A MongoDB White Paper

MongoDB: Bringing Online Big Data to Business Intelligence & Analytics

December 2013

mongoDB

Table of Contents

Introduction

Business Intelligence (BI) and analytics provides an essential set of technologies and processes that organizations have relied upon over many years to guide strategic business decisions.

With the emergence of new data sources such as social media, mobile applications and sensor-equipped "Internet of Things" networks, organization can extend BI to deliver real-time insight and discovery into such areas as operational performance, customer satisfaction and competitor behavior.

However, these new data assets, often referred to as "big data", challenge many of the previous assumptions around data storage and processing for BI and analytics applications. Not only do organizations have to manage much higher volumes of information, the data itself arrives at much faster rates, and is more complex and dynamic than existing transactional sources. To be useful, the data must be analyzed and visualized in shorter intervals than traditional reporting cycles.

While many BI systems will be able to adapt to these new requirements, the underlying databases may not afford the same flexibility. Organizations need to explore alternative technologies that augment their systems to fully integrate and benefit from big data.

With its rich document model, powerful analytical capabilities over high volumes of multi-structured data sets and broadest integration with leading BI and analytics tools, MongoDB provides a foundation to evolve BI to support real-time analytics for big data applications.

The Big Data Challenge for Business Intelligence & Analytics

In traditional BI platforms, the flow of data - starting with its acquisition from source systems through to transformation, consolidation, analysis and reporting - follows a well-defined sequential processes, as illustrated in Figure 1.

Operational data from multiple source systems is integrated into a centralized Enterprise Data Warehouse (EDW) and local data-marts via Extract Transform Load (ETL) processes. Reports and visualizations of the data are then generated by BI tools. This workflow is optimized for users, enabling the deep historical analysis

Figure 1: Traditional BI Process

used to inform strategic decision making at senior levels within the organization. Databases and reporting queries are predicated on a number of assumptions:

1. **Predictable Frequency.** Data is extracted from source systems at regular intervals - typically measured in days, months and quarters
2. **Static Sources.** Data is sourced from controlled, internal systems supporting established and well-defined back-office processes
3. **Fixed Models.** Data structures are known and modeled in advance of analysis. This enables the development of a single schema to accommodate data from all of the source systems, but adds significant time to the upfront design
4. **Defined Queries.** Questions to be asked of the data (i.e., the reporting queries) are pre-defined. If not all of the query requirements are known upfront, or requirements change, then the schema has to be modified to accommodate changes
5. **Slow-changing requirements.** Rigorous change-control is enforced before the introduction of new data sources or reporting requirements
6. **Limited users.** The consumers of BI reports are typically business managers and senior executives

EVOLVING BI AND ANALYTICS FOR BIG DATA

Businesses want to harness new data sources and fast time-to-insight in new and compelling ways. Examples include:

- Retailers tracking user preferences, web clicks and social sentiments to identify and automatically target geo and device-aware personalized content and promotions
- Utilities capturing household energy usage levels to predict outages and to incent more efficient energy consumption
- Governments detecting and tracking the emergence of disease outbreaks via social media signals
- Oil and gas companies taking the sensor output from their drilling equipment to make more efficient and safer exploration decisions

However the availability of new data sources generating big data is challenging the previous assumptions of data management and reporting within the BI platform.

The Need for Speed & Scale

Time to value is everything. For example, having access to real-time customer sentiment or logistics tracking is of little benefit unless the data can be analyzed and reported in real-time. As a consequence, the frequency of data acquisition, integration and analysis must increase from days to minutes or less, placing significant operational overhead on BI systems. In a growing number of cases, source data needs to be analyzed in place in order to provide the responsiveness demanded by the business.

The availability of new data sources drives an explosion in the amount of data organizations must manage, with analysts estimating a doubling in volumes every 12 to 14 months. Not only do BI databases have to handle much higher ingestion rates (often referred to as “data velocity”), there is also the challenge of how data is moved through the BI pipeline, from source systems to the EDW, data-marts and into analytical and reporting processes.

Agile Analytics and Reporting

With such a diversity of new data sources, business analysts can not know all of the questions they need to ask in advance. Therefore an essential requirement is that the data can be stored before knowing how it will be processed and queried.

The Changing Face of Data

Data generated by such workloads as social, mobile, sensor and logging, is much more complex and variably structured than traditional transaction data from back-office systems such as ERP, CRM, PoS (Point of Sale) and Accounts Receivable.

The existing relational databases used by these back-office systems are designed to model cleansed and neatly structured data into tabular row and column formats with defined values, enforced by rigid schemas. They were never designed for the polymorphic, semistructured or unstructured data that is now typical in many of today's big data applications.

Higher Uptime Requirements

The immediacy of real-time analytics accessed from multiple fixed and mobile devices places additional demands on the continuous availability of BI systems. Batch-based systems can often tolerate a certain level of downtime, for example for scheduled maintenance. Online systems on the other hand need to maintain operations during both failures and planned upgrades.

Taking BI to the Cloud

The drive to embrace cloud computing to reduce costs and improve agility means BI components that have traditionally relied on databases deployed on monolithic, scale-up systems have to be re-designed for the elastic scale-out, service-oriented architectures of the cloud.

