

Big data – from Concept to Reality with IBM System z and DB2 for z/OS

Big data—a concept that in the past few years has progressed from IT forums to the mainstream media—is the result of this incredible growth of information. Data stemming from a growing number of new sources continues to create volumes so large that they can easily move beyond our ability to understand and manage them. In response to this trend, big data technologies have emerged to help gain control over vast data quantities. Although some organizations have already adopted these technologies to help manage, analyze and leverage their information for business insights, big data projects are the wave of the near future for many others, surging to the top spot on their priority lists. One recent IBM study found, in fact, that more than two-thirds of chief marketing officers plan to implement new strategies and technologies for managing big data.

It's all about reducing the cost and risks of doing business

Taking advantage of this wealth of available information, big data projects can uncover new business opportunities to generate revenue as they help reduce business risk. By enabling organizations to perform historical or deep analytics on enormous volumes and varieties of data—often streaming at a high velocity—big data technologies can provide new insights to help organizations efficiently achieve business goals.

To make an effective big data platform a reality, organizations need a way to move their analytics systems closer to where the core transactional data like DB2 for z/OS is processed and stored. This approach can help improve business response with faster access to results produced by analytics. An efficient strategy can also help organizations leverage their existing analytics systems to work with their transactional and data warehouse systems to reduce IT costs.

Making Big Data a reality with IBM DB2 for z/OS

As many organizations are already performing analytics on core transactional data—many running them on the System z and IBM z/OS® platform with an IBM DB2® for z/OS database—expanding the analytics capabilities of the mainframe environment can be an ideal starting point for a big data project. Since System z is commonly used to run data warehouse functions and transaction and business analytics, it offers an optimal consolidation platform for a big data project. By expanding the types of data sets used by the analytics and decision engines on the System z platform to include the potentially vast amounts of differently structured data, organizations can create a powerful and cost-effective big data project.

One of the most efficient ways to do this is to run the big data collection and analysis functions separately on a third-party commodity platform such as Apache Hadoop, which can analyze and adapt the differently structured data to a structured format that can be used by z/OS. For example, since System z, DB2 for z/OS and IBM Information Management System (IMS™) often form the primary hub for all of an organization's transaction systems, performing prerequisite analysis for a big data project can spare the transaction systems on z/OS from heavy additional workloads. A commodity server solution such as Hadoop can be used to perform large batch jobs—such as analyzing archive data for the past 10 years—without impacting the mission-critical transaction environment of System z.

Although the combined use of mainframes and commodity servers can help organizations maximize efficiency and return on system investment, it also demands a key shift in system interaction—one that requires moving the commodity platform's analytics functions closer to the System z data. By connecting the System z data warehouse and decision-making functions with the big data analytics environment, organizations provide better, faster business responses.

In order to integrate the big data analytics environment of Hadoop with the DB2 for z/OS database environment on System z, the differently structured data needs to be adapted to be used by DB2, which uses predominantly structured data access. This requires connecting the data from Hadoop to z/OS and enhancing DB2 for z/OS to interact with the Hadoop environment.

Today, within DB2 for z/OS, organizations can use any of the standard Hadoop query languages such as Jaql, Hive or Pig to invoke a predefined Hadoop query. Once completed, DB2 for z/OS is notified and an SQL job finds the results of the Hadoop query on the big data platform and populates a DB2 table in a specified format.

To extend the functionality of DB2 for z/OS, organizations could utilize the emerging MapReduce standard. In this process, a MapReduce query is composed in SQL, and when the query is submitted, DB2 for z/OS recognizes the MapReduce query and offloads the Hadoop portion of the query to a big data environment, then retrieves the results and populates the target database tables.

One process that enables organizations to transparently share data across different environments begins with Hadoop gaining access to the DB2 catalog, which facilitates greater interchangeability. Hadoop could then understand how DB2 databases and tables are configured and structured, knowing where and how to store the results of queries with much less manual intervention.

In support of these processes, System z provides an optimal hub for big data. And by using the z/OS Distributed Data Backup feature of the IBM System Storage® DS8000 series of disk storage devices, it also can accelerate bulk data movement from other sources.

Think Big – Think System z

As data continues its unabated growth, the opportunities for big data will continue to increase, as well. Whether you are adopting a new big data initiative or plan to build upon an existing project and adapt to evolving data sources and structures, IBM System z provides the key component to a successful big data strategy. Due to its high availability, resilience and security, the System z platform is where a large proportion of worldwide data used for decision-making processes currently resides. By connecting analytics technologies with this data, you can efficiently and cost-effectively analyze data on a large scale and generate new business insights. IBM also offers the solutions you need to integrate, manage, secure and govern data throughout its lifecycle.

Virtually every organization has big data and has the opportunity to capitalize on it. As big data has moved into the spotlight, now is the time to take action, put your information to work and begin to reap the benefits.

To find out more listen to the [webcast](#) and download the [white paper](#)