



In-Memory or Live Data: Which is Better?

AUTHOR:
Ellie Fields, Director Product Marketing,
Tableau Software

DATE:
July 2011

The short answer is: both. Companies today are using both to deal with ever-larger volumes of data. For the business user, analyzing large data can be challenging simply because traditional tools are slow and cumbersome. Some companies are dealing with this by investing in fast, analytical databases that are distinct from their transactional systems. Others are turning to in-memory analytics, which lets users extract set of data and take advantage of the computing power on their local machine to speed up analytics and take query load off their transactional systems. This is especially important for SQL Server, which is often a critical part of the IT environment.

So which approach is better? There are times when you need to work within the comfort of your own PC without touching the database. On the other hand, sometimes a live connection to your SQL database is exactly what you need. The most important thing is not whether you choose in-memory or live, but that you have the option to choose.

Let's look at some scenarios where in-memory or live data might be preferable.

In-Memory Data is Better...

When your database is too slow for interactive analytics

Not all databases are as fast as we'd like them to be. If you're working with live data and it's simply too slow for you to do interactive, speed-of-thought analysis, then you may want to bring your data in memory on your local machine. The advantage of working interactively with your data is that you can follow your train of thought and explore new ideas without being constantly slowed down waiting for queries.

When you need to take load off a transactional database

Your database may be fast, it may be slow. But if it's the primary workhorse for your transactional systems, you may want to keep all non-transactional load off it. That includes analytics. Analytical queries can tax a transactional database and slow it down. So bring a set of that data in-memory to do fast analytics without compromising the speed of critical business systems.

When you need to be offline

Until the Internet comes to every part of the earth and sky, you're occasionally going to be offline. Get your work done even while not connected by bringing data in-memory so you can work with it right on your local machine. Just don't forget your power cord or battery—you'll still need that!

A Live, Direct Data Connection is Better...

When you have a fast database

You've invested in making your SQL Server implementation blazing fast. Why should you have to move that data into another system to analyze it?

You shouldn't. Leverage your database by connecting directly to it. Avoid data silos and ensure a single point of truth by pointing your analyses at a single, optimized database. You can give business people the power to ask and answer questions of massive data sets just by pointing to the source data. It's a simple, elegant and highly secure approach.

When you need up-to-the minute data

If things are changing so fast that you need to see them in real-time, you need a live connection to your data. All your operational dashboards can be hooked up directly to live data so you know when your plant is facing overutilization or when you're experiencing peak demand.

And the Best Choice? Both, of Course.

Even better is when you don't have to choose between in-memory and live connect. Instead of looking for a solution that supports one or the other, look for one that supports choice. You should be able to switch back and forth between in-memory and live connection as needed. Scenarios where this is useful:

- You want to use a sample of a massive data set to find trends and build your analysis. You bring a 5% sample of the data in-memory, explore it and create a set of views you want to share. Then you switch to a live connection so your reports are working directly against all the data. Publish your views, and now your colleagues can interact with your analysis and drill down to the part of the data most relevant to their work.
- You're flying to New York and want to do some analysis on the plane. You bring your entire data set, several million rows, into your local PC memory and work with it offline. When you get to New York, you reconnect to the live data again. You've done your analysis offline and in-memory, but you are able to switch back to a live connection with a few clicks.

Flexibility to Choose How You Work With Data



Direct connect to fast databases



In-memory accelerates slow databases

Neither in-memory nor live connect is always the right answer. If you're forced to choose, you'll lose something every time. So don't choose—or rather, choose as you go. Bring your data in-memory, then connect live. Or bring recent data in-memory and work offline. Work the way that makes sense for you.

The Tableau Data Engine

Tableau's Data Engine provides the ability to do ad-hoc analysis in-memory of millions of rows of data in seconds. The Data Engine is a high-performing analytics database on your PC. It has the speed benefits of traditional in-memory solutions without the limitations of traditional solutions that your data must fit in memory. There's no custom scripting needed to use the Data Engine. And of course, you can choose to use the Data Engine, or don't—you can always connect live to your data.

Data Engine to live connection—and back

The Data Engine is designed to directly integrate with Tableau's existing "live connection" technology, allowing users to toggle with a single click between a direct connection to a database and querying an extract of that data loaded into the Data Engine (and back).

Architecture-aware design

The core Data Engine structure is a column-based representation using compression that supports execution of queries without decompression. Leveraging novel approaches from computer graphics, algorithms were carefully designed to allow full utilization of modern processors. The Tableau Data Engine also has been built to take advantage of all the different kinds of memory on a PC, so you can avoid the common limitation that data sets must fit into your computer's RAM memory. This means you can work with larger data sets.

True ad-hoc queries

The Data Engine was designed with a query language and query optimizer designed to support typical on-the-fly business analytics. When working with data at the speed of thought, it is common to need to run complex queries such as very large multi-dimensional filters or complex co-occurrence queries. Existing databases generally perform poorly on these types of queries, but the Data Engine processes them instantly.

Robin Bloor, Ph.D. and founder of technology research firm the Bloor Group, writes: "Because Tableau can now do analytics so swiftly and gives people the choice to connect directly to fast databases or use Tableau's in-memory data engine, it has become much more powerful in respect of data exploration and data discovery. This leads to analytical insights that would most likely have been missed before. "

Flexible data model

One of the key differences of the Tableau Data Engine compared to other in-memory solutions is that it can operate on the data directly as it's represented in the database on disc. So there's no required data modeling and no scripting that needs to be done to use the Data Engine. One of the things that's so powerful about the Data Engine is you can define, just as with any other relational database, new calculated columns or you might think of it as sort of ad hoc data modeling at anytime.

Instance load and connection time

The Data Engine is unique in that once your data is loaded into the Data Engine, it has very fast start-up time. It only needs to read in that portion of the data which your queries actually touch. You might have a lot of data in the database that's not relevant to a particular analysis, but you are never going to wait for the Tableau Data Engine to read that data.

“Because Tableau can now do analytics so swiftly and gives people the choice to connect directly to fast databases or use Tableau’s in-memory data engine, it has become much more powerful in respect of data exploration and data discovery. This leads to analytical insights that would most likely have been missed before.”

- Robin Bloor, Ph.D and founder of The Bloor Group

The Customer Perspective

National Motor Club uses Tableau to understand membership behavior, reduce costs and do a variety of other analyses. Matt Krzysiak, COO of National Motor Club, said, “Our largest data set has about 70 million rows. Generally speaking we’re working with data sets with hundreds of thousands or just a few million. We use SQL server as the back end database throughout our organization, and we lay Tableau on top of that. It seems to work very well for us.”

“We used to do analysis the old fashioned way. We would ask the IT department for information, and they would give us results, usually through customer queries. ”

“With our previous tools trying to get an answer was a very iterative process that took lots of time. We would ask IT or an analyst for information, wait for the results, and look at the results. What used to take us hours, or days, or even weeks, now in Tableau with drag-and-drop features we can just rip through that stuff very quickly.”

Recommendations

Because of the variety of ways people need to work with data, we recommend that when choosing a business intelligence solution:

- Ensure it provides the option to take data in-memory to speed up your analytics.
- But don’t limit yourself to a solution that requires you to bring all data in-memory before you can analyze it.
- Evaluate how easy it is to switch back and forth between in-memory and live connect: if you need to do custom scripting, it may take away most of the flexibility you got from your choice.
- Examine carefully the data size limitations of your in-memory solution. Choose an in-memory solution that lets you deal with very large data sets.

¹ “Analytics at the Speed of Thought,” Robin Bloor, Ph.D, July 2011.