WHITE PAPER

INTELLIGENT BUSINESS STRATEGIES

Interactive Data Exploration Using a *Context-Aware* Analytics Engine

By Mike Ferguson
Intelligent Business Strategies
March 2020

Prepared for: Qlik LEAD WITH DATA

Table of Contents

SELF-SERVICE VISUAL ANALYTICS TOOLS – HOW THEY WORK

Interactive dashboard and visual data discovery tools have had the greatest business impact in recent years

In last five years we have seen broad adoption of interactive self-service BI tools in many companies aimed at line of business and departmental business analysts so that they can explore data on their own and quickly produce insight without the need for IT.

These tools offer a lot of capability. This includes access to multiple data sources, self-service data preparation, in-memory data caching, query and reporting, interactive ad hoc analysis and visualisation, development of dashboards and storytelling and server-side publishing of BI content for consumption via web browser and mobile device. The focus of self-service BI has been to shorten time to value while also trying to drive up consumption of BI across the enterprise. Yet, despite its success, advances are still occurring to further speed up development, to open these tools up to lesser skilled users and increase the percentage of employees across the enterprise consuming BI to better manage and optimise every-day business operations. Before we look these advances, let's first look at how self-service BI products work.

Interactive dashboard tools with visual data discovery capability can pre-load data into memory

Figure 1 shows two different architectures for self-service BI tools. Type 1 is an interactive visual data discovery and exploration tool that provides the ability to integrate data from one or more data sources and pre-load it into memory via some type of ETL or scripting capability. This approach needs to ensure advanced data compression to enable rapid loading of millions of data points. Data sources can include real-time streaming data, data at rest in one or more relational and/or big data stores or a combination of both. It may not be possible to load all data from sources such as a data lake in cloud storage, a Hadoop system or a NoSQL database due to data volume. Nevertheless, it should still be possible to access this data efficiently in a seamless way. Once the data is in memory, analysts and power users have the freedom to explore, analyse and visualise data in any way they want in order to identify new insights. The most important insights created can then be brought together to create managerial dashboard 'applications' that offer up integrated, actionable information to monitor key aspects of the business. Once built, in-memory data makes it possible to share these dashboard applications among large numbers of concurrent business users who can then interact with centrally managed, governed and trusted data at the speed of thought. A key capability with this type of architecture is that consumer users still have significant power to analyse data beyond the capabilities set out in dashboard applications that have been built for them as all the data loaded into memory is still available. Therefore, conducting further 'non-linear' analysis beyond the application to gain deeper business understanding remains possible.

Data from multiple data sources can be used to create visualisations and dashboards

Another self-service BI tool architecture is shown in Type 2. Here, the business analyst works more independently to build visualisations and insights. They do this by connecting to one or more data sources, either local files or corporate sources, and are guided by the tool as to what dimensions and measures are discovered in the data. From there, they can choose what data to load into memory (e.g. by generating SQL), start to mash together, and visualise data to produce new insight with further queries being issued to underlying databases

as and when needed. Unlike the multi-user dashboard and visual data discovery environment in Type 1, this approach is more of a single user data exploration and data visualisation environment where data requirements are less well known up front.

In-memory data allows many concurrent users to interact with data

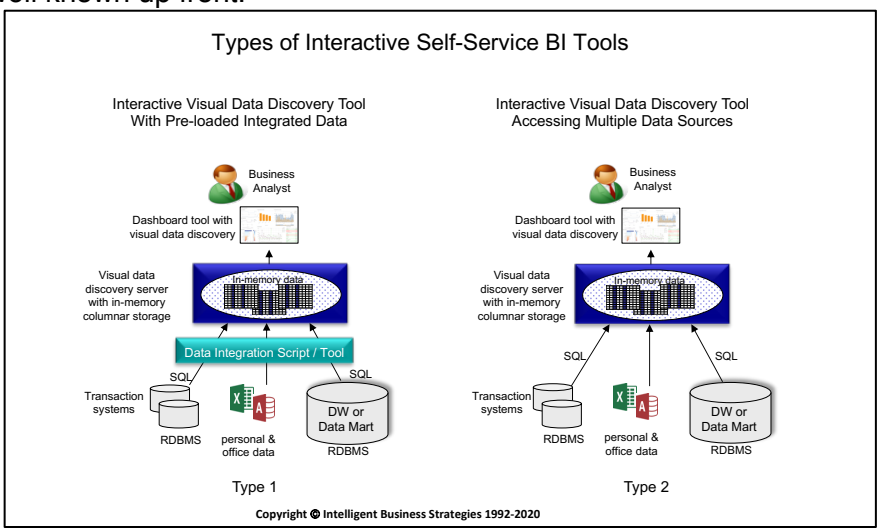


Figure 1

Some visual data discovery tools don't pre-load data into memory and are more aimed at single user data exploration

This approach also pushes complexity onto each individual user and requires them to know how to clean and integrate data from multiple data sources as well as analyse it. Business analysts must create their own personal data preparation and integration logic which is dependent on the number and complexity of data sources and on their skillset. The other problem is the potential unnecessary re-invention. Of course, multiple business analysts can be working concurrently but it is not like in Type 1 where multiple users are sharing the same integrated data on the BI platform. Nevertheless, once new insights have been identified, they can again be put in a dashboard and shared with information consumers who can then access the dashboard via the web or mobile device. However, the consumer has less power as these dashboards typically only support data as selected by the business analyst who created them.

