ANALYSIS AND AUTOMATION TO IMPROVE AN EXISTING BI DASHBOARD

by

Sacha RENAULT-CRISPO

APPLICATION PROJECT, M. ENG.
PRESENTED TO ÉCOLE DE TECHNOLOGIE SUPÉRIEURE IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER'S IN INFORMATION TECHNOLOGIES

MONTREAL, MARCH 22 2020

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

BOARD OF EXAMINERS (THESIS M. ENG)

THIS THESIS HAS BEEN EVALUATED

BY THE FOLLOWING BOARD OF EXAMINERS

Prof. Alain April, Thesis Supervisor
Software Engineering and IT at École de technologie supérieure

Prof. Abdelaoued Gherbi, President of the Board of Examiners
Software Engineering and IT at École de technologie supérieure

ANALYSIS AND AUTOMATION TO IMPROVE AN EXISTING BI DASHBOARD

Sacha Renault-Crispo

ABSTRACT

The PYX4 company has consistently grown over the past several years. It has used a dashboard for reporting and BI purposes that was quickly put together in order to provide stakeholders with a better understanding of where the company is going. The current dashboard requires a lot of manual effort to update the data from its data sources and cannot be used for real-time purposes. The company will be used as a case study for this technical project.

After a short literature review of current issues with BI, the current situation at PYX4 is presented, including a BI maturity assessment. The BI needs for the company is briefly analyzed in order to prepare for improvements and the automation of parts of its current dashboard.

A script was developed in order to automate the updates from one new data source (HubSpot). Two new dashboard reports, inspired by PYX4's current reports, were developed using elements from this automated data source. PYX4 can use these two reports to enhance decision making in real-time, which was not possible in the past. Additionally, these dashboard reports were developed for usage by both operational managers and upper management.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

Abbreviation	**Explanation**
BI	Business Intelligence
CRM	Customer Relationship Management
TDWI	The Data Warehousing Institute
IT	Information Technology
ERP	Enterprise Resource Planning
CSV	Comma-separated Values
HP	Hewlett Packard Enterprise
OAuth2	Open Authorization Protocol
API	Application Programming Interface
SME	Small and Medium-sized Enterprises
DW	Data Warehouse
ETL	Extract Transform Load

INTRODUCTION

Since the late 1990s, companies in all industries have been using IT and software applications to gain a competitive edge over rival firms. Software for ERPs and CRMs helped automate several processes within a company resulting in increased efficiencies and cost savings. Given that today most companies have implemented these software applications, the mere fact of operating these solutions can no longer be considered a competitive advantage. At this point, some even consider IT to be a commodity (Carr 2003).

Companies are thus required to look elsewhere to obtain this much needed competitive edge over their competitors. Corporations had been accumulating data for several years and needed to transform this data into actionable intelligence. This is when decision support and business intelligence (BI) systems emerged and increased in popularity. The objective now became how to use this data to help and guide decision making, and thus create a positive impact on revenue and increase operational and organizational efficiency. As mentioned in (Watson 2009), business intelligence is a broad category of applications, technologies, and processes for gathering, storing, accessing and analyzing data to help business users make better decisions.

Business intelligence is complex and to this day, many companies still struggle to maximize its efficiency and implement it across their organization effectively. Issues arise at many levels–data extraction at the data warehouse level as well as at the presentation level (dashboard). The impacts are also hard to measure immediately (Jaklič 2008), so some companies have issues justifying the implementation costs of BI systems or view BI as a “nice to have” rather than an essential.

This report is presented according to the following structure. Firstly, Chapter 1 will discuss related work in business intelligence and dashboards in the form of a short literature review. Chapter 2 will then introduce the current situation at PYX4, the company that will be analyzed as part of a case study. In addition, an assessment of PYX4's current business intelligence maturity is discussed. Chapter 3 presents proposed changes and optimizations to the current BI environment. Some of these modifications are then implemented, improving the business intelligence solution for decision making. Design and development decisions are discussed. Chapter 4 concludes by summarizing the work that has been accomplished and provides insight on some potential improvements to PYX4's business intelligence environment in the future.

CHAPTER 1

Literature Review

1.1 Context

Very little research has been done on the value of business intelligence for decisions making. As mentioned earlier, most companies are aware that a competitive edge can be gained through these types of systems, yet some organizations are hesitant to invest resources for the proper implementation of BI systems. As with most IT related investments, the benefits are sometimes intangible (Jaklič 2008).

The following literature review has identified three of the biggest challenges when it comes to gaining value from BI. The main issues are data quality and integration with other systems. In order to generate significant advantages in terms of decision making, most companies will be forced to deal with these two issues at some point.

1.2 Data Quality

Research shows that data quality is a crucial variable in enabling proper decision making. Over the years, data problems have been cited as the most common challenge companies face in managing high-performance BI systems. In over half of cases where a BI implementation projects fails, data quality issues are at the forefront (Işık, Jones, and Sidorova 2013).

