


OVERVIEW OF MICROSOFT'S CHANGE DATA CAPTURE

Chris Skorlinski
Microsoft SQL Escalation Services
Charlotte, NC
<http://blogs.msdn.com/ReplTalk>




What data has changed?

Loading of Data Warehouse updates

- Append New Static Type 0 Dimensions
 - Add New and Replace Current Type 1 Dimensions,
 - Track History of Type 2 Dimensions
- 

Database Auditing of changed data

- Need to only record “deleted” data
 - Keep complete history of who changed what
 - Feed “legacy” database after migrating
- 

Need ability to “rollback” selected changes without performing a complete Database Restore

What are some Techniques used to identify changed data?

Time Stamp or Row Version Column




- Easy to implement using DataModified column
 - Overhead on user table to query changes
 - Not get deleted data or previous values
- 

Table Triggers

- Can see both Deleted and Previous data.
 - Customizable to include User or InputBuffer
 - Big overhead on highly transactional tables
- 

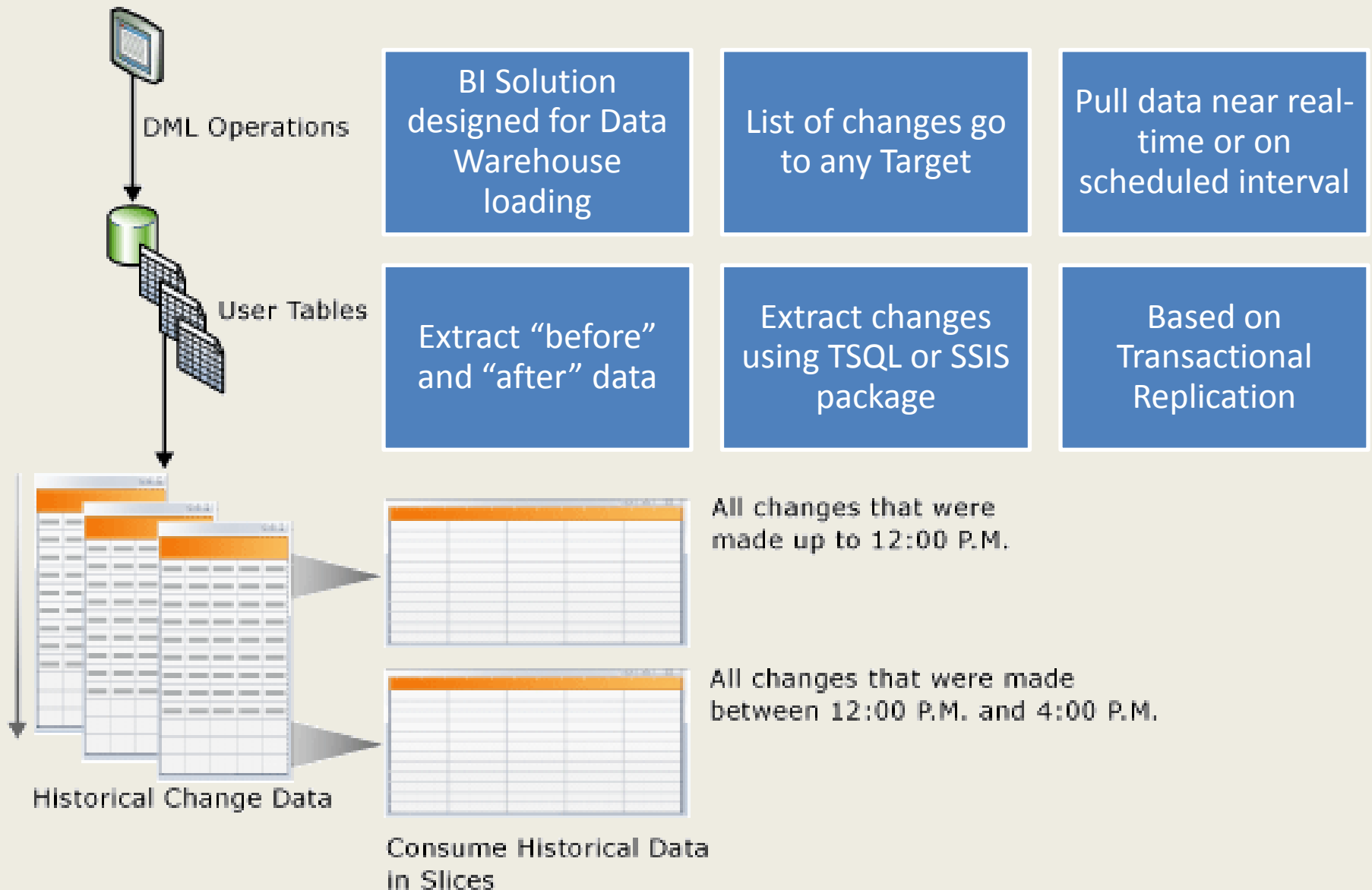
Database Snapshot (new for SQL 2005)

