

WHITE PAPER

Multiplying the Value of Big Data

How Qlik gives you an edge

TABLE OF CONTENTS

Summary	2
Introduction	2
The growing need for Big Data analytics	3
How Big Data goes from raw to analysis-ready	4
Using Big Data: Make it relevant and keep it in context	6
Qlik's options for handling your Big Data	8
Side-by-side comparison: Options for handling your Big Data	12
Qlik and Big Data connectivity	13
Going the last mile with Big Data	13

SUMMARY

- You can't get all of Big Data's benefits until you make it easy for all your business users to analyze it.
- The key to empowering business users to analyze Big Data is to make it easy for them to access this information and then show them only what's relevant to them.
- There could be multiple methods for making Big Data relevant, and they depend on the volume and complexity of your data.
- Qlik has a variety of options for enabling all your users to explore your Big Data and make discoveries within it — fast.

INTRODUCTION

Big Data is no longer a new term, it's the new reality. Many organizations are already putting it to use, storing and analyzing vast quantities of data to discover new opportunities and solve challenging problems.

Other organizations are still figuring out how to best work with Big Data. How is it relevant to what they do? What are their best options for making it useful? And how can they use it to get an edge on their competitors?

This paper begins with a quick take on some Big Data basics — what your business users need from it, and how to take it from its raw state to an analysis-ready form. Then we walk through Qlik's approach to making Big Data accessible to your business users, including a series of options for keeping your data relevant and in-context.

The growing need for Big Data analytics

In the past, the only people using Big Data were data scientists. To help solve major challenges in academia, government, and the private sector, they ran very complex algorithms on massive computing clusters. They're still doing it today. But now there's a new, larger need. To lead in the digital era, businesses everywhere want to empower all their users to work with Big Data.

In the old, algorithmic model, analysts mine through all the data available, trying to find the needle in the haystack. That approach doesn't make sense for business users, who have specific questions about slices of data relevant to their work. They want answers they can act on to questions like these:

- How have products sold since our last promotion?
- How effectively is the sales team cross-selling our products?
- Which products are not selling well?
- How much do sales vary by region or team?
- What happens if a natural disaster cuts off one of our primary suppliers?
- Does our service call history show a pattern of customer satisfaction or dissatisfaction?

Of course, business users were asking these questions long before the advent of Big Data. But without easy access to key data sets, they couldn't get reliable or actionable answers. Now we have the technology to give them those answers. Qlik supplies flexible, intuitive analytics that everybody in the business can use.

And that's not all. Qlik also covers the back end of the Big Data challenge, giving you the ability to integrate data from multiple sources (for example, Hadoop repositories, data warehouses, departmental databases, and spreadsheets) into one single, interactive analytics layer.

How Big Data goes from raw to analysis-ready.

Imagine a piece of metal ore that begins life buried in the earth and is extracted from a mine, transported to plants, and refined into solid gold. Big Data goes through a similar process, traveling from its raw form to an analysis-ready state and becoming more refined as it goes:

Step 1: Gathering

Business-oriented Big Data usually comes from one of three places:

1. Machines or IoT — data streams, server logs, and RFID logs
2. Transactions — website activity, point-of-sale data
3. The cloud — stock ticker prices, social media feeds

This data is often:

1. Unstructured — strings of text or images, or
2. Semi-structured — log data with a timestamp, IP address, and other details

It usually has:

1. High volume — terabytes to petabytes
2. High velocity — many terabytes of new data per day
3. High variety — hundreds of servers and applications, each creating data in their own format

Step 2: Initial Processing

As Big Data is gathered, it's copied into a repository. Tech vendors like Amazon, Google, Microsoft all offer their own platform with a suite of related services. Open source tools like Apache (Hadoop) offer distributed, scalable, and portable file systems designed to run on commodity hardware which many businesses use. All of these tools give you some ability to aggregate and manipulate data, but only as a first step toward making it ready for analysis.

