Lightweight DWH Data Analysis for SMEs

Lukas Dötlinger, Manuel Penz, Markus Reiter & Stephanie Widauer

January 7, 2021

Abstract

Contents

1 Introduction

A *data warehouse* (*DWH*) is a special type of database system, that focuses on reporting and analysing of it's data. Implementing such a system reduces the complexity to access business data in an analytical way and is an important step to achieve *business intelligence (BI)*. Enterprises typically bundle the data of all their operational databases together in one data warehouse. All departments within a company still use their own database for day to day production, as they don't use it for analysing and reporting of their business process. Hence the DWH system is mostly used in the management layer of an enterprise, since they deal with internal reporting and business analytics.

With increasing digitisation of business processes and communication, the amount of data, that companies are collecting, is increasing rapidly across all business sectors. Therefore a data warehouse system is becoming more interesting for many small- and medium-sized enterprises (*SMEs*), as they require a systematic approach to analyse their business data in a productive way. In the context of this research, companies with at most 250 employees are considered a SME, also known as *small and medium business (SMB)*.

The increasing demand for such systems, has lead to an increase in development for specific data warehouse solutions targeting SMEs. This paper aims to compare and analyse such systems for their suitability in the context of a small or medium enterprise. Furthermore, a comparison of different systems is used to give a general baseline for a lightweight implementation of a DWH in a SME.

2 Related Work

As the amount of digital data is ever increasing for all companies, the effort to manage it grows exponentially. To structure that data, most business have the option of implementing a data warehouse. Although there are many existing DWH implementations, only some are actually applicable to the setting of a smaller business. This is due to the fact that SMEs tend to lack certain expertise and have a limited budget, as well as a small amount of spare employees for IT. This further aggravates the use of many traditional DWH solutions, as they are tailored towards big enterprises and include many features, which are not relevant for a SME [1].

Furthermore, on-premise data warehouse systems require a certain level of storage and computing power, to fully utilise the advantages of the software. Those solutions often have a high up-front cost, making them a less ideal solution for a SME. Therefore, many vendors of DWH solutions offer their product as a cloud-based subscription service, which is well received by customers [2].

Cloud-based data warehouses enable smaller business to fully utilise all needed features of the DWH technology, as they require considerably less time and expertise to set up and configure. Additionally, there is no up-front cost, as there is no local infrastructure required. Many popular vendors also offer a *pay-as-you-go* subscription, which gives customers flexibility to try out their system for a very low fee. The reduction in cost and a static monthly payment model is the main reason why smaller companies can even consider implementing a DWH in their business process, as this was previously the main barrier, [3].

Business Intelligence is a kind of privilege which has been used from larger companies longer than a decade ago but during the last 10 years more and more SMEs choose to use this technology since it has been developed rapidly. These data analysis tools are now more lightweight and accessible for smaller businesses and therefore used to turn data into informed decisions in order to face main competitors. Small Business Analytics is a technique and practice which measures a specific performance of a small company on an operational or strategic level. This technique is used on small datasets to gain insights on company processes. These insights can serve as key factors determining crucial decision-making processes. In most cases when organisations approach small data, they often overlook these insights. There are different reasons why small data should be treated seriously:

- Focus on target - While big data sees the performance, small data is more focused on improving results. Key performance indicators (KPI) need to be identified and people should get one indicator assigned to track the development.

- Actionable - Big data serves information on every metric of each department. The problem is that all this data can get too general and overwhelming which forces data analysts to make strategic and organisational changes.
- All about what is happening now - Getting information from small data is quite easy and the data-source acts immediately compared to big data. If one needs a historical insight or wants to combine old data with current data, it is not possible without big data.
- Delivered ready to be served - Small data serves information in a strategic way which makes it easier to manage it and coworkers are more likely to utilise reports that will deliver clearer insights on the data.

Even though small data is a part of big data, both can be used separately, dependent on the quantity of departments in the business. Every business, even SMEs need a clear overview of where they stand on the market to achieve business goals. [4]

3 DWH Needs for SMEs

The data generated and captured by an *SME* is the most important asset available for the company itself. Since the amount of available data is constantly growing the only solution to not get into data management problems is to use a specific *Data Warehouse System (DWH).* It might not seem that every *SME* needs such a storage solution from the very first minute but there are different signs which show a business why it would be more efficient to switch over or start with a data warehouse system.

Heavy reliance on spreadsheets is for example one critical sign why *SMEs* should use a DWH System. The spreadsheet itself is a very common used file type in pretty much every business and its different departments to track data. While in most cases it seems to be pretty universal, a lot of these spreadsheets can grow to immense size and can become unmanageable. Combining the fact of growing sheets across all departments, combining these files to create a manual report takes a lot of time, not to mention the fact that every department can also rely on different sheets.

Spreadsheets are designed to take a specific amount of data divided into rows and columns. Repetitive data adding can lead to *spreadsheet overwhelming.* The file itself can handle either sluggish or just prevent the user from adding rows and/or columns. Therefore a data warehouse system can definitely increase the productivity, especially if multiple different sheets get combined.

If employed in different departments work on these sheets and one person needs to wait on specific information to create a report or analyse data, it *takes too much time just to wait* on other employees. If the data person needs to create a report gets added directly into one business centralised data source, analysing can be done at any minute. Furthermore other members in the same department also don't have to wait for data, due to an employee being too busy at the moment.