- Easy to implement for “undo”
 - Harder for “changed” data
- 

SQL Utilities and 3rd Party Tools

- Backup/Restore and TableDiff
- Lumigent Technologies Log Explorer
- Attunity Stream

CHANGE DATA CAPTURE (CDC) SOLUTION



Key CDC Components

Change Tracking Tables

- Based on schema of tracked table
- Copy of “before” and “after” data for each record changed

Log Reader Job

- Near real-time capture of transactions
- Same internal code as Transactional Replication
- Transactions tracked by LSN

Clean Up Job

- Purges tracking tables of data older than 72 hours (3 days)

CDC Created Functions

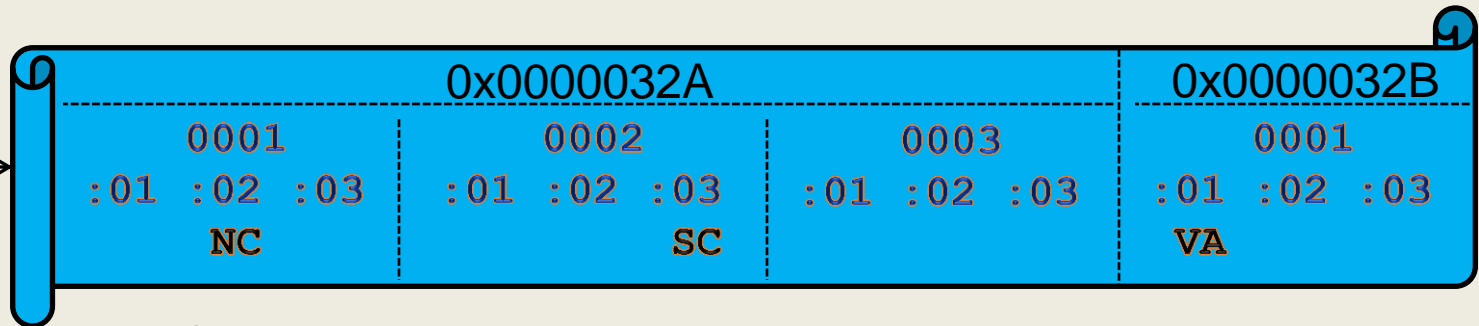
- sys.sp_cdc_help_change_data_capture
- cdc.fn_cdc_get_all_changes
- cdc.fn_cdc_get_net_changes

LOG SEQUENCE NUMBER (LSN)

- LSN uses to identify transactions written to the database transaction log
- Each transaction is given a unique LSN number.
- Often used to RESTORE to a specific transaction number.

Customer
Table

3 part number often represented in HEX
Ex: 0x0000032A000004CB0003



```
set @begin_lsn = sys.fn_cdc_get_min_lsn('Customer')  
set @end_lsn = sys.fn_cdc_get_max_lsn()
```

```
select * from cdc.fn_cdc_get_all_changes(@begin_lsn, @end_lsn)
```

How to enable Change Data Capture (CDC)

--Demo based on the [AdventureWorksLT2008] database

-- <http://msftdbprodsamples.codeplex.com/releases/view/45907>

USE [AdventureWorksLT2008]

GO

-- Create Sample table used for CDC Demo

```
CREATE TABLE [SalesLT].[CustomerCDCDemo](
    [CustomerID] [int] IDENTITY(1,1) NOT NULL,
    [FirstName] [dbo].[Name] NOT NULL,
    [LastName] [dbo].[Name] NOT NULL,
    [Phone] [dbo].[Phone] NULL,
    [CreditLimit] Money NULL,
    [ModifiedDate] [datetime] NOT NULL,
    CONSTRAINT [PK_CustomerCDCDemo] PRIMARY KEY CLUSTERED
(
```

