

Managing Data Warehouse Incremental Loads WITH SQL SERVER CHANGE TRACKING

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

1

Agenda

- ▶ Data warehouse change detection concepts
- ▶ Essentials of SQL Server change tracking
- ▶ Change tracking in DW loads

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

2

Tim Mitchell

- ▶ Business intelligence consultant – Dallas, TX
- ▶ Microsoft SQL Server MVP
- ▶ TimMitchell.net
- ▶ twitter.com/Tim_Mitchell

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

3

Upcoming Events

- ▶ Building Better SSIS Packages (full day):
 - ▶ Dallas: bit.ly/1NQ17x3
 - ▶ Boston: bit.ly/1Yo9xiH
 - ▶ Seattle (PASS Summit): bit.ly/1Wap1ok
- ▶ Advanced SSIS Class (4 days)
 - ▶ Washington, DC, Dec 7-10: bit.ly/1RxvMwa

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

4

Data Warehouse Change Detection Concepts

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

5

Data Warehouse Change Detection

- ▶ Why change detection?
 - ▶ Performance
 - ▶ Historical record

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

6

Data Warehouse Change Detection

- ▶ Detecting changes on the source
 - ▶ Update timestamp



CustomerID	TransactionType	TransactionAmount	EmployeeID	TransactionTime	Row Timestamp
BN144	REFUND	74.17	37	2012-09-26 19:15:51.827	2012-09-27 20:18:56.373
PK248	SALE	1021.11	84	2012-09-26 20:37:11.127	2012-09-27 21:18:57.450
RR100	SALE	92.90	84	2012-09-26 15:04:23.527	2012-09-27 21:18:57.450
LM87	SALE	172.13	60	2012-09-26 15:24:03.150	2012-09-27 21:18:57.450

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

7

Data Warehouse Change Detection

- ▶ Detecting changes on the source
 - ▶ Change data capture
 - ▶ Full history (not just delta)
 - ▶ SQL Agent
 - ▶ Asynchronous
 - ▶ Requires Enterprise Edition

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

8

Data Warehouse Change Detection

- ▶ Detecting changes on the source
 - ▶ Change tracking
 - ▶ Delta detection (not full history)
 - ▶ Synchronous

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

9

Data Warehouse Change Detection

- ▶ Detecting changes at the destination
 - ▶ Row hashing
 - ▶ Stores calculated hash in destination
 - ▶ Efficient comparison of binary hash

AA011239
John
Smith



0x54 0xa4 0x9b...

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

10

Data Warehouse Change Detection

- ▶ Detecting changes at the destination
 - ▶ Brute force
 - ▶ Column by column comparison
 - ▶ Least efficient
 - ▶ Fewest technical assumptions

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

11

Data Warehouse Change Detection

- ▶ Change detection tools
 - ▶ Plain T-SQL
 - ▶ T-SQL merge
 - ▶ SSIS slowly changing dimension wizard
 - ▶ SSIS data flow components
 - ▶ Third party tools

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

12

Data Warehouse Change Detection

- ▶ Facts or dimensions?
 - ▶ Change detection often needed on both
 - ▶ Change history usually reserved for dimensions

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

13

Essentials of SQL Server Change Tracking

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

14

SQL Server Change Tracking

- ▶ Overview
 - ▶ Introduced in SQL Server 2008
 - ▶ Any edition of SQL Server

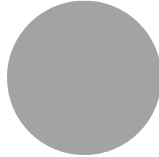
Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

15

SQL Server Change Tracking

- Overview
 - Enable on database
 - Turn on table-by-table
 - Retention



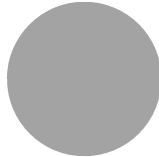
Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

16

SQL Server Change Tracking

- Technical info
 - Tracks inserts, updates, deletes
 - Column-level (optional)



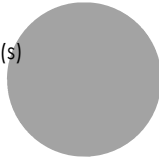
Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

17

SQL Server Change Tracking

- Technical assumptions
 - Metadata change on source table(s)
 - Primary key required
 - Storing the last sync version



Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

18

SQL Server Change Tracking

- ▶ Common functions
 - ▶ `CHANGE_TRACKING_CURRENT_VERSION()`
 - ▶ `CHANGETABLE()`
 - ▶ `CHANGE_TRACKING_IS_COLUMN_IN_MASK()`

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

19

Demos

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

20

Final Thoughts

- ▶ Change tracking
 - ▶ Fast, efficient way to track source changes
 - ▶ Based on version stamp
 - ▶ Requires source metadata changes
 - ▶ Relatively easy maintenance

Managing Data Warehouse Incremental Loads with SQL Server Change Tracking

Tim Mitchell | TimMitchell.net

21

Final Thoughts

Thanks!

- ▶ Slide deck on website
- ▶ Newsletter: TimMitchell.net/Newsletter

