



QLIKVIEW DEPLOYMENT FOR BIG DATA ANALYTICS AT KING.COM

QlikView Technical Case Study Series—Big Data



June 2012

qlikview.com



Introduction

This QlikView technical case study focuses on the QlikView deployment addressing big data with Hadoop at King.com—one of the largest online gaming companies in Europe. King.com is a worldwide leader in casual social games with over 40 million monthly players and more than 3 billion games played per month globally. King.com offers over 150 exclusive games in 14 languages through its premier destination, King.com (www.king.com), mobile devices, Google+, and Facebook, where it is a top 10 Facebook developer. The company is the exclusive provider of online games for leading global portals, websites and media companies.

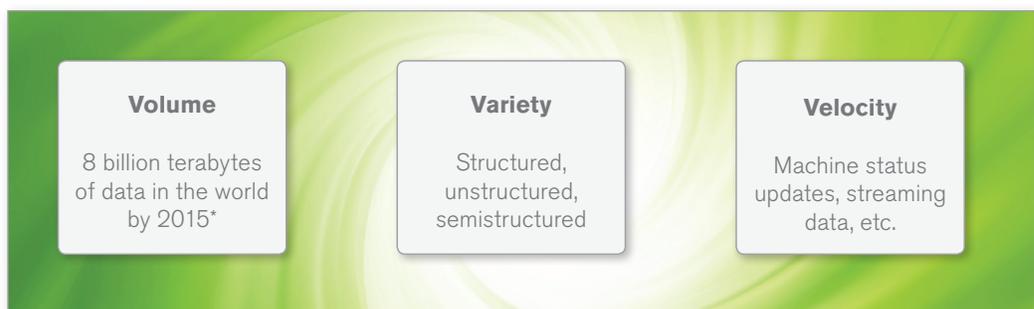
King.com is driven by passionate game developers and leading digital executives, and has offices in London, Hamburg, Stockholm, Malmo, Milan, Bucharest, San Francisco and Malta.

This paper explains how the QlikView Business Discovery platform solved the challenges King.com faced with big data and exposed its real value. It talks about the technical details of the QlikView deployment with the Hadoop solution and the business value that the solution provided.

Executive Summary

“Big data” is a term applied to data sets whose size is beyond the ability of commonly used software tools to capture, manage, and process the data within a tolerable elapsed time. Every day, 2.5 quintillion bytes of data are created. According to IDC, the volume of digital content in the world will grow to 8 billion terabytes by 2015. One aspect of the big data that makes its challenge unique is its three dimensions; Volume, Velocity and Variety (Figure 1.)

Figure 1. What is Big Data?



* IDC Predictions 2012: Competing for 2020, December 2011

On the other hand, big data is more than simply a matter of size; it is an opportunity to find insights in new and emerging types of data and content, to make businesses more agile, and to answer questions that were previously considered beyond the reach. Nowadays, every organization is challenged to find the solution to harvest this opportunity. King.com is no exception.

King.com overcame this challenge by using QlikView together with a Hadoop-based big data system to provide rapid insight into customer behaviors captured through their use of the games.

The main reason King.com chose QlikView was its flexible, associative technology, providing speed-of-thought, self-service analytics with big data, which was critical King.com's competitive success.

King.com is an analytics-driven organization. Every business user requires data to make decisions on a daily basis. These business users are from many areas of the business including managers from product, business, marketing, and advertising sales, and from games design and customer service. As these users started to gain a deeper understanding of the structure of the big data and got comfortable with using it, they required more self service capabilities where they can remix and reassemble the big data to gain new insights with the hundreds of dimensions available. That is when the IT department needed an analytics solution that can handle very large amounts of data in a self-service manner. The QlikView analytics applications at King.com contain millions rows of data enabling the business user to harvest the value of the big data. Although King.com has over 100 QlikView applications in daily use for the big data analysis, some departments leverage the ease and speed of development to quickly create “disposable” applications for rapid prototyping. Applications are often developed at a small scale initially and then rapidly scaled as required.

At King.com, QlikView also empowered the IT department with its flexible ETL capabilities. Mats-Olov Eriksson, who is the main architect of the analytics system at King.com explained how it uses QlikView to understand the data relationships that exist in the big data. “We use QlikView for explorative purposes when the data extraction and the transformations are designed and verified from the big data. The script editor, together with ODBC drivers by MapR and Hive as the infrastructure provider, allows us to create a substantial part of the ETL in the QlikView apps. We also use the QlikView Publisher as the tool to schedule the data reloads and to manage the system.”

