Why DBMSs Matter More than Ever in the Big Data Era

Having the right database infrastructure can make or break big data analytics projects.
Big data has become big news — and big business. In a recent survey, 83 percent of CIOs said they intend to invest in analytics and business intelligence over the next three to five years, and 94 percent of CMOs said they rely on advanced analytics to help them do their jobs.

Companies recognize that their vast and growing stores of data contain valuable insights. The race is on to harvest that information and to act on it.

In their quest for big data insights, companies are investing heavily in new tools such as Hadoop and NoSQL databases. But in their hurry to take advantage of the next big thing, too many are failing to pay enough attention to the underlying infrastructure that supports their big data efforts. In fact, when asked about their infrastructure, 55 percent of CIOs admitted they lack a scalable and extensible information foundation with which to manage big data.

The DBMS Matters
In most cases, analytics systems are fed by database management systems (DBMSs) that store an organization’s transactional data. It’s easy to overlook these foundational databases because they aren’t as trendy as the hot buzzwords of the day. But transactional data continues to be the No. 1 type of data for big data projects.

To stay competitive, organizations need to respond to changes faster than ever. To do so, they need a DBMS that can provide accurate data to analytics and business intelligence (BI) systems in real time. At the same time, data stores are constantly growing, so organizations need a DBMS that can scale to handle increasing data volumes — without breaking the bank. And because technology is always advancing, organizations must have a DBMS designed for today’s needs and beyond.

THE BIG DATA DBMS CHECKLIST
To ensure that their databases can handle the challenges of today’s big data environments, enterprises need a DBMS with these characteristics:

- Extremely fast processing
- Minimal downtime
- The ability to grow with the business
- Actionable compression technology
- Advanced security
- Simplified management
- Integration with the latest technology

83% of CIOs intend to invest in analytics and business intelligence over the next five years.
In addition to meeting these requirements, a big data database has to be affordable. IT managers continue to face pressure to do more with less. A DBMS that can support big data projects as cost-effectively as possible can help meet these demands.

Requirements for a Big Data DBMS

Because the DBMS plays a foundational role in big data projects, organizations should look for a database with the following characteristics:

**Extremely fast processing.** When it comes to big data analytics, speed is the name of the game. Advancements such as columnar architecture, dynamic in-memory processing, parallel vector processing, data skipping and others allow modern databases to reach new performance heights. Combining these new technologies, modern databases can achieve performance gains up to 1,000 times over traditional DBMSs.

**Minimal downtime.** With 24/7 online operations the norm, a single hour of downtime can cost enterprises from $100,000 to $1 million or more annually. As data stores grow and the pace of transactions accelerates, organizations need to be sure that they have a database architected to handle more data and more transactions without failure. That means having a DBMS with capabilities such as advanced clustering, automatic failover and disaster recovery, transparent scaling and support for rolling maintenance updates.

**The ability to grow with the business.** Over time, every business changes. To meet evolving needs, enterprises need a database that scales to meet growth demands and provides maximum flexibility. Ideally, a big data DBMS should be able to support transactional (OLTP), analytic (OLAP) and mixed workloads with ease. It should support multiple platforms and programming languages and offer a variety of deployment options, including both cloud-based and traditional options.

**Optimized storage.** As data grows, compression technology becomes very important as a way to control storage hardware costs. In the past, compressing data could result in slower performance because the data had to be uncompressed before it could be used for analysis. However, the latest big data DBMSs have found ways to solve that problem by being able to perform queries on data while it's still compressed — without the overhead of decompressing the data. As a result, companies can see both cost savings and performance improvements.

**Advanced security.** Every organization wants to minimize the risk of costly data breaches, and that means selecting a DBMS with industry-leading security. Strict authentication and authorization features, multilevel access control, built-in encryption capabilities and confidential data masking are necessary. In addition, because many organizations are facing tougher compliance requirements, the DBMS should provide a clear audit trail that can be traced to a specific point in time.

