Posted by Jim Tommaney
Created on Tue 13 December 2011 08:14 - Last updated Tue 14 February 2012 17:06

Since the launch of InfiniDB last year, we've been seeing InfiniDB enabling tremendous customer successes. We've seen hundreds of customers use InfiniDB to power their most impactful analytics projects.

We'd like to describe InfiniDB's architecture and explain what makes it so scalable, fast and simple. In a three-part blog series, we'll be covering how InfiniDB differs from other database systems.

Compared to relational databases, InfiniDB has three key benefits: I/O, parallelism and ease-of-use.

First, let's start with I/O. Traditionally, the bottleneck in database processing has been I/O. When you're moving large data volumes, small differences in I/O start to add up. You can imagine the phrase "being bitten to death by ducks"; a growing series of small I/O penalties make traditional relational database systems unusable for analytics on large datasets (i.e. typically data volumes over 500 GB or more, often described as "big data").

Traditionally, database technologies have found ways to alleviate -- but not fix -- the pain. For example, Netezza (an IBM company) allows for scaling the scan rate to overcome the I/O bottleneck. However, such solutions remain costly as they require large investments in proprietary hardware.

We set out to change all of that with InfiniDB.

One of the most impactful ways to alleviate the I/O bottleneck is to align the way that data is stored with the way that it's used. For analytics, this suggests 'columnar databases' or column-stores. Unlike a traditional row-based database, which store data in rows ("First Name", "Last Name", "Age") columnar databases store data in columns (i.e. for the column "Age": '32', '44', '65').

For analytics, specific columns tend to be pulled frequently (i.e. Average of Ages '32', '44', and '65') which makes columnar databases typically the tool-of-choice. And, since columnar databases like InfiniDB can be deployed on commodity hardware, this solution tends to be much less expensive as well.

Once installed, our customers often can't believe the performance that they gain with InfiniDB. They're even more excited when they learn that InfiniDB is priced by the core and doesn't tax their growing data volumes.

What happens when your data volumes increase further and you need to scale out? Stay tuned for our next post on Parallelism where we describe InfiniDB's industry-leading scale-out functionality.