

OPERATING AND ANALYSING OF BIG DATA OVER PENTAHO BUSINESS ANALYTICS, TABLEAU AND QLIK

Mariana Kovacheva

асистент, доктор, катедра „Информационни технологии и комуникации“, УНСС
e-mail: mkovacheva@unwe.bg

Abstract

Technologies and digitalization are two of the biggest driving forces for generating, gathering, working, and analysing Big Data. This paper examines three analytical tools – Pentaho Business Analytics, Tableau and Qlik, their characteristics, advantages, and disadvantages for working with Big Data – structured, semi-structured and unstructured.

Keywords: Big Data, structured data, semi-structured, unstructured data

Introduction

The digitalization and digital transformation are two of the most important processes happening in the last decade and will continue to be so in the near future. Both are a reason for generating big amounts of data, known also as Big Data. All the data that is gathered can be from one of the three types – structured, semi-structured and unstructured, and shouldn't be only gathered but also analyzed and used for taking decisions.

Tools for analyzing data

In today's world with the vast number of different types of digital devices – from smartphones, tablets, laptops, smartwatches, other types of smart devices that we are using in our everyday life, sensors, etc. which are gathering data every second of their existence, the problem does not end with the data being gathered, but only starts with it. All the types of data – structured, semi-structured and unstructured – after being gathered, need to be stored properly, processed then analysed and after that conclusion to be made, based on it. Big Data presents many advantages for the people and organizations using it, but also many challenges as working with different sources of data, which from side need new means and methods for processing the data, but also new skills and vision for working with it. [1]

Nowadays, there is large selection of different tools for analysing data. Ranging from simple spreadsheets software (like Excel, Google Spreadsheets, etc.) to very advanced statistical or machine learning platforms (like Apache Spark, SAS, etc.). The choice of a tool for analysing data depends highly on many requirements and questions that a company should ask itself before choosing. [2]

Part of the things that need to be considered when choosing an analytical tool are the shown in Table 1.

Table 1 Requirements for choosing analytical tool

Questions	Variations
What is the type of data that is gathered?	There is a possibility for the company to work with one of the following types structured, semi-structured or unstructured, or even with more than one.
What is the size of the data?	Is the company generating small files of MBs or maximum GBs, or is generating Big Data – PB, ZB, etc.
Who will use the analytical tool?	Experts like data analysts or data scientists, or marketers and salespeople.
What types of visualizations are needed?	Some analytical tools offer visualizations that another cannot make and vice versa. For this

	reason, is important for the company to clear what type it will need.
What is the budget?	Another important question is what the price of the tool be. There are different options from free, open source, paid tools. This choice varies depending on the previous questions – data types that are going to be gathered and analysed, who will use the tool, size of the data, are specific visualizations needed.

Based on the questions in the table above and their answers, it becomes easier for a company to make a proper choice for analyzing tool for their needs.

Comparison of tools for analyzing data

This paper examines Tableau, Pentaho Business Analytics and Qlik as analytical tools, which according to Gartner are in the Top 20 of their classification for Analytics and Business Intelligence Platforms Reviews and Ratings. [3]

Tableau

Tableau is a data visualization tool and business intelligence software that allows the users to connect, visualize and share data interactively. The instrument is used to show data in interactive and understandable way. [4]

Tableau is used for [5]:

- 1) **Data visualization** – transforming the data into easily understandable and readable charts, maps, and graphs.
- 2) **Business intelligence** – insights from data for the business to make informed decisions.
- 3) **Data analysis** – for exploration and analysing big sets of data.
- 4) **Dashboard creation** – for creating interactive real-time dashboards.
- 5) **Operational reporting** – for creation of reports.