Don't leave data behind

It's called a Big Data repository, but that doesn't mean it contains all the data you need to analyze. Often, the repository hosts the supporting evidence for a discovery made in operational data or even in a spreadsheet. For example, a spreadsheet containing customer satisfaction survey results may be the basis for an analytic inquiry, and the data within the Big Data repository allows the user to correlate a customer's support history with their satisfaction scores. If you restrict your analysis to the data in your repository, you won't get the full picture.

That's why your analytics solution should have the capability to ingest data from a wide range of sources — including data warehouses, databases, and spreadsheets — and not just Big Data repositories.

Step 3: Refinement

Depending on your environment, you may also need to use an enterprise data warehouse (EDW) as an intermediary step to store structured data. EDWs don't just store data; they also extract, transform, and load it, so they complement Big Data repositories. And because data in EDWs is structured and not raw, it is easier to query; in other words, it's one step closer to analysis-ready form.

Step 4: Analysis

The final step is bringing your Big Data into an analytical tool, and different tools have different capabilities. The typical business user needs to integrate data from multiple sources without worrying about where it comes from or how it's organized, and they need data modeling to be fast. This type of environment not only reduces the burden on IT to keep up with business demands; it also gives business users the power to incorporate additional data in their analysis as they need to.

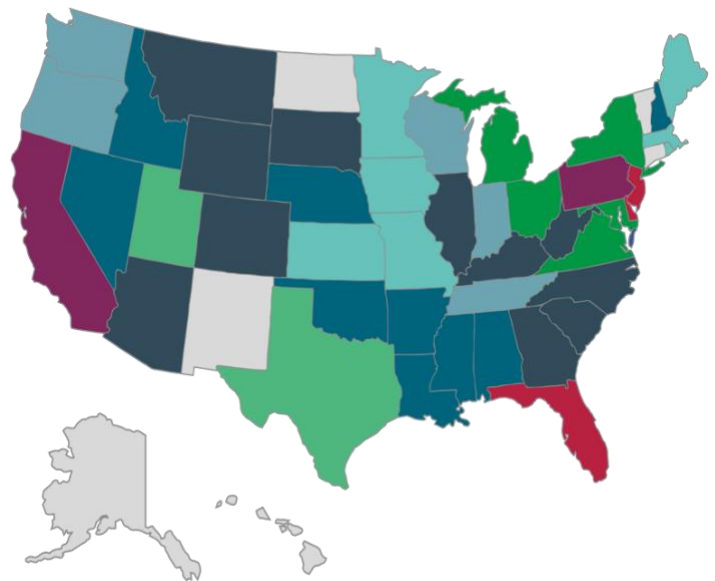
Using Big Data: Make it relevant and keep it in context

Most analytics solutions are designed for specialists, and they require advanced skills. Business users need their own way to navigate through massive amounts of data, find what's relevant, and get answers.

Two important aspects to consider when you're giving business users the ability to work with Big Data: the relevance and the context of the information.

Relevance: The right information to the right person at the right time

Qlik's approach is to start with what users need. And for users, access to the right data at the right time is more valuable than access to all the data, all the time. For example, a local bank branch manager would only want to understand the sales, customer intelligence, and market dynamics in their branch—not in the nationwide network. With that one consideration, we go from data overwhelm to data relevance.



Context: Allowing users to explore, dive deeper, and pivot

Qlik® products are different from other analytics solutions. Our one-of-a-kind Associative Engine dynamically associates every piece of your data with every other piece, in all your data sources. That makes it possible for users to dive in, explore freely, and always be surrounded by context. And whenever a user makes a selection, Qlik instantly updates all analytics and all associations to reflect that selection.

Let's look at an example. On the Qlik platform, a Sales by Region chart will be surrounded by related information — a Sales by Product chart or an interactive list of categories including date, location, customer, and sales history. If the user selects one region in the Sales by Region chart, the Sales by Product chart updates to show sales from the selected region. That makes it incredibly easy to focus on (for example) a particular product in a particular geography sold to a particular customer.

This associative experience becomes even more useful where there are hundreds or thousands of products, customers, and geographies. Rather than scrolling through thousands of items, users can slice massive datasets with a few clicks.

