

The next evolution of data warehousing: Smart consolidation

Driving a new level of agility and performance



For those who understand the power of an analytics-driven organization, this is the most exciting time in years. The opportunities are limitless: customers, prospects, suppliers and the business itself are creating endless geysers of data. Analytics tools are inexpensive, widely available and so easy to use that they make business sense in almost any situation.

And yet, many business analytics initiatives are stumbling, slowing to a crawl or simply failing because their traditional data warehousing architectures cannot keep up. The large, centralized data warehouses that organizations have depended on for years were not designed to handle the sheer volume of data, the need to integrate information from wildly different sources, and the new and practically infinite variety of formats of data, including unstructured formats.

Looking beyond just the warehouse, organizations are also realizing that one-size-fits-all data warehousing and analytics solutions simply cannot deliver complex analytics on big data while simultaneously meeting the service-level agreements (SLAs) required by core business processes and operational business reporting. These workloads place vastly different requirements on the underlying systems, requiring IT departments to take a wider, more holistic view of both their data and their systems architecture.

Meanwhile, demand for analytics is rising fast, placing tremendous pressure on IT organizations. Business analysts want to run highly complex analytics on massive amounts of data. Functional managers and executives insist on immediate access to operational reports, dashboards and scorecards.



If any of these demands are not met, users have proven perfectly willing to take matters into their own hands. In organizations everywhere, users are creating local, ad hoc, ungoverned data marts (see Figure 1). In some cases, departments even set up their own data warehouse, bypassing the central IT infrastructure completely.

These local efforts at building data marts and data warehouses can present serious problems. The groups that create them often find it nearly impossible to maintain data quality, consistency and currency. The analysis and insights that result are worse than useless—while they may appear to be trustworthy and well-founded, they can be incredibly inaccurate and ultimately lead to poor business decisions.

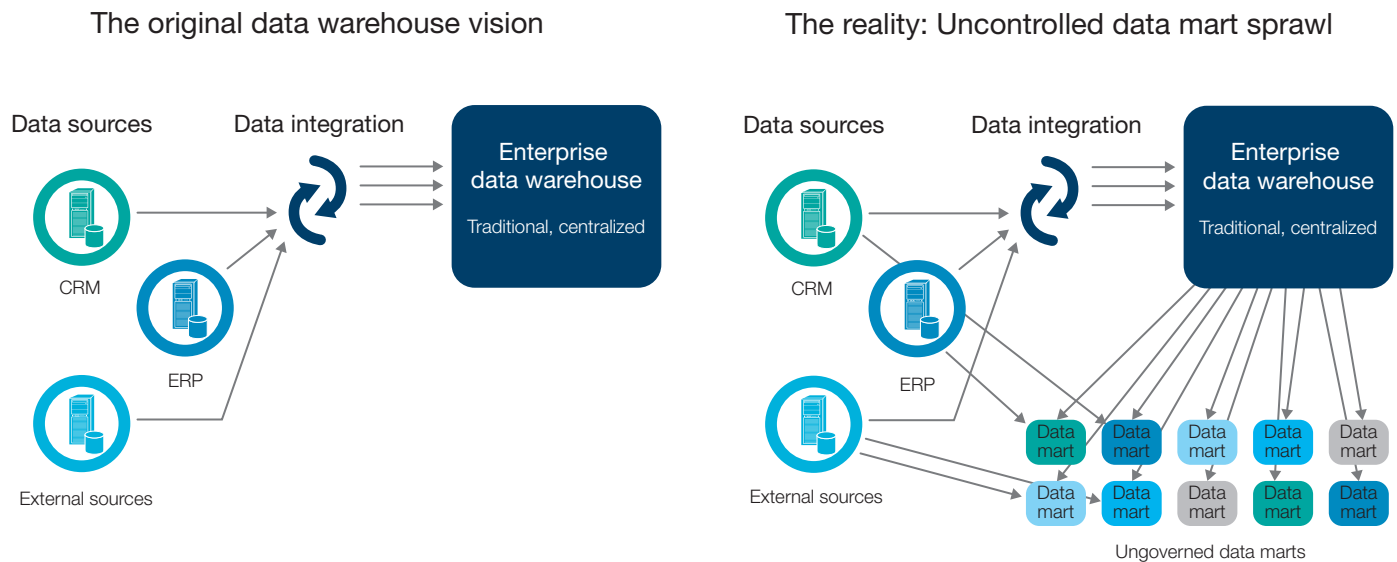


Figure 1: Demand for fast analytics is pushing data warehouses away from their original vision, leading to ad hoc data marts and other hard-to-control setups.

