IBM InfoSphere System z Connector for Hadoop

Bringing the Power of Hadoop to the Mainframe

Highlights

- Reduce cost and improve flexibility by shifting select workloads to Hadoop
- Easily transfer data from z/OS without the need for programming of special skills
- Protect sensitive data by optionally running Hadoop on the mainframe
- Make faster, more informed decisions and accelerate the delivery of new applications

For many, IBM® z/OS® mainframes form the back-bone of mission-critical business applications, securely and reliably storing and processing massive volumes of data day after day. Faced with spiraling data volumes and new business demands, organizations are looking for cost efficient ways to get more value out of critical mainframe data while ensuring that security and data integrity is maintained. For some types of applications, running additional extract, transform and load (ETL) operations on the mainframe is not practical.

IBM InfoSphere® System z Connector for Hadoop provides fast and seamless data connectivity between a variety of mainframe data sources and IBM InfoSphere BigInsights™, IBM's enterprise-grade Hadoop offering.

Customers can easily extract data from z/OS sources including DB2®, IMS™, VSAM and other file formats without the need for mainframe-based SQL queries, custom programming, or specialized skills. Once data is in Hadoop, clients can use the rich capabilities of InfoSphere BigInsights to quickly and cost-efficiently process and analyze data. Hadoop processing can take place on an external cluster connected to the zEnterprise® mainframe, or directly on mainframe Linux partitions using the System z Integrated Facility for Linux (IFL) for added security.

By extending the capabilities of IBM System z with IBM InfoSphere BigInsights and the IBM InfoSphere System z Connector for Hadoop, customers enjoy the best of both worlds - a hybrid transaction and analytic processing platform capable of managing mixed workloads and coping with the unexpected.

Big Data - It's defined by the fast velocity in which it is acquired, the broad variety of data types, and the great volume of accumulated data. Innovations around big data are delivering dramatically lower price points for traditional analytics and creating new opportunities for insights. Organizations are using big data to solve problems and outpace the competition.

However, analytics are only as good as the data brought to bear. In the case of accessing mainframe data, including databases, transactions, and application logs, this has traditionally required heavy-weight extract, transform, and load (ETL) processes and custom software.

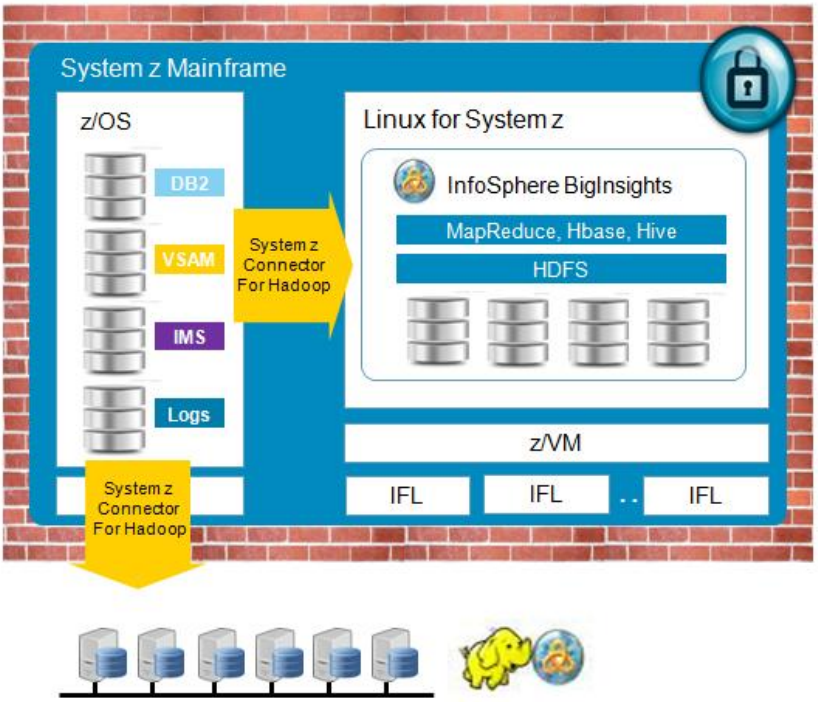
Traditional ETL can have drawbacks, including:

- Delays and costs for custom COBOL programming for Copybooks
- Costs of storage for staging during data transformation
- Consumption of batch windows

Many data scientists are turning to a new model: Agile extract and load, with transformation performed on the raw data in existing platforms or more flexible platforms like Hadoop.

By changing the traditional ETL process to EL-T (extract, load, and then separately transform), users gain faster data movement that delivers data with the details preserved for better analytics. This can unlock the power of Big Data. For example, faster access to more data can enable fraud detection algorithms to disable an account before additional charges can be made. This new agile extract and load model enables unified repositories ("data lakes") and next generation analytics.

The IBM System z Connector for Hadoop makes this possible today. Through an intuitive GUI, users can browse and select data from within z/OS on an IBM mainframe server, then copy —or schedule for copy—that into IBM InfoSphere BigInsights or third party Hadoop distributions on or off the mainframe.



Easy and secure access to mainframe data

IBM InfoSphere System z Connector for Hadoop provides native support for multiple mainframe data sources including IBM DB2, VSAM, QSAM and long form sequential data. It also supports SMF (System Management Facility) and RMF (Resource Management Facility) log file formats as well as system log files and operator log files.

By using IBM InfoSphere System z Connector for Hadoop to access mainframe data, customers can avoid the cost and complexities of writing programs to extract data from z/OS data sources, and can avoid the overhead of running SQL on the mainframe to extract data from DB2.

The Business Edge

Users benefit from IBM InfoSphere System z Connector for Hadoop by gaining a more comprehensive big data repository.

- Lower costs or improve batch windows by moving batch workloads to Hadoop or other big data platforms
- Make more informed decisions with critical information
- Make faster decisions with near real-time access
- Leverage the rapid innovation and talent in the Hadoop ecosystem

© Copyright IBM Corporation 2014

IBM Canada
3600 Steeles Ave East
Markham, Ontario
L3R 9Z7

Produced in Canada
October 2014
All Rights Reserved

IBM, the IBM logo, ibm.com, BigInsights, Cognos, DB2, System z, InfoSphere and z/OS are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at **ibm.com**/legal/copytrade.shtml

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

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

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product and service names may be trademarks or service marks of others.

References in this publication to IBM products and services do not imply that IBM intends to make them available in all countries in which IBM operates.

Please Recycle