-- Turn on Change Data Capture at the database level

```
exec sys.sp_cdc_enable_db
```

-- To turn off Change Data Capture

```
exec sys.sp_cdc_disable_db
```

How to enable a table for CDC

`sys.sp_cdc_enable_table`

```
@source_schema = N'SalesLT'  
, @source_name = N'CustomerCDCDemo'  
, @role_name = N'cdc_admin'  
, @capture_instance = N'CustomerCredit' -->>Name used in CDC functions  
, @supports_net_changes = 1  
, @index_name = NULL  
-->>Captured only required columns  
, @captured_column_list = 'CustomerID,FirstName,LastName,Phone,CreditLimit,ModifiedDate'  
-->>Store "captured" data on different disk  
, @filegroup_name = NULL
```

-- You should see the 2 SQL Agent Jobs automatically created and starting.

Job 'cdc.AdventureWorksLT2008_capture' started successfully.

Job 'cdc.AdventureWorksLT2008_cleanup' started successfully.

- cdc.<database>_capture scans the transaction log and copies changes from enabled tables to individual CDC change tables.
- cdc.<database>_cleanup purges individual change tables after 3 days.

You can view the default values used in these SQL Agent jobs by executing the [sys.sp_cdc_help_jobs](#) stored procedure. You can change the default values by executing the [cdc_jobs](#) stored procedure. By default the cleanup job removes change data after it is 3 days old.

get_ALL_changes

Once enable, use CDC Functions to retrieve history of changes.
Function can return "current" values or both "current" and "previous"

fn_cdc_get_min_lsn: to calculate starting LSN of changes in the table

fn_cdc_get_max_lsn() to calculate highest possible LSN for changes.

fn_cdc_get_all_changes Retrieve NEW or PREVIOUS and NEW

```
declare @begin_lsn binary(10), @end_lsn binary(10)
set @begin_lsn = sys.fn_cdc_get_min_lsn('CustomerCredit')
set @end_lsn = sys.fn_cdc_get_max_lsn()

select [__$start_lsn],[__$operation],[__$update_mask]
      ,[FirstName],[CreditLimit]
from cdc.fn_cdc_get_all_changes_CustomerCredit(
      @begin_lsn, @end_lsn, 'all'/'all update old')
```

get_NET_changes

Row changed multiple times, but I only want to see final "combined" changes.

Use CDC "net_changes" functions to retrieve "final" set of changes.

Special feature, more overhead, must be enabled when CDC is enable.

```
sys.sp_cdc_enable_table  
    ,@supports_net_changes = 1
```

```
declare @begin_lsn binary(10), @end_lsn binary(10)  
set @begin_lsn = sys.fn_cdc_get_min_lsn('CustomerCredit')  
set @end_lsn = sys.fn_cdc_get_max_lsn()
```

```
select [__$start_lsn],[__$operation],[__$update_mask]  
    ,[FirstName],[CreditLimit]  
from cdc.fn_cdc_get_net_changes_CustomerCredit(  
    @begin_lsn, @end_lsn, 'all'/'all update old')
```

DEMO TIME

1. Enable Database for CDC
2. Enable Customer Table for CDC
3. Update User Tables
4. Use CDC functions to retrieved changes

<http://blogs.msdn.com/b/repltalk/archive/2010/09/30/walk-through-of-microsoft-sql-server-change-data-capture.aspx>

CDC BEST PRACTICES

- Avoid INSERT followed by UPDATE
- Avoid capture tables that have frequent large batch updates.
- Capture to only the columns you really need to track by specifying the `@captured_column_list`.
- If you do not require support for net changes, set `@supports_net_changes` to 0.
- Spread IO, specify `@filegroup_name` in `sys.sp_cdc_enable_table`.

Wrap Up

What feedback can you provide me about today's presentation?

Is there a project you have in mind for CDC?

Can you think of a topic of interest for a SQLSaturday or online Webcast?

<http://blogs.msdn.com/ReplTalk>

- Change Data Capture Best Practices
- References and Links on Change Data Capture
- Performance Troubleshooting
- BING CDC or ReplTalk