Decades of experience have proven that allowing IT investments to be driven by local needs is unsustainable, as it almost guarantees inefficiency, redundancy, data currency and quality issues, and increased management costs. Furthermore, a locally driven information infrastructure makes data governance nearly impossible at precisely the time that auditable data quality is a nonnegotiable business requirement.

It is time to acknowledge the challenges to be faced and realize that the single, centralized, traditional data warehouse has reached a point of overload. To move forward, organizations need a strategy that delivers on several focused business requirements:

1. **Operational management:** Accelerate time-to-market to meet business SLAs for new and existing business processes, operational analytics and business intelligence (BI).
2. **Creative investigation:** Perform complex and ad hoc analytics on large amounts of data to discover new insights, trends and relationships.
3. **Big data:** Leverage unstructured data, social media, metering data and other “big data” information sources to gain more insights from more data—without affecting the business SLAs.
4. **Predictive analytics:** Forecast future trends and analyze risks and potential outcomes.

What to do?

Many IT organizations are adopting a strategy called *smart consolidation* that reconciles the need to simultaneously distribute data warehousing, analytics capabilities and infrastructure while centralizing management. Smart consolidation is a method for evolving an existing data warehouse architecture to meet today's demanding analytic needs, such as big data, streaming data and unstructured data. It involves thinking beyond the traditional warehouse structures that have provided great success with structured data, basic reporting and analysis.

Smart consolidation is driven by four goals:

1. Consolidate and govern enterprise data.
2. Optimize workloads for performance and SLAs.
3. Simplify the delivery of analytics by leveraging appliances.
4. Flexibly extend analytic capability as needed.

The basis for smart consolidation is to completely optimize the analytics architecture by placing the *right workload* against the *right data*, in the *right place*, at the *right cost* and the *right performance level*.

Smart consolidation acknowledges that an organization might require multiple, different types of databases, analysis tools and data formats. Organizations need traditional data warehouses, data warehouse appliances and operational BI systems that can accommodate different types of workloads. It also needs systems based on advanced technologies that can

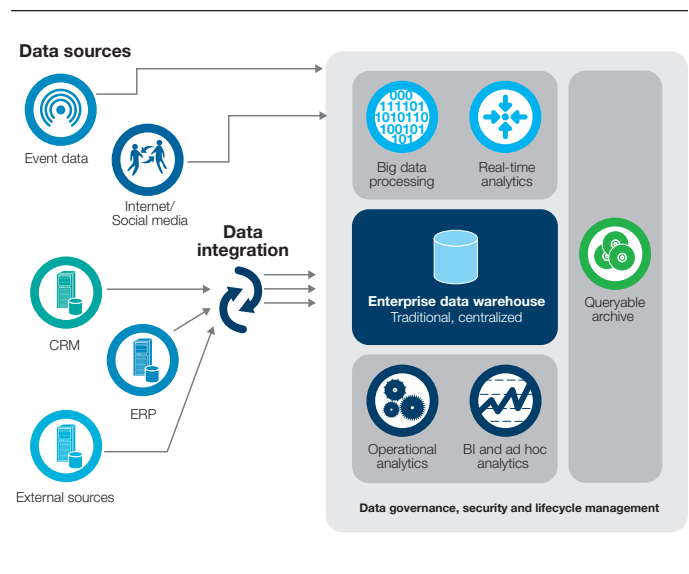


Figure 2: Smart consolidation is an evolutionary strategy that joins specialized processing elements into a data ecosystem that delivers information as well as broad, cross-organizational governance and resource management.

efficiently handle data that is moving extremely quickly as well as large volumes of data that do not change frequently.

