

Delivering information you can trust

April 2008



IBM **Information Management** software

IBM InfoSphere Change Data Capture for z/OS single scrape capabilities

Contents

- 3 *Single scrape: how it works***
- 5 *The multiple subscription advantage***
- 6 *Benefits***

Part of IBM's data integration family, IBM® InfoSphere™ Change Data Capture for z/OS is a high performance, peer-to-peer data replication solution that integrates data in real time, bi-directionally between DB2® running on z/OS® and different databases on other computing platforms. Using log-based Change Data Capture (CDC) technology, IBM InfoSphere Change Data Capture for z/OS can access and integrate mainframe data with lower latency than batch loading, query or message-based systems.

IBM InfoSphere Change Data Capture for z/OS is deployed quickly and easily using a graphical user interface and can be used to solve many different business problems including:

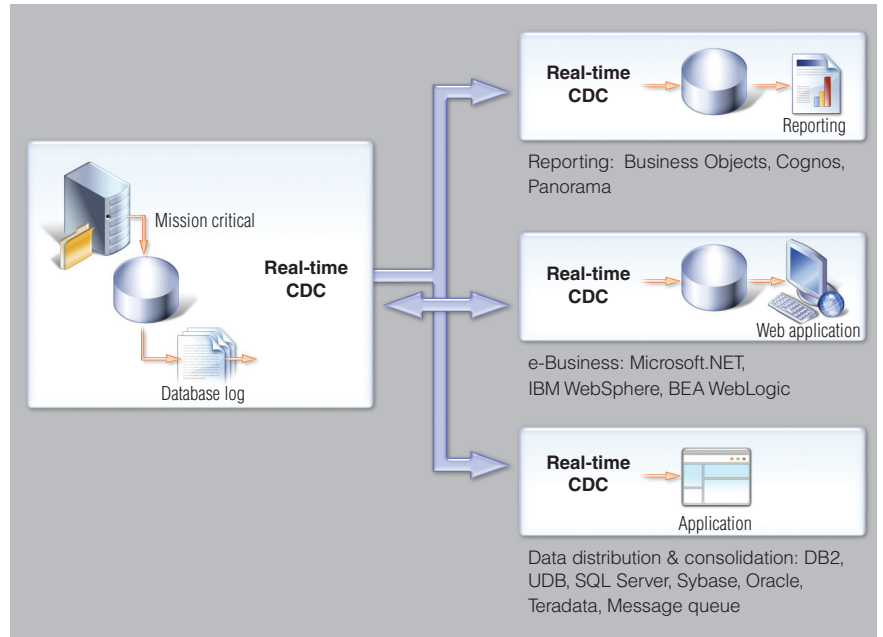
- *Loading a data warehouse in real-time so users can make operational and tactical business decisions based on the latest information*
- *Distributing information between different applications to ensure accurate and reliable data across the enterprise*
- *Integrating information between mission-critical mainframe applications and Web applications so employees, customers and partners have access to real-time information*
- *Consolidating data across systems in different regions, departments, or business units*
- *Improving the operational performance of systems already running at maximum capacity*

Single scrape: how it works

IBM InfoSphere Change Data Capture for z/OS uses a robust and configurable cache mechanism to support the Single Scrape function. As IBM InfoSphere Change Data Capture reads, or “scrapes” the DB2 database logs to capture changes that have occurred in the database the changes are cached for access by a subscription or integration process. Each integration process then accesses the cache to access the relevant data.

By employing this single scrape function it is possible to ensure that DB2 logs are only accessed once, regardless of the number of integration processes running concurrently. This results in reduced demand on DB2 and the DB2 logs, ensuring that IBM InfoSphere Change Data Capture for z/OS continues to place only a minimal demand on system resources.

Figure 1: Architectural overview of IBM InfoSphere Change Data Capture



The multiple subscription advantage

Single scrape greatly increases the efficiency of the integration process for customers who operate multiple subscriptions replicating from within a z/OS DB2 environment.

Each subscription within the IBM InfoSphere Change Data Capture environment is independent of other subscriptions with potentially different tables being replicated, independent start and stop times, different target system performance characteristics and different throughput rates.

If all subscriptions were to operate directly from a read of the DB2 logs, then all subscriptions would be constrained to operate at the pace of the slowest subscription. As a result, each subscription currently reads the DB2 logs independently of every other subscription. In large environments with many subscriptions, this can lead to scalability issues because the level of demand on the DB2 logs increases to the point that the read process can impact the source system.

Single Scrape functionality is designed to address this issue. This feature; which operates independently of all subscriptions; runs as a separate task within the IBM InfoSphere Change Data Capture for z/OS address space. The task reads the DB2 logs and places the retrieved data into a VSAM cluster called the Single Scrape cache. The cache is maintained in a circular fashion, so that the newest data to be added to the cache replaces the oldest data currently within the cache. The size of the cache is configurable, and is limited only by the system limits on the size of VSAM clusters.

Subscriptions access the cache to determine if the log data they require next is contained within the cache. If the data is within the cache then it is provided to the subscription and the DB2 logs are bypassed. If the data that the subscription requires is not within the cache, the subscription will wait until the cache contains the required data. If the data that the subscription requires is no longer within the cache, then the subscription will revert to reading the DB2 log directly and resume reading from the cache as soon as possible.

Benefits

- *Maximum performance with minimum demand on the system – Due to the structure and organization of the cache, only a single read of the DB2 log the necessary information for multiple subscriptions. This eliminates the need to read the log for each process and, as a result, overall product performance will improve.*
- *Easy administration – By using a circular cache, the size of the cache is well known, pre-defined and pre-allocated so there is never any danger of cache overflow that could cause process errors. Using a circular cache also means no new additional maintenance procedures are required because the oldest entries in the cache are automatically purged when overwritten.*
- *External cache – Leveraging the highly optimized VSAM file system means system demand is kept at a minimum.*
- *Resilient – In the unlikely event of a total loss of the cache, IBM InfoSphere Change Data Capture for z/OS will automatically rebuild the cache and allow all subscriptions to revert to DB2 log scraping until they catch up to the cache and are able to benefit from it once again.*

For more information

For more information about IBM Information Server, contact your IBM marketing representative or visit ibm.com/software/data/integration



© Copyright IBM Corporation 2008

IBM Software Group
Route 100
Somers, NY 10589

Printed in the United States
April 2008
All Rights Reserved.

IBM and the IBM logo, InfoSphere, DB2 and z/OS are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries or both.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc., in the United States, other countries or both.

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

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

All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

***TAKE BACK CONTROL WITH* Information Management**