Using relevance and context, Qlik transforms what looks like a Big Data problem into a manageable experience for users — no special skills required.

King.com and Qlik

“Implementing Qlik has cost less than 20% of the alternative solutions. The payback period was just a few months.”

— Mats-Olov Eriksson, Main architect of the analytics system

Background:

- Worldwide leader in casual social games
- 150 games in 14 languages
- 40 million monthly players
- 2 billion rows of new log data per day

Use case:

- Analyze ROI of marketing campaigns
- Track updates of new game offers

Technology:

- Logs stored in 14-node Hadoop cluster
- Batch processing creates KPIs and aggregates in Hive
- Qlik connects via ODBC (open database connectivity) to Hive

Qlik's options for handling your Big Data

How you structure, access, and use your Big Data differs, of course, depending on what you do and what type of data you're working with. Qlik offers a range of built-in capabilities and multiple techniques that you can use individually or in combination:

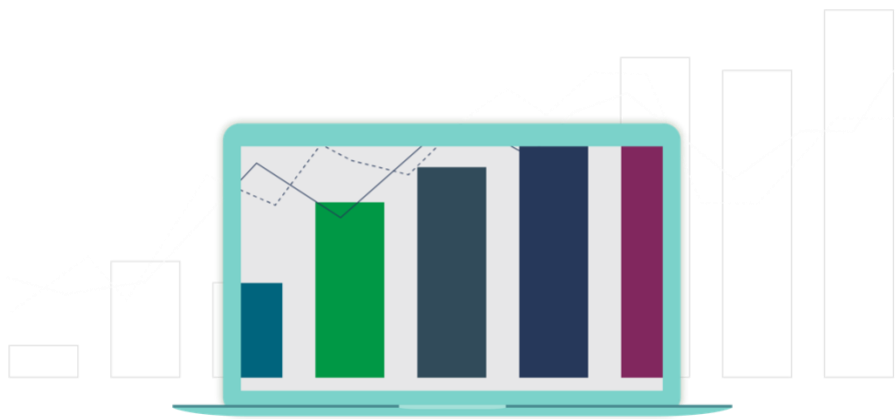
- In-memory
- Segmentation
- Chaining
- On-Demand App Generation
- Associative Big Data Index
- Custom and other methods

In some cases, one method will be enough. In others, you'll want to use several methods. As always, you'll have to navigate some tradeoffs between flexibility; user performance; and data volume, variety, and velocity.

In-memory

The Qlik Associative Engine optimizes the speed of exploration by automatically compressing data to 10% of its original size and storing it on the server, not the local machine — giving you fast response rates even with millions of rows. For many Qlik customers, this capacity alone is enough to satisfy their Big Data requirements.

And because the memory on standard computer hardware continues to grow in size and decrease in price, Qlik can handle larger and larger volumes of data. For example, a single 512GB server can handle uncompressed data sets near 4TB in size. With Qlik's compression scheme, the more redundancy in the data values, the greater the compression.



Unlike technologies that simply “support” multi-processor hardware, Qlik takes full advantage of all the power of multi-processor hardware. It distributes number-crunching across all available processor cores, maximizing performance (and your hardware investment). In a clustered environment, Qlik apps can be hosted on different servers. For example, an app containing a smaller amount of data can run on a server with less memory, while an app with larger amounts of data can run on a larger server. All of this, of course, is invisible to the user.

Another option is to deploy Qlik so that one server runs in the background, extracting and transforming large amounts of data, while another runs the user-facing app, free from the burden of back-end tasks. This multi-tiered architecture gives IT an added benefit: The transactional data only has to be extracted once and can then be reused in multiple Qlik apps.

Administrators can also configure Qlik to load only new data or data that’s changed since the last load. That greatly reduces the bandwidth required from any data source.

Segmentation

Segmentation is the process of dividing one Qlik application into multiple apps to optimize performance, security, scalability, and maintenance.

Users can segment data by region or department, or they can segment a small dashboard or summary app from another app containing the detailed data.