No single data system could efficiently serve all these requirements and perform well for both transactional and analytical workloads. Under the smart consolidation strategy, multiple specialized elements use industry standards to communicate and join together to form a fluid, agile data ecosystem that delivers business insight, cross-organizational data governance and centralized IT resource management (see Figure 2).

In this evolving ecosystem, a data infrastructure is built and managed on the premise that different business units require different types of data from many sources, and that the organization as a whole needs to integrate that data for analysis. Once integrated, operational analytics for key business processes and daily decision making are performed against consolidated data using systems that are optimized for mixed workloads—transactional analysis plus simple and complex analytical queries.

When the centralized warehouse becomes performance-constrained, the data needed for large, complex, time-

consuming analytics or specialized projects is intelligently replicated onto systems and appliances that are optimized for massively parallel analytical processing, such as data warehouse appliances or data marts. Through this separation of workloads, organizations get the best of both worlds: fast, flexible reporting and operational analytics to meet SLAs plus separate, dedicated systems for big data, complex analytics and creative investigation. Because different applications and analytical workloads can be assigned to optimal execution elements, organizations can create a framework of service levels, increasing overall efficiency.

Smart consolidation in action: What does it look like?

Consider the example of a credit card company. When a customer applies for a credit card, the sales department collects the customer's details and financial history, and then compares it to historical data from third-party reporting agencies to determine the customer's ability to manage and repay debts. The customer data flows to the marketing department, where it is analyzed for trends and compared with opinion content collected from the Internet to make decisions on promotional campaigns. Eventually, the customer might request a credit-line increase, at which time the customer service system will recommend up-sell opportunities and the lending department uses the customer's payment history to evaluate the request.

Meanwhile, the company's online transaction processing (OLTP) systems are fielding millions of transaction authorization requests per minute. Real-time analytics systems are looking for anomalies that may indicate fraud by comparing the streams of transaction data to patterns developed by analyzing customers' purchasing histories. As all this data ages and becomes more static, it shifts to archival systems and is stored using specialized technologies such as Apache Hadoop—yet it remains available for instant auditing and long-term trend analysis.

At the same time, the marketing department is investigating a new customer segmentation model to use in an upcoming product launch. Marketing has been busy analyzing its complete customer database to determine online banking trends as well as smartphone and mobile banking adoption rates. After many iterations of their segmentation model, the marketing team believes they have identified the data elements and customer behaviors that define a financially sophisticated and technologically savvy customer segment. Now, several months before the launch, the product manager is running predictive models to test the business case on combinations of marketing messages and user adoption rates. The team is free to test and retest assumptions, even though queries take a long time to execute, because those queries are running on an analytics-optimized data warehouse system—not the primary operational system.

The credit card company is taking advantage of a distributed data and distributed workload architecture. By intelligently decentralizing data and separating workloads, the company is able to creatively analyze data to identify new business models, test assumptions for new paid services and optimize launch and execution plans without impacting the daily, hourly and up-to-the-minute operational needs of its core business.

By allowing many different elements to serve specialized needs, smart consolidation also enables organizations to accommodate the endless variety and rapidly growing ocean of semi-structured and unstructured data. A data ecosystem provides a framework that supports specialized analytical tools for complex data such as statistics, as well as vertical industry data. It gathers data in motion and data at rest. And it enables organizations to apply elements of governance (for example, stewardship and auditability) to ensure that data is accurate, IT can determine the source of data and data is managed throughout its lifecycle. By taking a smart consolidation approach and implementing a data ecosystem, organizations can present a single, cohesive, logical face to users and applications, which do not need to know where a given piece of data physically resides. Analysts call this a *logical data warehouse*.

The IBM data warehouse portfolio

IBM has developed an information management portfolio that can help organizations build and evolve their data ecosystems at the right pace for their business. Moving beyond a one-size-fits-all mentality, IBM offers a diverse set of solutions—from hardware and software to accelerators and expert integrated systems—so organizations can meet their departments' unique data warehouse needs and workload requirements. Solutions from the IBM portfolio are designed to work together seamlessly, using standards-based metadata, definitions, data models and mapping to enable the frictionless flow, management and auditability of data across the enterprise.