IMPACTS TO TRADITIONAL BI DATABASES

The relational databases underpinning many of today's traditional BI platforms are not well suited to the requirements of big data:

- **Semi-structured and unstructured data** typical in mobile, social and sensor-driven applications cannot be efficiently represented as rows and columns in a relational database table
- **Rapid evolution of database schema** to support new data sources and rapidly changing data structures is not possible in relational databases, which rely on costly ALTER TABLE operations to add or modify table attributes
- **Performance overhead** of JOINs and transaction semantics prevents relational databases from keeping pace with the ingestion of high-velocity data sources
- **Quickly growing data volumes** require scaling databases out across commodity hardware, rather than the scale-up approach typical of most relational databases

Relational databases' inability to handle the speed, size and diversity of rapidly changing data generated by modern applications is already driving the enterprise adoption of NoSQL and Big Data technologies in both operational and analytical roles.

Integrating Online Big Data with BI and Analytics

Big Data can take both online and offline forms. It is important to differentiate these forms in terms of performance, availability and data usage requirements to better understand where technologies such as MongoDB and Hadoop can be used within a modern BI system.

DIFFERENTIATING BETWEEN ONLINE AND OFFLINE BIG DATA

Online Big Data refers to data that is created, ingested, transformed, managed and/or analyzed in real-time to support operational applications and their users. Big Data is born online. Latency for these applications must be very low and availability must be high in order to meet SLAs and user expectations for modern application performance. This includes a vast array of applications, from social networking news feeds, to analytics to real-time ad servers to complex CRM applications. Examples of databases powering online big data applications include MongoDB and other NoSQL databases with the capability to handle the real-time, richly structured, rapidly changing, high volume data sets now available to the business.

Offline Big Data encompasses applications that ingest, transform, manage and/or analyze data in a batch con-

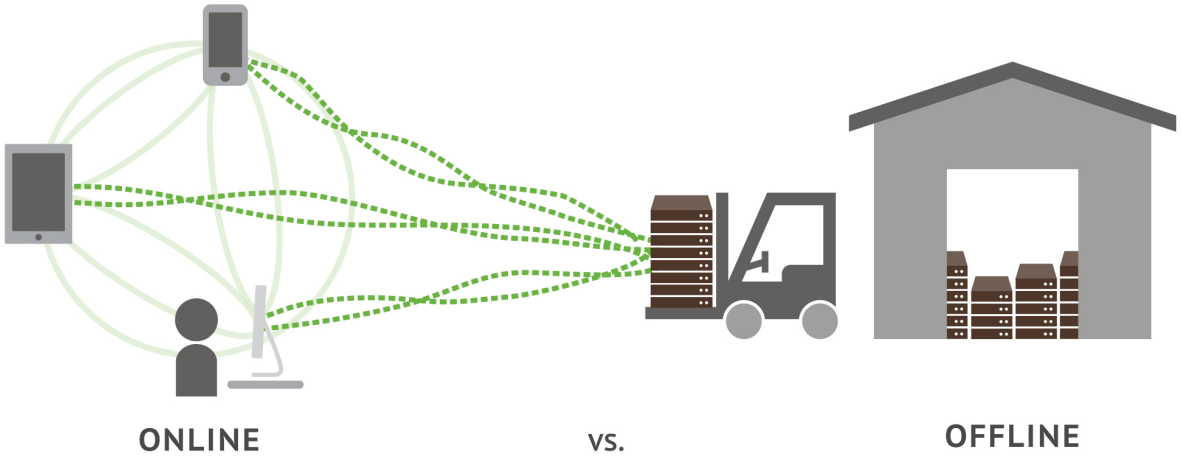

Figure 2: Big Data with MongoDB and Hadoop

text. They typically do not create new data. For these applications, response time can be slow (up to hours or days), which is often acceptable for this type of use case. Since they usually produce a static (vs. operational) output, such as historical reports or dashboards, they can even go offline temporarily without impacting the overall goal or end product. Examples of offline Big Data applications include Hadoop-based workloads and enterprise data warehouses.

Organizations evaluating which Big Data technologies to adopt should consider how they intend to use their data. Those looking to build applications that support real-time, operational use cases and analytics will need an operational data store like MongoDB. For those that need a place to conduct long-running analysis offline, solutions like Hadoop can be an effective tool.

Organizations pursuing both use cases can do so in tandem, and they will sometimes find integrations between online and offline Big Data technologies. For instance, MongoDB provides integration with Hadoop, which is discussed later in the paper.

To illustrate how MongoDB and Hadoop are used together, consider a major web company. They use MongoDB as the operational data store, storing and tracking rich user data (i.e., not just login information, but also online behavior). Additionally, it performs real-time analytics to dictate automated machine behavior. By contrast, the company uses Hadoop to perform more complex analysis offline. It pipes the data from MongoDB into Hadoop, where it groups user sessions, segments users by behavior and performs regression analyses to determine correlation and improve predictive models. Finally, it pipes the enriched data back into MongoDB, which further refines real-time analytics.

MONGODB FOR INTEGRATING ONLINE BIG DATA

With its rich document model, powerful query functionality, scalable architecture and integration with leading BI and analytics tools, MongoDB can be deployed as a key database component both within, and as an extension of, a BI platform, including:

- A conventional data source for regular ETL processes integrating data into the EDW.
- A "data hub," replicating and consolidating data from operational and EDW sources, allowing for cross-function, complete 360-degree view reporting and visualization. This is illustrated in Figure 3.
- A data store enabling real-time analytics and dashboards to be generated against live, operational data. This is shown in Figure 4, alongside the traditional EDW data flow.