This case study covers the technical deployment details King.com took for its QlikView deployment to analyze the data that is stored in its Hadoop cluster where 2 billion new rows are produced each day.

QlikView's unique associative technology and interactive user interface enabled the non-technical business users to navigate through the massive amounts of data to find what is relevant to them. The business users at King.com are able to search and analyze the big data in any way they want to. The solution provided rapid insight into customer behaviors, where more than 3 billion games are played per month globally. With the QlikView Business Discovery environment, the business users are able to analyze 40 million customers' gaming behavior to target new games and new customers. With QlikView, King.com was able, for the first time, to analyze the return on investment for marketing campaigns. For example, it can track uptake of new games offers. Business users were able to track and measure metrics such as the number of players, the number of games played, and time played instantly.

Technical Details

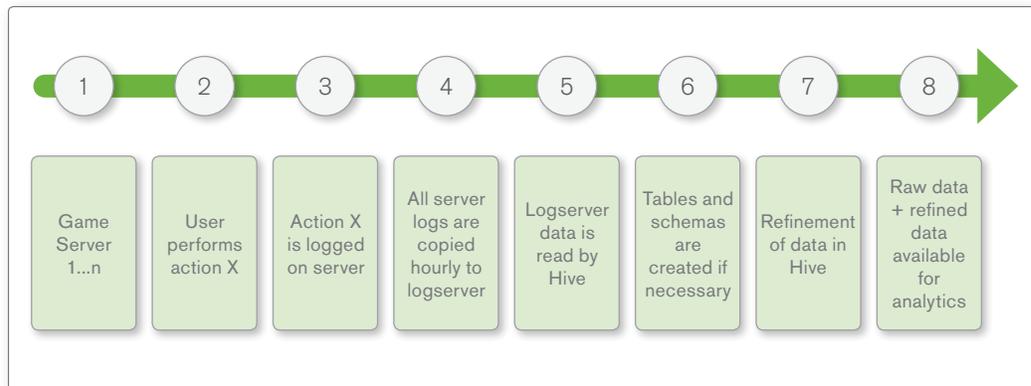
King.com uses a Hadoop-based big data solution to store massive amounts of gaming activity and customer data. Hadoop is a software framework which uses a distributed file system (normally HDFS) where the data is stored as flat files across several nodes. Usually inexpensive local hard drives are used in a Hadoop environment providing cheaper data storage and processing solution.

Hadoop provides the Map-Reduce framework to store and retrieve data, which creates one of the main limitations to extract data from Hadoop. For each query, a program should be developed by using the MapReduce framework. In most of the Hadoop environments, Hive, which is a data warehouse system for Hadoop, is used to run ad-hoc queries, and the analysis of large datasets.

King.com's technical infrastructure includes game servers, log servers, the Hadoop environment and the QlikView environment. It utilizes a 14-node cluster to host its Hadoop environment. "As the data size grows, we will be expanding the Hadoop infrastructure" says Mats-Olov Eriksson. Each user's 'event' is first logged locally on the game servers and then the information is copied hourly to a centralized log server. The log server files are then copied to the Hadoop environment and processed with MapReduce programs.

