

Ocena użyteczności współczesnych narzędzi IT w analizie danych Big Data

Rafał Wasiluk*, Piotr Muryjas

^a Politechnika Lubelska, Instytut Informatyki, Nadbystrzycka 36B, 20-618 Lublin, Polska

Streszczenie. W niniejszej publikacji poruszono temat współczesnych narzędzi IT w analizie danych Big Data. Przeprowadzone badania dotyczyły liderów rankingu Gartnera z 2017 roku, czyli Tableau Desktop, Qlik Sense i Microsoft Power BI. Ocena użyteczności tych narzędzi, została przeprowadzona zgodnie z wybranymi kryteriami, m. in. ceną i rodzajem licencji, rodzajami połączeń z bazą czy obsługiwany platformami operacyjnymi.

Słowa kluczowe: analiza danych Big Data; Tableau; Qlik; Microsoft Power BI

* Autor do korespondencji.

Adres e-mail: wasiluk.rafal@gmail.com

The assessment of usefulness modern IT tools of data analysis Big Data

Rafał Wasiluk*, Piotr Muryjas

^a Institute of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract. This publication discusses the topic of contemporary IT tools in Big Data data analysis. The research involved the leaders from Gartner 2017, Tableau Desktop, Qlik Sense and Microsoft Power BI programs. The utility rating of these tools has been evaluated according to the selected criteria, among others. The price and type of licenses, types of database connections, and supported operating platforms

Keywords: data analysis; Big Data; Tableau Desktop; Qlik Sense; Microsoft Power BI

* Corresponding author.

E-mail address: wasiluk.rafal@gmail.com

1. Wstęp

W dzisiejszych czasach codziennie masowo generują się duże ilości różnorodnych danych cyfrowych. Ten wzrost danych spowodował konieczność wprowadzenia nowych technik analizy danych, a także wywarł duży wpływ na działalność zarówno małych, średnich, jak i dużych przedsiębiorstw. Zarządy i rady nadzorcze wielu firm sięgają po narzędzia IT, które są w stanie usprawnić działania.

Pojęcie Big Data na przestrzeni lat interpretowane było wielorako. To zagadnienie najczęściej opisywano przez pryzmat cech, z których się składa. Początkowo o Big Data mówiło się tylko w oparciu o koncepcję trzech cech w modelu „3V”, takich jak objętość (ang. *volume*), różnorodność (ang. *variety*) i szybkość przetwarzania (ang. *velocity*). Z kolejnymi latami pojęcie Big Data zaczęło być definiowane w kontekście kolejnych cech: zmienności (ang. *variability*) i złożoności (ang. *veracity*), wartości informacyjnej (ang. *value*) oraz wizualizacji (ang. *visualization*).[1]

Bankowość, telekomunikacja, turystyka, ubezpieczenia, e-biznes i energetyka – to tylko niektóre z branż, w jakich współcześnie obecna jest analiza Big Data. W przypadku banków, nowoczesne narzędzia pozwalają na przykład na zbadanie wieku klientów, najczęściej korzystających z kart kredytowych. Jeżeli chodzi o telekomunikację, analiza danych Big Data pozwala choćby na zbadanie średnich miesięcznych rachunków klientów, którzy zrezygnowali z usług.

Na rynku można znaleźć bardzo dużo programów służących analizie, jednak celem pracy badawczej, opisanej w niniejszym artykule, była ocena tych narzędzi IT, które znajdują się na pierwszych miejscach w najnowszym rankingu Gartnera z 2017 roku [2]. Oceniono użyteczność Tableau Desktop, Qlik Sense i Microsoft Power BI, które dodatkowo zajmują czołowe miejsca wśród najbardziej wymaganych programów przez pracodawców na portalu Pracuj.pl.

Pierwsza część dotyczy obszarów zastosowań Big Data i wyzwań dzisiejszego świata dla wielkich danych. Pojęciem pokrewnym do analizy danych Big Data jest gromadzenie danych, związane z takimi zagadnieniami jak hurtownie danych, Hadoop i NoSQL. Temu zagadnieniu poświęcono drugi artykuł. Z kolei trzeci artykuł dotyczył współczesnych narzędzi IT do analizy danych. Opisano w nim wybrane programy, a także przedstawiono kryteria i metody oceny. W ostatniej części zaprezentowano ocenę użyteczności wybranych programów.

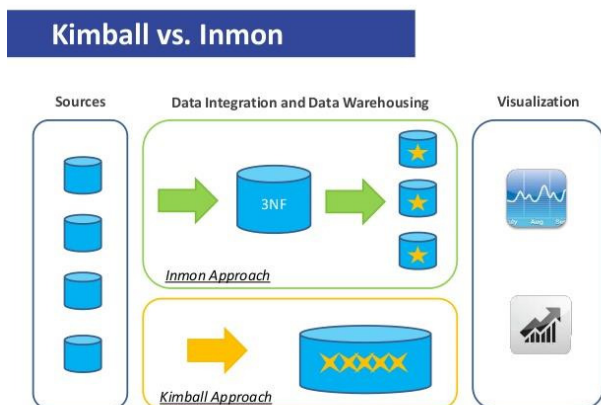
Celem badań była ocena użyteczności programów do analizy danych Big Data: Tableau Desktop, Qlik Sense i Microsoft Power BI zgodnie z przyjętymi kryteriami, m.in. takimi jak cena i rodzaj licencji, rodzaje połączeń z bazą, obsługiwane platformy operacyjne, czy możliwość analizy ad-hoc, in-memory czy w chmurze. Celem oceny było wskazanie najlepszego narzędzia IT do analizy danych z punktu widzenia, młodego analityka, kończącego studia i wchodzącego na rynek pracy.