ADVANCES IN SELF SERVICE BI – ARTIFICIAL INTELLIGENCE

In the last few years there have been major advances in self-service BI. In particular, through the introduction of artificial intelligence (AI) within the software. This is more frequently referred to as *Augmented Intelligence* and has emerged in several ways. One of these, has seen the introduction of natural language interfaces to broaden the reach self-service BI tools to many more users. Figure 2 shows how an AI engine can enable functionality like a natural language chatbot interface can be added to the type of self-service BI tool described in earlier in Figure 1.

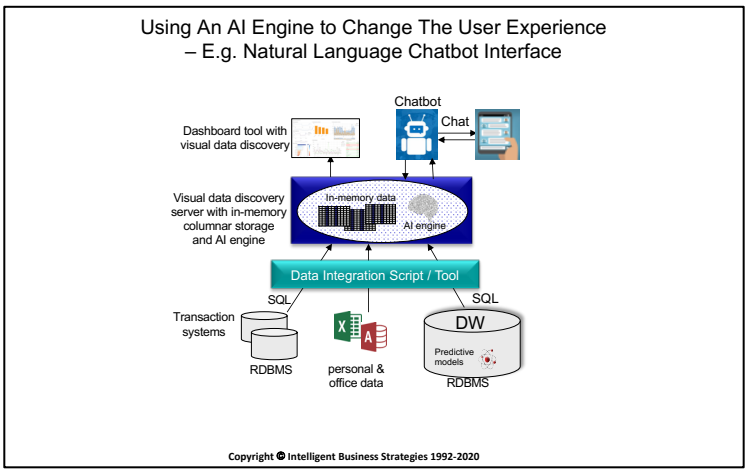


Figure 2

Another term for this is *conversational analytics* because the users can ask questions in a language such as English and gets an answer back also in English accompanied by an automatically generated visualisation. In addition, further questions can then be asked based on initial insights rendered. Therefore, this is not just natural language processing (NLP) to convert a question into a query on data. It is also natural language generation (NLG) by the self-service BI tool to answer the question and keeping context within a conversation. In addition to a natural language interface, we are also seeing the emergence of voice-based conversational analytics which enables users to just ask questions. The software then converts voice into natural language query on data and results are rendered using natural language generation.

There are many other ways in which artificial intelligence in self-service BI is helping to improve productivity and shorten time to value. Examples include:

- Providing visualisation and analysis recommendations
- Auto-generation of visualisations based on search and current analytical context
- Key driver analysis and automated outlier detection
- Auto-identification and recommendation of dashboard or report KPIs
- Automation of data preparation and visualisation steps
- Auto-invocation of built-in and 3rd party predictive and prescriptive analytics e.g. Automated forecasting
- Intelligent alerting to support management by exception
- Learning from a history of user analysis interactions

There is a lot going on and it is easy to get overwhelmed by it all. But if you are looking to select a self-service BI product what are the requirements that need to be defined to support interactive data exploration and analysis in a modern enterprise where data can be held in multiple stores?

KEY REQUIREMENTS FOR INTERACTIVE DATA EXPLORATION AND ANALYSIS

When you contrast the two types of architecture shown in Figure 1, you realise that users who need the Type 1 kind of architecture want more agility than just running ad hoc SQL queries.

They have many more requirements than need to be met in order to support interactive data exploration especially when more and more concurrent users are coming online making use of conversational analytics in addition to interactive dashboards. These include both user-driven and augmented intelligence requirements which are detailed in the Appendices. Some of the key ones include:

Need automated and recommended visualisations to see insights in the best possible way to shorten time to value

- Automated data profiling on data in selected data stores
- Guided data preparation
- Advanced data compression and high-performance load into memory
- Conversational natural language user interface
- Changing filters, sort order, visualisations, hierarchies, groups on the fly
- Automatically suggest contextual insights based on data selected and *not* selected ("peripheral vision") during analysis
- Automated visualisation recommendations based on analytical context
- Automatically switch from search to auto-generated visualization for a deeper user experience
- Real-time data visualisation for situational awareness
- Provide a scalable way to analyse big data without having to move it from where it may be stored
- Support open APIs to enable extensibility and embedding of insights and recommendations into applications

These requirements go a long way to democratising the use of self-service BI and analytics taking it beyond business analyst producers of analytical content to *consumers* who can become more analytically savvy by making use of augmented intelligence capabilities such as conversational analytics, smart search, auto-generated visualisations and suggested insights to help them to be more effective with self-service BI in their day job.

A smart underlying engine is needed to support growing numbers of users interacting with and doing their own lightweight analysis on data

When you look at it like this, you quickly realise that automated inference of associations in data together with augmented intelligence has a huge part to play in maximising the value of visual data discovery and interactive dashboard applications. In the modern business, where reducing time to value is always important and improving *effectiveness* among the masses is becoming mission critical, the requirement is as much about having an underlying 'engine' that is context-aware and supports all the aforementioned requirements as it is about SQL.

With that in mind it is worth looking at what such an engine can and should deliver.

SMART INTERACTIVE DATA EXPLORATION USING QLIK'S ASSOCIATIVE AND COGNITIVE ENGINES

Qlik offers a number of products in data and analytics

Having understood the requirements, this section looks at how Qlik steps up to meet these requirements. Qlik is a provider of data management and analytics software. It offers several products including Qlik Sense with Qlik Insight Bot™, QlikView, Qlik Data Catalyst and Qlik Big Data Index.

