The Associative Difference™

Freedom from the limitations of query-based tools

September, 2017

Table of Contents

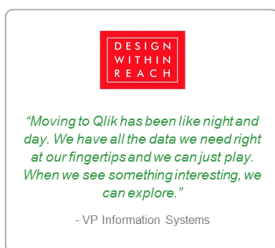# Introduction

This white paper describes the unique associative technology that underpins Qlik® products. It provides an overview of the Associative Difference™ - the value that only Qlik can deliver, outlines the differences between Qlik's Associative Engine and query-based BI tools, and explains how users of all skill levels can improve the effectiveness of their analysis to achieve a deeper, more complete understanding. The target audience for this paper includes business users and analysts, and technology decision makers who are evaluating and comparing data analytics software products.

Qlik's Associative Difference™

Simply put, relational databases and SQL queries were not designed for modern analytics. While it's true that SQL is required to pull data from many sources, most analytics tools depend on SQL and query based approaches as their fundamental architecture for modeling data and supporting interactivity. This is a major flaw – resulting in restricted linear exploration and analysis on partial subsets of data. Data sources must be brought together using SQL joins, and assumptions must be made in advance about what types of questions users will have. All other data is left behind. If a user wants to pivot their analysis based on something they discover, they will likely have to re-build complex queries, which often means going back to more experienced data experts. We call this the "ask, wait, answer cycle". Every new type of question has a waiting period.

DESIGN WITHIN REACH

"Moving to Qlik has been like night and day. We have all the data we need right at our fingertips and we can just play. When we see something interesting, we can explore."

- VP Information Systems

The Qlik Associative Engine is designed specifically for interactive, free-form exploration and analysis. It fully combines large numbers of data sources and indexes them to find the possible associations, without leaving any data behind. It offers powerful on-the-fly calculation and aggregation that instantly updates analytics and highlights associations in the data, exposing both related and unrelated values after each click. This means people are free to search, explore, and pivot based on what they see, without limitations and without having to go back to experts and wait. That is why Qlik users consistently discover previously unforeseen insights which have been missed by query based tools, driving tremendous value.

That's the Associative Difference - which only Qlik can deliver.

<div style="sidebar">

Qlik at a Glance

Founded in Lund, Sweden in 1993

Approximately 45,000 customers in more than 100 countries

Solutions driving leadership in the visual analytics market

- **Qlik Sense®**
  Self-service data analytics
- **Qlik Analytics Platform®**
  Custom and embedded analytics
- **QlikView®**
  Guided analytics and dashboards
- **Qlik NPrinting®**
  Centralized reporting
- **Qlik GeoAnalytics®**
  Mapping and geographic analysis
- **Qlik Sense Cloud®**
  Visual analytics online
- **Qlik DataMarket®**
  Third party data-as-a-service
- **Qlik Connectors®**
  Broad data connectivity options

More than 1,700 technology, solution, OEM, consulting and system integrator partners

More than 2,000 employees worldwide

Recognized leader by industry analysts including Gartner, BARC, Forrester, Butler Analytics, and Ventana Research

Headquartered in Radnor, PA, USA

</div>

Query-based tools – limitations and blind spots

Visualization tools that rely on queries to analyze information ultimately create blind spots in understanding. These tools use a structured, linear approach to visualize partial subsets of data, instead of offering the freeform exploration and search across all data that is offered by Qlik's Associative Engine. Often these tools attempt to simulate Qlik's exploratory experience, but they quickly hit limits in terms of flexibility and performance. The only way for these tools to answer questions that were not anticipated is to re-build complex SQL queries, which is difficult for the average business user and usually requires an experienced data expert.
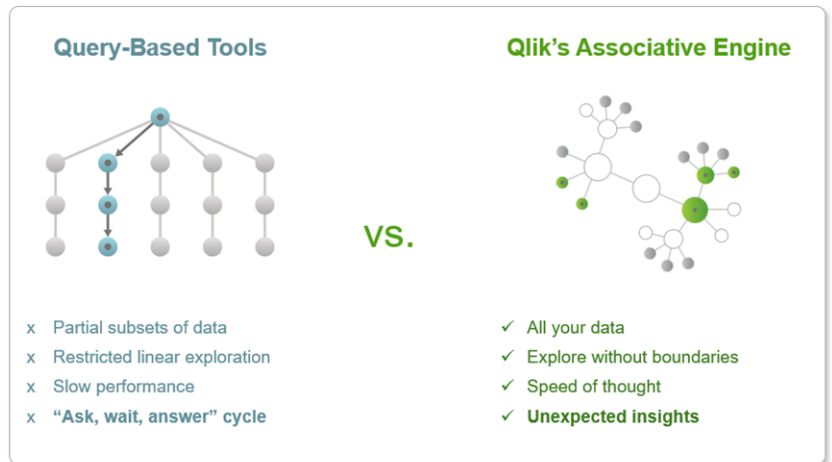
The drawbacks of query-based tools