2. Gromadzenie danych

Zanim nastąpi proces analizy danych Big Data, wcześniej wielkie dane gromadzone są w hurtowniach danych. Z tym zagadnieniem wiążą się także pojęcie Hadoop i model bazy danych NoSQL.

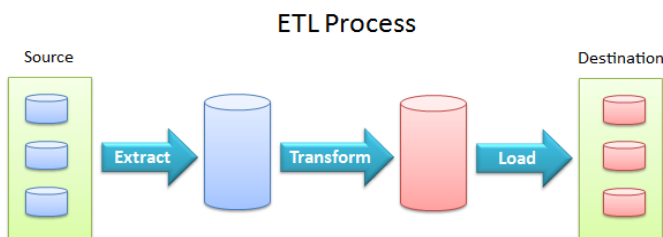
2.1. Hurtownie danych

Hurtownie danych, czyli *data warehouses*, w literaturze opisywane są zarówno w szerszym, jak i węższym ujęciu. Dla Inmona jest to jedynie baza danych o określonych cechach, podczas gdy Kimball wskazuje obecność dodatkowych elementów, takich jak ekstrakcja danych, ich analiza i narzędzia lub ich zestawy, umożliwiające realizację tych zadań (rys 1.).



Rys. 1. Porównanie definicji hurtowni danych Kimballa i Inmona [3]

Hurtownie danych zasilane są w procesie ETL, którego zadaniem jest integracja danych, pochodzących z różnych źródeł. Ten proces wizualizuje rys. 2.



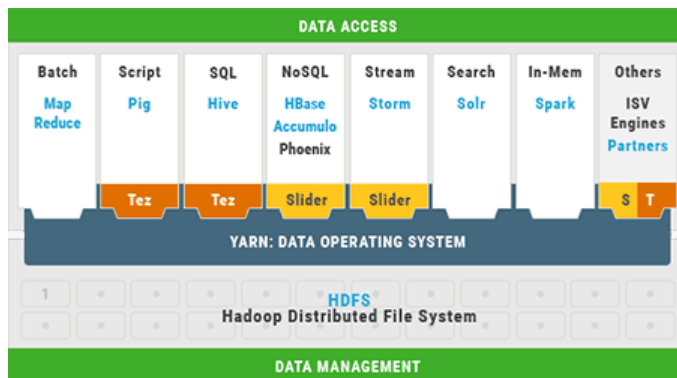
Rys. 2. Proces ETL [4]

Na pojęcie ETL składają się trzy procesy: pobieranie (ang. *extract*), przekształcanie (ang. *transform*) i ładowanie (ang. *load*) [4].

2.2. Hadoop

Przed zdefiniowaniem pojęcia Hadoop, warto wcześniej wspomnieć o MapReduce – systemie, przeznaczonym do tworzenia aplikacji, działających jednocześnie na tysiącach komputerów. Hadoop jest jednym z narzędzi Big Data, służącym do przetwarzania danych, głównie w postaci nieustrukturyzowanej. Apache Hadoop został napisany w języku Java. Hadoop, jako nazwa, stał się synonimem rodziny aplikacji do pracy z Big Data. Apache Hadoop tworzą 4 części: biblioteki bazowe Hadoop, Hadoop File System, Yet Another Resource Negotiator i Hadoop MapReduce. [5]

Wśród najbardziej znanych firm, korzystających z Hadoop, są Cloudera, Hortonworks i MapR. Na rysunku 3 przedstawiono wykorzystanie Apache Hadoop przez firmę Hortonworks [5].



Rys. 3. Wykorzystanie Apache Hadoop przez firmę Hortonworks [5]

Z roku na rok Hadoop znajduje swoje zastosowanie u coraz większej liczby klientów, głównie wśród gigantów branży informatycznej, takimi jak Facebook, Google, IBM, YAHOO!, Amazon, Twitter czy LinkedIn [6].

2.3. Model bazy danych NoSQL

NoSQL, czyli Not Only SQL, jest nierelacyjnym modelem bazy danych. Jego zastosowanie ma ze swojej definicji pozwolić na lepsze skalowanie, niż sam SQL. Mechanizm przechowywania danych NoSQL to „magazyn” danych, zapewniający przechowywanie i pobieranie danych w nieograniczony sposób [7]. Różni się od powszechnie stosowanego modelu relacyjnego. Dzięki jego wdrożeniu można osiągnąć wiele korzyści, takich jak poprawa zrozumienia danych, możliwość gromadzenia danych niestrukturalnych, skalowalność i elastyczność bazy oraz potencjalna możliwość tworzenia unikalnych modeli biznesowych. Porównanie NoSQL i SQL przedstawia tabela 1.

