An Evaluation of Major Vendor BI Platforms

vs

Open-Source Languages

Josh Cubero

Golden Gate University

MSBA 320

Introduction

Big data and analytics are a massive global market estimated at $206.95B in 2020 and is expected to grow to $549.73B by 2028. As of 2020, there were at least 15 major vendors in the analytics space such as Microsoft, Oracle, Qlik, and Salesforce.com. With the deluge of analytics options available, how does an entity choose the best solution? The answer is simple, they should understand what the problem is they're trying to solve, then conduct research, and select the best solution to solve the problem. This study is an example of such an exercise, where the researchers conduct research on available analytics solutions, perform a compare and contrast analysis, then make a closing recommendation.

SAS

SAS, short for Statistical Analysis System, is a commercially available statistical software package used in data analytics and business intelligence. The platform is written in C programming language and was first created in the early 1970's at North Carolina State University. SAS offers users the ability to conduct analytics through their programming language or a Graphical User Interface (GUI). Users can perform several tasks with SAS such as data entry, statistics and mathematic analytics, financial planning & analysis, even project management.

The SAS platform has components that perform different functions such as SAS/Graph for visual presentations, SAS/STAT for statistics, and SAS/OR to conduct operations research. SAS contains four tools that are essential to the platform: Data Access Tool, Data Management Tool, Data Analysis Tool, and Data Presentation Tool. The data access tool allows SAS to work with relational and non-relational data, while the data management tool ingests raw data and transforms the data to create the SAS dataset. Additionally, SAS can be used for data analysis such as descriptive statistics, ANOVA, student's t, and F-test. Lastly, SAS can be used to visualize data via tables, lists, and graphs.

As with any technology platform, SAS has its advantages and disadvantages. The SAS programming language overall is easy to learn and can process large datasets. Another advantage is that SAS makes debugging easy and has algorithms that are stringently tested prior to release, so no user-beta testing is required. Next, because SAS is a revenue generating platform, it comes

with outstanding customers in the event the user experiences an issue. Lastly, SAS' GUI is well liked for its ease of use in creating graphs and plots.

One of the first disadvantages found in this research is that SAS is proprietary and not open source. User licenses are required and some of the libraries and packages require an additional expense. Next, because SAS is closed-source, its algorithm development lags some of the open-source platforms in use today. Furthermore, text mining is difficult in SAS, and is an additional monetary charge to the user.

Power BI

The next BI and analytics platform in this study Microsoft's Power BI, which is the company's foray into the business intelligence space. Power BI users can employ the platform via a local download called Power BI Desktop or via SaaS on the Power BI Service. Users can download and develop reports on Power BI Desktop without incurring a charge. However, Power BI Service requires a fee-based subscription. Users can access a free 60-day trial of Power BI service.

Power BI has several strengths, with one being that the tool is interactive and allows users to pull data from a multitude of data sources from SQL Server to Excel, Mongo DB, and approximately 50 others. In terms of custom analytics, Power BI users have a choice of employing a scripting language called Data Analysis Expression (DAX) or M Language in the Power Query Editor. Once the metrics have been created, users can utilize the metrics to create a wide variety of visuals.

While Power BI has some advantageous features, it also comes with some disadvantages, notably, the interface. Power BI's UI is extremely bulky, with filter, visualization, and field panes that combine to consume nearly half of the UI. The bulk of the UI makes it difficult to see visuals on even standard sized monitors. Next, Power BI uses in-memory analytics, which means that the platform is very costly in terms of memory usage. Users lose significant RAM with each instance of Power BI that's operating. Lastly, Power BI's is extremely rigid in the application of relationship management between tables in the data model. Therefore, Power BI will not render a visual if data is from multiple tables that do not possess an active relationship.

SPSS

SPSS, or Statistical Package for the Social Sciences, is a proprietary statistical package offered by tech giant IBM. It was first developed in 1968 and is typically used for statistical analysis. SPSS is employed mostly in healthcare, academic research, marketing research, and data mining. SPSS wasn't always an IMB solution, with the latter acquiring the former in a $1.2B acquisition in 2009. The SPSS platform has three primary product offerings: IBM SPSS Statistics which has business and research uses such as ad hoc analysis, predictive analytics, hypothesis testing, and geospatial analysis. Next, there's the IBM SPSS Modeler that enables the user to deep-dive data sets with a pre-coded set of algorithms and models. Finally, there's the IBM SPSS Modeler in Cloud Pak for Data. As the name would imply, this offering is cloud enable, affording users the ability to containerize data and execute predictive models in the cloud.