IBM data warehouse solutions can also help accelerate the time-to-value, reduce management complexity and control costs. For example, expert integrated systems offer complete, pre-integrated, workload-optimized systems that draw on years of IBM expertise from thousands of successful infrastructure and application implementations. Each system is ready for immediate use and designed to simplify the entire system lifecycle, from acquisition to retirement.

With deep support for open industry standards, IBM solutions work with many systems already in place, allowing organizations to leverage existing investments while providing a catalyst for evolving from a monolithic data architecture to a more effective, efficient ecosystem. Organizations can start with only a few elements of smart consolidation, and then grow and scale as business demands change.

IBM InfoSphere Warehouse

IBM® InfoSphere® Warehouse is a complete, multipurpose environment that allows organizations to access, analyze and act on real-time operational data as well as historical information, whether structured or unstructured. It helps simplify the selection, deployment and maintenance of the information management infrastructure while offering the flexibility for dynamically integrating data and transforming it into actionable insights. With data volumes growing rapidly, InfoSphere Warehouse incorporates capabilities such as multi-temperature data management and Adaptive Compression that help sustain outstanding performance and maximize storage efficiency.



With InfoSphere Warehouse and other offerings such as InfoSphere Information Server and InfoSphere Master Data Management, organizations can more easily create a “single version of the truth” that enables line-of-business (LOB) executives to analyze multiple facets of their organization and discern ways to improve revenue, efficiency and productivity. IBM Cognos® is built into advanced editions, helping organizations customize data warehouses for data mining, modeling, scoring and more.

InfoSphere Warehouse leverages other software from IBM, such as Cognos and IBM SPSS®, to facilitate multidimensional analysis, discovery mining and predictive mining. It also integrates with the unstructured data of traditional flat-file databases to analyze structured, semi-structured and unstructured data within the same context.

IBM PureData System for Analytics

Integrated systems and appliances have quickly become popular by offering complete solutions that are purpose-built for specific situations. Through the integration of highly intelligent software and highly reliable hardware, these systems provide a dedicated resource that simplifies deployment and management while reducing the total cost of ownership. Without the need for in-depth system or query tuning, integrated systems and appliances free up IT staff to focus on other strategic activities.

The IBM PureData™ System for Analytics is a workload-optimized, expert integrated system that makes key contributions to the data ecosystem. The PureData System for Analytics allows data to be managed and exchanged throughout a wider architecture. In addition, it enables IT to place a compact system—optimized for analytical processing and designed for flexible growth—precisely where high-performance analytics are needed.

The PureData System for Analytics offers the speed, simplicity, scalability and intelligence that organizations require:

- **Speed:** The patented integration of IBM Netezza® software and hardware accelerators delivers performance that is typically much faster than traditional warehousing solutions.
- **Simplicity:** This fully integrated system can be ready for data loading in less than four hours. The integrated Netezza technology eliminates the need for complex database management tasks such as defining and optimizing indexes, and manually administering storage.
- **Scalability:** Available in various size configurations, the PureData System for Analytics can be expanded to handle petabytes of user data.
- **Intelligence:** The PureData System for Analytics is able to run advanced analytic algorithms in minutes and provides in-database support for leading analytics applications such as Cognos, SPSS and SAS.



Integrated systems for operational analytics from IBM

Complementing the fully integrated PureData System for Analytics and the highly customizable InfoSphere Warehouse software are the integrated systems for operational analytics from IBM. Within the data ecosystem, these systems combine data from dozens of systems and multiple platforms, whether operational or analytical. Once the data is integrated, organizations can take advantage of a highly automated system with a lower total cost of ownership.

Available in multiple configurations, the systems are designed to handle a wide range of workloads while delivering high-speed performance. Support for several software options, including Cognos BI software and a variety of third-party applications, facilitates numerous types of analysis.

As pre-built systems that include software and hardware, the integrated systems for operational analytics can be deployed in a matter of days, rather than months. As a result, LOB executives can quickly gain a consistent view of information and generate new insights from the organization's data.

