MOBILE

BIG DATA
R E P U B L I C | Transform Your Business With Data

ABOUT US   REGISTER   LOGIN

HOME   BLOGGERS   MESSAGES   POLLS   WEBINARS   RESOURCES   VIDEOS | STRATEGY   ANALYTICS   **TECHNOLOGY**   FB   TW   LI   G+   RSS

FINANCIAL SERVICES   HEALTHCARE   GOVERNMENT   EDUCATION   RETAIL

BIG-DATA PLATFORMS & TOOLS

In-Memory Databases: Fact or Fiction

Robert Plant, Associate Professor, School of Business Administration, University of Miami
3/6/2013
Comment
2 comments

Login

50%   50%

The modern era of "in-memory" databases dates back to the early 1980s when the Transputer (from a conjunction of "transistors" and "computers") was designed in the UK by David May and Robert Milne, and manufactured by Inmos.

Splitting tasks for over 30 years
The goal of the Transputer was to overcome the limitations of sequential processing, whose performance ability was plateauing. Their approach was visionary: to develop chips that would be able to break a problem down into pieces and perform computations in parallel. Then, the different parts of the problem could be run on different processors, potentially on different machines, but overall working together to optimize the resource utilization and achieve very fast computational results.

This was a non-trivial task and a quantum leap from the consumer computing world of the time, which was marveling at running WordStar or Word Perfect on 80286 Intel processors! However, the approach also needed a new theoretical base. This was pioneered by researchers at the Programming Research Group, Oxford University, headed by Professor Sir Tony Hoare FRS, which led to the development of the process algebra mathematics for Communicating Sequential Processes (CSP), the specification system for the a new concurrent programming language, Occam.

Unfortunately, like many great steps forward, the world was not quite ready, and those of us who were graduate students at the PRG during this period can attest to the difficulties of developing code on this platform, even in a lab setting. As a result, the technology went into deep "research-only" mode for several decades, only to reemerge recently as "in-memory" databases (IMDB) when the entire ecosystem needed to support its operational performance was significantly more mature. Key to this was the ability to manipulate databases in columnar forms, closer to Ted Codd's original intent, rather than in the rows that have dominated since the 1980s.

Analyzing on the fly
The commercial success of Inmos was unfortunately limited, but with the fall in price of DRAM and the increased performance of processors, the IMDB approach has made their construction attractive for the mainstream vendors. A leading proponent is Hasso Plattner, who at Sapphire Now in 2012 stated, "New applications will have to deal with big data. We have to analyze it on the fly, so we have to have a system that is transactional and analytical at the same time. We cannot have a multi-stage system. This is too slow for modern applications." He said: "You're not going to wait for more than three seconds for something."

However, he also noted "that we cannot abandon 99 percent of our customers who are sitting on [IBM] DB2, Oracle, NextDB, Microsoft, etc. SQL is not dead. The world is living on SQL. If you pull SQL out of the enterprise world, the world will break down." SAP's answer is HANA, which in 2012 was demonstrated at Sapphire as a system capable of working on a 100Tb in-memory database running on IBM System x3850 servers.

Oracle is also in the game, even if, as reported in PC World, Larry Ellison in 2012 famously asked for the name of the pharmacist used by the folks at SAP if they don't think you need a relational database, and that "this is nonsense, there is no in-memory technology ready to take the place of a relational database." However, in the same presentation he states they have the "leading in-memory database TimesTen."

MORE BLOGS FROM ROBERT PLANT

A Review: Big Data in 2013
39 comments
Data scientists, NSA, C-suite buy in... Big data changed a lot in the last 12 months.

What Would Seymour Cray Think of Big Data?
13 comments
How would the tech legend respond to the 4 Vs?

The 360 Degree View of a Customer
56 comments
Big data is helping companies like Netflix see what you really want (and even when you go to sleep).

Lessons for Big Data From President Obama's Healthcare Implementation
24 comments
Three big data lessons for CEOs from HealthCare.gov's bumpy start.

Is Big Data Ruining Sports?
29 comments
From sailing to the NBA -- big data could increase the gap between sporting haves and have-nots (while ruining the magic in the process).