Advantages and Disadvantages of Tableau

Tableau is one of the most popular business intelligence tools on the market. It has user friendly interface which allows the user to make the needed dashboards and reports with minimum effort. With the fact that Tableau supports different data sources like spreadsheets, databases, big data platforms make it easy to connect and analyse data. The tool offers great variety of interactive visualizations, which enable the user to customize its needs and demands easy. Tableau is scalable and can work with small and big data sets, to meet the needs of every type of business. Finally of its advantages, shown on Figure 1, is the fact that Tableau can be integrated with different data sources and with other business intelligence tools. [6][7]

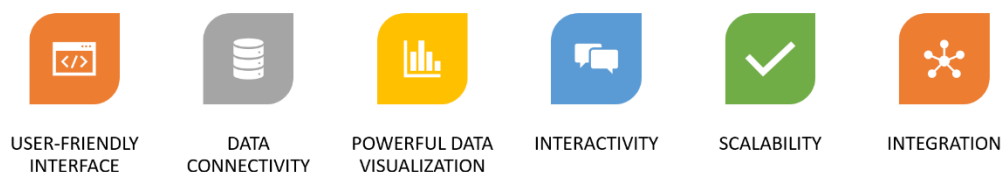


Figure 1 Advantages of Tableau

As every tool on the market, Tableau has not only positive sides, but also negative which are shown on Figure 2. The BI tool can be expensive for larger organizations and the cost may include licenses, support, training, etc. Even that Tableau is scalable and can work with different sources, when it's working with extremely large datasets, some issues on the performance may arise. On the other side, some of the complex features of Tableau to be used, may require additional learning for the user to maximize the potential of the software. One of the other downsides of Tableau is that the data needs to be pre-processed – cleaned and transformed in other tools before coming to the BI tool. Also, Tableau doesn't support statistical analysis and some features for advanced analytics and interactivity are limited or unavailable when Tableau is offline. [6] [7]



Figure 2 Disadvantages of Tableau

Pentaho Business Analytics

Pentaho Business Analytics is a business intelligence platform which is an open source and provides data integration, reporting, data visualization and analytics.

Pentaho Business Analytics is used for [9]:

- 1) **Data Integration** – extracting data from various sources, transforming it into usable format and loading the data into a system that will work with it. This is also known as ETL or Extract, Transform and Load. This process is very important for organizations that want to gather data from different data sources.
- 2) **Business Intelligence Reporting** – this tool allows creation of reports based on the data loaded.
- 3) **Data Visualization and Dashboards** – the BI tool provides instrument for creating dashboards and visualizations which are interactive. The users can design their dashboards and communicate the insights within the organization.
- 4) **Predictive Analytics** – allows to discover trends and patterns inside the data.
- 5) **Aggregation and Performance Optimization** – designs and manages aggregate tables, optimizes queries performance in the data warehouse, which is very important when talking about large datasets.
- 6) **Ad hoc Reporting and Analysis** – the tool allows creation of reports on the go, which is important for companies that need fast visualization and analysing of data.
- 7) **Data Warehousing** – can be used for easier ETL processing.
- 8) **Open-source customization** – as the platform is open-source can be customized by the specific needs and demands of the organization, which is very useful for companies with specific data formats.
- 9) **Integration with other systems** – Pentaho Business Analytics integrates good with different databases, sources, tools, and different infrastructure.

Advantages and Disadvantages of Pentaho Business Analytics

Pentaho Business Analytics has many advantages shown on Figure 3. One of the biggest pros of the BI tool is that is open source, which means that because the source code is open, it can be modified according to the needs of the company and can be fitted in budget constraints. The tool is designed to be scalable, which means that with the growing of a small company to big one, Pentaho can still be the software that is used, because can also work good with large datasets. In addition to that, Pentaho has good integration capabilities and work with different data sources, formats, databases, etc. [10] [11]



