

NatCDC

The

Change Data Capture Solution For ADABAS™

Overview.....	2
Processing Overview	3
Features.....	3
Benefits.....	4
NatCDC for Data Warehousing	6
Integration with Extraction, Transformation and Loading (ETL) Tools	6
General Information.....	8
Software AG Product Dependencies	8
Supported Source Environments	8

Overview

Change Data Capture (CDC) is a term that describes the ability to take the transaction log of a database (a log which records each and every transaction made against the database) and then process this data to produce a file that details the “delta” of changes. For every unique record handled and no matter how many times the unique record was handled within the transaction log – the “delta” produced would contain a single record that contains the last logical image of the record from the transaction log time span.

When CDC is applied to ADABAS, the mechanism that provides the source of Change Data is the ADABAS Protection Log (PLOG). The data that a PLOG contains has a wide-range of uses to an organization such as:

- Updating Data Warehouses
- Audit reporting
- Data mining
- Troubleshooting
- Selective data recovery
- Data monitoring
- A Source for Data Replication

The primary purpose of the PLOG with an ADABAS database is the PLOG’s use in providing data recovery in the event of a failure in either hardware or software. To this end, the data that a PLOG contains is readily usable to the ADABAS data recovery utility. It is not however immediately usable for its other potential uses, due to several challenges:

1. **Raw PLOG Data Contains Transactions Against All ADABAS Files**

Files of interest must be “split out” for file specific processing

2. **PLOG Data is Stored in Compressed Form**

This requires record “decompression”, an expensive processing activity for the CPU

3. **PLOG Data is Stored as Variable-Length Records**

Record lengths vary from file to file, and also within the same file due to recurring fields

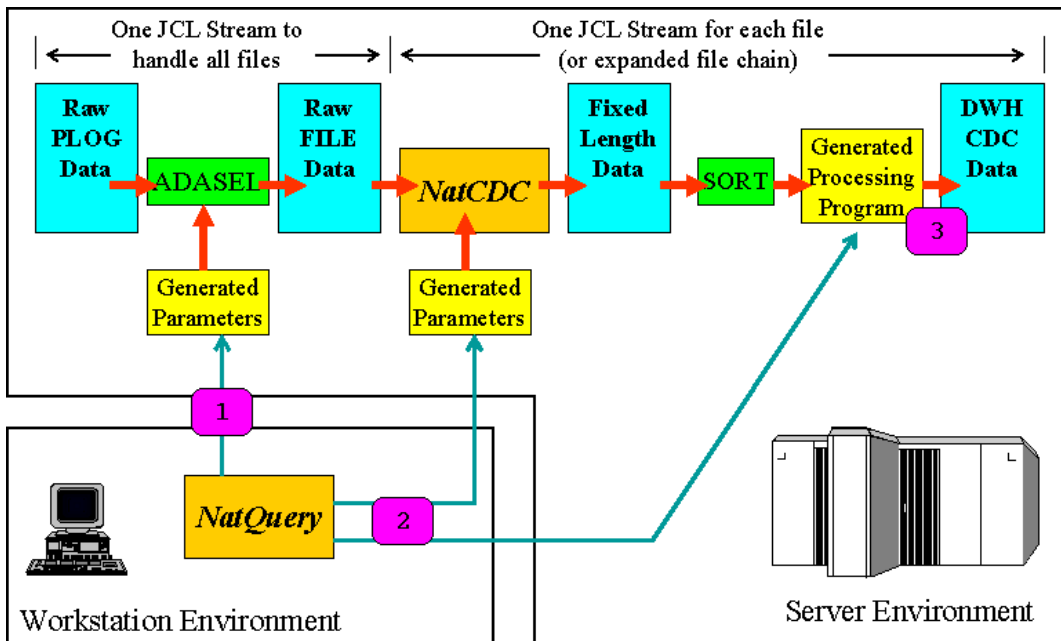
4. **PLOG Data, After Decompression, May Need Further Transformation**

Internal ADABAS formats such as Date (D) and Time (T) need to be transformed into external equivalents

To address these challenges we have designed NatCDC, a single Natural program that is a highly optimized PLOG processing engine. The combination of NatQuery, NatCDC and ADABAS utilities offers the most straightforward, flexible, easily understood and cost-effective solution to tapping the data available in a PLOG. It is also easily one of the best performing solutions, both in terms of CPU usage and Clock Time.