**Simplified management.** Operational management, administration and
maintenance can eat into IT budgets. Enterprises should look for time-saving management features, such as automatic optimization, built-in monitoring, self-healing functions, autonomic memory tuning and simplified configuration tools.

**Integration with the latest technology.**
To be effective, a DBMS must interact with a wide variety of tools. A big data DBMS should integrate with NoSQL tools, sensors, social technologies, mobile solutions and, of course, business intelligence and analytics tools. It should also support virtualization and cloud computing environments. In short, it should be flexible enough to meet all of today's needs while allowing an organization to prepare for the future.

**Keeping Costs Under Control**
Besides providing a solid foundation, databases play another role in big data projects: cost control. In some cases, the operational efficiencies and other financial savings provided by a DBMS can improve the investment case for big data initiatives.

DBMSs designed to support big data provide these financial benefits in several ways.

Load-and-go simplicity, as well as topology and database patterns, can dramatically accelerate new project deployments and reduce project risks. This allows business users and applications to reap rewards much more quickly.

Mixed workload support and optimization for both transactional and analytics workloads provides the versatility to adapt to changing business requirements without changing the underlying technology.

Continuous ingesting of new data and fast query processing allow organizations to capitalize on time-sensitive business opportunities. This can make the company as a whole more agile and competitive, leading to increased revenue and profits.

Utilizing data compression technologies can reduce the amount of storage needed for user data and help organizations put off the purchase of additional storage. And multitemperature data management can dynamically and seamlessly optimize data movement among storage groups to postpone storage upgrades.

The right DBMS can eliminate the need for “rip and replace” server upgrades. By choosing a database that can improve performance while still using legacy hardware, enterprises can avoid the costs associated with upgrading the hardware platform.

A DBMS that utilizes advanced cluster technology, failover protection and disaster recovery helps keep mission-critical systems up and running. And keeping cluster topology transparent to applications allows businesses to adjust scalability and processing power to handle peak demands without application changes.

A DBMS can also make existing IT staff more efficient. A modern database architected to meet the needs of big data may actually require less tuning and administration than older, less full-featured databases.
It can free up personnel to focus on more important IT projects that have a bigger impact on profitability. A next-generation DBMS makes a big data project more likely to succeed. Without the right infrastructure, analytics and BI projects may never lead to the types of improvements that their supporters envisioned. But with the right DBMS in place, big data initiatives can go on to achieve more impressive results than anyone imagined.

**IBM is the world’s largest information technology company, with 80 years of leadership in helping clients innovate.** Drawing on a breadth of capabilities and best practices from across IBM and our extensive partner ecosystem, we offer clients within every industry a wide range of services, solutions and technologies that can help them improve productivity, respond rapidly to the needs of their business and reduce development and operations costs. [www.ibm.com](http://www.ibm.com)

---

**KEY DATABASE ACCELERATION TECHNOLOGIES**

The challenge of big data lies not only in the size of the data stores but also in the speed with which data is being generated or changed. To keep up, the latest databases include a number of technologies that can provide speed-of-thought performance.

**Dynamic in-memory processing.** Random access memory (RAM) is much faster than hard disk storage. One way databases can speed processing is by keeping the database in memory. That works very well until the size of the database exceeds the size of the RAM available. To overcome this problem, the next step forward is dynamic in-memory processing, which uses intelligent prefetch methods to achieve in-memory speeds even for extremely large databases. This capability helps keep performance high while not breaking the bank on memory costs.

**Data skipping.** For a given query, the system probably doesn’t need to examine every single block of data in a database. Some modern databases can identify and skip over data that is irrelevant to the question at hand. Skipping data not needed for a query can significantly reduce the processing burden and improve query performance.

**Columnar store.** Traditionally, databases have stored all information in a row-based format. However, for some workloads, a column-based format provides significant performance gains, and there is no need for indexes of columnar data. Look for databases that can provide the best of both worlds — the option of storing and querying data in rows or columns, depending on what will provide the best performance.