For example, say a retail company has a massive set of data, and they want to enable their executives, their business users, and a few power analysts to explore it. They can use segmentation to “break up” the data into chunks that serve those different groups. Each of these groups can use their app without the RAM and CPU required for the full version. And with Qlik, the segmented versions require almost no maintenance or overhead.

Chaining

Chaining is the linking (or jumping) from one Qlik application to another while maintaining “states” or selections that the user has already made.

Separate Qlik apps, even running on different servers, can share selection states. For example, a CRM application includes several customer subject areas, each corresponding to a department. Qlik can be configured to have a dashboard and comprehensive app of the overall customer base. These apps are then linked or chained to subject-area apps specific to each department. So, like segmentation,

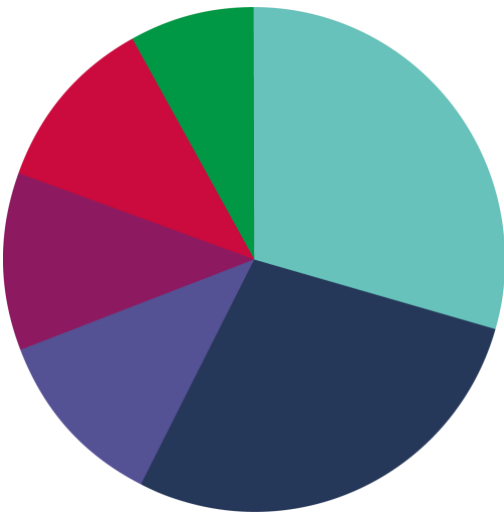
chaining is a method that allows you to manage multiple, streamlined apps that work together as opposed to one giant app.

Segmentation and Chaining

Segmentation and chaining can be used together. First, you segment multifaceted views into subject-specific views, and then you chain the subject-specific views together. For example, if a user selects a certain product code in one app and then jumps to another app, all of the visualizations and analytics in the new app reflect the previously selected product code.

On-Demand App Generation

On-Demand App Generation (ODAG) is a function that automatically creates an analysis app when the user selects a “slice” of data.



The vast majority of users don't want to analyze the entire Big Data source, so they need a way to slice the data into relevant sections. Segmentation (or segmentation and chaining) could work, but only if the user already knows what they need — and often, they don't. Plus, if different users want to look at different segments, the process becomes too cumbersome.

Segmentation also limits users' ability to explore freely in any direction. If they want to pivot to an area outside the selected segment, they'll have to start over with a new app.

ODAG solves both of those problems. First, the user works within a selection app, choosing from a “shopping list” of company-defined subsets like Time Period, Customer Segment, or Geography. This triggers the immediate creation of a detailed app containing the data selected. Within this detailed app, like any other Qlik app, the user can explore that data in any direction or even include other data sources. And since these apps are governed by the standard Qlik security rules, your administrator can control who can access the detailed data versus summary information.

With ODAG, users have the freedom to “fail fast,” investigating different slices of the data source without having to develop a new app every time they pivot. At the same time, administrators can give users broad access to massive troves of Big Data, since only the requested slice is being managed in-memory at any time.

Associative Big Data Index

The Qlik Associative Engine brings together very large quantities of data, fully indexes them, and keeps them in-memory, so users can freely explore and find all possible associations. The Qlik Associative Big Data Index uses the same proprietary technology to deliver the same associative experience — now applied to the biggest data repositories an organization can have.

The Associative Big Data Index fully combines multiple data sources and indexes them to find all possible associations, without leaving any data behind. Users are free to search, explore, and pivot based on what they see, with no limitations and no waiting. They have full access to all underlying details in their data, with no need to prepare the data elsewhere and transfer it for analysis.

Rather than bringing data to Qlik, the Qlik Associative Big Data Index brings Qlik to the data. Discoveries are always current, because the index is immediately updated with any updates to the data sources. The expensive and time-consuming step of transferring data into an intermediary repository is gone. And multiple apps can use the same indices, so it's easier for app developers and power users to coordinate and streamline development projects.