There are several advantages to using SPSS, particularly if the user is not technically inclined. SPSS is simple to use, with little to no learning curve. The platform can easily handle both quantitative and qualitative data. Additionally, users can choose and match an appropriate graph for the selected distribution. Because SPSS has no code, the likelihood of errors has been greatly reduced. SPSS can conduct regression, loglinear, classification, and even neural networks.

SPSS has several disadvantages, with one very common, cost. The platform is expensive to license, so its use is limited to those researchers that have a specific need for the tool. Next, because SPSS has little to no code, it's functionality is limited and lacks flexibility. Finally, SPSS' functionality is like Microsoft Excel, so researchers could be overspending for a product they could very well use for free.

Tableau

Tableau was first developed in 2003 by Pat Hanrahan, Christian Cabot, and Chris Stolte with the intent of making databases more interactive. Prior to Tableau, analyzing data was the job of IT staff who would analyze data and build reports. Tableau was one of the first companies to create a no-code user-friendly analytics platform that is easy to learn. Tableau has several offerings: Tableau Desktop, Tableau Server, and Tableau Online. Users can create reports in Tableau desktop, then publish the reports to the centralized Tableau Server for user consumption.

One of the advantages of Tableau is that users can analyze disparate data to create various visualizations. Its drag and drop capabilities enable users to create interactive visuals in just a few minutes. Additionally, Tableau's interface is very versatile, but also can warn the user about creating charts that violate analytics visualization best practices. Tableau can be employed with a little training to create a litany of visually appealing charts and graphs. The platform boasts drag and drop capabilities, can be contained to many data sources, and provide real-time analytics.

Tableau is a for-profit vendor, so the first disadvantage is that the platform is not open-source or free. While not as expensive as some of its competitors, Tableau's licensing fees could prove quite costly for small to medium-sized businesses. While Tableau is simple to use for basic charts, there are many situations where SQL is required, which once again leaves business users at the mercy of the IT department to create queries. Additionally, Tableau is not able to perform large-scale reporting or build data models and relationships, and therefore is not an end-to-end BI solution. Finally,

R Programming

R Programming is a popular programming language used in statistical programming and visualizations. The language is open source, with a GNU operating system. R programming provides users with the ability to create a litany of statistical analytics procedures such as linear and non-linear models, hypothesis testing, classification, time-series analysis, and clustering. R programming language is indeed an environment unto itself, allowing users to define and implement new functions, and implement statistical techniques.

The first, and very important advantage of R programming is that it's free and open source. Because R is open source, anyone can contribute to R improvements, make changes to packages, or even develop new packages. Next, R can create detailed scientific graphs and charts with just several lines of code through its ggplot2 and plotly libraries. Another advantage of R is that it's compatible with other programming languages such as Java and Python, as well as database management systems such as Hadoop.

R does have its disadvantages as well. R is indeed a programming language, so newer users will need to learn how to write programming algorithms, which some users haven't' the time or interest. Next, R programming algorithms have limited documentation, so if a user wants

to adapt a visual, they might not find much assistance or examples in the documentation. Finally, some of the R programming algorithms that exist are complex, leaving newer users with the task of mixing their code with the new algorithm.

Python

The final analytics tool in this study is the Python programming language, which is an interpreted, multi-use, object-oriented programming language. Python is a high-level language and has many applicable uses beyond analytics. The programming language is open source and has a myriad of libraries that enable developers' great flexibility to create Python applications. Python can be used to create games and web applications, in addition to its analytics uses. Because Python is a general-purpose, interpreted programming language and therefore has no compilation step, developers can easily debug and test their code.

Python programming language has several advantages in the analytics space. The language is free and open source, so users can expand libraries or create new libraries. Python's analytics libraries such as Numpy, Pandas, Matplotlib, and Seaborn provide developers with the ability to create fantastic visualizations that create value and insights. Python's multitude of libraries make it very flexible, enabling developers to gain new insights and views that have never been seen before. Beyond visualizations, Python has massive applications in machine learning and AI, and is heavily used in those fields. Lastly, Python is extremely popular in analytics and is the #1 most popular programming languages in 2022 according to a study performed by Northeastern University. Because of Python's popularity, there's a massive support community, users can easily find effective solutions to difficult coding problems.

Next, Python does have some disadvantages. The language is a legitimate computer programming language, which means that users must have at least a basic understanding of computer science principles. This fact will alienate many self-service, business-type users who do not possess computer science skills. Additionally, as Python is an interpreted language, it executes code line by line, which tremendously affects speed. This may not seem like an issue for a few thousand lines of code. However, if the developer is analyzing millions of lines of code with an algorithm running at $O(N^2)$ or worse time complexity, the algorithm's performance will surely be degraded.