As the most widely deployed NoSQL database with the fastest growing big data community and the most extensive partner ecosystem, including many of the leading BI tools vendors, MongoDB is at the forefront of integrating online big data to BI platforms. The following sections explore MongoDB's capabilities for this integration.

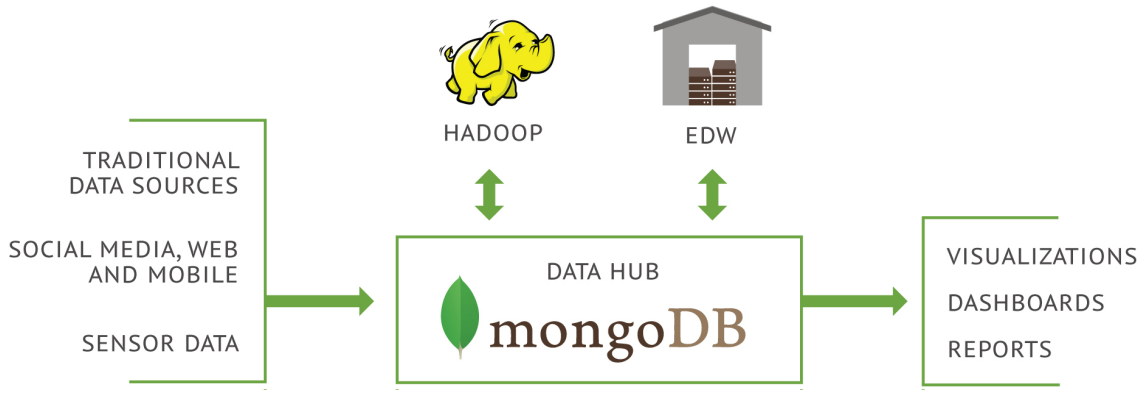

Figure 3: Data Hub Powered by MongoDB

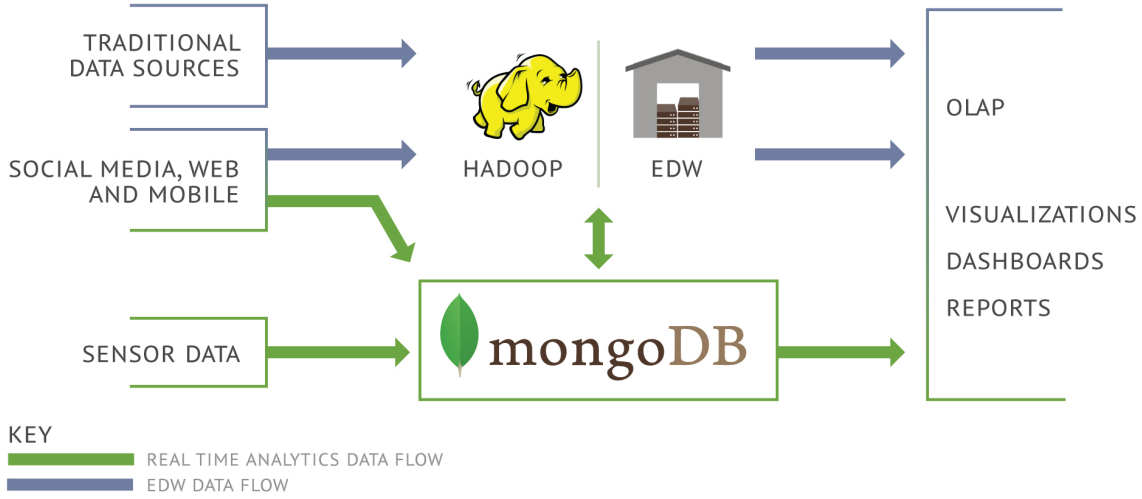

Figure 4: Real-Time Analytics Powered by MongoDB

MODELING COMPLEX DATA WITH MONGODB'S BSON DOCUMENTS

Rich data structures with dynamic attributes comprising text, geospatial data, media, arrays, embedded elements and other complex types are common in today's web, mobile, social and sensor-driven applications.

Unlike relational databases that flatten data into 2-dimensional tabular structures of rows and columns, MongoDB efficiently models and stores this rich data as documents (analogous to rows in a relational database) in a binary representation called BSON (Binary JSON). The BSON encoding extends the popular JSON (JavaScript Object Notation) representation to include additional types such as int, long, and floating point. BSON documents contain one or more fields comprising a value with a specific data type. Each field can be indexed and queried, giving additional flexibility to BI reporting.

With sub-documents and arrays, BSON documents also align with the structure of objects at the application level. This makes it easy for developers to map the data used in the application to its associated document in the database, reducing complexity and making it easier to bring new applications to market faster.

RAPIDLY EVOLVING THE DATA MODEL WITH DYNAMIC SCHEMAS

MongoDB's dynamic schema provides a major advantage for BI applications that need to ingest, store and process rapidly evolving data streams from new sources.

Collections (analogous to tables in a relational database) can be created without first defining their struc-

ture. Documents in a given collection need not all have the same set of fields. Users can adapt the structure of documents just by adding new fields or deleting existing ones, making it very simple to extend BI applications by adding new attributes for analysis and reporting.

REAL-TIME ANALYTICS FOR ONLINE BIG DATA

Through powerful query functionality and indexing, MongoDB enables users to run analytics in real-time directly against their data.

