

Informatica PowerCenter

Lesson 18: Incremental Aggregation

Lesson Objectives

- In this Lesson you will learn about:
 - What is incremental aggregation ?
 - Understanding Incremental Aggregation Process
 - Implementation – Sample mapping



18.1. Introduction To Incremental Aggregation

- Using incremental aggregation, we apply captured changes in the source data (CDC part) to aggregate calculations in a session.
- If the source changes incrementally and we can capture the changes, then we can configure the session to process those changes
- This allows the Integration Service to update the target incrementally, rather than forcing it to delete previous loads data, process the entire source data and recalculate the same data each time you run the session.
- INCREMENTAL AGGREGATION, may be helpful in cases when we need to load data in monthly facts in a weekly basis.
- Note : Applicable if CDC (Change data capture) is implemented
 - Use incremental aggregation when the changes do not significantly change the target

Use this mapping when you do not need to keep any previous versions / history of dimensions in the table.

18.2. Understanding Incremental Aggregation Process

- The first time we run an incremental aggregation session, the Integration Service processes the entire source.
- At the end of the session, the Integration Service stores aggregate data for that session run in two files, the index file and the data file in the cache directory specified in the Aggregator transformation properties.
- For next run Integration Service check if the CDC data from the source is existing in the index file and does the incremental aggregation and updates the index and data file accordingly.
- Each subsequent time we run a session with incremental aggregation, the Integration Service creates a backup of the incremental aggregation files.
- The cache directory for the Aggregator transformation must contain enough disk space for two sets of the files.
- The Integration Service creates new aggregate data, instead of using historical data, when we configure the session to reinitialize the aggregate cache, Delete cache files etc.
- When the Integration Service rebuilds incremental aggregation files, the data in the previous files is lost.
- Note: To protect the incremental aggregation files from file corruption or disk failure, periodically back up the files.

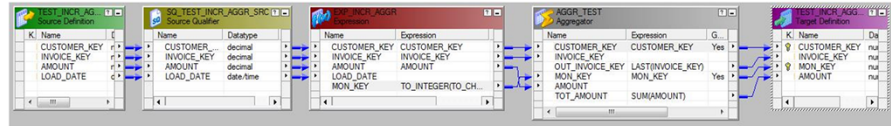


Copyright © Capgemini 2015. All Rights Reserved 4

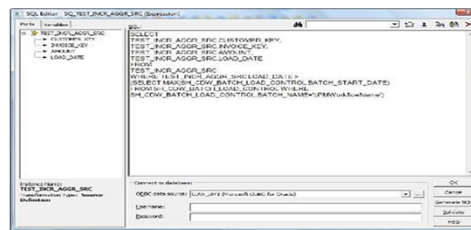
Use this mapping when you do not need to keep any previous versions / history of dimensions in the table.

18.3. Implementation of Incremental Aggregation

- Sample Mapping



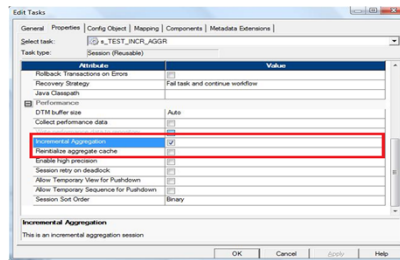
- CDC Source : Source query should fetch only the changed record .
In below case , Batch Load Control table is used



Use this mapping when you do not need to keep any previous versions / history of dimensions in the table.

18.3. Implementation of Incremental Aggregation

- Expression and Aggregate transform can have the desired logic
- Session Configuration : Click check box of Incremental aggregate



- If we want to reinitialize the aggregate cache suppose during first week of every month we will configure the same session in a new workflow at workflow level with the Reinitialize aggregate cache property checked in.

Use this mapping when you do not need to keep any previous versions / history of dimensions in the table.

18.3. Implementation of Incremental Aggregation

- Example with Data :
- Source data table

CUSTOMER_KEY	INVOICE_KEY	AMOUNT	LOAD_DATE
1111	5001	100	01/01/2010
2222	5002	250	01/01/2010
3333	5003	300	01/01/2010
1111	6007	200	07/01/2010
1111	6008	150	07/01/2010
2222	6009	250	07/01/2010
4444	1234	350	07/01/2010
5555	6157	500	07/01/2010

CUSTOMER_KEY	INVOICE_KEY	MON_KEY	AMOUNT
1111	5001	201001	100
2222	5002	201001	250
3333	5003	201001	300

- Target Table After 1st Weeks Load

Use this mapping when you do not need to keep any previous versions / history of dimensions in the table.

18.3. Implementation of Incremental Aggregation

- Now during the 2nd week load it will process only the incremental data in the source i.e those records having load date greater than the last session run date. After the 2nd weeks load after incremental aggregation of the incremental source data with the aggregate cache file data will update the target table with the following dataset:

CUSTOMER_KEY	INVOICE_KEY	MON_KEY	AMOUNT	Remarks/Operation
1111	6008	201001	450	The cache file updated after aggregation
2222	6009	201001	500	The cache file updated after aggregation
3333	5003	201001	300	The cache file remains the same as before
4444	1234	201001	350	New group row inserted in cache file
5555	6157	201001	500	New group row inserted in cache file

Use this mapping when you do not need to keep any previous versions / history of dimensions in the table.

Summary

- This Lesson gives knowledge incremental aggregation



Review Question

- Is a pre-requisite for incremental aggregation

A. SCD
B. CDC

Option B

- Cache file is recreated when incremental aggregation

- check box is checked

A. True
B. False

Option B