Tabela 1. Porównanie NoSQL i SQL (opracowanie własne)

	NoSQL	SQL
Przechowywanie	Obejmuje wiele typów baz danych	Relacyjna baza danych
Elastyczność	Schematy dynamiczne, informacja aktualizowana na bieżąco	Zapis zgodny z ustalonym schematem, brak możliwości aktualizacji danych dynamicznie
Zgodność ACID	Większość systemów posiada zgodność ACID	Domyślnie zgodne z ACID
Dostęp	Dostępny w dobrze zdefiniowanych wzorach	Brak wzorców, skupia się na definicji indeksu

Za przykład bazy danych, opartej na modelu NoSQL, może posłużyć Cassandra, HBase, Mongo DB, Voldemort, Riak i CouchDB [8].

3. Badania

Przedmiotem badań były współczesne narzędzia IT do analizy danych. Do przeprowadzenia porównania użyto trzech programów, zajmujących czołowe miejsca w najnowszym raporcie Gartnera, przedstawiającym najlepsze narzędzia Business Intelligence i platformy analityczne w lutym 2017 roku [2].



Rys. 4 Raport Gartnera – Magic Quadrant for Business Intelligence and Analytics Platforms [2]

Do oceny wybrano trzy programy: Tableau Desktop, Qlik Sense i Microsoft Power BI.

Tableau Desktop to narzędzie do analizy danych, które pomaga firmom wizualizować i zrozumieć swoje dane. W zależności od potrzeb, użytkownicy Tableau mogą wybrać zarówno darmową, jak i płatną wersję programu. To narzędzie uważane jest za standard w tworzeniu projektów wizualizacyjnych BI. Opis oferty Tableau znajduje się w tabeli 2 [9].

Tabela 2. Oferta Tableau [10]

Typ	Opis
Tableau Desktop	Pozwala na generowanie wykresów w trybie live lub offline. Możliwe jest tworzenie raportów na urządzeniach mobilne.
Tableau Server	Zapewnia kompatybilność oraz pełną obsługę chmury m.in. na Microsoft Azure.
Tableau Online	Wersja bezinstalacyjna /przeglądarkowa. Pozwala na publikowanie dashboardów w chmurze między zespołami i dla klientów.
Tableau Mobile	Wersja na smartfony i tablety, głównie do prezentowania gotowych kokpitów menadżerskich.
Tableau Public	Wersja na przeglądarki internetowe, służy do publikowania dashboardów na stronach internetowych.

Qlik Sense jest intuicyjnym systemem, którego prosty interfejs pozwala na tworzenie interaktywnych wizualizacji danych, bez skryptów czy złożonych zapytań SQL, dzięki czemu użytkownicy Qlik Sense nie muszą prosić o pomoc w specjalistów. To narzędzie pozwala na ujednoczenie źródeł danych, tworząc je bardziej czytelnymi. Qlika można używać w dowolnym miejscu z dowolnego urządzenia. Doświadczeni programiści mogą rozszerzać funkcjonalność tego systemu dzięki otwartemu API, a to pozwala im na nieograniczone działanie w nowych obszarach biznesowych (tabela 3) [11].

Tabela 3. Oferta Qlika [12]

Typ	Opis
Qlik Sense	Pozwala na analizę danych, budowanie elastycznych i interaktywnych wizualizacji, ułatwiających podejmowanie ważnych decyzji.
Qlik Sense Cloud	Zapewnia bezpieczny hosting oraz interakcję z Qlik Sense Enterprise, przy pracy nad wspólnym projektem.
Qlik View	Program mniej zaawansowany niż Qlik Sense. Umożliwia tworzenie biznesowych dashboardów, kontrolowanie ich i zarządzanie nimi.
Qlik Analytics Platform	Platforma do umieszczania gotowych dashboardów na dowolnych urządzeniach, zarówno mobilnych, jak i stacjonarnych.
Qlik NPrinting	Pozwala na łatwe i szybkie tworzenie raportów, w różnych popularnych formatach, korzystając z analizy Qlik Sense i Qlik View.
Qlik Data Market	Pomaga znaleźć, podłączyć i zarządzać danymi ze źródeł zewnętrznych.

Microsoft Power BI służy do analizy, wizualizacji oraz udostępniania danych w formie gotowych raportów. Ponadto, pozwala na tworzenie kokpitów menadżerskich, a także umożliwia bezpośrednie podłączenie do danych. Zbiór funkcji i usług online w Microsoft Power BI jest zorientowany w chmurze, przez co użytkownicy mają do nich dostęp w dowolnym miejscu i w dowolnym czasie (tabela 4) [13].

Tabela 4. Oferta Microsoft [14]

Typ	Opis
Power BI Desktop	Darmowy program, służący do tworzenia zaawansowanych raportów oraz publikowania ich w usłudze BI.
Power BI Pro	Płatny program, mający więcej możliwości niż darmowa wersja Power BI Desktop (m.in. w wersji rozszerzonej limity danych są zwiększone do 10 GB)

Tableau Desktop, Qlik Sense i Microsoft Power BI są narzędziami IT, przeznaczonymi nie tylko dla domowych użytkowników, ale także małych, średnich, a nawet dużych przedsiębiorstw. Każdy z tych użytkowników ceni sobie konkretne funkcje w wymienionych programach, które z jego punktu widzenia będą użyteczne. Użyteczność, o której mowa, dla każdego użytkownika wybranych programów, wiąże się z innym pojęciem.

