## *2011-03-25 - The LinkedIn Bitemporal Data Group*

### Topic: entity integrity (EI) and temporal entity intgrity (TEI). Part 2.

Before I turn to temporal referential integrity, let me emphasize that we have not discussed the bitemporal form of TEI. So far, we have ignored transaction time. The bitemporal form of TEI is easily stated, but perhaps not so easily understood. It is this:

Within any pair of clock ticks, one in valid time and one in transaction time, at most one row may represent one object.

This says that, at any point in time, we cannot assert more than one statement about an object at any point in time. The first point in time is a point in transaction time or, as we call it, assertion time. The second point in time is a point in valid time or, as we call it, effective time.

For example: I cannot claim, on 3/23, that the unit price for product P123 on 3/22 was $25, and also claim, on 3/23, that the unit price for product P123 on 3/22 was $22. That is a contradiction. Both statements can't be true.

Of course, I can claim, on 3/23, that P123 was priced at $25 on 3/22, and was priced at $22 on 3/21. In this case, we can see that there was a price increase from the one day to the next.

On the other hand, I can claim on 3/23 that P123 was priced at $25 on 3/22, and then claim on 3/24 that P123 was actually priced at $22 on 3/22. In this case, we can see that we corrected our earlier statement.

In this example, we can see the two temporal dimensions of bitemporality at work. The non-temporal statement "P123's unit price is $22" can describe P123 at different points in time. But once we put this statement into its first temporal envelope -- that of valid time -- then we have a statement like "P123's unit price, on 3/22, is $22".

Can we have two or more statements about P123's unit price on 3/22? We can as long as we claim, of only one of them, that it is true. In a bitemporal table, the rows that we currently claim are true are the ones whose transaction time is current, i.e. the ones whose transaction begin dates are in the past and whose transaction end dates are in the future, specifically set to 12/31/9999. Transaction time is a second temporal envelope within which we can put statements already located in a valid time temporal envelope. (The transaction time envelope can also contain non-temporal statements, but that's a (short) story for another time.)

In the example above, when we inserted the P123 on 3/22 row on 3/24, we gave it a transaction time period of [3/24/11 - 12/31/9999]. On 3/23, the P123 on 3/22 row had a

 

transaction time period of [3/23/11 - 12/31/9999]. As part of the logical unit of work that inserted that new row, we also overwrote the transaction end date of the previous row, ending that assertion on the same clock tick that we began its correcting assertion.

The result is two rows whose transaction times are, respectively, [3/22/11 - 3/23/11] and [3/23/11 - 12/31/9999]. These rows satisfy TEI because they do not share a single clock tick in transaction time even though they specify the same valid time clock tick for P123.

These rows also record a progression in time that is orthogonal to the progression in time recorded by valid time. A series of rows for the same object, on the same clock tick of valid time, is a series of corrections to mistaken or otherwise deprecated data. A series of rows for the same object, on the same clock tick of transaction time, records a series of states through which the object passed.

We began by discussing TEI within the context of valid time only. We have now extended our discussion of TEI to show how it works in bitemporal time. In bitemporal time, we have rows in our tables which tell us, not just what things were like at various points in time, but also what, at various points in time we said things were like at various points in time.

Next time, I'll move on to temporal referential integrity.

 