MongoDB users have access to a broad array of query, projection and update operators supporting native queries to data in-place within MongoDB, including:

- Key-value queries returning results based on any field in the document, often the primary key
- Range queries returning results based on values defined as inequalities (e.g., greater than, less than or equal to, between)
- Geospatial queries returning results based on proximity criteria, intersection and inclusion as specified by a point, line, circle or polygon
- Text Search queries returning results in relevance order based on text arguments using Boolean operators (e.g., AND, OR, NOT)
- Aggregation and MapReduce queries, discussed in more detail below
- Native Hadoop integration for deep, offline analytics.

With the combination of MongoDB's dynamic document model and comprehensive query framework, users are able to store data before knowing all of the questions they will need to ask of it.

Data Aggregation

The MongoDB Aggregation Framework generates aggregations of values returned by the query (e.g., count, minimum, maximum, average), similar in concept to a SQL GROUP BY statement.

Using the Aggregation Framework, documents in a collection pass through an aggregation pipeline, where they are processed by operators. Expressions produce output documents based on calculations performed on the input documents. The accumulator expressions used in the $group operator maintain state (e.g., totals, mins, maxs, averages) as documents progress through the pipeline.

Using new Pipelined Data Transformations in MongoDB 2.6, multi-step data enrichment and transformations can be performed in the database with a simple declarative interface, enabling processes such as lightweight ETL to be performed by MongoDB.

Result sets from the aggregation pipeline can be written to a named collection with no limit to the output size (subject to the underlying storage system). Existing collections can be replaced with new results while maintaining previously defined indexes to ensure queries can always be returned efficiently over rapidly changing data.

In-Database MapReduce

MongoDB provides native support for MapReduce, enabling complex data processing that is expressed in JavaScript and executed across data in the database. Google's V8 JavaScript engine, which is integrated into MongoDB, enables multiple jobs to be executed simultaneously.

For maximum flexibility, MapReduce operations can run across both single servers and sharded collections.

The MapReduce operation uses a temporary collection during processing so it can be run periodically over the same target collection without affecting intermediate states. This mode is useful when generating statistical output collections on a regular basis.

Integrating MongoDB and Hadoop

Like NoSQL, Hadoop usage is growing in the enterprise where it is often being deployed to handle ETL processes and Data Warehouses.

The MongoDB Connector for Hadoop presents MongoDB as a Hadoop data source allowing a MapReduce job to read data from MongoDB directly without first copying it to HDFS, thereby eliminating the need to move TB of data between systems. MapReduce jobs can pass queries as filters, thereby avoiding the need to scan entire collections and speeding up processing; they can also take advantage of MongoDB's indexing capabilities, including text and geospatial.

The Connector makes it easier for Hadoop users to integrate real-time data from MongoDB with Hadoop for deep, offline analytics. As well as reading from MongoDB, the Connector also allows results of Hadoop

jobs to be written back out to MongoDB, to support real-time operational processes and ad-hoc querying.

The Connector supports MapReduce, Pig, Hadoop Streaming (with Node.js, Python or Ruby) and Flume jobs with MongoDB. Support is also available for SQL-like queries from Apache Hive to be run across MongoDB data sets.

OPTIMIZING ANALYTICS QUERIES WITH INDEXES

MongoDB provides a number of different index types to optimize performance of real-time and ad-hoc BI queries across highly variable, fast moving data sets.

Indexes can be created on any field within a document or sub-document. In addition to supporting single field and compound indexes, MongoDB also supports indexes of arrays with multiple values, short-lived data (i.e., Time To Live), sparse, geospatial and text data, and can enforce constraints with unique indexes. Refer to the documentation for the full list of index types.

The MongoDB query optimizer selects the best index to use by periodically running alternate query plans and selecting the index with the best response time for each query type. The results of this empirical test are stored as a cached query plan and are updated periodically.

MongoDB also supports covered queries - where returned results contain only indexed fields, without having to access and read from source documents. With the appropriate indexes, BI workloads can be optimized to use predominantly covered queries.

SCALING MONGODB FOR NEW DATA SOURCES

As data volumes grow, it is essential for users to select a database that will grow with them.

MongoDB provides horizontal scale-out of data sets using a technique called sharding. Transparent to the underlying applications, sharding distributes data across multiple physical partitions and nodes. Sharding allows MongoDB deployments to address the hardware limitations of a single server, such as storage and RAM or bottlenecks in disk I/O, without adding complexity to the application. Sharding enables MongoDB to quickly scale both capacity and performance as new data sources are added to the BI platform.

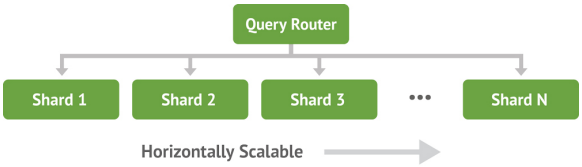

Figure 5: Automatic Scaling Across Commodity Compute Clusters

MongoDB automatically balances the data in the cluster as the data grows or the number of nodes in the cluster increases or decreases, making it ideal for cloud-based deployments.

MongoDB Integration with BI and Analytics Tools

To make online big data actionable through dashboards, reports, visualizations and integration with other data sources, it must be accessible to established BI and analytics tools. MongoDB offers integration with more of the leading BI tools than any other NoSQL or online big data technology, including:

- Actuate
- Alteryx
- Informatica
- Jaspersoft
- Logi Analytics
- MicroStrategy
- Pentaho
- Qliktech
- Talend

The following sections profile Informatica, MicroStrategy, Pentaho and Qliktech integrations with MongoDB. For others, please refer to the MongoDB Software Partners page.