Qlik Sense is a next generation self-service smart visual data

Qlik Insight Bot is an add-on tool to Qlik Sense that provides converational analytics

Qlik Sense is Qlik's next generation platform for modern analytics, supporting a variety of use cases including business analyst led visualization and analysis, centrally deployed interactive dashboards for information consumers, conversational analytics (with Qlik Insight Bot), mobile analytics, embedded analytics and reporting (with Qlik NPrinting). It is based on Qlik's Associative Analytics Engine, and more recent versions also include a cognitive engine supporting AI-powered insight generation, automation, and natural language interaction. It also includes smart data preparation, collaboration, and data storytelling to help information consumers step through insights. It is available in Enterprise and Business editions, and deployed in combinations of SaaS, private cloud, and on-premises sites based on a modern, multi-cloud architecture. **Qlik Insight Bot™** adds conversational analytics to Qlik Sense (currently English only), using NLP and NLG to broaden the use of Qlik Sense by opening it up to less skilled employees across the enterprise.

QlikView is Qlik's first generation tool for building interactive dashboard applications

QlikView is Qlik's first generation Windows based in-memory product aimed at IT and power user business analysts who need to build interactive dashboard applications for rapid deployment across the enterprise. It also runs on the Qlik Associative Engine.

Qlik Data Catalyst is data fabric and data catalog software to create ready made data assets for use in Qlik Sense and in data science

Qlik Data Catalyst is enterprise data management fabric used to produce business-ready data available a marketplace (a data catalog) for on demand consumption. It can be deployed on premises or in the cloud and includes publishing data directly to Qlik Sense.

Qlik's Big Data Index provides access to big data without the need to load it all into memory

Qlik Big Data Index is a value-add product that provides an associative binary index of data stored in sources such as cloud storage data lakes and Hadoop®. It allows the Qlik Associative Engine to store information about resident big data without the need to load all the data into memory.

Qlik Sense and QlikView are both built on the same platform

Qlik Sense and QlikView include a set of open and standard APIs and management tools. The open APIs enable development of mashups, custom analytics applications, embedded analytics, and direct engine-level integration with third-party predictive models built in Python, R and DataRobot, supporting in-context, real time user exploration.

Qlik's Cognitive Engine brings AI and machine learning to the platform

Qlik Sense can make use of the Associative Engine and Cognitive Engine together to provide automation and contextual insight suggestions

Qlik analytics are driven by the Qlik Associative Engine which supports interactive, contextual exploration and search. Qlik Sense also offers a new Cognitive Engine which brings artificial intelligence and machine learning capabilities to the platform. Qlik Sense brings the Associative Engine and the Cognitive Engine together to add contextual insight suggestions and automation to the user experience. The combination of these two engines together is referred to by Qlik as The Associative Difference.

Let's look first at the Qlik Associative Engine and then at the Qlik Cognitive Engine to see the impact the Qlik Associative Difference makes to the overall user and analytical experience when the two engines work together as part of Qlik Sense.

THE QLIK ASSOCIATIVE ENGINE AND WHAT IT ENABLES YOU TO DO

The Associative Engine supports rapid exploratory analysis of in-memory data

The Qlik Associative Engine is a multi-threaded high-performance analytics engine, built to support interactive, contextual analytics that can fully exploit multi-core processors and make use of several methods to optimize data and analytics calculations based on the user's exploratory context. It operates differently from BI tools that rely on SQL or ad-hoc queries against an underlying DBMS by enabling:

It also supports concurrent users accessing interactive dashboard aplications

- Visualisation and analysis of data from an optimized, in-memory data cache comprised of data from multiple sources
- Large numbers of concurrent users accessing and interacting with dashboard applications, supporting exploration and search through dynamic calculation and association

ANALYTICAL APPROACH

The Associative Engine supports a stepwise approach to producing business insight

With respect to the former, the Associative Engine allows organisations to implement a stepwise, iterative, rapid prototyping agile approach to producing insight involving business users from the outset.

Loading the data

Data from multiple sources is loaded into memory and converted into compressed binary format for performance

Qlik's self-service data preparation facility generates scripts to prepare and load data

The first step is to identify and define the data to be loaded. This is a step where the Qlik Data Catalyst Data Catalog can be used to identify trusted data. IT or a data analyst then connects to underlying system(s) and loads data into memory, either using visual data preparation or scripting for more complex problems. This effectively creates an in-memory data mart, optimized in a compressed binary format – something unique to Qlik. Qlik's drag and drop self-service data preparation capability includes association (linking) recommendations, data profiling, and automation, generating a load script in the background. Running the script loads the data, which the Associative Engine converts into compressed binary format (more on this later).

Qlik's Big Data Index provides a scalable way to index and analyse big data in cloud storage or Hadoop without the need to load it all into memory

Set-level SQL queries can be used in scripts to run against each table (or view) in an underlying data store without doing any joins. This means you can bring in all the data you need into memory. In addition, the Qlik Big Data Index (QBDI) can also be used to access big data without moving it into memory. QBDI is a Qlik Associative Engine scaled up to handle billions of rows of structured data held in Parquet files in Amazon S3, Amazon Elastic File System (EFS) or Hadoop HDFS the. QBDI runs in the environment where data resides. This is done by running multiple Associative Engine instances in pods on a Kubernetes cluster that can run on any cloud or on-premises with each pod in Kubernetes pointing to the data. Qlik Data Catalyst can also be used to create the files that QBDI ingests. A partitioned index made up of indexlets (IDX files) is then built on every column in every Parquet file and persisted on high performance solid-state disks (SSDs). Relationship associations in the data are also kept. Qlik Sense can then connect to QBDI directly to use the index without moving the

data or On-Demand App Generation (ODAG) can be used to take a slice of data into memory.

