STARCOUNTER VMDBMS TECHNOLOGY

Background

New generations of enterprise databases and other data engines will play an important part of our future. The fundamental principles of database technology have roots from the early 1970s. This is good and bad. The good thing is that database technology has become a science with solid mathematical grounds and is well researched. The bad thing is that many of the underlying assumptions are no longer true.

The ongoing RAM revolution

Databases use computer memory to store things. Random access data stored in RAM is 100 000 times faster than to random access of data stored on disk. The main media for the last 40 years has been the disk drive. The reason for this is not, as many believe, that RAM loses its data when power is cut; as disks can also fail. The real reason why RAM is not the primary operating medium of databases has always been cost and availability. When the SQL database was born, one megabyte of RAM was over 160,000 US dollars. Now the same amount RAM is about 1 cent. For data security and recovery, enterprise databases have transaction logs. Such logs can efficiently be written to disk since sequential writes do not require the mechanical arm of the disk drive to move and writing speed can be faster than the rate of change in a RAM database.

The first fix – big database caches

What do you do as a database vendor when RAM is suddenly cheap and a 64-bit CPU can address it in tera bytes? As you cannot reinvent and rewrite your product in an instant, the easiest option is to increase the size of your caches. Today, many traditional databases operate entirely in RAM if enough of it is available. While speed is much faster than if the RAM was not present, the entire architecture of the DBMS has not change and the overhead that used to consume a fraction of the execution time when the cache was small now consumes almost all the time.

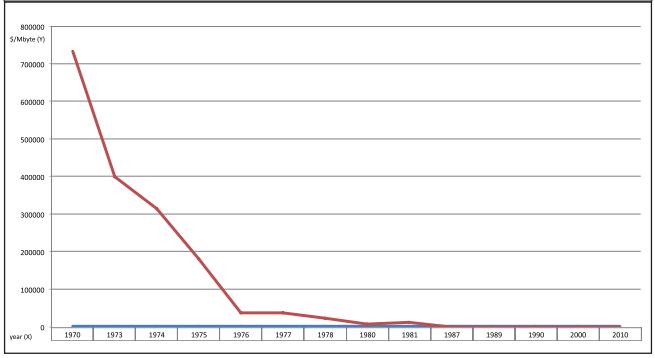
The first generation RAM databases

While increasing the cache speeds things up it will not match the speed of a true RAM database. If you adapt your database engine and assume that the data is always in RAM, you can make a database engine much faster than a traditional database. A RAM database is typically 10 times faster than a big-cache database. Leading vendors include IBMs SolidDB and Oracles TimesTen.

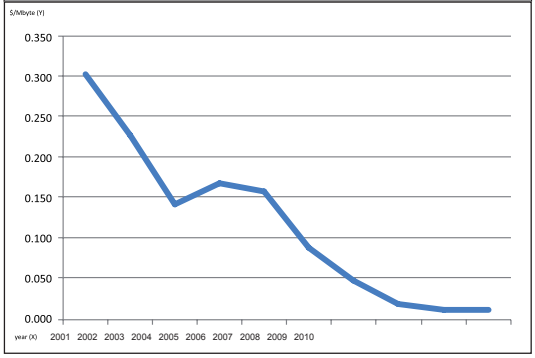
The second generation RAM databases

The modern computer is nowadays a complex cluster of CPUs and memory. Taking advantage of the NUMA architectures where all memory is not equal can radically improve the performance of the first generation RAM databases. A CPU cache is a hundred times faster than normal RAM memory. Taking advantage of this and fully exploring the physical differences between RAM and Disk, a second generation RAM database can be 10 times faster than a first generation database and a hundred times faster than a big-cache database.

RAM prices through the decades



AM prices in the last ten years



The third generation ram database

If a second generation RAM database can be a hundred to a thousand times faster than a traditional database, a new opportunity opens up. If the database is as fast as your object heap in your computer language, why would you move data between the database and the heap when you are operating on your objects, rows or tuples? When we are measuring database operations in nanoseconds instead of milliseconds, this is not an unlikely question to ask.

One reason why you would like to move data from the database to the object heap of languages such as Java or C# is that you have a snapshot copy of the data. Nobody will see your changes until you decide to save them or serialize them. The downside is that query languages such as SQL will not be able to sort or query your local changes together with the database data until they are saved, making coding on your business object more difficult. But what if your computer language used the database as the heap? Transaction isolation would make the changes invisible to others until you wanted to make them public. This is the A and I in ACID - atomicity and isolation. So you keep the effect that you can work on a transaction isolated from other transactions, but you gain the possibility to see your changes in your SQL queries. If you are adding a line item to a purchase order with existing items and have not yet saved your changes, you can still use SQL to sort all your rows, new and old. While in an atomic change, different threads of Java or C# object code would see the same object in different states at the same time.

The need for moving data back and forth using serialization and deserialization would be redundant. The solution would be super easy for the developer and a lot faster than the second generation RAM databases. Such a merger between the object heap of the virtual machine (VM) and the transactional control by the database management system (DBMS) we will refer to as a VMDBMS.