Informatica Corporation (Nasdaq:INFA) is a leading provider of data integration software. Organizations around the world rely on Informatica to realize their information potential and drive top business imperatives. Informatica Vibe, the industry's first and only embeddable virtual data machine (VDM), powers the unique "Map Once. Deploy Anywhere." capabilities of the Informatica Platform.

Worldwide, over 5,000 enterprises depend on Informatica to fully leverage their information assets from devices to mobile to social to big data residing on-premise, in the Cloud and across social networks.

Product Overview

Informatica + MongoDB is a powerful combination that increases developer productivity up to 5x, enabling them to build and deploy big data applications much faster. Informatica provides access to virtually all types of data from modern and legacy systems at any latency, processes and integrates data at scale, and delivers it directly into MongoDB.

With Informatica, companies can unlock the data in MongoDB for downstream analytics to improve decision making and business operations. Using the Informatica PowerCenter Big Data Edition with the PowerExchange for MongoDB adapter users can access data in MongoDB, parse the JSON-based documents and then transform the data and combine it with other information for big data analytics - all without having to write a single line of code.

Capabilities

- **High Productivity Development Environment.** Provides a visual metadata-driven development environment so developers can achieve up to 5x productivity over hand-coding data integration from scratch.
- **Universal Data Access.** Provides access to virtually all types of data including RDBMS, mainframe, ERP, CRM, social media, machine and sensor device data, cloud applications, and industry standards data (e.g., FIX, SWIFT, HL7, HIPAA, ASN.1, EDI).
- **High-Speed Data Ingestion and Extraction.** Accesses, loads, transforms and extracts big data between source and target systems and MongoDB at high speeds using high-performance connectivity.
- **Metadata Discovery for Flexible Data Models.** Automatically discovers schemas by sampling records in a collection to provide a representative collection schema. Users can edit the discovered schema by adding or removing columns required for analysis.
- **Embedded Entity Access.** Creates pivoted columns so users can easily access data from embedded documents and arrays and integrate data independent of the application data modeling in MongoDB.
- **JSON Document Handling.** Interacts with MongoDB's BSON serialization format with the ability to ingest complex JSON structures directly into MongoDB as documents or extract selected JSON elements for analysis.
- **Distributed Data Access Through Read Preferences.** Provides "Read Preferences" that allows users to choose which members of the MongoDB Replica Set to use for data integration. This gives users a choice to distribute data integration jobs to secondary nodes in a replica set, so as to not affect performance of the primary node and the main application load.
- **Map Once, Deploy Anywhere.** Includes the Vibe Virtual Data Machine so users can build data integration pipelines once and deploy them anywhere such as distributed computing platforms like Hadoop, on-premise or in the Cloud.
- **Unified Administration and High Availability.** Automatically schedule and coordinate data integration workflows with high availability and reliability using a unified administration console.

Customer Benefits

- **Discover Insights from Big Data Faster.** The power of big data means organizations can access and analyze all of their data. A growing number of the worlds' top companies and government agencies use MongoDB for applications today which means more and more data is being stored in MongoDB. Using Informatica, data analysts can discover insights by combining data in MongoDB with data from other operational and analytical data stores across the enterprise.
- **Run Better Applications with Better Data.** The ease of getting data into MongoDB from other enterprise and third-party systems is critical to big data applications delivering trusted, complete, and relevant information to business users. Informatica en-

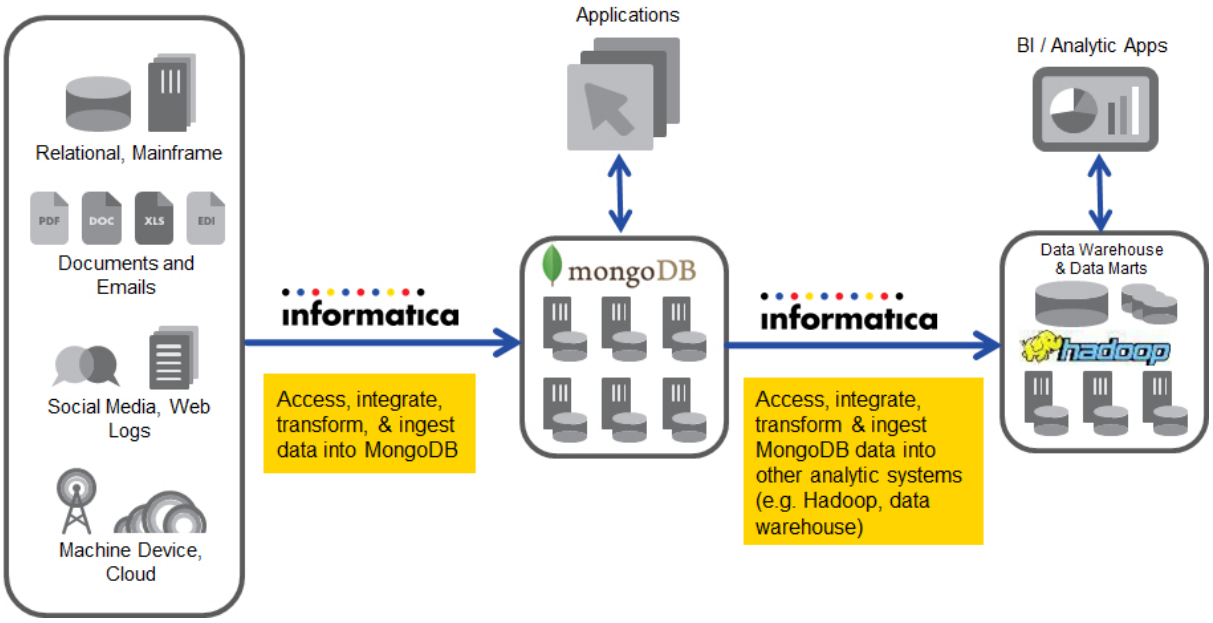

Figure 6: Informatica for MongoDB

sures that users can access and integrate all of their data with MongoDB and other enterprise data to provide a complete picture.