Data governance can be controlled to drive consistency across in-memory data

Note that in an exploratory environment, calculations may not be known ahead of time. For this reason, the Associative Engine has been written to support on-demand calculations after underlying data has been queried and loaded. Because joins and pre-defined aggregates are not encouraged during script development, it opens up a broader set of in-memory data for freeform exploration and interactive analysis.

Also, several things can be done for data governance during this process including:

Qlik's Smart Data Load facility can be used to create reusable data preparation services

- Using Qlik's self-service data preparation capability to create 'data preparation services' that can be published for reuse by other business analysts. This means a catalog of data preparation services can be kept for business analysts to find and reuse while enforcing data governance.

Trusted dimension data can be taken from MDM systems

- Commonly defined dimensional data (e.g. customer, product, supplier etc.) can be created by accessing trusted, integrated master data in your underlying master data management (MDM) system. This practice guarantees consistent dimension data with common definitions across all scripts and Qlik in-memory data marts.

- IT design the in-memory data model including access to an MDM system for common dimension data.

- Already clean and integrated data can be taken from a data warehouse.

Trusted 'base' data can shared across projects to avoid re-invention and improve consistency

- Creating reusable 'base' datasets that can be governed and shared with others to avoid people re-inventing scripts to do this repeatedly in potentially inconsistent ways. With Qlik Data Catalyst, this can be done by creating popular datasets (e.g. dimension data and transaction data) in advance and publishing them in the Qlik Data Catalyst data catalog for business analysts to use. This creates a 'layer' of re-usable base datasets which then become input to the Smart Data Load facility.

- A combination of these approaches

Associative exploration and visualisation development

Business analysts can focus on producing insight, building visualisations and interactive dashboards

Because data is staged in the process above, business analysts and power users can focus more on producing insight. To do this, users explore and refine context by making selections and / or searches. After each interaction, the associative engine determines the new context (selection state) and applies that globally to all metrics, recalculating them on the fly (dynamic calculation). It also determines data values associated and unrelated to the new context, keeping the user aware of possible hidden insights at all times as they analyse the data. Producing visualizations can happen by business analysts going through a traditional build process, with assistance and automation from AI in Qlik Sense, to produce interactive, dashboard applications with guided analysis aimed at a community of users in a specific part of the business. Or, with the new Insight Advisor capability, driven by Qlik's Cognitive Engine (more on this later), users can auto-generate visual insights and charts and add them to dashboards, based on data set and refined using natural language search.

Information consumers make use of interactive dashboard applications and interact with data

Insight consumption and guided analysis

The third step is the information consumer who can make use of the dashboard applications created for them to interact with the data. In a similar way to the business analysts, they too utilise the Qlik Associative and Cognitive engines while interacting with the data to guide their analyses and keep them aware of hidden insights. This provides large communities of business users with a means to interactively search and explore, to get insights and answer follow up questions. In addition, depending on their skill set, business users can also clone, edit and create new visualisations or utilise the Insight Advisor to request further metrics using natural language search.

Multi-Table Data Model

Information consumers make use of interactive dashboard applications and interact with data in traditional and big data stores

The Qlik Associative Engine takes advantage of a multi-table in-memory data model created from one or more underlying data sources including those indexed using Qlik Big Data Index. By bringing data into memory and resolving ambiguities in the data through unique column naming, the Associative Engine can automatically determine relationships in the data prior to any exploratory analysis taking place. Note that nothing is pre-calculated which means that any calculation can be performed on demand. In addition, because the relationships are known, the Associative Engine will always calculate an aggregation in the right table thereby avoiding any risk of incorrect aggregation errors such as double counting. Avoidance of any risk of incorrect results also applies to bi-directional relationships that have been known to cause problems in other tools. Furthermore, re-calculation can also be performed on-demand of all in-memory aggregates created on top of the multi-table data model at the speed of thought.

Aggregation errors such as double-counting and those caused by incorrect navigation of bi-directional relationships are automatically avoided

In-Memory, Compressed, Binary, Columnar Data

All data is converted into compressed binary format with columnar storage

A key differentiator of the Associative Engine is its patented approach to compressing data. Data from each source data table is converted into two types of in-memory data structure:

- A set of 'symbol' tables (one for each column in the original table)
- A compressed binary data table that replaces the original table

Only the distinct values in a column are kept

A symbol table contains a binary value for each distinct value in a column in the original table. The binary value then replaces the original column value in the original data table.

This means that much more data can be loaded into memory

Qlik's patented approach to compressing data into binary indexing allows the Associative Engine to load more data into memory faster than other tools

By doing this for every column, the original data table is converted, column-by-column to a compressed binary data table with each row/column cell value replaced with a binary reference. These binary references, act like foreign keys pointing back to a corresponding distinct value in a columnar symbol table. The compressed binary data table therefore contains the same number of columns and rows as the original uncompressed data, but is now full of binary references that make it much more compact. Also, if two fields have the same name in two different tables (i.e. a relationship) then they have the same symbol table and binary representation. The result of this compression technique is that there are no data types for columns in the compressed data table. Absolutely everything is binary. It is the symbol tables that give the Associative Engine its compressed columnar and binary indexing capability so that it can support concurrent interactive data exploration, on-demand calculation and concurrent interaction with dashboard applications at the speed of thought. From a loading perspective, the conversion to binary makes it very fast to load millions of rows.