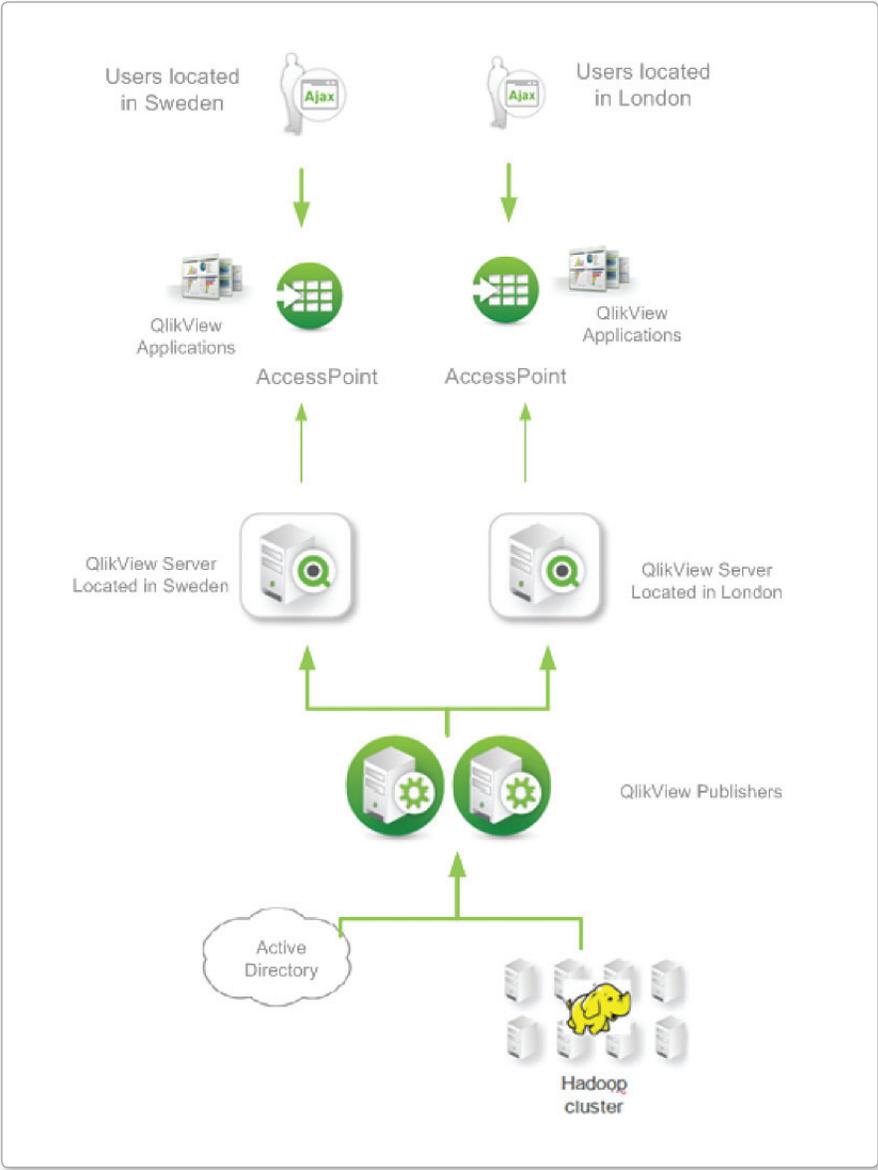
It has hourly processing of the data to populate a limited dashboard with KPI's like game installs, revenue and game play to provide near real time analytics to its business users. The main batch processing for analytics happens daily to create the KPIs and aggregated views in HIVE, which is then made available to QlikView applications for analysis via an ODBC connector connecting to HIVE.

Figure 2. Big Data Processing and Analytics Creation



King.com's QlikView environment consists of two QlikView Servers and two QlikView Publishers. Its QlikView servers are geographically dispersed and it uses QlikView Publisher for the distribution of QlikView applications across QlikView Servers. One of its QlikView Servers is located in London and the other one is located in Sweden. QlikView Publisher is also used to schedule the data extracts from HIVE into the QlikView environment. Further refinement of data for different QlikView applications is also done through QlikView Publisher.

Figure 3. QlikView Production Environment Architecture



Analytical Challenges

The main analytical challenge for King.com was to create a self-service BI environment for its business users. This challenge was exacerbated by the volume and the velocity of its big data. Using Hadoop and MapReduce was only the first piece of the solution for the big data challenge. The real value it was seeking was to give everyday business users access to the relevant parts of the big data and enabling them to derive insights from it.

The query performance was also another challenge. Although it used HIVE to query the big data in the Hadoop system, the performance of it was very slow as HIVE is not designed for OLTP workloads and does not support real-time queries.

King.com's gaming system generates 2 billion rows of data per day and this volume will continue to increase in the future. Analyzing the data without disturbing the game load was an important performance requirement. Another requirement was to have a simplistic analytics and reporting system in order for King.com to have game development teams geographically separated from the platform development. And finally, as its business grows very fast, the requirements for analytics were more sophisticated than ever. Having the data available for complex queries and analytics with fast performance was a necessity.

The IT department was challenged to empower business users with self-service analytics and give them an excellent experience that will keep them coming back for more. Business users wanted to explore the data on their own and able to slice and dice the data by many permutations of the hundreds of dimensions available in the big data.

Solution

King.com chose the QlikView Business Discovery platform because of its unique associative technology and interactive user interface enabling the daily business users to navigate through the massive amounts of gaming data to find what's relevant to them. With QlikView, business users can navigate and interact with the big data any way they want to; they are not limited to just following predefined drill paths defined on the HIVE queries. More importantly, they can literally see relationships in the big data with QlikView's unique associative capability and leverage all of the dimensions of the big data with any combinations during its analysis.