Custom and other methods

When you take advantage of the 1700+-member Qlik partner network and our open APIs, you have even more options for working with your Big Data. For example, you could develop a custom app using Javascript and the same APIs used by our ODAG apps, where user selections would trigger the creation of filtered data sets for analysis. The development of custom apps requires more advanced technical skills, but it gives you the freedom to precisely tailor the UI and functionality to your user's needs.

Side-by-side comparison: Options for handling your Big Data

There's no one right way to access and analyze your Big Data. Take a look at your user requirements and data sources, and decide which method or combination of methods makes the most sense.

	In-Memory	Segmentation & Chaining	On-Demand App Generation	Qlik Associative Big Data Index	Other Methods
What It Does	Highly compresses data into memory.	Users move among multiple related segmented apps (e.g., by region)	User selections trigger the creation of an app for analysis	Qlik's associative experience on top of Big Data	<ul style="list-style-type: none"> ● Partner technology ● Custom-built app in which user selections trigger the generation of a filtered data set for analysis via: <ul style="list-style-type: none"> ○ Multiple APIs in Qlik Sense ○ QMS API/EDX in QlikView ● Other custom solution
When To Use It	<ul style="list-style-type: none"> ● The compressed data source fits into server memory ● Only aggregated or summary data is needed ● Only record-level detail over a limited time period is needed 	A data source that's too unwieldy to be managed in server memory and <i>can</i> be split into pre-defined segments	A data source that's too unwieldy to be managed in server memory and <i>can't</i> be split into predefined segments	<ul style="list-style-type: none"> ● Want to directly utilize Qlik's Associative Engine against the entire Big Data repository ● Allow multiple apps to access a single index 	When you want functionality that goes beyond what the other options offer
Data Volume	100's of millions to billions of rows	100's of millions to billions of rows per segmented app	Billions of rows	Billions of rows	Billions of rows

Qlik and Big Data connectivity

Qlik is an open platform, and it gives you a number of built-in and third-party options for connecting to Big Data repositories.

- **ODBC Connectivity.** Out-of-the-box, Qlik includes drivers for connecting with Apache Hive, Cloudera Impala, and other software. You can also access other Big Data tools through the vendor's ODBC Connector.
- **Data-source specific connectivity.** We've partnered with multiple vendors to be certified on their ODBC drivers. Two examples: We're certified with MapR for Apache Drill and with SAP for their HANA ODBC driver.
- **Partner-developed connectivity.** A number of Qlik partners have developed additional connectors designed to work with data sources or applications that fall outside current Qlik connectivity. See this growing list at market.qlik.com.

Going the last mile with Big Data

Big Data analytics shares a common challenge with the telecom industry — the last mile.

Bringing telephone, cable, and Internet service to end points in customers' homes is expensive. Service providers have to fan out the network from the backbone, rolling out trucks, digging trenches, and installing lines. In some cases, providers pass high installation costs to the customer, or they neglect to go the last mile at all.

Today, most technology providers working on Big Data are focused on processing it — dealing with the backbone, not the last mile. They're not delivering analytics platforms that empower business users to make real discoveries with real business outcomes. That's what Qlik does.

Qlik gives every user the power to access Big Data, collaborate on it, use it in combination with traditional data sources, and explore it to make discoveries that make a difference. Businesses around the world are using Qlik to increase the value of their Big Data investment — by making sure it's not restricted to data scientists alone.



About Qlik

Qlik's vision is a data-literate world, one where everyone can use data to improve decision-making and solve their most challenging problems. Only Qlik offers end-to-end, real-time data integration and analytics solutions that help organizations access and transform all their data into value. Qlik helps companies lead with data to see more deeply into customer behavior, reinvent business processes, discover new revenue streams, and balance risk and reward. Qlik does business in more than 100 countries and serves over 50,000 customers around the world.

qlik.com

© 2020 QlikTech International AB. All rights reserved. All company and/or product names may be trade names, trademarks and/or registered trademarks of the respective owners with which they are associated

BIGDATA042720-M