It also means a large numbers of concurrent users interacting with data at the speed of thought

It also makes it fast to analyse and from a server perspective, a 64-bit operating system with a large amount of memory will mean more data can be loaded into memory. This fits with a symmetric multi-processing (SMP) scale-up architecture.

LOGICAL INFERENCE AND CALCULATION BASED ON CONTEXT

The Qlik Associative Engine support logical inference and on-demand calculations

One of the greatest challenges in interactive data exploration is to be capable of supporting many information producers and information consumers all concurrently accessing data while changing their minds 'at the speed of thought'. Requirements like changing query data selections on the fly, changing filters on the fly, changing sorts on the fly, changing hierarchies on the fly, adding your own calculations on the fly and repeatedly re-doing complex calculations beyond the capabilities of SQL aggregate and analytic functions are all a challenge. So, besides compressed binary, columnar data, how does the Associative Engine support this? The answer is by using a two-step process as each user interacts with in-memory data.

The Qlik Associative Engine supports logical inference and on-demand calculations

Step one is logical inference where the Associative Engine understands and manages a global context. Before a user selects anything, the assumption is all that data is in play. However, every time a user clicks one or more values in a field (selects data), or performs a search, the engine immediately calculates what distinct values in all related tables qualify as being relevant to the data selection i.e. the engine logically infers a new context. If you were in a crowded room of people who are all standing and holding a ticket with a number on it and the first digit of the number is selected and read out, then logical inference is like saying "would everyone who does not have that first digit on their ticket please sit down". It is about relevance. But it is not happening just on one table. The Associative Engine knows the relationships between data tables. Therefore, it effectively determines (infers) and flags all distinct values in all the tables required for the analytics view, as either associated (relevant) to or unrelated (irrelevant) to the global context as defined by the data selected. The reason for doing this is so that it can expose to the user the data that is *both related and unrelated* to the question at hand, and also allow the engine to ignore unrelated values when making calculations for analytics on just relevant data satisfying the user's query. The powerful thing about logical inference is that it continually minimizes the data needed to satisfy a user's question with every click. Subsequent calculations therefore only ever take place on relevant data. If the user changes his or her mind, the engine simply re-determines the relevant data subset based on the new context. Logical inference allows the Associative Engine to always know which data values in which tables to use, and to always optimize calculations on that data.

"Would all data that is not needed please sit down"

The Associative Engine continuously minimises the data needed to satisfy a user's query with every click the user makes

It also continuously determines the shortest path to the data needed

All calculations are done on-demand on relevant in-memory data highlighted by logical inference

Step two is dynamic calculation, where the engine performs all calculations and aggregations on-demand on the relevant data set, based on the new global context as determined by logical inference. Each element of a calculation on each relevant table is calculated before calculating final results. Formulae can be nested and filters added. Every time a selection is made, the engine totally recalculates the hypercube. Remember, because the data is in memory in compressed *binary* format (or indexed by the QBDI), and has been minimised by logical inference, calculations are extremely fast. Attempting to support high numbers of concurrent users interacting at the speed of thought, when

performing dynamic calculations on uncompressed data on disk in a row–oriented database using SQL queries would take too long.

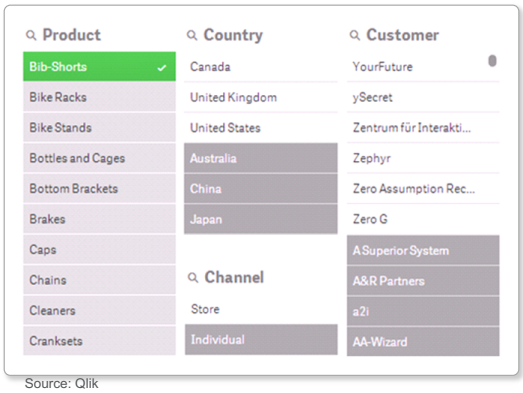
Also because calculations are on-demand and in-memory, users get freedom to ask for any views of data they would like. There is no need to pre-calculate aggregates or complicate the data model with aggregate in- data structures that need to be pre-loaded with data. It is all done at memory speed.

THE POWER OF GREY

The Associative Engine also knows about data a user selected but which was then excluded due to a subsequent selection

A unique benefit of logical inference as described above is the ability for people to see not just data that is relevant to their selections, but also the excluded data set – data that is unrelated to their current selections. This so called 'peripheral vision' is unique to Qlik, and very hard if not impossible to replicate using SQL queries, as the relational DBMS would filter these values out of query result sets. An example of this is shown in Figure 3 where the user has selected a Bib-Shorts product and the engine immediately knows from logical inference the countries where it is not selling and the customers that have not bought it.

The Associative Engine can highlight insights that a user may be totally unaware of by allowing them to also see data they did not select



Source: Qlik

Figure 3

The point about this is that a user can see not only what is related to selections but also what is unrelated (excluded), whereas in the case of a SQL query, they would only see the related data relative to the filters in a WHERE clause and would not see what was excluded. The only way to see this would be to run another query with inverse criteria. However, the problem is that if you don't see what you don't know, you may never know to even ask.

Another example is shown in Figure 4 where a business analyst has looked at sales of a set of high margin low sales products in a given region. Here you can see that even though the selection is relevant to specific channels, the channels the products **did not sell through** are also highlighted in grey on the lower left.

