Multi-year Data Capture Session Notes

Thursday, April 12, 2018

2:15 PM - 3:15 PM	Expanding Multi-year Data Capture: 1 year later
	Ballroom AB
	A year ago, the Alliance launched a set of development activity to expand the ability to capture multi-year data. Through this effort and other effort in the community, we have also benchmarked a number of "temporal" data capture technologies. This research will be presented in the context of what kinds of temporal data management each solves.

> ⚠ These broad session notes attempt to capture the spirit of the discussion and should not be interpreted as a transcript. Although Ed-Fi Alliance staff were involved in capturing these observations, the notes below should not be construed as official, complete, or 100% accurate.

Presentation by Chris Moffatt

The slide deck will be posted. These notes will just cover highlights and discussion.

Key question: incorporate T-ODS work in to the Core, or release on the Exchange? Need community feedback on this. Looking for 6 members to participate in special interest group.

Loading temporal data

- Generally need to be able to export Ed-Fi supportable (XML?) from source systems and bulk load into the ODS instead of using API
- Could use API on a clean database to load old data in stages, building up one year after another

Questions on SQL Server

- Using SQL Snapshots? Part of the import process can work with Snapshots if you have the license for it, otherwise using backup and restore.
- Thus when speaking of T-ODS snapshots, it is not equivalent to SQL Server Snapshot.

Temporal API

- For now, deferring on building a public API for updating temporal data
- But there is a proof of concept of an API
- Source systems pushing to API probably don't support temporal parameters anyway, so bulk loading would be needed

Some existing models

- Wisconsin
    - Separate ODS for each year
    - Collection_ODS - load each year into the Collection instance
        - Run validation on it
        - Publish snapshots for reporting
- Arizona
    - Ed-Fi ODS-like database, with addition of date range columns
    - Removes SQL keys to allow for duplication on natural keys
- Data Vault architecture (Certica) - immutable snapshots
    - Table-valued functions bring back "Ed-Fi-like" data structures)
    - Load data marts from data vault as needed
    - Each year clear out your ODS, but the old data are in the vault (for example)
    - Could move data more frequently to get more granularity, but there is a cost

SQL Server System Versioned Temporal Tables

- Prefer to avoid more vendor lock-in
- T-ODS records snapshot of old and current value in a table, but SQL Temporal only has old data in the temporal table and current data in the main table.
- You could turn on SQL Temporal without much effort
- T-ODS lets you decide what the range is - you can set an Effective End Date on your own. This is based on whenever you decide that you want a snapshot. Effective End Date could thus be null.
- Whereas SQL Temporal automatically captures when the data change.