3.1. Metody i przebieg badań

Do oceny opisanych narzędzi IT (Tableau Desktop, Qlik Sense i Microsoft Power BI) metodą porównawczą, wybrano osiem kryteriów. Lista wykorzystanych kryteriów oraz opis użytych metod została przedstawiona w tabeli 5.

Tabela 5. Kryteria oceny użyteczności wybranych programów

Lp.	Kryteria oceny	Metody
1.	Cena i rodzaj licencji	Sprawdzenie i zestawienie cen programów w celu zbadania, który z nich jest najdroższy, a który najtańszy.
2.	Rodzaje połączeń z bazą	Sprawdzenie i zestawienie wybranych programów pod względem ilości i rodzajów obsługiwanych baz danych.
3.	Obsługiwane platformy operacyjne	Sprawdzenie na jakich platformach operacyjnych obsługiwane są wybrane programy i zestawienie wyników.
4.	Target firmy	Sprawdzenie dostępności produktów zarówno dla dużych, jak i małych i średnich firm.
5.	Narzędzia do przygotowywania danych	Sprawdzenie i zestawienie informacji, dotyczących wbudowanych narzędzi do przygotowywania danych na wybranych programach.
6.	Dokumentacja	Sprawdzenie dostępu do samouczków i dokumentacji producenta. Zbadanie dostępu do forów użytkowników.
7.	Funkcje dla wysoko wykwalifikowanych użytkowników	Sprawdzenie i zestawienie wybranych narzędzi IT pod względem obecności zaawansowanych funkcji (analizy lub rozszerzalności narzędzi).
8.	Analiza ad-hoc, in-memory oraz w chmurze	Sprawdzenie, czy wybrane narzędzie ma możliwość analizy ad-hoc i analizy in-memory. Zestawienie wybranych programów pod względem możliwości analiz Business Intelligence w chmurze.

Użyteczność danego programu, dla konkretnego użytkownika, oznacza co innego. Przykładowo, właściciel jednoosobowej firmy, może skorzystać z funkcji tworzenia interaktywnych raportów, które następnie może przedstawić na dowolnym urządzeniu mobilnym podczas spotkania z klientem. Taki użytkownik nie ma potrzeby zarządzania dużymi wolumenami danych czy wynajmowaniu serwerowni, w odróżnieniu do dużych przedsiębiorstw. Zarządy czy rady nadzorcze korporacji cenią sobie programy, które są w stanie wykonać analizy danych Big Data w szybkim tempie. Ponadto, zwracają także uwagę na brak jakichkolwiek ograniczeń, przykładowo w ilości danych. Dlatego też, przy ocenie użyteczności wybranych narzędzi IT, na pierwszym planie stawiano użytkownika i jego potrzeby, które determinują pojęcie użyteczności Tableau Desktop, Qlik Sense i Microsoft Power BI.

3.2. Wyniki badań

Cena i rodzaj licencji

Pierwszym kryterium, zgodnie z którym dokonano oceny użyteczności wybranych programów, była cena i rodzaj licencji. Badania skupiły się na określeniu różnic między ceną a czasem licencji w poszczególnych wersjach programu. Dane, dotyczące ceny i rodzaju licencji Tableau, obrazuje tabela 6.

Tabela 6. Tableau – licencja abonamentowa [10]

Wersja	Cena	Liczba użytkowników	Czas licencji
Tableau Desktop	-	1	14 dni
Tableau Desktop	35\$	1	1 miesiąc
Tableau Desktop Professional	70\$	1	1 miesiąc
Tableau Desktop Professional (akademicka)	-	1	1 rok
Tableau Server	35\$	1	1 miesiąc
Tableau Online	42\$	1	1 miesiąc

Każda z cen, przedstawionych w tabeli, może być wyznaczona indywidualnie, w zależności od użytkowników, długości trwania umowy czy wielkości baz danych.

Drugim badanym programem, pod względem ceny i rodzaju licencji, był Qlik. Zgodnie z cennikiem licencji, przedstawionym na stronie producenta Qlika, comiesięczny abonament, z którego może skorzystać jeden użytkownik, zależy od wersji programu. Przykładowe dane, dotyczące kosztów, związanych z konkretną wersją Qlika, zostały przedstawione w tabeli 7.

Tabela 7. Qlik – licencja abonamentowa [12]

Wersja	Cena	Liczba użytkowników	Czas licencji
Qlik Sense Desktop	-	1	Bez limitu
Qlik Sense Enterprise	Koszt licencji ustalany indywidualnie		
Qlik Sense Cloud Basic	-	1-5	Bez limitu
Qlik Sense Cloud Business	25\$	1	1 miesiąc
Qlik View	-	1	Bez limitu
Qlik Analytics Platform	Koszt licencji ustalany indywidualnie		

Produktami darmowymi Qlika są Qlik Sense Desktop, Qlik Sense Cloud Basic i Qlik View. Użytkownicy mogą korzystać z nich bez limitu. Natomiast jeżeli chodzi o Qlik Sense Enterprise i Qlik Analytics Platform, cena jest ustalana indywidualnie, w zależności od czasu licencji i liczby użytkowników. Z kolei z wersji Qlik Sense Cloud Basic, bez limitu czasowego może bezpłatnie korzystać do 5 osób.

