

# *The Beginner’s Guide to In-Memory Technology*



**By: Elad Israeli, Founder at SiSense**

## **Table of Contents**

- Why In-Memory Technology Is Important for Business Analytics..... 2
- In-Memory Technology Basics..... 2
- In-Memory Associative Technology ..... 3
  - Illustration: Disk/RAM utilization when querying 2 fields..... 4
- Elastic In-Memory Technology ..... 5
  - Illustration: Disk/RAM utilization when querying 2 fields..... 5
- Comparison Table – RDBMS, Associative and Elastic In-Memory ..... 6
- A Case Study for In-Memory Technology: Large Tables (WeFi) ..... 7
- A Case Study for In-Memory Technology: Disparate Sources (Target) ..... 8

## Why In-Memory Technology Is Important for Business Analytics

According to Gartner Research, one of the biggest issues with Business Intelligence (BI) adoption is poor query performance. Business users want immediate results using fast and responsive applications, and anything longer than a typical Google search is simply too long to wait.

Query performance is a challenging issue because analytics queries are executed frequently, and by multiple people simultaneously. This takes a toll on modest hardware resources.

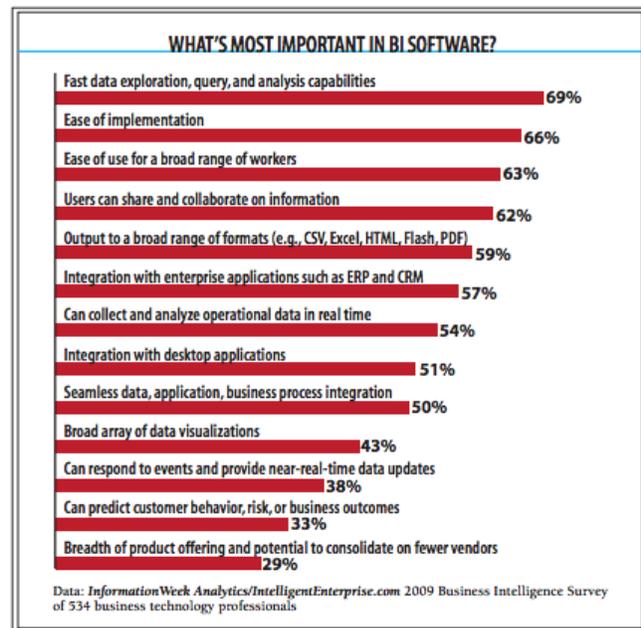
The challenge is compounded by the fact that business requirements constantly change and are generally unpredictable. It is nearly impossible to build a database that will perform well regardless the type of query business users will throw at it.

Traditionally, the prominent type of database used for storing, manipulating and querying data was known as “RDBMS” (Relational Database Management System). Examples of such databases include Oracle MySQL, Microsoft SQL Server and IBM DB2.

The RDBMS's wide ecosystem and significant backing by the software giants both drove it to be extensively deployed as the back-bone of BI/analytics solutions. But RDBMS's were originally designed to efficiently store and manage transactions (rows) on a disk drive, which has proven to perform poorly when used to execute typical analytic queries involving grouping, aggregations and large tables joins. RDBMS simply access the disk too often to be able to process those queries fast enough.

Several methodologies were developed to work around this limitation, including the creation of a data warehouse with aggregations, star schema models and/or pre-calculated OLAP cubes. Yet, over time these methodologies have been proven to be complex, costly and could only partially answer business requirements because of their lack of flexibility.

Ultimately In-Memory technology became a great alternative approach for delivering best performance on analytical queries while requiring fewer resources to deploy.



## In-Memory Technology Basics

Generally speaking, a computer has two types of data storage mechanisms – disk (often called a hard disk) and RAM (random access memory).

The important differences between them are outlined in the following table:

DISK	RAM
Abundant	Scarce
Slow	Fast
Cheap	Expensive
Long Term	Short-Term

Most modern computers have 15-100 times more available disk storage than they do RAM. My laptop, for example, has 8GB of RAM and 300GB of available disk space. However, reading data from disk is much slower than reading the same data from RAM. This is one of the reasons why 1GB of RAM costs approximately 320 times that of 1GB of disk space.