Processing Overview

The following diagram graphically demonstrates NatCDC processing.



Description of Process:

1. NatQuery is used to generate parameter cards for ADASEL that allows ADASEL to properly process raw PLOG data. ADASEL decompresses the raw PLOG data for specific files of interest.
2. NatQuery is used to generate parameters cards for NatCDC, and at the same time generates a file-specific processing program that performs required data handling and format conversions.
3. The NatCDC / Sort / Processing JCL stream is run, which delivers data into usable form.

Features

NatCDC offers the following general features:

- **100% Data Integrity**
 - PLOG are processed using SAG's own PLOG utilities
- **Simple Mainframe Installation / Minimal Mainframe Footprint**
 - One primary Natural object program (NatCDC) installed via and automated NATLOAD
 - Required Parameters, JCL and Natural programs are all generated on a workstation and then FTPed onto the mainframe server
- **Full Handling of all ADABAS Data Structures Including:**
 - Multi-Valued Fields (MUs), Periodic-Group Fields (PEs) and MUs in PEs

- Automatic format translations including Date, Time and Timestamp
- Proper sign handling of all numeric-based fields
- Capture of C* values for all MUs, PEs and MUs in PEs
- **Efficient and Trusted Decompression**
 - SAG-supplied utilities are utilized to handle Decompression
- **Full Conversion of PLOG Variable-Length Records to Fixed-Length**
 - Final data record layout is user-specified
- **Full PLOG Record Header Translation**
 - **STCK time, binary and packed number conversion)**
- **Full Handling of All Transactions**
- **Full Ability to Select / Skip Specific Fields in Final Output**
- **Full Support for Expanded Files, with Physical to Logical ISN Conversion**
- **Transactional Data is Supplied with Standard “Transaction Header”**
- **Full Generation of All Objects Required for PLOG Handling**
 - Fully Graphical User Interface (GUI)
- **User-Selectable Output Options:**
 - Delivery of “Delta”, or the Logical “Last” image (flagged as **S**Store, **D**Delete or **U**Update)
 - Logical “First and Last” images (flagged as first **B**efore and Last **A**fter)
 - All Transactions (thus providing a means to audit transactions when used with the NatQuery)
- **Reporting:**
 - Exception reporting for recurring field handling
 - Total count of all Before and After images
 - Total of all records handled
 - Total number of Stores, Deletes, Updates and Store / Deletes
 - Total number of unique records handled

Benefits

NatCDC offers an organization the following benefits:

- **Minimal Install Footprint on Mainframe Server**
 - Only Natural programs are physically installed into the mainframe
- **Efficient Use of CPU**

- Utilization of SAG's own Decompression yields excellent processing speeds
- **Cost Effective**
 - Leveraging existing Software AG utilities reduces cost
- **Flexible Handling of Transactional Data**
 - Used in conjunction with the ADABAS utility ADASEL, NatCDC delivers transactions that precisely reflect ADABAS' own optimistic transaction approach to transaction handling (this is the NatWorks recommended approach)
 - Used in conjunction with the ADABAS utility ADACDC, NatCDC delivers transactions that can reflect only committed transactions
- **Individual File Processing Can Occur Concurrently**
 - Through the use of either ADASEL or ADACDC, all transactional images occurring against files of interest can be written to a single disk file, with all file specific processing being able to then occur concurrently.
- **PLOG Data is Obtained Directly from the PLOG**
 - Minimal ADABAS access required to process data (the only ADABAS access required is to process a given file's FDT)
- **Zero Programming Effort Required to Obtain Change Data Capture**
 - No manual coding is required to use NatCDC effectively

NatCDC for Data Warehousing

When a data warehouse is first implemented, data extraction is performed against one or more source environments to secure the required data. This data is then applied to the target data warehouse environment.