Trzecim badanym programem, pod względem ceny i rodzaju licencji, był Microsoft Power BI. Przykładowe dane, dotyczące kwot, związanych z konkretną wersją programu, zostały przedstawione w tabeli 8.

Tabela 8. Microsoft Power BI – licencja abonamentowa [14]

Wersja	Cena	Liczba użytkowników	Czas licencji
Power BI Desktop	-	1	Bez limitu
Power BI Pro	9,99\$ / 99,99\$	1	1 miesiąc / 1 rok
Power BI Premium	Koszt licencji ustalany indywidualnie		

Jak wynika z tabeli, licencję na Power BI Desktop użytkownik może uzyskać bezpłatnie na nieograniczony okres. Za Power BI Pro jeden użytkownik musiałby za miesiąc zapłacić 9,99 dolarów, co w skali roku daje 99,99 dolarów. Natomiast koszt licencji do Power BI Premium ustalany jest indywidualnie, w zależności od ceny, liczby użytkowników i czasu licencji.

Rodzaje połączeń z bazą

Wybrane narzędzia IT zbadano także pod względem rodzajów połączeń z bazą. Część informacji na ten temat, odnośnie Tableau Desktop, wizualizuje tabela 9.

Tabela 9. Tableau – wybrane rodzaje połączeń z bazą [10]

Wersja	Rodzaj połączenia
Podstawowa	<ul style="list-style-type: none"> • Microsoft Excel • Pliki tekstowe (np. pliki .csv) • Microsoft Access for Windows • JSON • Tableau Data Extract
Profesjonalna	<ul style="list-style-type: none"> • Actian • Amazon • Anaplan • Microsoft • MySQL • Oracle • Hortonworks Hadoop Hive • MapR Hadoop Hive • Pliki tekstowe (np. pliki .csv) • Tableau Data Extract

Jak prezentuje tabela 9, w wersji podstawowej Tableau Desktop użytkownicy mają możliwość skorzystania zaledwie z pięciu rodzajów połączeń, podczas gdy wersja profesjonalna udostępnia znacznie więcej możliwości rodzajów połączeń z bazą.

Qlik, w odróżnieniu do Tableau, niezależnie od wersji programu, korzysta z tych samych rodzajów połączeń z bazą (tabela 10).

Tabela 10. Qlik – wybrane rodzaje połączeń z bazą [12]

Wersja	Rodzaj połączenia
Qlik	Bazy danych: <ul style="list-style-type: none"> • Apache Hive, • ClouderaImpala, • MySQL Enterprise, • Teradata, • Oracle.
	Pliki: <ul style="list-style-type: none"> • CSV, • HTML, • XML, • XLSX, • Natywne pliki programu Qlik – QVD i QVX.

Dla każdej wersji Qlika, dostępne są te same rodzaje połączeń, zarówno z bazami, jak i z plikami. Podobnie jest z rodzajami połączeń w przypadku Microsoft Power BI (tabela 11).

Tabela 11. Microsoft Power BI – wybrane rodzaje połączeń z bazą [14]

Wersja	Rodzaj połączenia
Microsoft Power BI	Pliki: <ul style="list-style-type: none"> • Excel, • Text/CSV, • XML, • JSON, • SharePointFolder.
	Bazy danych: <ul style="list-style-type: none"> • SQL Server Database, • Access Database, • Azure, • IBM DB2 Database, • Impala.
	Serwisy internetowe: <ul style="list-style-type: none"> • SharePointOnlineList, • Microsoft Exchange Online, • Dynamics 365 (online), • Salesforce Objects, • Google Analytics.
	Pozostałe: <ul style="list-style-type: none"> • Web, • OData Feed, • Spark (Beta), • R Script, • ODBC.

Dla każdej wersji Microsoft Power BI są dostępne te same rodzaje połączeń.

Obsługiwane platformy operacyjne

Wszystkie wersje Tableau są dostępne na platformy operacyjne Windows, Mac, Android, iPhone/iPad czy przeglądarki internetowe. W zależności od konkretnej wersji programu (Tableau Desktop, Tableau Server czy Tableau Online), określone są inne wymagane parametry. Dla przykładu, Tableau Desktop i Tableau Online mogą być obsługiwane na platformach operacyjnych iMac/MacBook, a w wersji Tableau Server nie ma takiej możliwości.

W wersji Tableau Desktop minimalne wymagania sprzętowe wynoszą 2 GB RAM, podczas gdy w Tableau Server potrzeba minimum 8 GB RAM. Tableau Desktop nie obsługuje przeglądarek internetowych, podczas gdy zarówno Tableau Server, jak i Tableau Online działa m.in. pod Chrome, Apple Safari, Internet Explorer 11 czy Mozilla Firefox [10].