Figure 3 Advantages of Pentaho Business Analytics

As every tool on the market, Pentaho Business Analytics also has some disadvantages, which depending on the company and the case, may become crucial for the selection of the tool – Figure 4. The analytics capabilities of Pentaho are not as advanced as of the specialized tools and especially in the advanced predictive analytics or machine learning. The BI tool has limited visualization capabilities, even that provides reporting and dashboards, the quality and advancement is not as good as this of the specialized tools. Pentaho is built on Java and if the organization is using as a tool, there might be need of Java knowledge. Some companies may prefer other technologies. Pentaho requires bigger amounts of resources as computing power and memory; this also highly depends on the task and on the size of the dataset. Some of the features of the tool are not available in the open-source edition and if the company needs some advanced features, which are only available in the commercial version, this will add additional cost to the organization and directly becomes disadvantage. [10] [11]



Figure 4 Disadvantages of Pentaho Business Analytics

Qlik

Qlik is a business intelligence tool and data visualization platform. It is known for its associative data model which enables the users to explore the relationships between the data. According to Gartner Magic Quadrant, Qlik is the leader in the sphere in data integration, quality and analytics solutions which are offered by the tool [12].

Qlik is used for [13]:

- 1) **Business Intelligence** – this tool is used for analysing and interpreting the organization's data. Qlik has associative data model which allows to explore the data relationships and to make more informed decisions for the business.
- 2) **Data Visualization** – interactive and visually appealing dashboards and reports can be created. Through the creation of different types of dashboards, graphs, charts, etc. users can make more informed decisions. Qlik allows the customization of dashboards and reports as in this way the KPI's are well communicated through the organization.
- 3) **Data Integration** – allows connection and integration of data different sources, which allows companies to view their data connected at one place.
- 4) **Collaborative Analytics** – users can collaborate in between each other as they can share dashboards and visualizations.
- 5) **Predictive Analytics** – Qlik has abilities for predictive analytics, even it is not the tool main aim, but the users are able to use it for forecasting trends and to make predictions.

Advantages and Disadvantages of Qlik

As every tool, Qlik has its own advantages and disadvantages, which will be examined in the following paragraphs. The first advantage is that Qlik is scalable, which means that its power can grow depending on the needs of the company, which usually means bigger datasets to work with. The data visualizations of the BI instrument are very powerful, which enables the users to create easily interactive and understandable dashboards. The associative model of Qlik makes it easy and flexible for the users to look and explore the relationships in the dataset. Qlik supports integration of data from various sources, which makes it easier for companies to work with unified data sources – databases, files and many more repositories. It is very user-friendly and easy to use, only with drag and drop the user can create dashboard. [14]

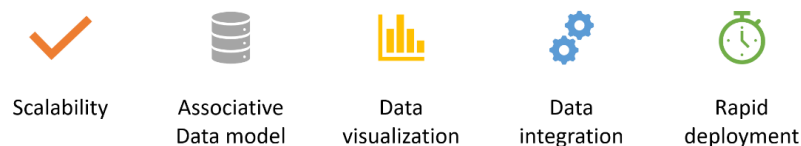


Figure 5 Advantages of Qlik

Qlik has limited predictive analytics and organizations which need advanced predictive modelling will need specific tools for this aim. Another disadvantage is the price of Qlik, because of licensing and another costs, especially for bigger companies the price can be very high. Another thing that can make organizations choose a different tool is that Qlik is very resource intensive, which depends highly on how big the data sets are and how complex should the visualizations be. The training of

personnel who is using Qlik may be needed to fully use the tools capabilities. Finally, Qlik highly depends on the data quality. If the data is incomplete, inaccurate, or misleading in some way, the results and analysis won't be accurate. [15]



Figure 6 Disadvantages of Qlik

Comparison of Tableau, Pentaho Business Analytics and Qlik

Based on the previous points and the detailed information that we looked in this paper we can make a comparison table of the three analytical tools in Table 2 [16] [17].

Table 2 Comparison of Tableau, Pentaho and Qlik

	Tableau	Pentaho	Qlik
Data Integration	More focused on visualizations.	Focus on ETL processes.	Data integration
Ease of Use	Most user-friendly	Less user-friendly than the three	User-friendly
Cost	More expensive with licenses and everything needed for its use.	Free open-source version, the enterprise one is expensive.	More expensive with licenses and everything needed for its use.
Scalability	Less scalable than Qlik	Less scalable than Qlik	Most scalable of the three, because of its in-memory processing and associative data model.
Support	Largest and active community.	The support may be dependent on enterprise edition.	Big and active community.
Customization	The tool provides customization in terms of visualizations and dashboards.	It's customizable, but Pentaho requires more technical knowledge.	The tool provides customization in terms of visualizations and dashboards.