Compare and Contrast

The analytics platforms in this study all have their advantages and disadvantages, now we'll quickly discuss their similarities. First, Power BI and Tableau, the leading competitors in the data visualization space according to Gartner's Magic Quadrant for Analytics, with the edge going to the former. Both platforms are closed source, can be used locally or on the cloud, and have drag and drop capabilities. Tableau has an edge over Power BI in that the former is usable on Mac and PC, where the latter is PC only. Next there's SAS and SPSS which are enterprise level solutions with greater applicability than Tableau and Power BI. Lastly, there's R and Python programming languages. R and Python are more suited to users with computer science skills, they're simply not suitable for users not willing to learn coding. However, Python and R enable far greater analytics capabilities than the other analytics solutions in this study, especially Power BI. In Power BI there must be a relationship between fields to create a visual. Whereas Python and R, the developer has no relationship limitations.

Recommendations

What analytics solution from this study should the reader choose? Really, it depends on what outcomes the user requires and who is going to build the analytics. If the user needs to customize the analysis, then R or Python would likely be the choice. In organizations with a BI OLAP cube, they'll likely employ Power BI or Tableau. Next the user's skillset must be accounted for in selected an analytics solution as well. If the BI architecture is self service and intended for business users, then avoiding R and Python would be best. The recommendation of this study is to encourage the reader to select an analytics tool that best suits the needs of analysis, as all the solutions in this study are great at what they do. One solution isn't inherently better than the other, but all are different and are suited for different purposes.

References

What is SAS | Key Concept And Advantages of Statistical Analysis System. (2019, April 5). EDUCBA. https://www.educba.com/what-is-sas/

DataFlair Team. (2018, March 17). *Advantages of SAS | Disadvantages of SAS Programming - DataFlair*. DataFlair. https://data-flair.training/blogs/disadvantages-and-advantages-of-sas/

SAS Advantages | SAS Disadvantages - Javatpoint. (n.d.). Www.javatpoint.com. https://www.javatpoint.com/advantages-and-disadvantages-of-sas

15 power tips for Microsoft Power BI. (n.d.). CIO. Retrieved April 1, 2022, from https://www.cio.com/article/228018/power-tips-for-microsoft-power-bi.html

Monster Merger: IBM Buys SPSS For Approx. $1.2 Billion In Cash Deal. (n.d.). TechCrunch. https://techcrunch.com/2009/07/28/monster-merger-ibm-buys-spss-for-approx-12-billion/

IBM. (2019). *SPSS software*. Ibm.com. https://www.ibm.com/analytics/spss-statistics-software

What is SPSS? | Features, Types, and Statistical Methods Of SPSS. (2019, June). EDUCBA. https://www.educba.com/what-is-spss/

What Is SPSS Data Analysis & How Does It Work? (n.d.). Small Business - Chron.com. https://smallbusiness.chron.com/spss-data-analysis-work-74542.html

Advantages of using SPSS in your dissertation analysis: How to do a regression analysis? Home: Author. (n.d.). Canvas.becker.edu. https://canvas.becker.edu/eportfolios/69775/Home/Advantages_of_using_SPSS_in_your_dissertation_analysis_How_to_do_a_regression_analysis

Sewall, A. (2019, December 5). *Stata vs. R vs. SPSS for Data Analysis | Commons Knowledge - University of Illinois at Urbana-Champaign*. http://publish.illinois.edu/commonsknowledge/2019/12/05/stata-vs-r-vs-spss-for-data-analysis/

History of Tableau - javatpoint. (2016). Www.javatpoint.com. https://www.javatpoint.com/history-of-tableau

Absent Data. (2017). *Advantages and Disadvantages of Tableau - AbsentData*. AbsentData. https://www.absentdata.com/advantages-and-disadvantages-of-tableau/

Tableau Software Review: Pros and Cons of a BI Solution for Data Visualization. (2019, June 19). *Pros and Cons of Tableau Software for Data Visualization [Review] | SaM Solutions*. SaM Solutions. https://www.sam-solutions.com/blog/tableau-software-review-pros-and-cons-of-a-bi-solution-for-data-visualization/

R: What is R? (2019). R-Project.org. https://www.r-project.org/about.html

Moore, C. (2021, December 25). *Pros and Cons of R Programming Language*. Nannu Says. https://nannusays.com/tech/pros-and-cons-of-r-programming-language/

Python.org. (2019). *What is Python? Executive Summary*. Python.org; Python.org. https://www.python.org/doc/essays/blurb/

DataFlair Team. (2018, January 2). *Advantages and Disadvantages of Python - How it is dominating programming world - DataFlair*. DataFlair. https://data-flair.training/blogs/advantages-and-disadvantages-of-python/