Qlik z kolei, obsługiwany jest m.in. na systemie operacyjnym Windows 7 lub nowszym. W Qlik Sense wymagany jest co najmniej czterordzeniowy procesor, podczas gdy do Qlik View jedynie dwurdzeniowy. Zarówno Qlik Sense, jak i Qlik View, działają pod Chrome, Apple Safari, Internet Explorer 11 czy Mozilla Firefox. [12]

Natomiast Microsoft Power BI, obsługiwany jest jedynie pod platformami Windows. Wszystkie rodzaje tego narzędzia IT (Microsoft Power BI Pro i Microsoft Power Professional) mają te same wymagania sprzętowe (minimum 4 GB RAM oraz 1,5 GB wolnej pamięci przestrzeni na dysku). Oba programy są obsługiwane przez Internet Explorer 9 lub nowszy [14].

Grupa docelowa firmy

Tableau, kieruje swoje produkty do zarządów i rad nadzorczych w dużych i średnich przedsiębiorstwach z różnych branż (np. bankowość, komunikacja, ubezpieczenia). Tableau tworzy także analizy w innych obszarach, choćby takich jak rząd, edukacja czy organizacje non-profit.

W przypadku Qlika, grupy docelowej firmy są nie tylko duże i średnie przedsiębiorstwa, jak wspomniane było w przypadku Tableau, ale także małe, nawet jednoosobowe firmy.

Microsoft Power BI, podobnie jak Tableau, tworzy swoje produkty z myślą o dużych i średnich przedsiębiorstwach, działających na różnych obszarach i w różnych branżach.

Narzędzia do przygotowywania danych

Zarówno Tableau, jak i Microsoft Power BI nie posiadają wbudowanych narzędzi ETL do przetwarzania danych. Ich transformacje ograniczają się jedynie do zadawania zapytań, w języku zrozumiałym dla źródła. Qlik natomiast posiada wbudowane narzędzia ETL do przetwarzania danych [15, 16].

Dokumentacja

Na stronie producenta Tableau, Qlik oraz Microsoft Power BI, znajdują się odnośniki do dokumentacji, w której zawarte są samouczki oraz przydatne wskazówki. Jedynie Qlik posiada polskojęzyczną wersję dokumentacji. Natomiast wszystkie trzy narzędzia posiadają fora użytkowników w języku polskim.

Funkcje dla wysoko wykwalifikowanych użytkowników

Produkty Tableau dostosowano do wymagań wysoko wykwalifikowanych użytkowników, znających język R. Tableau posiada funkcje takie jak analiza trendów, modelowanie regresji czy korelacje.

Qlik Analytics Platform (QAP), który ma otwarte API aplikacji, pozwala programistom na tworzenie własnych obiektów (pisanych w języku skryptowym JavaScript).

Microsoft Power BI z łatwością tworzy proste wizualizacje. Jednak gdy chodzi o zaawansowane analizy, potrzebny jest ekspert w języku DAX do analizy danych. [17]

Analiza technologii ad-hoc, in-memory i w chmurze

Tableau Desktop pozwala na wykonywanie obliczeń ad-hoc, dzięki połączeniu Live lub Offline. Dzięki połączeniu Live, można np. tworzyć zapytania ad-hoc, mając wciąż aktualizowane zbiory danych. Podczas połączenia Offline wykonywane są obliczenia in-memory.

Tableau posiada analizę danych in-memory. Jest to wykorzystanie pamięci operacyjnej komputera do analiz, dzięki czemu ograniczenia przepustowości sieci przestają wpływać na prędkość analiz (wszystkie analizowane dane zostają pobrane do pamięci podręcznej).

Analiza danych w chmurze, w Tableau, dostępna jest jedynie w wersji Professional i Server. W tańszej wersji programu nie ma możliwości połączenia z chmurą.

Nowa odsłona Qlik Sense oferuje użytkownikom możliwości analizy ad-hoc, in-memory oraz w chmurze. Qlik posiada wbudowany silnik asocjacyjny (QIX), dzięki któremu możliwe jest generowanie nowoczesnych raportów oraz analiza ad-hoc. Każde kliknięcie na wykres spowoduje wykonanie analitycznego obliczenia, takiego jak zwijanie, filtrowanie itp. Qlik obsługuje także analizę in-memory, która dzięki wykorzystaniu pamięci operacyjnej RAM komputera, wykonuje obliczenia. Inwestycja w dużą ilość RAM, pozwoli użytkownikowi na obliczanie milionowych danych w czasie rzeczywistym.

Analizy danych, przy wykorzystaniu pamięci operacyjnej RAM komputera, są dostępne w Power BI Desktop. W każdej wersji programu można wykonywać zarówno analizy ad-hoc, analizy in-memory i analizy w chmurze. [18]

4. Ocena użyteczności wybranych narzędzi IT

Wybrane programy zostały ocenione zgodnie z przyjętymi kryteriami i metodami, w skali od 1 do 5. Celem oceny było wskazanie najlepszego narzędzia IT do analizy danych z punktu widzenia młodego analityka, kończącego studia i wchodzącego na rynek pracy. Tabela 12 prezentuje ocenę użyteczności wybranych programów.

