Data Virtualization Reduces Synchronization Volumes

Data synchronization evolved roughly in parallel to data integration. As soon as there were two applications that were inter-dependent from a data perspective, the data had to be shared (integration) and the shared data needed to be the same in both applications (synchronization). The problem was addressed several ways over time, initially with bulk uploads, later by attempting to rewrite all applications to use a shared database, and then with messaging systems such as publish/subscribe technologies. All kinds of techniques such as Change Data Capture (CDC) have been devised and the technical approaches are improving slowly, but certainly not keeping up with the growing need most businesses. Finally, with the adoption of Data Virtualization (DV), we are seeing the beginning of a dramatic wave of probably the greatest improvement ever in reducing synchronization volumes, particularly where bi-directional DV is supported. With the decreased volumes also comes significant latency reduction, fewer databases to support, and overall simplification of the infrastructure.

Root Causes

It may be obvious what drives the need for data synchronization. In order for an application, dashboard, or service to use data from multiple different sources, the solution has historically been to design a database for the application and write code or use utilities or one-to-one data integration and transformation tools to populate and periodically update the database. While the application may update the database with new data, some of that data may be needed for other applications. Right now, with all the dashboards, business analytics, cloud apps, and custom applications, most companies have a proliferation of these for-purpose data stores/marts. More than any IT manager readily acknowledges.

Data Virtualization Comes of Age

Data Virtualization is readily available now with powerful but easy to use tools to assist in configuring integrations across multiple disparate sources without ever moving the data. Any of the applications that now have for-purpose databases can access the data directly from the multiple sources, aligned appropriately, as if it were in a physical database designed just for the application. In the case of Enterprise Enabler®, the virtual data model is built implicitly with design, testing, and deployment in a single Integrated Development Environment. The data can be packaged with the push of a button as ODBC, JDBC, OData, SOAP or REST web services, and for SharePoint 2010, or 2013 BCS Custom Connector access. Full CRUD capabilities are generated, so that the application can also write data securely back to the original sources. Enterprise Enabler is the only tool that handles write-back natively, without considerable custom coding.

What's Different when Data Virtualization Enters the Picture?

With the above picture in mind, it is clear that many, if not most, of the for-purpose databases can be eliminated. Synchronization is simply no longer necessary. Updates made in the application are written directly back to the sources. Data generated by the application can be propagated back live not only to the sources, but can also be updated or written to any other data stores that use the data, thus eliminating the need for synchronizing that change multiple places.

In the end, the impact can be tremendous. Imagine a source of record populating all the copies of the data automatically whenever it is generated or changed. Latency is all but eliminated, and the security of the data is heightened by not making copies. Cloud apps that access data on-demand via Data Virtualization can get on-premise data from behind the firewall without the data ever moving.