This is a unique and very powerful capability that gives the user peripheral vision

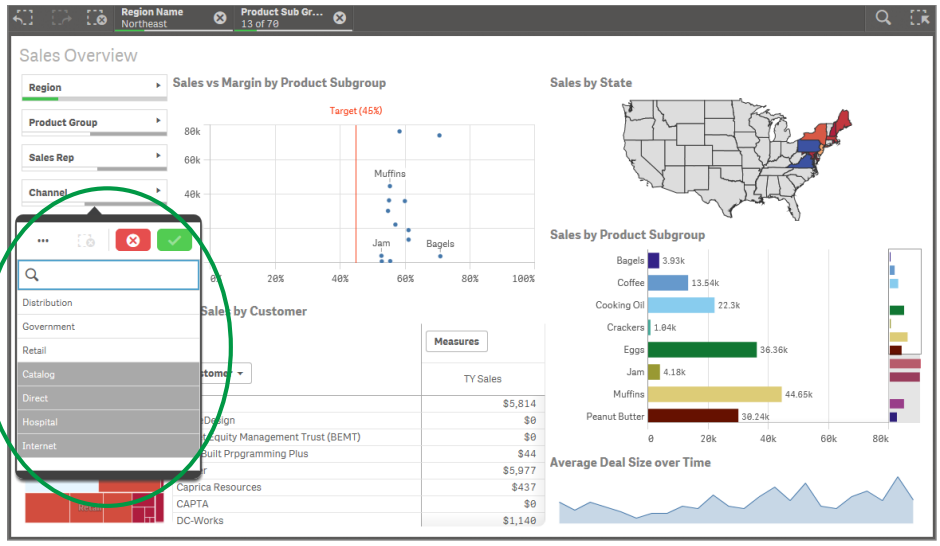


Figure 4

Users can discover insight they were totally unaware of even when not looking for it during an analysis

This 'power of grey' is a very powerful concept. This is about surfacing insight you may be totally unaware of even when you are not looking for it. This is something unique to Qlik and is very valuable in interactive data discovery. It is made possible by the Associative Engine maintaining two independent statuses for every field value. These are:

- An input state - which is either 'selected' or 'not selected'
- An output state – determined by logical inference which is either 'possible (associated)' or 'not possible (unrelated)'

In terms of input state, a field value is either selected or not. If selected, logical inference then determines the relevant data. Therefore, by tracking what is selected (the input state), the Associative Engine can determine what is and is not possible (the output state).

ENABLING AUGMENTED ANALYSIS USING QLIK'S COGNITIVE CAPABILITIES

Qlik has added a Cognitive Engine and conversational analytics to its platform to improve productivity, make better decisions and reduce time to value

Having understood how data is loaded data into in-memory compressed, binary, columnar format, and how using logical inference and the power of grey within the Qlik Associative Engine helps to analyse data, what then is the impact of the Qlik Cognitive Engine and the AI-powered conversational analytics capabilities offered by Qlik? The best way to describe this is to revisit the stepwise approach to acquire, prepare, load and analyse data and show how the Cognitive Engine and Qlik Insight Bot work with the Associative Engine to improve the analytical process and experience. The impact of these is shown in the table below:

Accelerated data preparation

Automated insight suggestions and automated highlight of most relevant hidden insights

Step	Cognitive Capability Added	Business Benefit
Data preparation	▪ Automated data profiling & outlier detection ▪ Automated highlighting of data issues ▪ Relationship recommendations – Automated suggestions on how best to link data together ▪ Automated transformations and transformation recommendations	▪ Lower skills bar ▪ Accelerated data preparation ▪ Shorter time to value
Data exploration and analysis	▪ Auto generated and recommended KPIs ▪ Automated suggestions of insights ▪ Automated suggestions of associative insights - highlighting the most relevant hidden insights (in the excluded, grey values) ▪ Auto invocation of analytical models developed in Python, R or on 3rd party ML platforms ▪ Conversational analytics via NLP / NLG	▪ Deeper insights available quickly ▪ Shorter time to value ▪ 'Always on' peripheral vision ▪ Predicted impact awareness ▪ Better decisions ▪ Extensible cognitive capability ▪ Explainable insights

Natural language search and auto-generation of explainable insights

Automated visualisation

Self-learning analytical platform

	▪ Full natural language search with auto generated insights ▪ Auto switch from search to auto visualisation	▪ Many more users guided by and acting on insights ▪ Smarter business operations
Data visualisation	▪ Visualisation suggestions ▪ Automated generation and rendering of visualisations	▪ Better understanding ▪ Shorter time to value ▪ More timely decisions
Action	▪ Automated alerts ▪ Recommended actions	▪ More timely decisions ▪ Better optimized business
All steps	▪ Observation of behavior and interaction history	▪ Self-learning analytical platform ▪ Progressively smarter business

Looking at these capabilities, there are many things that stand out. Firstly, there is data augmentation and accelerated data preparation leading to better decisions and shorter time to value.

Built-in insights advisory service

AI-driven peripheral vision automatically analysing the impact of hidden insights

Second is integration of the Associative Engine and the Cognitive engine to provide a built-in *insights advisor service* that automatically suggests insights in the context of the analysis on selected data being performed by the user in real time. It can also automatically suggest context aware hidden insights (areas to investigate 'in the grey')- a unique differentiator. This Associative Insights advisory service coupled with auto-generated visualisations means the software is trying to ensure the user doesn't miss anything important. The ability of this AI-driven peripheral vision to show predicted hidden insights is even more powerful. In addition, the ability to invoke models developed with DataRobot, R or Python *in context* by passing over only the data needed gives business users on-demand near real-time access to predictive models during associative exploration and analyses. This provides an extensible cognitive capability exploiting models already built by data scientists.

