



Business School
UNIVERSITY OF COLORADO **DENVER**

Information Systems Program

Module 4

Data Integration Concepts, Processes, and Techniques

Lesson 2: Change Data Concepts



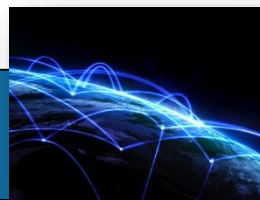
Lesson Objectives

- Explain the types of data sources involved in data integration
- Provide examples of typical data quality problems encountered during data integration
- Reflect on the relationship between type of change data and data quality

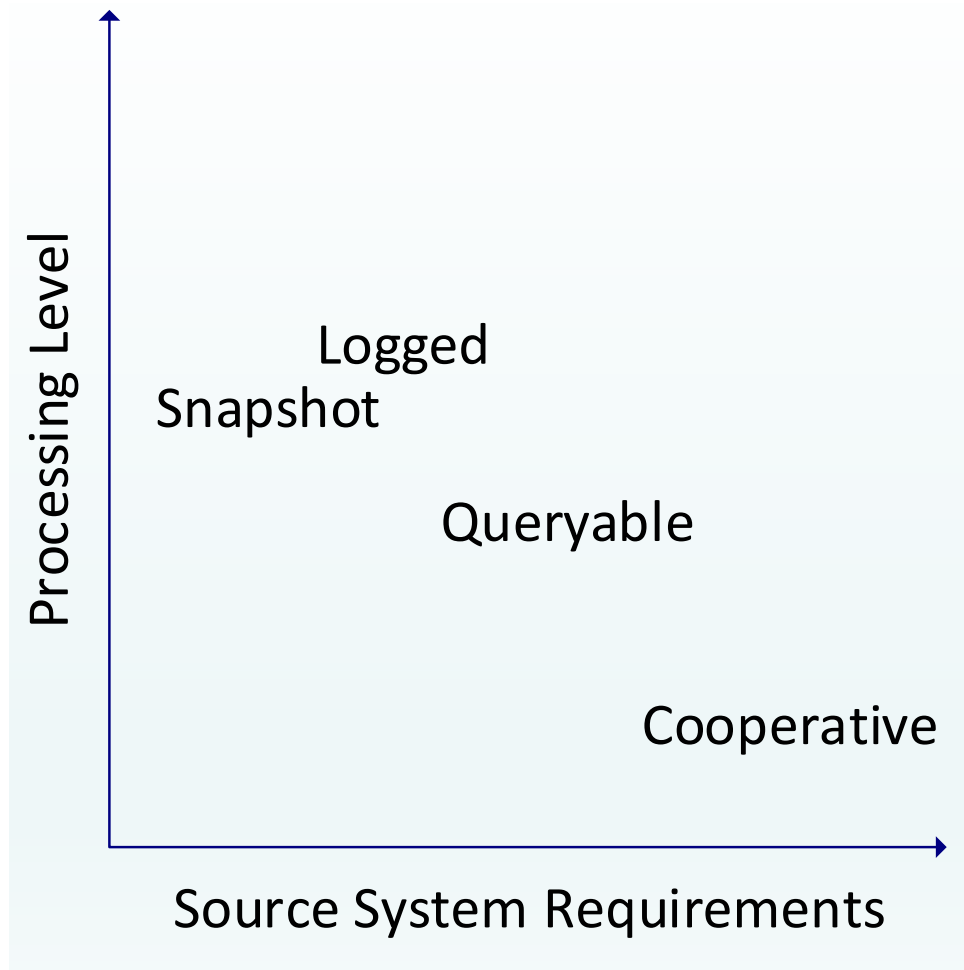


Basics of Change Data

- Derived from internal and external data sources
- Used to populate and refresh a data warehouse
 - Insert rows in fact and dimension tables
 - Update rows in dimension tables
- Challenges
 - Difficult to change to source systems especially external systems
 - Lack of SQL access and descriptive (meta) data especially for legacy data