The IBM PureData System for Operational Analytics—one of the integrated systems for operational analytics—helps deliver real-time insights into business operations. Designed to handle more than 1,000 concurrent operational queries,¹ the PureData System for Operational Analytics can scale up to a petabyte of data capacity.² Like other expert integrated systems from IBM, it can be ready for data load in just four hours and offers integrated system maintenance that streamlines administration.

The IBM Smart Analytics System is also an integrated system designed for operational analytics. In particular, the IBM Smart Analytics System 5710 is an all-in-one, single-server solution that is ideal for growing companies and enterprise departments looking to quickly and affordably deploy warehousing, analytics and BI capabilities. With deep integration of server, storage and InfoSphere Warehouse and Cognos BI software, it is a flexible, compact BI-in-a-box solution that is easy to implement and can deliver immediate value.

A variety of organizations are already using the integrated systems for operational analytics from IBM:

- An auto parts dealer identifies the most profitable customers and products, analyzes assortments-based consumer shopping behaviors, and optimizes inventory at the store level while supporting more users and delivering better response times.
- An Asian telecommunications company supports more than 1,000 concurrent users accessing 88 TB of data from 51 integrated data sources, including approximately 2,000 tables.
- A UK telecommunications provider tracks dropped calls, monitors customer usage patterns and identifies potentially fraudulent activity.
- A Brazilian financial institution analyzes client buying behaviors, identifies cross-sell and up-sell opportunities, and uncovers new market trends.

Specialty capabilities and applications

The new data ecosystem also encompasses a variety of capabilities targeting specific needs, including needs relating to particular industries and user responsibilities. The following examples represent ways that IT can serve specialized needs with industry-standard capabilities, essentially using the ecosystem as a foundation for addressing these needs without deploying highly specific point solutions.

Governance

Data quality, consistency and structural integrity are paramount for good governance. IBM solutions can exchange data with a variety of BI applications through standard application programming interfaces (APIs). Optimized BI tools from IBM facilitate a quick and easy exchange of data and metadata. Data quality, lineage and integrity are maintained and managed to ensure the system can determine that the data source and structure is accurate, appropriate and does not duplicate existing data.

IBM InfoSphere BigInsights

The data ecosystem must be able to accommodate “big data”—the tremendous volume and variety of data available today. The ecosystem must aggregate structured data from traditional databases, as well as unstructured data from other sources—including text, email, clickstreams, log files, social media, audio, video and images. Only by efficiently managing and analyzing all types of data can organizations generate accurate, comprehensive and up-to-date insights.

IBM InfoSphere BigInsights™ accommodates the open-source Hadoop framework for the distributed processing of large data sets across multiple computers. It also supports query languages (such as Jaql for JavaScript) that analyze both structured and unstructured data types.

IBM InfoSphere Streams

While data warehouses and InfoSphere BigInsights handle data at rest, IBM InfoSphere Streams tackles data in motion that is produced by a wide range of sources, from web traffic and financial transactions to device monitoring systems and sensors. Drawing from work conducted by IBM Research, InfoSphere Streams provides a development platform and runtime environment that enables customers to develop and execute applications that ingest, filter, analyze and correlate massive volumes of continuous data streams. With InfoSphere Streams, organizations can conduct near-real-time analysis on data that previously was too difficult or voluminous to gather. With these newly available insights, organizations can garner intelligence about ongoing activities faster and more accurately than ever before.

IBM SPSS

Some queries go beyond traditional BI queries into the realm of statistical analysis. IBM SPSS software participates in the ecosystem by helping users identify more complex mathematical relationships. With SPSS software, organizations can identify highly profitable customers; conduct predictive analytics that help forecast future trends, outcomes and risks; and understand relationships among disparate customer or prospect groups. Like IBM Cognos software, SPSS software generates easy-to-read reports that enable business managers and executives to visualize and understand data so they can make better, faster decisions.