Discrepancies in data and reports can be the result of different departments creating their own data and reports. The difference in the results can be time consuming to sort out and for *SMEs* this can lead to costly mistakes. In most cases the reason is caused by adding different, sometimes not trustworthy data sources. If the point of data discrepancy is reached it may be time for businesses to sort out this problem by looking into a data warehouse system which ensures eliminating mistakes like duplicate data.

If the *time invested in creating reports* is too much, then *SMEs* should decide using a DWH System. Ideally such reports can be created with few clicks and prevent employees from going to different sources to check if the data is already updated. Since data warehouses consolidate data, all departments have to just turn to one source for data. Maintenance can be further simplified by using the ability of such systems to set up to automatically update if the source data gets

changed or updated and it is guaranteed that the data which departments rely on is always correct.

4 DWH Services for SMEs

For small and medium businesses, probably the most important aspect when choosing a data warehouse system is cost, both for the initial development and for the ongoing maintenance of such a system.

Nowadays, Software as a Service (SaaS) can provide many advantages over traditional services. The pay-as-you-go model is very friendly towards small businesses which could not otherwise easily justify the upfront cost for servers and related costs for hosting a data warehouse.

This means that, in many cases, SaaS is the most cost-effective and also the simplest solution for small businesses to opt for. When comparing them to traditional services, SaaS products virtually don't need any setup time and can be deployed instantly.

In the case of data warehouse systems, SaaS is also commonly referred to as Data Warehouse as a Service (DWaaS).

Given the advantages above, we focus in this section on some concrete DWaaS products and review what they have in common, how they differ and whether they are in fact suitable for small businesses.

4.1 Segment (by Twilio)

After first logging into Segment, the user can choose the team they are working on (Engineering, Marketing, Founder/Executive, Product, Analytics) and select the first data source, e.g. a website, programming language or HTTP API. Next, data destinations have to be selected, e.g. Google Analytics, Intercom, etc. Finally, the user gets to the dashboard, which provides an overview of all data sources and destinations and a way to add new ones.

In total, Segment supports 98 different data sources and 650 data destinations at the time of writing. Additionally, creating custom data sources and destinations by building JavaScript function that access the corresponding API. Also, by supporting programming languages as data source and webhooks as data destinations, virtually any software can be integrated.

Segment does not offer any Analytics capability on its own but is meant to simplify data collection and distribution by managing all data sources and destinations in a single place, therefore reducing complexity and increasing flexibility. For example, website analytics can be switched from Google Analytics [5] to GoSquared [6] without changing the website itself.

4.2 Panoply

Panoply is a data warehouse solution build on *AWS Redshift* and offers four plans: *LITE*, *STARTER*, *PRO* and *BUSINESS*. The last of them is specifically

aimed at SMBs, according to their own website. There is a free version available for testing which has the functionality if the *LITE* plan for a period of 14 days. After logging in for the first time, the website prompts a user to create a data warehouse, that is required to have a unique name. Next, a user is prompted with the possibility of adding a data source, which can also be skipped. Afterwards, the user has full access to the instance.

Panoply offers the integration of 122 data sources that have been integrated by the Panoply team. Additionally, there are 131 data sources, which are developed by partners, that can be added. In total, 253 different data source are supported. To analyse the collected data, *Panoply* offers the integration of 43 visualisation tools, of which 42 are *BI* tools.

Panoply is a full solution to sync, store and access a companies data, while also providing analytic features. Additionally to the supported visualisation tools, data can be structured and viewed in a traditional tabular form.

The different pricing tiers are differentiated by three main parts: amount of data sources, storage space and support. The suggested SMB solution, called the *BUSINESS* plan, includes 10 distinct data sources, 100 GB of storage and support with a reaction time of less than an hour. It also includes *Data Governance* features, yet the storage itself is based in the *USA*, without any other option. All plans offer an unlimited amount of users. More storage and data sources is possible for the *Enterprise* plan, which is adapted to a companies needs. This adds the possibility of storing the data in one of 19 different countries.

5 Performance Example

6 Conclusion

References

[1] Raghavendra Raj, Shun Ha Sylvia Wong, and A. Beaumont. "Business Intelligence Solution for an SME: A Case Study". In: *KMIS*. 2016.

[2] Alessandro Agostino, Klaus Søilen, and Bart Gerritsen. "Cloud solution in Business Intelligence for SMEs –vendor and customer perspectives". In: *Journal of Intelligence Studies in Business* 3 (Dec. 2013), pp. 5–28. DOI: 10.37380/jisib.v3i3.72.

[3] Sérgio Fernandes and Jorge Bernardino. "Cloud Data Warehousing for SMEs". In: Jan. 2016, pp. 276–282. DOI: 10.5220/0005996502760282.

[4] Sandra Durcevic. "Guide for Big Data, Business Intelligence & Analytics for Small Business". In: *datapine.com* (Nov. 2018). URL: https://www.datapine.com/blog/business-intelligence-for-small-business/.

[5] *Google Analytics*. URL: https://analytics.google.com/ (visited on 01/07/2021).

[6] *GoSquared*. URL: https://www.gosquared.com (visited on 01/07/2021).