Conclusion

Digitalization and digital transformations are two of the most important driving forces nowadays. All the data that is generated needs to be not only gathered, but to be properly processed and analysed for the companies to be able to use it for their business problems. In this paper we examined three of the most popular analytical tools, which are used not only for visualizations and dashboards, but for ETL processes, data integration and customizations for the need of the business to be met. A comparison of the most important features of the three tools was made.

References

- [1] S. Yordanova and K. Stefanova, 'Big Data Challenges - Definition, Characteristics and Technologies', no. 1, pp. 13–31, 2019.
- [2] J. Zhang, '10 Data Analysis Tools for Beginners and Experts', Medium. [Online]. Available: <https://towardsdatascience.com/10-data-analysis-tools-for-beginners-and-experts-2d083203b06e>
- [3] 'Analytics and Business Intelligence Platforms Reviews and Ratings'. [Online]. Available: <https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>
- [4] 'What is Tableau?', Tableau. [Online]. Available: <https://www.tableau.com/why-tableau/what-is-tableau>
- [5] 'What is Tableau? | Beginner's Guide to Key Features', DATADRIVE. [Online]. Available: <https://godatadrive.com/blog/what-is-tableau>
- [6] 'Advantages and Disadvantages of Tableau', JavaTpoint. [Online]. Available: <https://www.javatpoint.com/advantages-and-disadvantages-of-tableau>
- [7] 'What is Tableau? A Guide to Tableau Features and Advantages', BeyondKey. [Online]. Available: <https://www.beyondkey.com/blog/what-is-tableau/>
- [8] D. Tarnaveanu, 'Pentaho Business Analytics: a Business Intelligence Open Source Alternative', vol. III, pp. 23–34.
- [9] 'Pentaho Platform: Pentaho Business Analytics - Datasheet', Hitachi Vantara. [Online]. Available: <https://www.hitachivantara.com/en-us/web/datasheet/lumada-dataops-pentaho-business-analytics-datasheet.html>
- [10] 'Pros and Cons of Pentaho', Trust Radius. [Online]. Available: <https://www.trustradius.com/products/pentaho/reviews?qs=pros-and-cons#overview>
- [11] N. Singh, 'Pentaho and Big Data'. [Online]. Available: <https://www.codingninjas.com/studio/library/pentaho-and-big-data>
- [12] 'What is Qlik?', Qlik. [Online]. Available: <https://www.qlik.com/us/>
- [13] K. Schlegel *et al.*, 'Magic Quadrant for Analytics and Business Intelligence Platforms', Apr. 2023. [Online]. Available: <https://www.gartner.com/doc/reprints?id=1-2D773G95&ct=230411&st=sb>
- [14] L. Katerova, 'All You Need to Know About Qlik Cloud. Moving From On-Premise to Cloud.' [Online]. Available: <https://b-eye.com/blog/all-you-need-to-know-about-qlik-cloud/>
- [15] 'QlikView Advantages and Disadvantages | QlikView Benefits', Data Flair. [Online]. Available: <https://data-flair.training/blogs/qlikview-advantages-and-disadvantages/>
- [16] R. Kaur, 'Pentaho vs Tableau: Which Business Intelligence Software is the Winner?' [Online]. Available: <https://www.selecthub.com/business-intelligence/pentaho-vs-tableau/#:~:text=Tableau%20supports%20only%20row%2Dlevel,tool%20overall%20for%20this%20category.>
- [17] 'Qlik Vs Pentaho: In-Depth Comparison'. [Online]. Available: <https://6sense.com/tech/data-visualization/qlik-vs-pentaho>