Change Data Capture for SSIS in SQL Server 2012

Copper Blue
Consulting

Who am I?

- Allan Mitchell

- Joint author on 2005/2008 SSIS Book by Wrox

- Websites

  - www.CopperBlueConsulting.com

- StreamInsight and SQL Azure Advisory Boards

- Specialise in Data and Process Integration

- Microsoft SQL Server MVP

- Twitter: allanSQLIS

- E: allan.mitchell@CopperBlueConsulting.com

Why?

The Problem

Data, Data Everywhere

❖Everyone has it

❖Lots of it

❖Dispersed around our environment

❖Heterogeneous Sources

Copper Blue
Consulting

Problem with data – Part 1

❖What is data?

❖Raw

❖Doesn't mean a lot

❖Needs context

❖Needs enriching

❖Need to take next steps

❖Information

❖Knowledge

❖Wisdom

Problem with data – Part 2

❖ Timeliness

❖ When do people need it

❖ How do they need it

❖ "Out of date" data is next to useless

❖ Not everyone needs real-time

❖ Everyone should make "in date" their goal

Copper Blue Consulting

Problem with data – Part 3

❖Quantity

❖No longer only for big Corps

❖1TB of disk is ~ $50

❖Moving it around

❖Windows of operation

❖Only transfer what we need

Copper Blue
Consulting

What are people doing?

❖As soon as data in source changes

❖Extract

❖Cleanse

❖Conform

❖Load

❖Only moving what needs to be moved

❖Right?

WRONG

Copper Blue
Consulting

Why?

❖ Hard

❖ No tracking on Source Systems

❖ Costs money to design

❖ Need new skillsets

Copper Blue
Consulting

How?

CDC in SQL Server 2012

- Introduced in SQL Server 2008
- Not the easiest thing to code
- No integration with SSIS
- Uses Transactional replication Agent to read log LSNs

Copper Blue
Consulting

Setup

- Enable DB for CDC
  - *sys.sp_cdc_enable_db*
- Enable table for CDC
  - *sys.sp_cdc_enable_table*

Components

- **Control Flow**
  - CDC Control Task: Manages state and recovery of CDC Packages.

- **Data Flow**
  - CDC Source: Reads change data from capture tables
  - CDC Splitter: A Conditional Split for CDC

Copper Blue
Consulting

Packages

- **Initial**: This package does the initial "Full Pull"
- **Incremental:** This does the incremental extracts

Copper Blue
Consulting

CDC Control Task

- **Initial Load**
  - Mark initial load start
  - Mark initial load end
  - Mark CDC start
- **CDC**
  - Get processing range
  - Mark processed range
  - Reset CDC state

Copper Blue Consulting

CDC Source

- All
- All with old values
- Net
- Net with update mask
- Net with merge

Demo

Putting it together

Thankyou

e: allan.mitchell@copperblueconsulting.com
t: @allanSQLIS

Copper Blue Consulting