In a disk-based RDBMS, there are two things that cause heavy disk operations and therefore poor performance:

1. **Table Scans:** Loading of an entire table from disk to RAM (for calculations)
2. **Complex Data:** Querying data scattered across many tables and/or fields (joins)

In-Memory technology aims to address both these issues and by that provide out-of-the-box query performance faster by orders of magnitude than traditional RDBMS. By doing that, this technology enables a much faster time to value and significantly less effort and money invested in developing, setting up and maintaining analytics infrastructure.

All In-Memory technologies share the same premise that it is simply much faster to perform calculations over data that is stored in RAM than it is when that same data is stored in a table on a disk. These technologies also benefit from the fact that 64-bit computers today are a commodity and that it is relatively cheaper to add more RAM to a computer today than it used to be.

## In-Memory Associative Technology

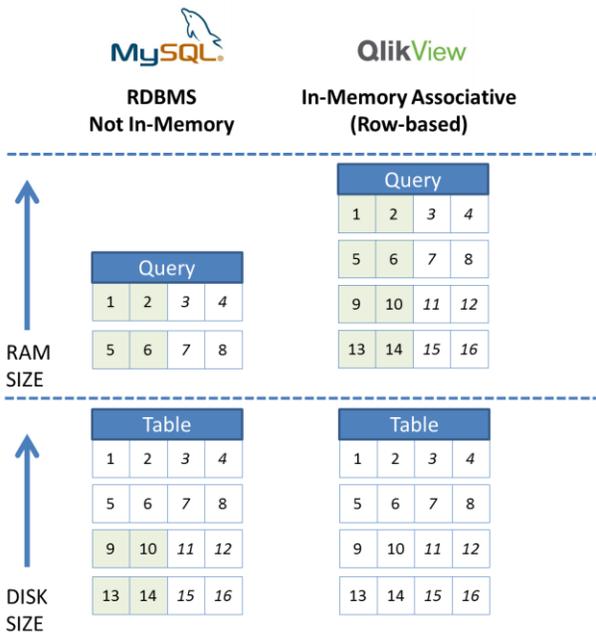
In-Memory Associative\* technology, by QlikTech, uses RAM-based storage instead of disk-based storage to avoid disk reads during query processing. It also compresses the data, and applies associations to avoid costly join operations in certain queries.

In-Memory Associative technology performs beautifully, at small scales. When datasets are simple and small, it enables speedy development compared to a solution built on top of an RDBMS. In-Memory Associative technology represents the first generation of In-Memory technologies to be introduced into the business analytics mainstream.

However, its main inhibitor to wide enterprise adoption has been scalability. The challenge it continues to face is that RAM, when used to store and analyze raw business data, tends to run out quickly and unexpectedly. As storage sizes go, RAM is tiny and many data sets these days are too large to fit. Moreover, each query to the database uses up additional RAM for intermediate calculations.

Complex scenarios still require that data be extensively modified, or even loaded into an RDBMS data warehouse, prior to being loaded into the memory-based storage. This can happen when data sets are complex and/or when there are many users querying the database simultaneously and repeatedly. In such cases, the added value of such technology is debatable and cost-saving benefits of using it become less significant.