Immediately after the initial load of warehouse data is accomplished, the data originally placed into the warehouse rapidly becomes “out-of-sync” with the source environment – due simply to the fact that production systems continue to change.

One solution to this problem is to periodically completely refresh all data in the warehouse with new data. This approach, while entirely workable, becomes cumbersome in situations where large exceedingly large amounts must be processed, simply because in most situations a large percentage of the data residing in the warehouse does not require any changes (I.E., the source data has not been changed).

To address this situation, a Change Data Capture approach provides significant advantages. Using this approach, the warehouse is periodically updated with only those data records that changed – not all data records – with huge savings in comparison to full reloading. These savings are seen in the time to capture changed data, the time to move these changed records from the source platform to the target platform, and the time required to apply these changes to the target platform.

It is in this situation where NatCDC can be utilized to achieve one of the best performing and most cost effective methods of handling ADABAS Change Data Capture.

Integration with Extraction, Transformation and Loading (ETL) Tools

While Extraction and Change Data Capture are two key processes in implementing a Data Warehouse or Data Mart, organizations that are approaching an implementation often look for additional capabilities such as automatic data transformations and automatic loading of captured source data onto the desired target.

While integration to ETL tools can occur in a number of ways – virtually all ETL tools allow for the immediate processing of NatQuery and NatCDC generated data as long as the physical layout of this data is described in a supported format. To facilitate the integrations of NatQuery / NatCDC into the ETL tool of the customer’s choice, NatQuery has the ability to generate two types of interface files; or files that describe the physical layout of a NatQuery / NatCDC produced data. These files then allow NatQuery-provided extract data to be further processed by the Extraction, Transformation and Loading (ETL) tools.

To support integration with most ETL tools, NatQuery supports the generation of COBOL File Definition (CFD) files – or files that describe the physical layout of NatQuery / NatCDC produced data in COBOL format. With CFD files generated, the data produced by NatQuery / NatCDC becomes immediately usable to any ETL tool that can import a CFD file.

Beyond the ability to generate CFD files that describe the output of NatQuery / NatCDC, NatQuery additionally supports the generation of DataStage export (DSX) files – a file format that is proprietary to Ascential Software’s DataStage product. In addition to being able to describe the layout of a NatQuery / NatCDC produced data, an additional capability of DSX file generation is that NatQuery can provide full metadata on each ADABAS field being extracted by imbedding all PREDICT information directly into the DSX file. For customers who are considering the use of DataStage – this is a significant advantage - as both the layout of the file and all metadata are generated into a single import file for DataStage.

With either the CFD or DSX approach, a NatQuery / NatCDC customer is able to immediately describe the layout of extracted data to the ETL tool of choice, thereby allowing that ETL tool to complete the transformation and loading of required data.

General Information

NatCDC is an add-on to the product NatQuery. Please refer to NatQuery documentation for any additional Product Dependencies and Requirements.

Software AG Product Dependencies

The following table summarizes the associated Software AG products and their relative dependencies on the server platform.

Product	Features/Comments
NATURAL	Required – version 2.3.1 or higher
ADABAS	Required – version 6.2 or higher. To use some features of NatCDC, ADABAS 7 or higher is required (I.E. the use of ADACDC)

Supported Source Environments

The following table summarizes the operational environments that support NATURAL™ and are therefore suitable for use with NatCDC.

Platform	Operating System
IBM	MVS
IBM	VSE
Siemens	BS2000
Fujitsu	MVS derivative
Hitachi	MVS derivative