Extensible 'in-context' cognitive capability by invoking models developed in other tools

Natural language conversational analytics democratises the use of self-service analytical tools and opens them up to many more users

Third is that NLP/NLG based conversational analytics provided by Qlik Insight Bot on top of Qlik Sense. In this case, the Associative Engine enables contextual awareness and exposes hidden insights via NLG. Natural language search makes it easy to use the software to ask questions while NLG ensures that insights are explainable building up user confidence. In addition, the ability to automatically switch from natural language interaction to visual exploration, in-context, is possible because both experiences run on the associative engine. Conversational analytics will likely cause democratisation of analytics resulting in a significant increase in the number of concurrent users. This makes it even more important to have a *context-aware* AI driven tool to improve user productivity.

Finally, the Cognitive engine goes further by building up an interaction history to enable Qlik Sense to observe user behaviour and become self-learning.

COMPLEX CALCULATIONS USING PARAMETERS AND DYNAMIC DATA SETS

All calculations are on-demand and in-memory and support parameterised variables

Set analysis dynamically determines the set of data to be used in a calculation and visualisation based on what a user selects

In addition to the associative and cognitive capabilities already discussed, the Qlik calculation engine also supports a wide range of aggregate, statistical and analytical functions that can be used in on-demand calculations on in-memory data. It also supports parameterised variables and set analysis. The former allows variables to be created that will be recalculated every time you click and make a selection. These variables can be used in other formulae and also for what-if analysis.

Set analysis can be used to dynamically determine the set of data used in specific calculations of visualisations based on what a user selects. This is done using an expression that can be defined for:

- A calculation used in an individual visualisation
- A part of a calculation used in an individual visualisation
- A calculation on another visualisation that includes all data except that selected for a specific visualisation
- Other combinations

It is often used to compare results of two different time periods even if a user selects only one period

It is used to dynamically restrict or extend the data set used in a calculation or part of a calculation

Set analysis is commonly used when you want to do a calculation outside your filter e.g. to compare results of two different time periods even if a user selects only one period. For example, if a user selects data to look at sales of a product category for the first quarter this year, set analysis could dynamically extend the dataset to calculate and visualise the sales of the same category for the same quarter last year to allow the user to compare results. Alternatively, it could invoke a custom predictive analytic built in Python, R or in DataRobot to extend the data set to allow a trend to be plotted beyond the current period selected. It could also be used to compare what has been selected against everything (e.g. sales of a product against total market sales), or a second subset of data.

The Associative Engine is aimed at supporting interactive data exploration and concurrent usage of guided interactive analytic applications

Multiple set analysis expressions can be used in the same visualisation by applying different filters on different aggregations within the same formula i.e. different parts of the formula are calculated using a different set of data.

Building on this, the Qlik Associative Engine supports alternate states, allowing user-driven comparative analysis. Dashboard creators can define multiple states, link UI objects to these states, and create analytics comparing values between the states. Users can then make selections to define sets to compare and evaluate the results in realtime.

Furthermore, the ability to use the calculation engine together with custom built machine learning models developed with DataRobot, R or Python invoked *in context* augments the calculation engine providing extensible analytical capability.

CONCLUSIONS

Working together, Qlik's Associative and Cognitive Engines differ significantly from BI tools generating SQL based queries. They are not aimed at replacing SQL databases, but instead allow organisations to go beyond this to highlight *and predict* hidden insights while also supporting interactive, and automated data exploration as well as large numbers of concurrent users leveraging conversational analytics and guided interactive analytic applications. The platform is designed to allow:

It allows business analysts to focus on rapidly building interactive dashboard applications assisted by AI driven suggestions and automation to shorten time to value

- IT to focus on data storage, data governance, performance optimization and complicated requests
- Separation of data modeling from user interface development so that IT and business can work together to iteratively design data models, govern data and rapidly integrate trusted data to load into memory
- Business analysts to focus on rapidly create their own interactive metrics, reports and dashboards, leveraging built-in Cognitive Engine suggestions and automation during development to shorten time to value

It also allows business analysts to explore subsets of data from multiple sources and answer tough business questions

- Large numbers of concurrent users in different parts of the business to consume applications, interact with data at the speed of thought using associative exploration and to do further lightweight analysis on the data to help them make timely operational and managerial decisions
- Further extend the reach and power of data through conversational analytics, allowing less sophisticated users to easily ask and answer questions through natural language
- Allow business analysts who have a good understanding of the data needed to explore subsets of structured data from multiple sources in-memory to discover and produce new insight as well as answer tough business questions

Highlighting insight you may be totally unaware of in combination with everything else makes it very well suited to interactive exploration and analysis

The combination of the Associative and Cognitive Engine in Qlik Sense is a real differentiator

Capabilities like data augmentation, automated data profiling, automated transformations, fast data loading, holding data in compressed binary, columnar format, dynamic logical inference for every user on every click, on-demand complex calculations on dynamic data sets, automated insight suggestions, automated visualisation and conversational analytics, allow Qlik to meet almost all of the requirements highlighted for interactive data analysis and exploration. However, doing all this plus providing the ability to automatically highlight and suggest the most relevant hidden insights you may be totally unaware of (by combining their Associative and Cognitive Engines in Qlik Sense) makes it stand out as a leader in helping organisations improve productivity to reduce time to value and empower the masses across an enterprise to easily consume and act on actionable insights. This helps organisations to be more effective in every-day business operations and more competitive in the marketplace.