### Illustration: Disk/RAM utilization when querying 2 fields



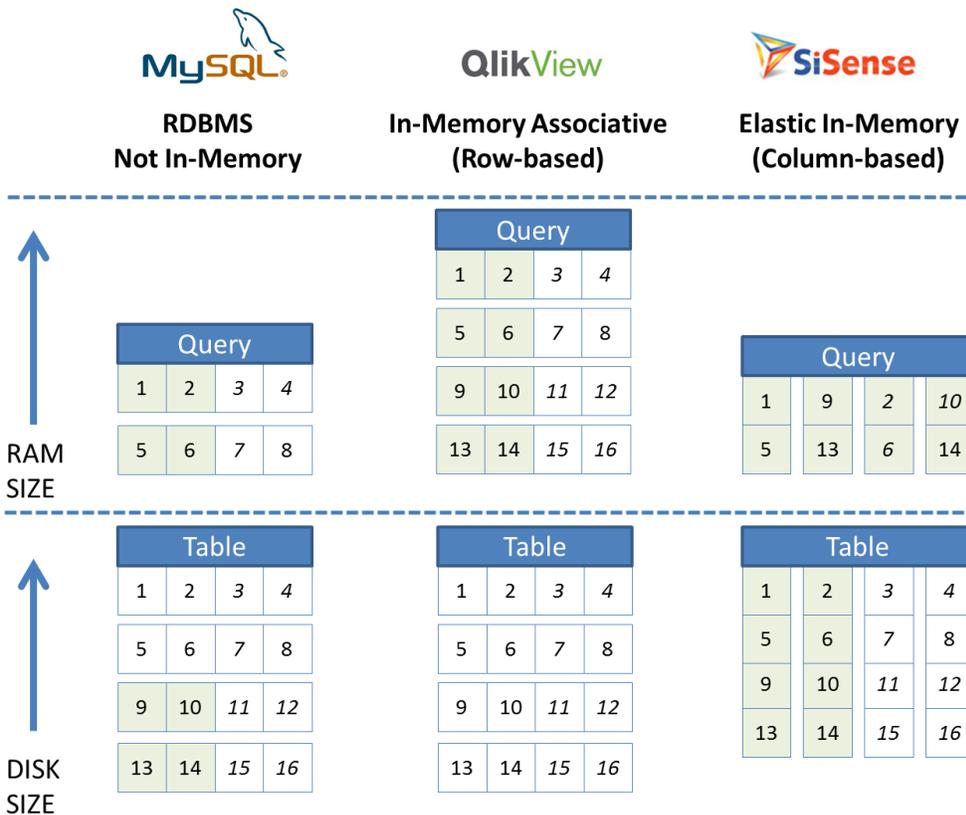
## Elastic In-Memory Technology

Elastic In-Memory is the latest generation of In-Memory technology for business analytics and sets itself apart by being fast as well as scalable. The name Elastic comes from its unique ability to stretch beyond the hard limitations imposed by older generation technologies.

It employs a disk-based columnar database for storage to provide fast disk reads and is able to load data from disk to RAM (and vice versa) when it is needed. The queries themselves are processed entirely in-memory without any disk-reads throughout. And most importantly, there is only a subset of the data physically stored in RAM at any given time, leaving more space for other operations to take place in parallel.

Elastic In-Memory technology also has a unique way of handling joins. Instead of joining tables, it uses columnar algebra to merge between fields. This way, the join operation can be processed entirely in-memory, fast.

### Illustration: Disk/RAM utilization when querying 2 fields



## Comparison Table – RDBMS, Associative and Elastic In-Memory

The table below compares between RDBMS technology, QlikTech's In-Memory Associative technology and SiSense's Elastic In-Memory Technology by a set of several technical aspects:

- **Columnar Storage:** whether the technology supposed storage of columns rather than tables.
- **In-Memory Query Processing:** whether the technology typically requires reads from disk during query execution
- **Out-Of-The-Box Performance:** Fast query response to queries involving joining, grouping and aggregating data – without custom work or specialized configuration.
- **Data Capacity:** Is there a cap on data capacity beyond what can be stored on a single hard disk (TBs of data).
- **Scalability Level:** The ability of the technology to support growing data volumes and concurrent usage without having to significantly modify/re-build the solution.