IBM InfoSphere Warehouse Packs and IBM Industry Models

IBM InfoSphere Warehouse Packs and IBM Industry Models—which work with the entire IBM data warehouse portfolio—help accelerate industry-specific, or vertical, capabilities as well as horizontal functional capabilities that apply to multiple industries. They contain physical or logical data models, data-mining algorithms and sample reports based on specific business issues that help reduce data warehousing project time, deployment cost, data modeling, report building and risk.

For instance, the InfoSphere Warehouse Pack for Customer Insight gives organizations the infrastructure necessary to analyze customer profitability across products and channels. The InfoSphere Warehouse Pack for Market and Campaign Insight helps analyze the elements of a successful marketing campaign or sales initiative. The InfoSphere Warehouse Pack for Supply Chain Insight enables organizations to gather, organize and leverage all relevant information about the entire supply chain, including vendors' activities, inventory, distribution channel performance, related financial forecasting and overall resource planning.

The InfoSphere Warehouse Packs are a subset of the IBM Industry Models. IBM Industry Models combine deep expertise and industry best practices in blueprints, which both business and IT communities can use to accelerate industry solutions. Part of the IBM portfolio, the IBM Industry Models are based on IBM experience with more than 500 clients and more than 10 years of development.

IBM Industry Models offer comprehensive data models containing data warehouse design models, business terminology models and analysis templates to help speed up development of BI applications. In addition, they include best-practice business process models with supportive service definitions for development of a service-oriented architecture (SOA), all helping to accelerate projects and reduce project risk.

IBM Informix TimeSeries for smart metering

IBM Informix® TimeSeries for smart metering offers an excellent example of how IBM solutions can help address industry-specific needs. Informix TimeSeries provides a highly reliable and high-performance metering solution for utility companies that quickly and precisely notifies technicians about outages, reports electricity usage more accurately, helps reduce electricity theft, improves operational efficiency and improves customer service through accurate billing and in-home feedback mechanisms.

A clearer view for better insight

Analytics-driven organizations thrive on data. The good news is that tremendous volumes of data are now available for the asking, along with analytics tools that can uncover previously undreamed-of insights. IBM can help organizations develop information strategies that match their evolving environment, and develop ecosystems that encompass the new multitude of information sources; data volume, variety and velocity; and complex business needs. With a smart consolidation strategy and IBM solutions, organizations can do more than keep pace with the data revolution—they can truly use it to their advantage.

About IBM data warehousing and analytics solutions

IBM provides a comprehensive portfolio of data warehousing, business analytics and information management solutions that can help organizations apply analytics to new business areas and generate new insights that facilitate faster, better decision making. Adopting IBM solutions can help organizations produce results rapidly and economically: IBM solutions are built to accelerate the time-to-value, reduce complexity and control costs.

The IBM portfolio includes software and hardware as well as expert integrated systems that draw on deep expertise to integrate database, server and storage resources into a single, easy-to-manage solution that requires minimal setup and ongoing administration. Workload-optimized for data, the expert integrated systems such as the IBM PureData System are designed to provide fast, consistent analytic performance.

IBM solutions support big data and new, emerging types of analytics workloads—including continuous analysis of data in motion. By enabling organizations to effectively and efficiently analyze the tremendous volume, variety and velocity of data available today, IBM solutions can help organizations maximize the value of information assets and optimize business outcomes.

For more information

To learn more about IBM data warehousing and analytics solutions, please contact your IBM sales representative or visit: ibm.com/software/data/infosphere/data-warehousing

To learn more about IBM expert integrated systems, including the PureData System for Analytics and the PureData System for Operational Analytics, visit: ibm.com/puredata



© Copyright IBM Corporation 2012

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
October 2012

IBM, the IBM logo, ibm.com, BigInsights, Cognos, Informix, InfoSphere, PureData and SPSS are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

Netezza is a registered trademark of IBM International Group B.V., an IBM Company.

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided. Actual available storage capacity may be reported for both uncompressed and compressed data and will vary and may be less than stated.

¹ Based on IBM internal tests of prior-generation system, and on system design for normal operation under expected typical workload. Individual results may vary.

² Total raw data capacity based on one Extra Large configuration with five full-rack data expansion add-ons.



Please Recycle