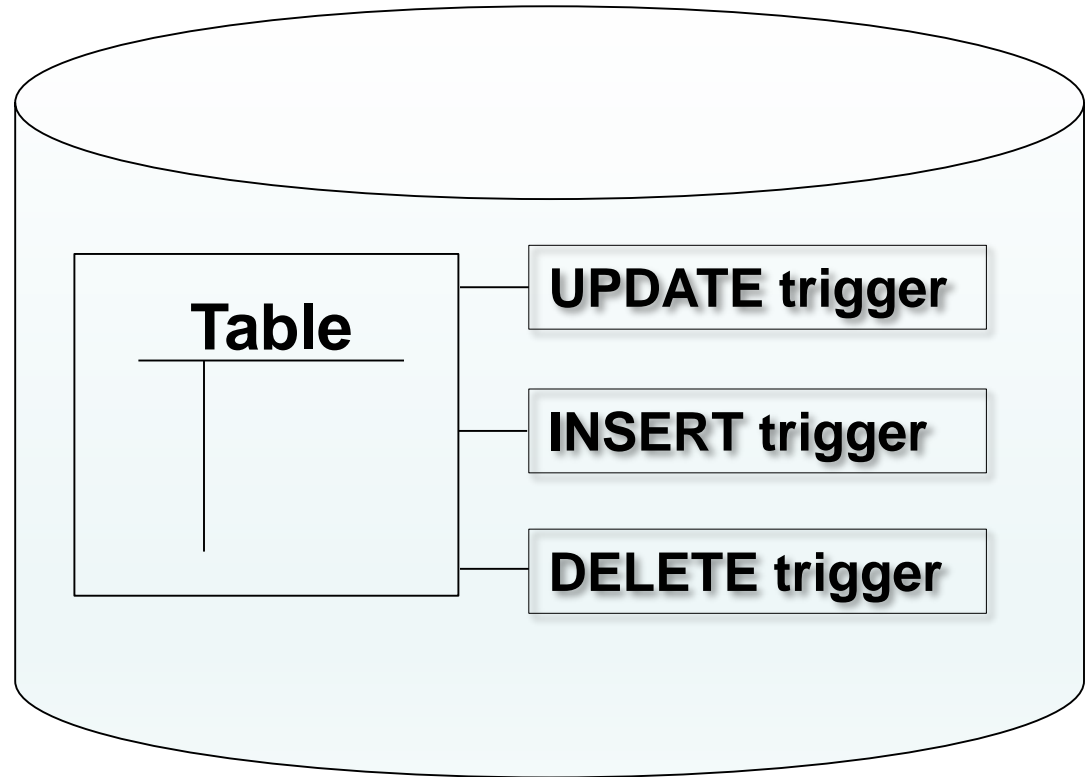
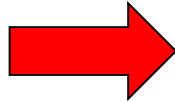
Change Data Classification



Cooperative Change Data

Applications

UPDATE ...
INSERT ...
DELETE ...



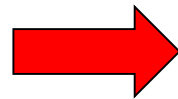
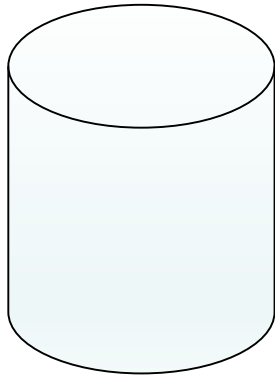
Logged Change Data

IP Address	111.111.111.111
Remote user	-
Authenticated user	-
Timestamp	[08/Oct/2014:11:17:55 -0400]
Access request	"GET / HTTP/1.1"
Status	200
Bytes	10801
Referrer URL	"http://www.google.com/search?q=log+analyzer&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a"
User agent	"Mozilla/5.0 (Windows; U; Windows NT 5.2; en-US; rv:1.8.1.7) Gecko/20070914 Firefox/2.0.0.7"



Queryable Change Data

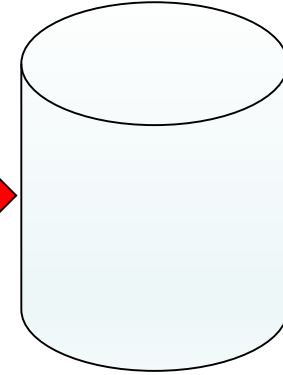
Event table with
date columns



```
SELECT ...  
FROM <EventTable>  
WHERE <event-cond>
```

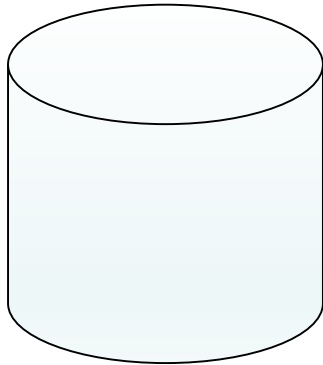


Recent events
table

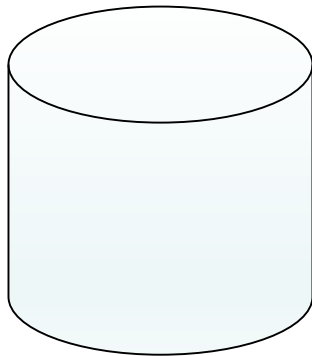


Snapshot Change Data

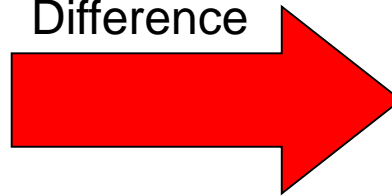
Previous Source File



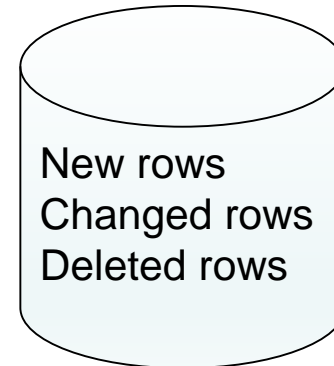
Current Source File



Difference



Delta



Data Quality Problems

- Multiple identifiers
- Different units
- Missing values
- Text data with different components and formats
- Conflicting data
- Different update times



Summary

- Change data used in data integration
- Understand source system requirements and processing level for each type of change data
- Data quality problems more prevalent with legacy systems