Partial views

Query-based tools either connect directly to underlying sources, or utilize some form of a data staging area. SQL joins must be used to combine sources, resulting in data loss or inaccuracy – especially when combining large numbers of sources. Data warehouses and staging areas must be fully modeled in advance, requiring significant effort. And to support user interactions, queries are run for every click / question, limiting flexibility and confining users to partial subsets of data.

A global IT and networking provider brought together 500 million records of complex customer data from multiple systems, and allowed their large salesforce to freely explore customer portfolios to find the best cross-sell opportunities for maintenance and additional products. The company generated 100 million in support renewals and 4 million in cost savings, and greatly improved customer satisfaction.

No context

Query based visualizations are discrete, disconnected entities, that don't stay in context with one another. This means that filtering a single visualization doesn't show the relationship or impact that selection has on the other objects within the app (or dashboard). Some tools try to create a unified context by allowing objects to be 'wired' together using front-end code, but this results in a lot of queries running at the same time and performance issues – the approach doesn't scale.

Limited flexibility

Visualizations and queries must typically be built by power users, limiting everyone else to the pre-conceived questions the developers thought of. There is no way for business users to search and explore to ask new questions, beyond simple filtering within an object. The only choice is to wait for a data expert to build a new query or visualization.

Data Loss

If information is loaded from multiple sources, query-based tools can risk data loss due to joins executed at load time. These tools typically require a primary data source to be defined, and only bring in subsets of data from secondary sources that match the values in the primary source. This can result in data loss, which business users may not even be aware of.

Incorrect results

Query based tools also run the risk of incorrect calculations, as values can be double or even triple counted if queries and joins are not defined correctly and data is denormalized. Getting things right typically requires users to have strong familiarity with the data model and skills to properly structure queries.

Poor performance

Query-based tools are at the mercy of the database(s) that support them, and no matter how robust the database, queries take time to execute. The problem is further compounded when wired dashboards execute multiple queries at the same time. As more and more concurrent users try to execute more and more queries, underlying databases simply can't keep up. A slow and unresponsive system disrupts a person's train of thought and they lose their ability to connect ideas and take next steps.

An analogy

Let's say the goal is to understand how an internal combustion engine works. With a query-based tool, you would see some of the individual parts of the engine in isolation. You would be able to evaluate one part at a time, but would be left on your own to attempt to understand the relationships (associations) between the parts, and how they fit together as a cohesive whole to create a working engine.

With Qlik's associative technology, however, you have the ability to analyze a complete working engine, with each part having a relationship to all of the other parts. You can tweak the throttle (make a selection) and see how this affects the fuel intake, carburetor and exhaust. You can watch the pistons pump and turn the crankshaft. You can deconstruct the engine at your leisure and look at each part in the context of the parts next to it. This is the power of Qlik's Associative Engine.

The Associative Difference means that when users look at information, they know precisely how it is related. If they want to narrow the context down to a particular set of selections, they can immediately see how the rest of the data across an entire application responds. They are not limited to seeing only a subset of the data contained in a query result set, or restricted in how they explore it.
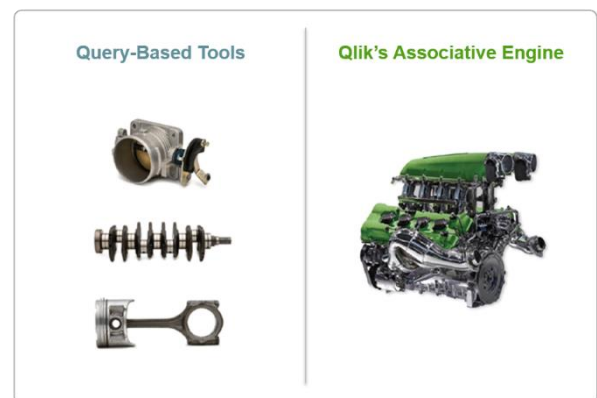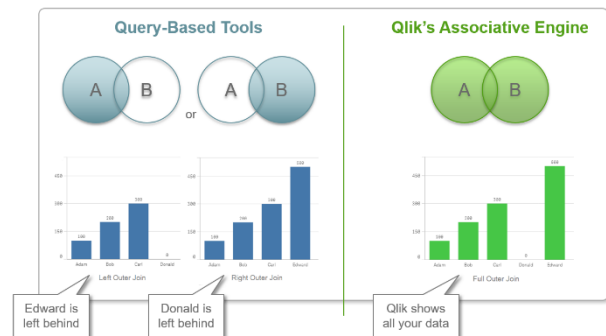
Qlik's associative technology A to Z