- Lower Costs of Data Integration. Up to 80% of the work in a big data project involves data integration and the resource skills to undertake the work are often in short supply. The good news is that organizations can staff big data projects with over 100,000 trained Informatica developers around the world. Informatica makes it easier to load data into and extract it from MongoDB using readily available resource skills thereby lowering ongoing operational costs.
- Deliver Business Impact with Rapid Deployment. MongoDB is the world's leading NoSQL database designed for ease of development and scaling. By providing a flexible schema, it makes application development truly agile. Informatica is a leading provider of data integration software. By providing a single, consistent data integration approach for all types of data including traditional relational structures, complex file formats and flexible dynamic schemas used in MongoDB, organizations are empowered to rapidly adopt MongoDB into their enterprise data management infrastructure for maximum business impact.

Learn More

PowerExchange for MongoDB supports codeless reading and writing of data to MongoDB including nested documents, arrays and arrays of nested documents. Review the following YouTube demonstrations to learn more:

- Part 1: Configuring PowerExchange to connect to MongoDB.
- Part 2: Create mappings that read and write to MongoDB
- Part 3: Write JSON documents directly to MongoDB.

For more information on Informatica PowerCenter and PowerExchange please visit the Informatica website. To purchase the Informatica PowerCenter Big Data Edition and PowerExchange for MongoDB please contact Informatica.

MicroStrategy

Founded in 1989, MicroStrategy Incorporated (Nasdaq: MSTR) is a leading worldwide provider of enterprise software platforms. The Company's mission is to provide the most flexible, powerful, scalable and user-friendly platforms for analytics, mobile, identity and loyalty, offered either on premises or in the cloud.

Product Overview

The MicroStrategy Analytics Platform™ is a comprehensive family of powerful, easy-to-use analytics solutions bringing together over 20 years of innovation in analytics and business intelligence. It contains three distinct products – MicroStrategy Analytics Enterprise, MicroStrategy Analytics Desktop and MicroStrategy Analytics Express.

Together, these products support a full spectrum of business needs with maximum flexibility and affordability, ranging from small and medium organizations to the largest global businesses; running at individual, team, departmental, or enterprise scale; delivered via web or mobile devices; supporting high performance analytics on gigabytes, terabytes or petabytes of data; and deployed on-premises or up to 100x more quickly and at lower cost in the MicroStrategy Cloud. The common user experience shared across the MicroStrategy Analytics Platform makes it easy for organizations to start small with self-service analytics on Desktop or Express and grow to take advantage of the comprehensive, enterprise-grade features of Enterprise.

Connecting to MongoDB

MicroStrategy connects in one of two ways to MongoDB

1. Using Simba® ODBC driver. This method allows for ease-of-use for users familiar with SQL and ODBC. The Simba driver will expose data in MongoDB in relational form to MicroStrategy.
2. Using the MicroStrategy Web Services connector. This method is more flexible and powerful in that it doesn't require any pre-defined data structures. Connectivity and queries are defined using the MicroStrategy web services query generator and editor.

Capabilities

- Parameterized Queries MicroStrategy can parameterize queries to MongoDB using prompts for flexibility. Run-time prompts provide report input parameters that control most aspects of a report, and give the user significant ad hoc reporting capabilities. Prompts enable the personalization of reports from a single report definition, reducing the number of objects stored in the metadata repository. Users can store personalized prompt answers for easy reuse.
- Combining Multiple Data Sources. MicroStrategy can combine data from MongoDB with data in other Enterprise or cloud data sources. This allows users to seamlessly report, analyze and monitor data across multiple data sources through a single multi-dimensional view of the business. Companies can get BI applications up and running in almost no time with minimum data engineering.
- Dashboards and Visualizations. Business users can create dashboards and interactive visualizations over the web. A Visual Analysis is a presentation of large data volumes as intuitive data visualizations. This format of information delivery is specifically designed for data discovery and is aimed to empower business users to get their business insights and spot outliers quickly without involving people in IT. Instead of looking through rows and columns of data, a user can see their data in a visual format and quickly derive insights by adding attributes and metrics onto the visualizations.

  MicroStrategy provides a large number of advanced visualizations to choose from such a Graph Matrix, Maps, Heat Maps and standard charts. A visual analysis is fully interactive and provides standard analysis functionality such as multi-key sorting, pivoting, drilling, adding new metrics and aggregations, ranking, page-by, thresholding as well as filtering.
- Reporting. MicroStrategy enables users to build information-rich mobile applications rapidly that combine multimedia, transactions, analytics and custom workflows, putting business reports, KPIs, documents and dashboards in the hands of an organization's employees and customers. MicroStrategy Mobile lets decision makers run an organization wherever they are and view business data on the iPad, iPhone and Android smartphones and tablets in multi-touch tables, graphs, charts, visualizations, maps, documents and dynamic dashboards. It securely connects to a MicroStrategy Mobile Server which manages the transmission of interactive reports, documents and dashboards to the Mobile client applications.