Tabela 12. Ocena użyteczności wybranych programów

	Tableau Desktop (pkt)	Qlik Sense (pkt)	Microsoft Power BI (pkt)
Cena i rodzaj licencji	4	5	5
Rodzaje połączeń z bazą	4	4	5
Obsługiwane platformy operacyjne	4	3	4
Target firmy	4	5	4
Narzędzia do przygotowywania danych	4	5	4
Dokumentacja	5	5	5
Funkcje dla wysoko wykwalifikowanych użytkowników	5	5	4
Analiza ad-hoc, in-memory oraz w chmurze	5	5	5
Podsumowanie:	35	37	36

Cena i rodzaj licencji

W skali od 1 do 5, zarówno Qlik, jak i Microsoft Power BI, oceniono na maksymalną ilość punktów. Tę decyzję uzasadniono tym, że nie trzeba być studentem, żeby mieć darmową wersję programu. Co prawda Tableau udostępniła studentom bezpłatną wersję akademicką Tableau Desktop Professional na rok i daje możliwość przedłużenia licencji po okazaniu legitymacji studenckiej, jednak nie ma on tych samych możliwości, co darmowe wersje Qlika i Microsoft Power BI. Kluczową funkcją, jaka zaważyła w ocenie, był brak możliwości dzielenia się dashboardami w Tableau Desktop Professional.

Rodzaje połączeń z bazą

Pod względem rodzajów połączeń z bazą, Qlik i Tableau uzyskały (4) punkty. Najwyższą notę (5) otrzymał produkt Microsoftu. Ta decyzja została podyktowana tym, że Qlik posiada najmniejszą ilość rodzajów połączeń z bazami danych, w porównaniu z innymi ocenianymi narzędziami. Tableau stracił punkt, gdyż w wersji darmowej ma zaledwie 10 rodzajów połączeń, bez możliwości łączenia się z danymi w chmurze. Zarówno Power BI, jak i Tableau, mają możliwość analizy danych w portalach społecznościowych. Qlik z kolei nie posiada takiej możliwości, za co odjęto mu punkt w powyższej ocenie. Microsoft Power BI w tym zestawieniu prezentuje się najlepiej, gdyż niezależnie od wersji programu, udostępnia te same rodzaje połączeń - z plikami, bazami danych, serwisami internetowymi czy innymi.

Obsługiwane platformy operacyjne

Jeżeli chodzi o obsługiwane platformy operacyjne, żaden z programów nie otrzymał maksymalnej ilości punktów. Tableau i Power BI otrzymały ex aequo (4) punkty zaś Qlik oceniono na (3). Powodem tej decyzji były następujące zależności:

- żaden z wyżej wymienionych programów nie jest dostępny na każdą platformę,
- Tableau i Qlik nie mają wersji na Windows Phone,
- Power BI i Qlik nie mogą być obsługiwane w systemie operacyjnym Mac (są dostępne tylko na iPhone'a/iPada), z kolei Tableau, jako jedyny z nich, posiada wersję na urządzenia dekstopowe Maca.
- Wymagania sprzętowe wszystkich narzędzi były porównywalne (wymagały m.in. 2-4 rdzeniowego procesora, 2-8 GB RAM, ~1.5 GB wolnej przestrzeni na dysku).
- Jeżeli chodzi o obsługiwane przeglądarki, Tableau Desktop w zestawieniu wypadł najgorzej, ponieważ nie obsługuje żadnej. Qlik i Power BI obsługują te same przeglądarki.

Podsumowując, pomimo że Tableau i Power BI otrzymały tyle samo punktów, wybierając Tableau nie trzeba się zastanawiać czy wszystkie komputery w firmie są pod systemem Windows czy może któryś ma system Mac.

Target firmy

Zestawiając wspomniane narzędzia IT pod względem targetu firmy, Qlik otrzymał maksymalną ilość punktów (5), natomiast Tableau i Power BI ex aequo po 4 punkty. Tableau, podobnie jak Microsoft Power BI, kieruje swoje produkty do dużych i średnich przedsiębiorstw, które posiadają już własne hurtownie danych i narzędzia do ich przetwarzania. Natomiast Qlik tworzy programy przeznaczone głównie do małych i średnich firm, dzięki zastosowaniu wbudowanych narzędzi ETL. Warto zauważyć, że istnieje wersja darmowa Qlik Sense Cloud, która zachęca właścicieli małych przedsiębiorstw do korzystania z tego narzędzia, gdyż oferuje możliwość tworzenia i udostępniania użytkownikom prostych raportów za darmo.

Narzędzia do przygotowywania danych

Wśród ocenianych firm, pod względem narzędzi do przygotowywania danych, najlepiej prezentuje się Qlik, który głównie za to, iż posiada własne narzędzie ETL, dostał 5 punktów. Dwa pozostałe programy w zestawieniu uzyskały o jeden punkt mniej, gdyż są uzależnione od zewnętrznych narzędzi ETL.

Dokumentacja

Wszystkie trzy programy posiadają bardzo bogatą dokumentację, zarówno w języku angielskim jak i polskim. Tableau, Power BI i Qlik mają samouczki w postaci filmików, przykładowe zastosowania programów oraz dodatki, wzbogacające możliwości analitycznie. Tak bogata

oferta sprawiła, że każdy z programów oceniono maksymalną ilością punktów (5).