Qlik delivers an associative user experience that allows users of any skill level to explore and refine context through simple searches and selections, starting anywhere and going anywhere. Every time the user clicks, the Qlik engine instantly responds, dynamically recalculating all analytics based on the new context and highlighting associations in the data across all data sources. This speed-of-thought feedback encourages users to think of new questions and continue their individual paths of exploration and discovery.
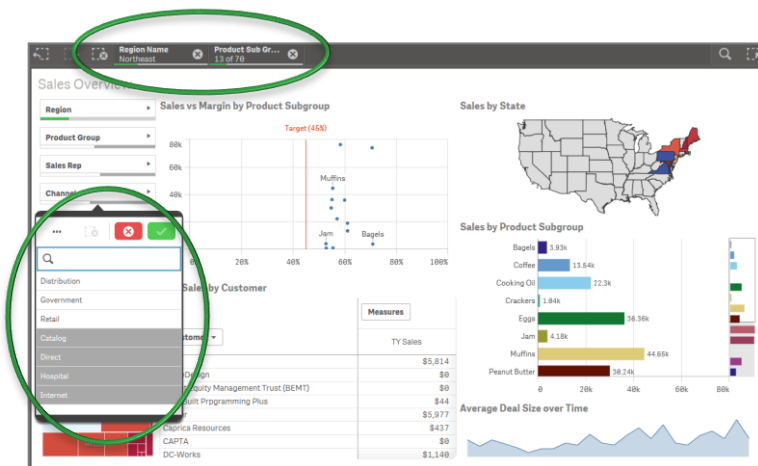
Understanding associations

An association is simply a relationship between a data value and another. For example, a product may have been sold in a certain country but not others. This product would be associated with the country it was sold in, and unrelated to the others. With Qlik's associative technology, if you select the product it will appear in green, the associated country will appear in white, and the unrelated countries will appear in gray.

Associations can be positive in nature — a value that is related to another, or negative — a value that is unrelated to another. It is important to note that the unrelated (gray) values provide as much insight as the positive (white) ones – often indicating new opportunities or areas of risk.

Selections appear in green, associated values in white, and unrelated values in gray


All data values have associations with others in the data set. For example, a product could be associated to the customers that bought it, the channels it was sold through, and individual transactions it was included in. Associations can work in any direction, extend across multiple relationships, and apply to combinations of values — such as regions associated with a set of selected products and countries.

Users won't always know what associations exist or which are important in advance. A query-based tool could easily report that a product was sold in a region. However, noticing that this product was not sold through certain channels in the geography may not be as easy. And by missing critical insights like this, users only get part of the story.

All your data

Qlik's Associative Engine fully integrates data from multiple sources, without suffering data loss from executing SQL joins at load time. The Qlik Associative Engine achieves what is technically known as a "many to many full outer join", meaning that users have access to all their data from all their sources without leaving any data behind. This allows people to get a complete understanding without blind spots or inaccurate insights due to missing or double counted data.

For more insight on bringing together data from multiple sources, check out the blog post entitled "Equal Rights For Your Data" on the Qlik blog.
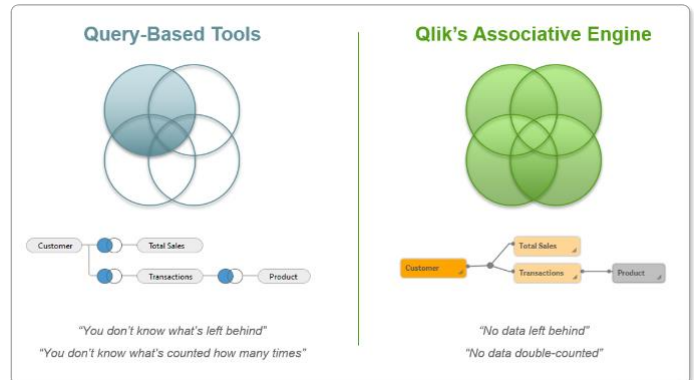
## Explore without boundaries