Customer Benefits

The integration of MicroStrategy with MongoDB provides two main benefits to joint customers:

- MicroStrategy allows business users to analyze structured and unstructured data stored in Mon-

goDB without the need to code. This allows business users to directly engage in the analysis of data without relying on IT staff with the necessary coding expertise.

- MicroStrategy allows blending data from multiple enterprise sources with data in MongoDB and enables interactive and visual analysis of that across a range of devices. By doing so MicroStrategy simplifies the process of turning the data captured in MongoDB into actionable information.

Learn More

To learn more about MicroStrategy, visit www.microstrategy.com and follow them on Facebook and Twitter. Contact information for all local MicroStrategy representatives can be found at www.microstrategy.com/about-us/contact-us.

pentaho

Pentaho is a leading independent provider of business and big data analytics and data integration to thousands of enterprises and software vendors worldwide. By introducing big data capabilities in early 2010, Pentaho addressed the first emerging big data use cases while continuing to deliver technology innovations through its open source heritage.

Pentaho answers both the business and technical challenges inherent in gaining actionable insights from incoming data. Pentaho's platform supports the entire big data workflow from data ingestion, preparation and integration, to interactive visualization, analysis and predictive analytics. Customers can achieve more comprehensive insights with the ability to accurately blend big data with other data at the source in real-time.

Graphical tools reduce complexity and eliminate the inherent coding associated with big data deployments. These capabilities help customers realize value from their big data up to 15X faster than with traditional approaches. An industry-leading adaptive big data layer supports leading Hadoop distributions, MongoDB and other NoSQL databases and high performance analytic databases to offer broad support for big data analytics and integration.

Product Overview

Pentaho Business Analytics 5.0 is purpose built for current and future analytics requirements providing a full spectrum of analytics regardless of data type and volume, IT architecture or analysis requirements. A simplified, modern user experience empowers all

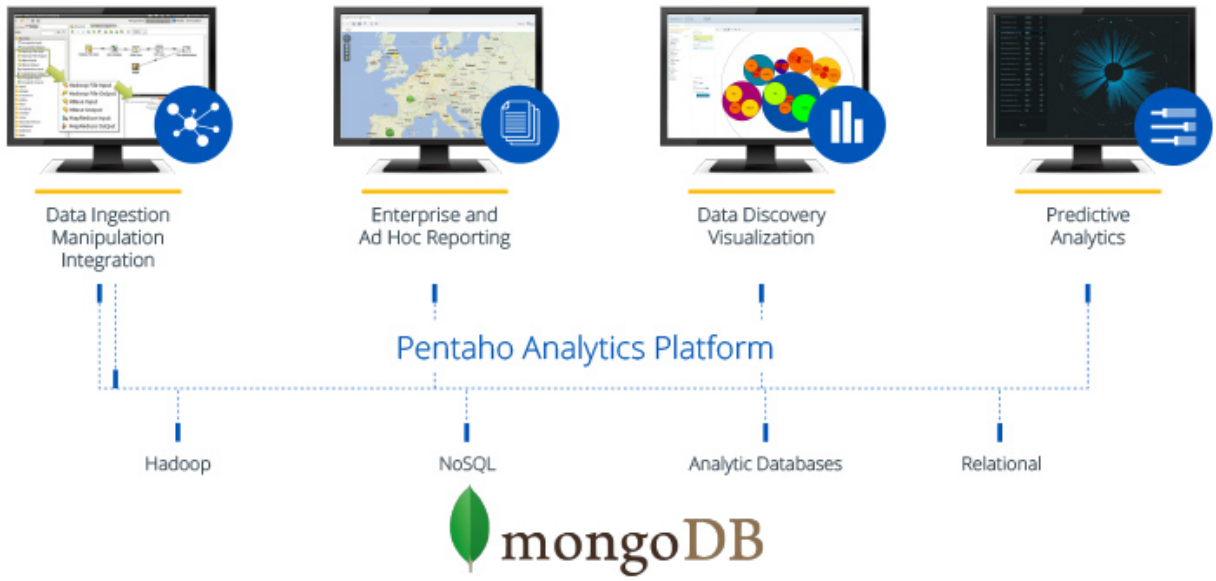

Figure 7: Pentaho for MongoDB

users to turn data into competitive advantage. Pentaho 5.0 includes:

- Comprehensive big data analytics on blended data for real-time, up to date, accurate insights
- Among the broadest and deepest big data and native NoSQL integration for analytics on any data source
- Simplified, enterprise-class administration, security, and management helps IT manage huge data volumes efficiently
- Streamlined analytics experience with powerful customization and predictive capabilities
- Open, extensible, embeddable platform to support any architecture
- Modern REST services for application developers to embed the complete platform easily

Pentaho 5.0 relieves the pain associated with deploying big data analytics such as automating complex scripting and programming jobs; providing capabilities to easily blend, visualize and analyze existing, new and future data sources; speeding query times; and providing simple, clean and intuitive interfaces for end users and administrators.

Capabilities

Pentaho and MongoDB are natural partners as open source innovators providing deep native integration with big data and NoSQL data sources. Pentaho Business Analytics 5.0 and MongoDB together provide a powerful, interactive big data analytics solution to meet the increased requirements for data driven businesses.