The solution also provided 'speed of thought' analysis on King.com's big data. As QlikView holds all the data needed for analysis in memory, the relevant piece of the big data is available for immediate exploration by the business users. They experience zero wait time as QlikView performs the calculations needed to deliver the analysis that users request on the big data.

Every day, aggregated data are extracted from the Hadoop environment to the QlikView environment. For analytical purposes, it was not necessary for the business users to extract every single record level of the data that is stored in Hadoop. The business users have a high degree of confidence that the aggregated data is highly representative of the entire data set within Hadoop as the main aggregation level is the marketing channel. King.com's marketing department has a very granular level of channel definitions that group the players based on many permutations of the demographic dimensions. The granularity that exists in the marketing channels reduces the statistical impact of the aggregation effect. Another reason for the aggregation was to provide consistency across different analytics because of the high velocity of its data.

King.com created many QlikView applications with different data aggregation levels, providing complete Business Discovery capabilities to the business users. Different levels of data is used on each QlikView application providing a granular level of analysis to the business users as they ask more detailed business questions during their analysis.

The Business Discovery solution provided analytical flexibility where the business users can, for the first time, analyze the return on investment of the marketing campaigns, measure daily average revenue per user, continuous second day retentions, daily and monthly active users and many more KPIs by many dimensions. Before deciding to implement QlikView, marketers would ask games developers to write scripts to track, for example, the number of games played per channel. This was not sustainable, so King.com decided to search for a solution that would handle its big data needs, as well as the diverse requirements of numerous departments and business users. After selecting QlikView, the first application was in production within a week. Commenting on the return on investment, Mats-Olov said: "Implementing QlikView has cost less than 20% of the alternative solutions. The payback period was just a few months."

"Another big advantage of using QlikView was to be able to extract data from different data sources" says Mats-Olov Eriksson. QlikView allowed customized metadata to be applied as external tables from various sources and to be used with the big data extracted from the Hadoop system. According to Eriksson, QlikView's associative data model and its capability to extract and merge data from different data sources helped his business intelligence team become a metadata driven business intelligence unit. Eriksson emphasized the importance of metadata when analyzing the big data and said: "QlikView enabled us to add new grouping, external metadata and improve the nomenclature of the big data without having to do anything else but reloading the data into QlikView applications. QlikView automatically associates the metadata that is stored in a separate system with the Hadoop Big Data during the data reloads. This provided great flexibility to us as we can change and add the metadata information without changing anything in the Hadoop system, QlikView takes care of merging the new metadata with the big data."

Conclusion

Analysts and business users at King.com were constantly challenged to efficiently access, filter, and analyze massive gaming data for better decision making. By implementing the Hadoop solution, it was able to solve the efficient storage and processing challenge of the big data but making this data easily accessible and relevant for the business users was still a challenge until it deployed QlikView.

QlikView's unique associative technology and interactive user interface enabled the daily business users to navigate through the massive amounts of data to find what's relevant to them. The business users at King.com are able to search and analyze the big data in any way they want to. The solution provided rapid insight into customer behaviors, where more than 3 billion games played per month globally. With QlikView Business Discovery environment, the business users are able to analyze 40 million customers' gaming behavior to target new games and new customers. With the QlikView solution, the return on investment of the marketing campaigns were achieved for the first time and the business users were able to analyze and measure metrics such as the number of players, the number of games played, time played, daily average revenue per user, continuous second day retentions, daily and monthly active users and many more KPIs by many dimensions.

QlikView also enabled the IT department to understand the structure of the big data with its flexible ETL capabilities. With QlikView, the business intelligence team at King.com became a metadata-driven business intelligence unit. By leveraging QlikView's associative data model, it was able to load external metadata information and merge it with the big data. This simplified the big data classification process and helped the team provide the relevant part of the big data to the appropriate business users' groups.

References

QlikView and Big Data

<http://www.qlikview.com/us/explore/resources/whitepapers/qlikview-and-big-data>

Big Data Meets Trade Promotions

<http://www.qlikview.com/us/explore/resources/whitepapers/big-data-meets-trade-promotions>

QlikView Google BigQuery Integration

<http://www.qlikview.com/us/explore/resources/whitepapers/qlikview-google-bigquery-integration>

QlikView Development and Deployment Architecture Technical Brief

www.qlikview.com/.../global-us/direct/datasheets/DS-Technical-Brief-Dev-and-Deploy-EN.ashx

QlikView Architecture and System Resource Usage Technical Brief

www.qlikview.com/.../DS-Technical-Brief-QlikView-Architecture-and-System-Resource-Usage-EN.ashx