Associative Exploration

Qlik offers an associative user experience that allows people to explore freely, across all their analytics, without restrictions or boundaries. Users can interact anywhere, within visualizations, charts, graphs, filter panes, even a global selections interface. After each click, the Qlik engine instantly recalculates all analytics and highlights associations in the data. And because the engine dynamically calculates based on a full set of record-level data, users can pivot in their analysis to new ideas or data, making any selections they want, at any level of detail, without being limited by predefined questions or hierarchies. The process builds on itself as users ask more questions, add more context, and become more informed at every step along the way.

1 Ask a question
   – Use interactive selections and keyword searches
   – Ask any questions, not just pre-defined ones
   – Interact with any visualization or chart, in any order
2 Get instant feedback
   – All analytics and KPIs dynamically recalculated
   – Associations revealed to the user in green, white, and gray
   – New context (selection state) applies across the entire app
3 Evaluate results, make discoveries, and formulate 'the next question'
   – Spot insights in all surrounding visualizations
   – Understand what values are associated with current selections
   – Think of follow-up questions based on what you see
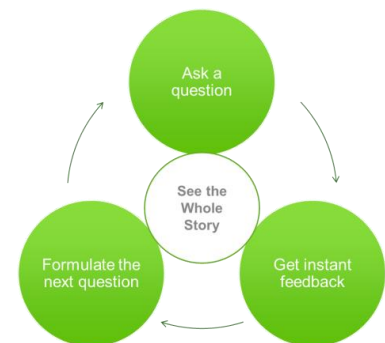
An example of exploration and discovery
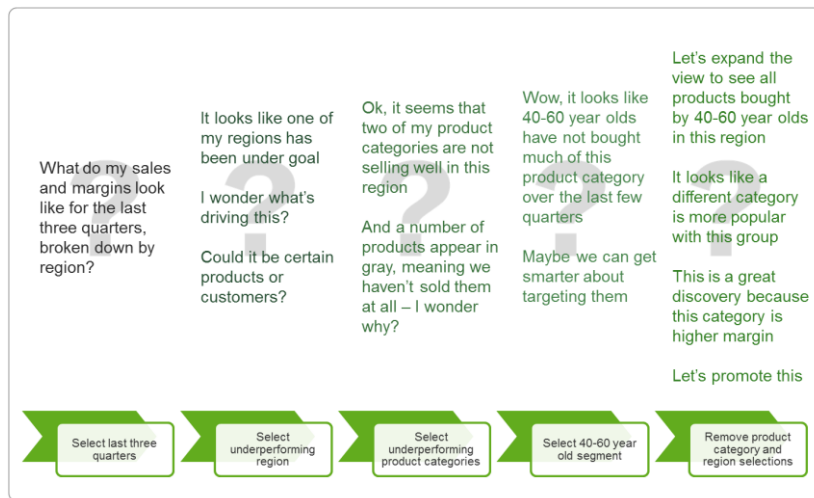
In this example, the user starts with an open-ended question and ends up asking a series of questions that lead to a better understanding of the business as a whole. Insight is generated at every step along the way, with key insight about a particular customer segment in a certain region ultimately revealed, prompting the user to take action. The next user might start with a different question and end up taking a different path, but either way, questions are answered without the need to build additional queries or visualizations. The value adds up as more users make more discoveries across the business.

Smart search

Smart Search allows for a simplified approach to asking questions, when a person does not necessarily know where to look for the right information. People can search across all their data using keywords, and will get immediate, ranked feedback on where their values match. Furthermore, if multiple values are entered, results will include not only matching dimensions but also associations that exist between the values, ranked intelligently based on strength of association. And search also includes the metadata within charts and graphs themselves, returning visual thumbnails for easy navigation to the right analytics. This powerful capability provides an easy way for people to ask questions and get immediate insight, broadening access and value to even more users.
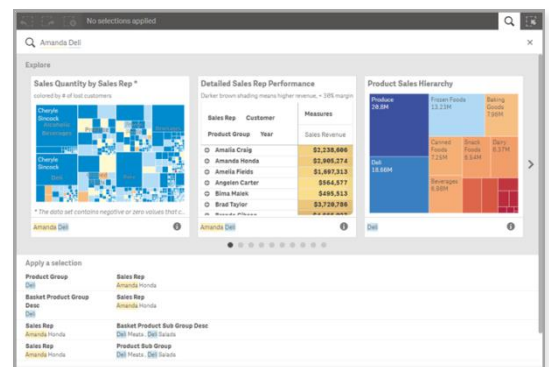
The power of gray