APPENDIX - DETAILED REQUIREMENTS FOR INTERACTIVE DATA EXPLORATION AND ANALYSIS

Augmented intelligence requirements driven by built-in artificial intelligence are shown in *italics* within each category

Category	Requirements
Data requirements	• Support access to traditional, big data and streaming data sources • Access to a data catalog to search for and find relevant data across in data warehouses, data marts and other data stores across the distributed and multi-cloud data landscape • *Automated data augmentation suggestions of additional relevant data to add to data manually selected for integration and analysis*
Data preparation and integration	• *Automated data profiling and statistics generation on data in selected data stores* • *Guided data preparation via automated transformation and integration recommendations during data integration development* • *Assisted data joining to correctly join relational and non-relational data* • Direct integration with 3rd party advanced analytics engines to invoke predictive models (built by data scientists using other tools / languages) on-demand to produce scores (results) that can be integrated with other data prior to loading into memory and subsequent analysis • Advanced data compression and high-performance load into memory • Automated and scheduled refresh of data
Exploration and Analysis	• Analyse user-defined subsets of data • Add / remove data sources as and when required • *Conversational analytics using natural language processing (NLP) or voice together with natural language generation (NLG) to query and analyse data* • *Discover data and associations using global search* • Make selections in all objects • Change selections on the fly, including in a non-linear fashion • Change filters on the fly • Analyze all objects in-context together (selection in one object updates all objects) • Change sort order on the fly • Change visualisations on the fly • Create your own hierarchies • Throw away hierarchies • Create your own groups e.g. from a set of products (binning) • Add your own calculations • Do complex calculations beyond the capabilities of SQL functions • Test hypothesis using what-if analysis • Perform statistical analysis • *Automatically identify and recommend dashboard and report KPIs* • *Automatically rank key driver metrics that lead to business value* • *Automatically suggest contextual insights based on data selected by the user during analysis*

Automated suggestions of additional relevant data

Automated data profiling, transformations and guided integration to accelerate data preparation

Need to change data selections, filters, sorting and visualisations all at the speed of thought

Conversational analytics to simplify usage and broaden the reach to more users

Provide a built-in automated insight advisory service

Add new hierarchires, groups and calculations

Provide peripheral vision to ensure the user sees everything

Provide 'in context' access to custom built analytics

Leverage powerful visualisations

Need automated and recommended visualisations to see insights in the best possible way to shorten time to value

Natural language generation to clearly explain the business impact of insights

Create and publish interactive dashboards for consumption

Support large numbers of users concurrently interacting with data via conversational analytics

Open APIs are needed to integrate with 3rd party tools and to embed analytics in applications and processes

	• Discover insights from the excluded data set ("peripheral vision") • *Automatically suggest contextual insights based on data **you <u>didn't</u> select** ("peripheral vision") during analysis* • *Perform predictive analysis and automated forecasting on data for algorithmic and predictive insights* • *Direct integration with 3rd party machine learning engines for on-demand invocation of predictive models (built by data scientists using other tools / languages) in the context of the analysis being performed*
Visualisation	• Leverage powerful visualisations to quickly see problems on a lot of data ▪ Heat maps, graph matrix, network diagrams, geographic maps, small multiples etc. ▪ Spot outliers (dots outside of clusters on scatter plots) ▪ Identify major problems shown up on a heat map • Use single step query and visualisation with *automated charting* • *Automated visualisation recommendations and rendering based on analytical context in both a traditional self-service tool user interface and natural language conversational analytics user interface* • *Automatically switch from search to auto-generated visualization for a deeper user experience* • *Automated interpretation and explanation of visual insights using natural language generation (NLG)* • Real-time data visualisation for situational awareness • Create user assembled dashboards that include: ▪ Multiple visualisations ▪ Multiple data sets ▪ Connected visualisations that work together in context ▪ Ability to alert from within the dashboard (with mobile in mind) ▪ Ability to act from within the dashboard (with mobile in mind) • Publish dashboards to the web and mobile devices (including offline use)
Performance	• Persist data in memory for multi-user interactive analysis on published dashboards • Support multi-threading and automated scale out to handle masses of concurrent users utilizing conversational analytics and mobile BI • Provide a scalable way to analyse big data without having to move it from where it may be stored
Embedding	• Support open APIs to enable: ▪ Extensibility of the software product ▪ Insights and recommendations to be embedded in applications ▪ Integration with advanced analytics during data loading & analysis

About Intelligent Business Strategies

Intelligent Business Strategies is a research and consulting company whose goal is to help companies understand and exploit new developments in business intelligence, analytical processing, data management and enterprise business integration. Together, these technologies help an organisation become an *intelligent business*.

Author

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an independent IT industry analyst and consultant, he specialises in business intelligence, data management, big data and analytics. He is also Chairman of Big Data LDN. With over 38 years of IT experience, Mike has consulted for dozens of companies on data strategy, data architecture, data warehousing, data lakes, big data, data governance, master data management and analytics. He has spoken at events all over the world and written numerous articles. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates, an independent analyst organisation. He teaches popular master classes in Big Data Analytics, Data Warehouse Modernisation, Designing, Managing and Operating an Enterprise Data Lake, Enterprise Data Governance, Master Data Management and Machine Learning and Advanced Analytics.

Water Lane, Wilmslow
Cheshire, SK9 5BG
England
Telephone: (+44)1625 520700
Internet URL: www.intelligentbusiness.biz
E-mail: info@intelligentbusiness.biz

Interactive Data Exploration Using A Context Aware Analytics Engine