More from Robert Plant

FLASH POLL

All Polls

DIGITAL AUDIO

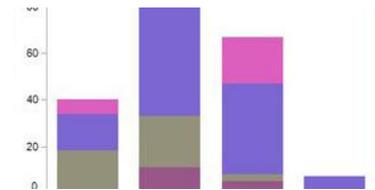### Big-Data Platforms & Tools Radio
Latest Archived Broadcast
Who Can See From Your Platform?
We discuss the best internal setups to make sure the right sets of eyes are watching the data development.
Listen

Full schedule   Archived shows   Download to calendar

DATA VISUALIZATION SHOWCASE

Visualization: What's Stopping Big Data Improvement?

Survey results show bigger companies still confident in spite of massive IT overhaul requirements. Explore this data here.

More Data Visualization Showcase

BDR IN YOUR INBOX

Enter email   SUBMIT

One of their current products aimed at the analytics market is the Oracle Exalytics appliance, which supports ten compute cores with four Intel Xreon E7-4800 processors with 1TB of memory and which can be clustered.

Of course, the picture would not be complete without Microsoft, which in late 2012 announced its Hekaton in-memory OLTP will be a part of its 2014 or 2015 SQL Server Database, along with its Hadoop for Windows product, HDInsight Server. Hekaton is designed to optimize the use of the most frequently accessed tables in the database by keeping them in-memory and the rest in a traditional database solution-architecture, thus speeding up the process 100X (Hekaton is Greek for hundreds).

So, the signposts are clear (even allowing for usual divertary smoke screens by Mr. Ellison): In-memory databases using column tables are on their way; they will be big enough to capture the biggest databases and will ultimately take over. As such, CIOs and managers need to start to prepare for this eventuality. They can do this in three ways:

- First, set up some computer science-style continuing education for your tech group, focused on columnar relational database theory. The big firms won't be offering this too soon, as they need to gear up and align their take on the theory with their product offerings. As in all of computer science, it's best to learn the theory yourself and apply it to the problem; that way you obtain insight from a neutrals perspective.
- Second, obtain a small in-memory device, and play with it. The technology is robust enough that with training, your team can commence building prototype systems. This will allow the team to grow in confidence with the technology for when the larger tasks are to be built out, and simultaneously lower the consulting bill for those deployments. Vendors are also starting to offer application developer boot camps, and these are fast ways to ramp up on the technology, build relationships, and network.
- Third, this area is moving fast; as such it is highly recommended that your data team attend the industry conferences of all the major players in this area, as well as user groups such as the SAP HANA Technology Meetup Groups.

While Inmos and the programming language Occam have faded into the past, in-memory database developers of the future would do well to live by William of Ockham's principle "Pluralitas non est ponenda sine neccesitate." (If he were texting that thought today to a big data developer, he would probably have written "Keep it simple.") In other words, don't over-complicate things unnecessarily.

Related posts:

- Hadoop 2.0 & Beyond: Reinventing Hadoop
- Hadoop 2.0 & Beyond: Bypassing MapReduce

— Robert Plant, *Associate Professor, School of Business Administration, University of Miami*

Email This   Print   Comment

COMMENTS

Newest First | Oldest First | Threaded View

Edwin Willems, User Rank: Exabyte Executive
3/16/2013 | 5:42:02 AM

Login

👍 👎
50%   50%

Re: In-Memory Databases: Fact or Fiction
Robert, great to start a topic on in memory databases. I'm so surprised about the few comments on this topic. However I think it's getting overhyped, and many people will think so what? There's however one aspect that I find pretty interesting wrt in memory databases, and that's "associative". While most vendors just talk about in memory databases as a performance thing, I see other companies such as QlikTech to talk about in memory not as a performance thing in the first place, but rather as a capability to start doing things differently, in their case associative. Why is it that most vendors don't take it beyond just performance...?

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Saul Sherry, User Rank: Blogger
3/6/2013 | 10:41:19 AM

Login

👍 👎
50%   50%

Passionate Displays
@Robert, interesting to see this is a concept which has been floating around for so long.

Why do you think there is such vehement disregard for In-Memory process sharing from certain quarters?

Reply | Post Message | Edit/Delete | Messages List | Start a Board

BIG DATA REPUBLIC