Pentaho engineers worked closely with MongoDB engineers over the past several years to develop a true, native integration that offers Pentaho customers full access to their MongoDB data. Pentaho delivers an industry-first integration with MongoDB including analytics capabilities with full support for MongoDB Replica Sets, Tag Sets and Read and Write Preferences. Pentaho 5.0 includes:

- Support for the MongoDB aggregation framework for powerful, efficient data summarization for reporting and analytics without coding.
- Metadata discovery – ability to sample documents and automatically infer a metadata layer.
- Improved control for inserting/updating data into MongoDB collections.
- MongoDB template for Instaview, Pentaho's big data discovery application, to broaden MongoDB data access to data analysts.

With enhanced Pentaho reporting and data integration users can:

- Create data transformations from source systems to MongoDB and from MongoDB to target systems, via a drag and drop interface. Data from MongoDB can be blended on demand with any other data source.
- Create pixel perfect reports directly against data stored in MongoDB.
- Quickly empower data analysts to visualize and explore data in MongoDB with Instaview.

Customer Benefits

With Pentaho, the value of data in MongoDB increases dramatically for customers as it can be accessed, blended, visualized and reported on in combination with any other data source for increased insight and operational analytics. By providing just-in-time data blending, end-users can have an extended 360-degree view of their customers.

The combination of Pentaho Business Analytics and MongoDB helps business users and developers:

- **Increase Data Value.** With Pentaho, MongoDB data can be accessed, blended, visualized and reported in combination with any other data source for increased insight and operational analytics.
- **Reduce Complexity.** Reporting on data stored in MongoDB is simplified, increasing developer productivity with automatic document sampling, a drag and drop interface and schema generation.
- **Accelerate Data Access and Querying.** With no impact on throughput, the integration builds on the features and capabilities in MongoDB, such as the Aggregation Framework, Replication and Tag Sets.

Pentaho and MongoDB customers such as Soliditet and Travian Games exemplify the value of the joint solution.

Soliditet, the Nordic region's leading credit ratings company freed up significant developer resources and brought credit ratings products to market 50% faster. With Pentaho and MongoDB, Soliditet empowers their

customers to utilize real-time analytics when seeking ratings information.

Travian Games, a top German In-Browser game developer, experiments daily with browser game innovations to set new gaming trends for the future. With Pentaho and MongoDB, data generated by 120 million gamers fundamentally drives product innovation daily, ensuring Travian maintains and increases their competitive advantage.

Learn More

- Learn more about Pentaho for Big Data
- Learn more about Instaview for MongoDB
- For a free evaluation, download Pentaho Business Analytics
- To learn more about Pentaho platform pricing contact Pentaho at (866)660-7555 or visit: pentaho.com/contact

QlikView

QlikTech (NASDAQ: QLIK) is a leader in Business Discovery — user-driven Business Intelligence (BI). QlikTech's Business Discovery solution bridges the gap between traditional business intelligence solutions and standalone office productivity applications.

Its QlikView Business Discovery platform enables intuitive user-driven analysis that can be implemented in days or weeks rather than months, years, or not at all. The in-memory associative search technology it pioneered allows users to explore information freely rather than being confined to a predefined path of questions. QlikTech has offices around the world serving approximately 22,000 customers in over 100 countries.

Product Overview

QlikView Business Discovery works with existing BI applications and adds new capabilities: insight for everyone, zero-wait analysis, mobility, an app-like model, remixability and reassembly, and a social and collaborative experience.

QlikTech worked with MongoDB to offer the QlikView Expressor MongoDB Datascript Module, which allows users to read from and write to a MongoDB instance.

Capabilities

QlikView Expressor 3.10 introduces read and write Datascript connectivity to MongoDB. The product includes a library that enables developers to write Datascript routines that can read data from and write data to MongoDB. The Expressor Dataflow is an ideal way to move data into or out of a MongoDB collection as it can easily process the nested data structures found in MongoDB documents. This is essential for pulling data into QlikView or simply migrating data from other sources.

Using QlikView Expressor Read Custom and Write Custom Operators and/or Datascript Modules, developers can create and share reusable operations that work with MongoDB. And, with the QlikView Expressor Extension SDK, these operations can be included in their own custom operators that can be packaged and shared with the QlikView Expressor Extensions manager.

Learn More

Practical examples of the integration are included in the QlikView Expressor MongoDB Module Documentation.

Conclusion

Big data is an essential extension of BI and analytics platforms, presenting new sources of operational insight and discovery to the business. However, the rate of data ingestion coupled with its complexity and volume are beyond the design constraints of many traditional databases used in BI systems today.

With its rich document model, powerful analytical capabilities over large, multi-structured data sets and integration with leading BI and analytics tools, MongoDB provides a foundation to integrate online big data with existing BI and analytics platforms.

About MongoDB

MongoDB (from humongous) is reinventing data management and powering big data as the leading NoSQL database. Designed for how we build and run applications today, it empowers organizations to be more agile and scalable. MongoDB enables new types of applications, better customer experience, faster time to market and lower costs. It has a thriving global community with over 5 million downloads, 100,000 online education registrations, 20,000 user group members and 20,000 MongoDB Days attendees. The company has more than 600 customers, including many of the world's largest organizations.

Resources

For more information, please visit mongodb.com or mongodb.org, or contact us at sales@mongodb.com.

Resource	Website URL
MongoDB Enterprise Download	mongodb.com/download
Free Online Training	education.mongodb.com
Webinars and Events	mongodb.com/events
White Papers	mongodb.com/white-papers
Case Studies	mongodb.com/customers
Presentations	mongodb.com/presentations
Documentation	docs.mongodb.org