With Qlik's associative technology, users can see not only the values that are associated with their selections, but also those that are unrelated. These values appear in gray, and often convey the most impactful insights such as areas of new opportunity or risk. For example, if you select a set of products, you may notice there were certain customer segments these products were not sold to. This capability is unique to Qlik and a great way to find unexpected insights.

> *A leading global bank and investment firm discovered nearly 20 million in mortgage pipeline that was not associated with any loan processor. These mortgages showed up 'in the gray'. The firm immediately pursued the lost pipeline, generating tremendous value from a single discovery that was not possible with query based tools.*

AGGREGATE
INDUSTRIES

"We discovered that some of our perceptions of who were our best customers conflicted with the information we started to see in the Qlik applications. Now, we can see who is paying consistently."

- CIO

BUTLER ANALYTICS

"… What really distinguishes Qlik® Sense is its extensibility and the associative data engine, which gives new meaning to the word 'discovery.'"

- Butler Analytics, Qlik Sense and Tableau Positioning, June 2015

bics

"With Qlik's Associative Search capabilities, our customers are now able to intuitively analyze over a terabyte of data with 'speed of thought' response, without being confined to a limited data set or a defined path of questions."

- Data Warehouse Manager

Context is key

Qlik's associative technology maintains a single context for all analytics across an entire application. When a user interacts with an object, making a selection or search to refine context, all analytics and data relationships are instantly updated to reflect this new context. This allows a user to explore across all their visualizations, at different levels of detail, at the 'speed of thought', to spot potential areas of interest and get a sense of where to go next.

No restrictions

Qlik's associative technology does not place any restrictions or boundaries on the user. People are free to explore and search all their data, in any direction they want, probing possible data relationships and key areas of concern as they follow their own paths to insight. This flexibility is critical when looking at information from many angles, and at many different levels, to ultimately get a better understanding of the situation as a whole.

Speed of thought

The performance of query based tools is at the mercy of underlying databases, which are generally not designed to support interactive, modern analytics. Inevitably queries will be slow, and this poor performance will result in delays which derail thinking.

The Qlik Associative Engine dynamically calculates analytics and highlights associations as quickly as the user can think of questions. This means business users can get the answers they need without having to wait - for the system or for other people - resulting in deeper insight and ultimately better decisions.

The reason our engine can provide unmatched speed and flexibility is simple – it's what it was designed to do. The Qlik Associative Engine combines several unique and patented approaches including compressed binary data storage, logical inference, and dynamic calculation, to deliver 'speed of thought' response for high numbers of users, analyzing large, complex data sets, who are asking questions that are not pre-defined. It is our core technology advantage, with over 15 years of innovation and investment, delivering value to over 45,000 customers worldwide.

Qlik's associative technology vs. query-based tools

The Associative Difference, the unique value only Qlik products can deliver, is summarized in this table.

	Query-based tools	Qlik's Associative Engine
All Your Data	Can't combine large numbers of different data sources	Combine any number of disparate data sources
	Must fully model data and hierarchies in advance	No need to fully model data or hierarchies
	Joins at load time leave data behind (or double count)	No joins at load time means no data left behind
	Can't handle dirty / incomplete / sparse data	Dirty / incomplete / sparse data is loaded and exposed
	Often require pre-aggregation for large data sets	No pre-aggregation, transaction detail always available
Explore Without Boundaries	Cannot support non-linear exploration / new questions	Built for interactive, free form exploration
	No global search or search based analysis	Global search for data, associations, and analytics
	Multiple objects don't stay in context without wiring	All objects stay in context together automatically
	Loss of unrelated (gray) values	Unrelated (gray) values exposed in analysis
	Static hierarchies, no on-the-fly calculations	Dynamic hierarchies with on-the-fly calculations
Speed of Thought	Cannot keep up with a user's thinking	Provides 'speed of thought' analysis
	Slow performance, especially with many users / queries	High performance dynamic calculation for large # users
	Need to query production systems or build DW	No effect on production systems, no need for DW
	Risk exposing sensitive data if database is not secure	Dynamic data reduction for data security
	Limited capabilities for big data sources	Scalability for big (and small) data sources

Additional resources

- For more information and to download Qlik products, head to qlik.com.
- To experience Qlik's associative technology in action, check out our online demos.
- To get started, register to use Qlik Sense Cloud® for free.