Feature	RDBMS SQL, Oracle, DB2	In-Memory Associative QlikView	Elastic In-Memory SiSense
Columnar Storage	Some	No	Yes
In-Memory Query Processing	No	Yes	Yes
Out-Of-The-Box Performance	Slow	Fast	Fast
Data Capacity	Unlimited	Limited (By size of RAM)	Unlimited
Scalability Level	Large Scale	Small Scale	Small/Large Scale

## A Case Study for In-Memory Technology: Large Tables (WeFi)

WeFi provides technology for optimizing and managing Wi-Fi networks. For mobile operators, WeFi acts as a data offloading tool in overloaded cellular networks; end users are better able to find and enjoy optimal wireless connections around the world. WeFi's tools allow its user community to map the global Wi-Fi network with minimal effort; over 75,000,000 access points have been identified around the world.



Millions of WeFi users were generating some seriously big data—the database team dealt with an average table size of more than 5 million rows for behavioral data, and over 500 million rows for activity records of some of the most active clients. Reporting on this activity was essential to helping the company's management team make decisions.

WeFi needed to perform advanced analysis on large amounts of data in three categories:

1. The behavior of millions of WeFi users, including retention activity and data acquisition activity (13 tables in one model, with average table size of more than 5 million rows)
2. The performance and activity of wireless networks to which its users are connected (the average table has more than 70 million rows)
3. The activity records of active clients (average table size now exceeds 500 million rows)

WeFi's RDBMS database team had been manually running SQL queries, but they struggled with performance to generate the reports that gave the management team crucial feedback.

WeFi carried out its initial Elastic In-Memory implementation by loading data as-is into a single data model; additional stores were added to the server later on. A dedicated PC machine serves as the central server, and today the WeFi team runs the majority of its business reporting and dashboards using Elastic In-Memory technology, using a total of six data stores. Business users today have the autonomy to create custom dashboards and reports, which dramatically reduces the time it takes to get new data insights; lightning-fast query response times mean answers to key strategic questions are available faster. Meanwhile, the database team can focus on crucial tasks without spending hours writing and running SQL queries.

More case-studies at <http://www.sisense.com/case-studies>

## A Case Study for In-Memory Technology: Disparate Sources (Target)

The Target Corporation, based in Minneapolis, MN, serves guests at 1,755 stores in 49 states and at Target.com, and also operates 37 distribution centers nationally. As a publicly-owned, global corporation, Target has a global presence with a secondary headquarters in India and sourcing offices around the world. Target also has more than 350,000 team members worldwide and is consistently recognized as an employer of choice.



With such a recognizable brand, it's only natural that Target would want to protect its customers and its revenues with theft- and fraud-prevention technologies that could scale to meet the challenges posed by a worldwide retail network. The key to understanding and reducing theft was in knowing where it was most likely to happen and which products were most likely to be stolen.

Theft-related information was available from several unconnected sources, but without a central database or tools to analyze, Target's management feared they might not be making the most of their data. The team tasked with finding a solution identified three major goals: Since organizing and aggregating data from various locations drained time and ate up the data team's resources, Target's business intelligence solution would have to reduce the workload associated with pulling theft data from multiple sources.

Far more theft-related data was available than the team really needed, so simplifying the information after it had been aggregated in one place was the next step. Theft-prevention measures were designed around very basic information: what is being stolen, and where is the theft occurring?

Theft prevention is a company-wide challenge, not one the data team would try to solve by itself. Because multiple departments in different locations would use theft information at various points, the team needed a solution that could standardize data consolidation, analysis, and distribution.

Fed up with the complexity of multiple channels, wasted resources, and indecipherable reports, Target's team started looking into business intelligence options that fit their needs and goals. After initial research, tools of different technologies were evaluated for long-term implementation, including: RDBMS, RDBMS/OLAP and both Associative and Elastic In-Memory technologies. Eventually, the tool running Elastic In-Memory technology was selected due to its ability to consolidate data from multiple sources and the ability to run fast multi-source reports that can then be shared with numerous users.

More case-studies at <http://www.sisense.com/case-studies>

\* QlikTech, QlikView and In-Memory Associative are trademarks of QlikTech International AB  
\*\* SiSense, Elastic In-Memory and ElasticIQ are trademarks of SiSense Ltd.  
\*\*\* QlikTech International AB is not associated with this document.