The VMDBMS defined

The bottleneck apparent in second generation RAM databases is the transferring of data between database and application code. Because data is physically stored only in one place, the VMDBMS eliminates data transfers between application and the database as well as transformation between different data formats. This is possible because the application can access data in the database as fast as its internal temporary data. The data is never moved, the data always resides in the database all

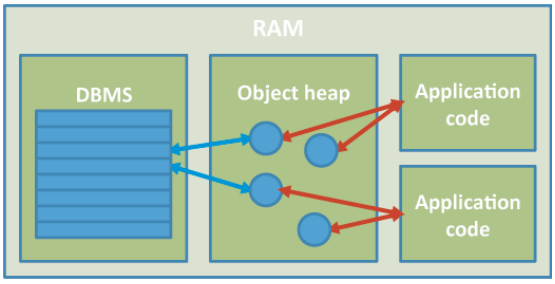
```
public class Quote : Persistent {
   public Person Who;
   public string Say; }

public class Person : Persistent {
   public string Name;

   public static void HelloWorld() {
      Transaction( delegate {
         Person p = new Person();
         Quote q1 = new Quote();
         Quote q2 = new Quote();
         p.Name = "Albert Einstein";
         q1.Say = "E=mc²";
         q2.Say = "Make things as simple " +
                  "as possible, but not simpler.";
         q1.Who = q2.Who = p; } );

      foreach ( Quote q in SQL(
         "SELECT q FROM Quote q WHERE Who.Name=?",
                              "Albert Einstein" ) )
         Feedback.WriteLine( q.Say ); } }
```

the time, and application directly accesses the data managed by the database management system. There is no copy whatsoever of the data local to the application. The “local” state of change in progress is no longer a product of a separate Java or C# object on the heap, but rather a consequence of the isolation and atomicity of the ACID engine managing the joint heap/database. As long as the log(s) are secured on disk, the database image can calmly be hibernated to disk in its own pace using asynchronous writes. Checkpoints and recovery works the same way as in your legacy database.



VMDBMS queries

A VMDBMS uses an enterprise database engine as a heap for languages such as Java and C#. As such, you can expect to be able to run queries on your objects. You class is a table and your class instances are your rows. Inheritance work in the way you expect them to. As opposed to environments where the database is remote your queries will have very little overhead and you can run millions of them per second on a single node. Also inheritance and path expressions (person.City.Country.Name) can be natively supported instead of having to be translated to a relational database resulting in poor performance.

Lock free ACID concurrency

As the VM and DBMS are merged, the database transaction can automatically restart a transaction that is in conflict with another transaction. This is known as optimistic concurrency control. Optimistic concurrency control does not compromise of concurrency control, but it benefits from the fact that if conflicts are unlikely, it is better to deal with them only if they occur. This is difficult to do if the code running the transaction is separate from the database engine. In the VMDBMs the code can declare a transaction scope (see above) and the transaction can automatically be restarted. In a third generation database, transaction running time is around a thousand times faster than using Java or C# on a remote database. This means that pessimistic concurrency control, which is based on locking and slows the database down even if there is no conflict can be avoided.

A bad performance spiral

The fuel of a space rocket is burned to produce thrust to accelerate mass of the rocket out of the gravitational force of the Earth. The more energy that is needed, the more fuel it takes. The more fuel it takes the more energy is needed. You are in a bad performance spiral. Transactions can be said to work in a similar way. The slower the transaction, the more you need to rely on pessimistic concurrency control (locking) as conflicts are more likely. The more the engine relies on pessimistic the slower the transactions, the slower the transactions, etc.

A good performance spiral

If the database is fast enough, it can be used as a plug in heap manager meaning that the data of the object is not copied to the memory of Java or C#. This means that transactions become much faster. This reduces concurrency conflicts. Also, as the language and the DBMS can cooperate, the DBMS can automatically restart transactions such that lock free optimistic concurrency schemes can be used. This also means faster transactions. The faster the transaction, the less likely the conflict. The less likely the conflict the more you can rely on optimistic concurrency control, meaning faster transactions. You are in a good self feeding performance spiral.

A billion objects

There is much more to a VMDBMS than to provide persistent objects. The normal Java or C# heap lacks the capabilities to host large number of objects. Already when they reach the millions, the memory manager and garbage collector begins to struggle. There is no transaction scope, so that changes are not atomic and transactions are not isolated. Your threads needs to rely on locking to be thread-safe, a method that is significantly slower than optimistic concurrence control. One of the most important features however is the possibility to query your database using query languages such as SQL and that you have the safety features of checkpoints or recoveries and other enterprise database features.

Where should the VMDBMS be used?

The availability of memory has been increased dramatically. This can be somewhat offset by the increase in information volumes. However, when you compare the number of people, products and sales transactions on the planet, and given the fact that the name ‘Bill’ still requires four character of storage, for many information systems, the availability of cheap computer memory outweighs the growth in entities to store information about. It is rather the real time and online exposure of the information that has changed dramatically in terms of structured storage. Unstructured information that is not transactional has grown much more rapidly with the birth of the internet than has the structured data involved in ACID transactions. The VMDBMS is ideal for applications such as ERP systems, online applications, finance and other areas were structure and transactions are needed.

Einstein’s general theory of relativity

In third generation RAM database is that a database access is measured in nanoseconds. The implication is that physical distance between interacting nodes becomes very important. It takes light three nanoseconds to travel one meter. As Einstein has showed, this is an absolute limit for electricity based or light based computers. By writing your software, be it an ERP system or a cloud service, the VMDBMS makes sure that the speed of light works in your favour.