Appendix: Let's geek out on the technology

Qlik users experience The Associative Difference from Qlik products which is only possible because of the unique capabilities of the patented Qlik Associative Engine. This powerful calculation and data indexing engine is Qlik's core advantage, with over 15 years of innovation and investment. It is built specifically to support interactive, free-form exploration and analysis, offering high-speed dynamic calculation and 'speed of thought' response for high numbers of users and data both big and small. In this appendix we will cover the key aspects of how our engine works and what makes it different.

Compressed binary indexing

Data assembly

The Qlik Associative Engine integrates a full set of record level data from multiple sources into its in-memory engine. Qlik provides a robust set of data preparation and integration capabilities for transforming and bringing together disparate data sources, including visual interfaces for loading and transforming data, smart data profiling of relationships and values, and powerful scripting for complex data integration scenarios. And because table joins are not executed at load time (they are performed dynamically as the user explores), all data from all sources is retained. The Qlik Associative Engine achieves the equivalent of a many to many full outer join, without suffering the data loss associated with one-sided SQL joins executed at load time.

For more insight on bringing together data from multiple sources, check out the blog post entitled "Equal Rights For Your Data" on the Qlik blog.

Binary indexing

Data is indexed and stored in a highly optimized, in-memory compressed binary format that optimizes performance and drives the associative user experience. Data relationships are managed by the engine, defined through common column names in the multi-table model. Instead of storing values repetitively, the engine creates binary pointers for each unique value, storing the actual values only once. Analytics are not pre-calculated, meaning that any calculation can be performed on demand — leading to near limitless flexibility in terms of both exploration and aggregation. And because table joins are executed by the engine dynamically as the user explores, calculations will always occur on the correct tables and the correct set of values, avoiding the risk of incorrect results.

This unique columnar, binary indexing capability is the foundation for the Qlik Associative Engine supporting interactive, associative data exploration and on-demand calculation, across high numbers of concurrent users and large data sets.

Logical inference and calculation

As high numbers of users explore information, it is a challenge to provide each of them with the flexibility and 'speed of thought' response they need. Users constantly interact, making selections on the fly, searching for new information, and executing complex calculations on different subsets of data that cannot be predicted in advance. The Qlik Associative Engine handles this difficult task through a two-step process every time a user interacts with an application.

Logical inference

Step one is logical inference — which essentially means determining data associations relative to current context. Before a user selects anything, all that data is in play. However, each time a user makes a selection, the engine immediately calculates what distinct values in all related tables are associated to the new context. By doing this, the engine can highlight for the user the important relationships in the data, revealing both associated and unrelated values. In addition, this process allows the engine to minimize the data needed to support subsequent calculations, maximizing performance. Logical inference means the Qlik Associative Engine always knows which data in which tables to use, and always takes the fastest path to the right data.

Dynamic calculation

Step two is dynamic calculation. Once the associated data set has been determined by logical inference, the Qlik Associative Engine calculates all analytics and aggregations in real-time — storing the results in a cached hypercube. Calculations are broken into pieces and performed on various tables as needed, and a number of advanced techniques are used to maximize performance. Every user interaction triggers the engine to recalculate the values stored in the hypercube, providing the user with answers specific to their unique questions. And because data is highly compressed and optimized in-memory, and data sets for calculations are minimized by logical inference, the calculation is extremely fast.

For more information on Qlik's patented engine technology, see the white paper entitled "Interactive Data Exploration With An In-Memory Analytics Engine", written by Mike Ferguson from Intelligent Business Strategies.

Queries can't keep up

Achieving this combination of flexibility and performance is extremely difficult if not impossible with query-based tools. There is simply no way to provide 'speed of thought' response for high numbers of concurrent users, analyzing large, complex data sets, who are asking questions that are not pre-defined. Even query-based tools that claim to have in-memory technologies for supporting ad-hoc queries are still limited by the very same query structure they rely on.

The unique combination of compressed binary storage of data, logical inference, and dynamic calculation allows the Qlik Associative Engine to make a highly complex task incredibly simple and transparent for the user. Our engine has evolved over fifteen years of innovation and investment, delivering unprecedented value to over 45,000 customers.