Funkcje dla wysoko wykwalifikowanych użytkowników

Pod względem obecności funkcji dla wysoko wykwalifikowanych użytkowników, ze wszystkich porównywanych programów, najlepiej wypadły Tableau i Qlik. Oba narzędzia IT uzyskały ex aequo maksymalną ilość punktów (5), gdyż funkcjonalność tych programów można rozszerzać znając język skryptowy (np. java script). Znajomość języka R również pomoże tworzeniu rozbudowanych analiz w przypadku wspomnianych narzędzi. W tym zestawieniu na ostatniej pozycji uplasował się Power BI, który stracił punkt za to, że możliwość korzystania z zaawansowanych funkcji, uzależnił od znajomości specjalnego języka dedykowanego DAX do analizy danych.

Analiza ad-hoc, in-memory oraz w chmurze

Każdy z opisywanych programów, niezależnie od wersji posiada możliwość analizy ad-hoc, in-memory oraz w chmurze. Dlatego zarówno Tableau, Qlik oraz Microsoft Power BI otrzymały w zestawieniu tą samą liczbę punktów (5).

5. Wnioski

Analizując otrzymane wyniki, program Qlik, na tle badanych narzędzi IT, nie pod każdym względem wypadł najlepiej, jednak mimo to, otrzymał najlepszy wynik (37 z 40 możliwych punktów). Microsoft Power BI łącznie dostał 36 punktów. Tym samym zajął drugie miejsce w zestawieniu oceny użyteczności współczesnych narzędzi IT do analizy danych Big Data. Na ostatnim miejscu, różnicą 1 punktu, uplasował się Tableau z wynikiem 35 punktów. Zgodnie z powyższym, najbardziej użytecznym programem do analizy danych, z punktu widzenia studenta i analityka biznesowego, wchodzącego na rynek pracy, okazało się narzędzie Qlik. Warto jednak zaznaczyć, iż różnica między pierwszym a ostatnim miejscem zestawienia, była minimalna. Wszystkie badane programy spełniają kryteria użyteczności pod każdym względem. Te tezę może potwierdzić fakt, iż w żadnym z kryteriów oceniania, nie pojawiła się ocena niższa niż 3 punkty. Hipoteza, postawiona na początku niniejszej pracy, została zweryfikowana pozytywnie.

Literatura

- [1] Marz N., Warren J.: Big Data. Najlepsze praktyki budowy skalowalnych systemów obsługi danych w czasie rzeczywistym, Helion, Gliwice, 2016
- [2] Howson C., Idoine C., Oestreich T., Richardson J., Sallam R., Tapadinhas J., Magic Quadrant for Business Intelligence and Analytics Platforms, Gartner, <https://www.gartner.com/doc/reprints?id=1-3TYE0CD&ct=170221&st=sb.>, [02.06.2017]
- [3] Rangarajan S., Data Warehouse Design – Inmon versus Kimball, <http://tdan.com/data-warehouse-design-inmon-versus-kimball/20300> [03.06.2017]
- [4] Mudrakola S., Manage your data with these 3 open source etl tools, <http://techgenix.com/open-source-etl-tools/> [04.06.2017]
- [5] Strona firmy Hortonworks, wykorzystującej Apache Hadoop <https://hortonworks.com/products/data-center/hdp/> [21.05.2017]
- [6] Siudziński M., Hadoop, czyli przetwarzanie rozproszone w open source, <http://itwiz.pl/hadoop-czyli-przetwarzanie-rozproszone-open-source/> [20.05.2017]
- [7] Czajkowski L., A Comparison of SQL and NoSQL to Simplify Your Database Decision, <https://blog.couchbase.com/comparison-sql-nosql-simplify-database-decision/> [11.06.2017]
- [8] Harazin J., Narzędzia Business Intelligence dla managera: QlikView i Qlik Sense, <http://www.jcommerce.pl/jpro/article/narzedzia-business-intelligence-dla-managera-qlikview-i-qlik-sense> [5.06.2017]
- [9] Baker P., Tableau Desktop, <http://www.pcmag.com/article2/0,2817,2491943,00.asp>, [6.06.2017]
- [10] Strona producenta Tableau, <https://www.tableau.com>, [04.06.2017]
- [11] Prichard J., What's the difference between Qlik Sense and Qlik View?, <http://optimalbi.com/blog/2016/01/18/whats-the-difference-between-qlik-sense-and-qlik-view/> [3.06.2017]
- [12] Strona producenta Qlik, <http://www.qlik.com/pl-pl/> [04.06.2017]
- [13] Ris O., Microsoft Power BI, PCMag, <http://uk.pcmag.com/microsoft-power-bi/74174/review/microsoft-power-bi> [26.05.2017]
- [14] Strona producenta Microsoft Power BI, <https://powerbi.microsoft.com/en-us/> [04.06.2017]
- [15] Rui R., Use Power BI Desktop as an ETL Tool, <https://ruiromanoblog.wordpress.com/2016/04/21/use-power-bi-desktop-as-an-etl-tool/> [30.05.2017]
- [16] Chadran S., Is Microsoft Power BI a data warehouse also?, Olympic Software, <https://www.olympic.co.nz/is-microsoft-power-bi-a-data-warehouse-also/> [28.05.2017]
- [17] Iseminger D., Measures in Power BI Desktop, <https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-measures/> [01.06.2017]
- [18] Deng S., Saxton A., Schiavon A., Wprowadzenie do usługi Microsoft Power BI Embedded, <https://docs.microsoft.com/pl-pl/azure/power-bi-embedded/power-bi-embedded-get-started> [29.05.2017]