

**SYBASE®**