Decisions based on a data set increase in reliability and accuracy when the data set is big, with multiple sources, and has been cleansed. With the advancement of technology, systems using large data sets and advanced decision making capabilities are accessible to small and medium sized companies as well. That being said, collecting and cleaning data can complicate the process of building a big data warehouse, as it is demanding and costly. Even if this Data warehouse is up and running, top level management will need to implement software and

processes, as well as assign people to be involved with this big data in order for it to be used effectively for decision making. Despite having the proper processes, software and people in place, it is possible that upper management does not trust the information on a dashboard in place. This was demonstrated in Silahtaroğlu and Alayoglu (2016) when top executives were interviewed. In this study, while they all agreed that using big data along side software tools for decision making was important, only 20% of the firms actively used BI for strategic decision making. There is a clear gap between the decisions made by top executives and the information that is provided to them through their dashboard tools.

Executives reject the use of BI to support their decisions for the following reasons (Silahtaroğlu and Alayoglu 2016):

- The company is too small to have proper business intelligence tools,
- The executives feel experience is enough to handle strategic decisions,
- Business intelligence systems do not apply to their industry.

Despite the claims of Silahtaroğlu and Alayoglu (2016), BI systems can be adapted to almost any business. Tools nowadays have become increasingly available and at very low cost (Işık, Jones, and Sidorova 2013). Large corporations like Facebook (Thusoo et al. 2010) and smaller companies (Leite, Pedrosa, and Bernardino 2019) are taking advantage of open source tools that can help reduce cost. Leite, Pedrosa, and Bernardino (2019) also demonstrate that implementing open source technologies are within reach even for SMEs. That being said, it is important that these companies have the proper resources and knowledge to implement such solutions.

In addition, if the right data is available, BI tools can be implemented in order to augment decisions in a wide variety of industries. Successful BI implementations have occurred in the library and information industry (Ng and Cheng 2018), tourism industry (Albusaidi, Udupi, and Dattana 2016), social media industry (Thusoo et al. 2010) and postal service industry (Leite, Pedrosa, and Bernardino 2019).

In Wieder and Ossimitz (2015), the authors investigate the direct and indirect effects of BI management quality on improved decision making at the managerial level. BI management capabilities refer to the resources and the processes required to combine BI software and BI strategy into a BI solution that is tailored to the needs of a company. One part of BI management is data quality. Proper BI management closely aligns IT and business throughout the whole BI solution life cycle implementation and is closely related with BI success. As stated by Wieder and Ossimitz (2015), decision making benefits of a BI application can only be achieved if IT resources and business requirements are aligned through proper BI management.

Interestingly, a similar PLS analysis conducted by Işık, Jones, and Sidorova (2013) shows that BI success is significantly related to data quality but is not positively related. This negative relationship was different than their hypothesis. The findings also indicate that while data quality is not as strongly correlated with BI success it is correlated to other BI capability variables. This might indicate the dependence of data quality on other variables such as the tool's flexibility or its integration with other systems. This contradicts the findings of Wieder and Ossimitz (2015).

1.3 Integration with Multiple Sources

The BI solution may be sourcing data from an ERP, CRM, third party data and other systems. Some companies still use flat files. The quality of the communication between the various systems providing data is important as it can ensure the reliability of the data as well as adequate performance (Işık, Jones, and Sidorova 2013). For example, a proper integration between BI and an ERP can leverage key advantages from both solutions. Firstly, the data management and clean data structure from the ERP can be leveraged. When this is properly

integrated with a BI tool, the decision making capabilities are subsequently more reliable and efficient (Nofal and Yusof 2013).

Companies are beginning to integrate a growing number and variety of data sources for their BI solutions, which results in increasing pressure given that complexities increase as more data sources are combined. This is one of the main reasons why integration is considered a critical success factor (Işık, Jones, and Sidorova 2013).

CHAPTER 2

Current Business Intelligence Environment at PYX4

2.1 Introduction

This chapter presents the current business intelligence situation at PYX4. First, the company analysed as part of this case study is introduced. Secondly, the status of the current reporting and dashboard functionality will be discussed in the form of an as-is analysis. Finally, a brief analysis of PYX4's current BI maturity level will be conducted based on two widely used maturity models.

2.2 Company Overview

PYX4 is the company that will be studied in the practical portion of this case study. The company has provided access to the necessary systems and data in order to make the applied part of this research paper possible. PYX4 operates in the computer software development and consulting industries. They offer a variety of solutions ranging from process modeling to quality management as well as risk management software. In addition, they offer consulting services relating to a client's business and software needs.

PYX4 was founded in Europe in 2008 and rapidly expanded into North America in 2013. They currently have over 10,000 users worldwide using the applications they develop in house (PYX4 2020). They employ over 50 employees globally and are still growing. They have recently invested in initiatives to increase their sales, particularly in North America.

2.3 As-Is Reporting Analysis

PYX4 currently uses Google Data Studio for their reporting needs and have attached a few external sources manually via spreadsheets (e.g. Google Sheets). In terms of their business intelligence solution, all tools used by PYX4 are free. Most of the data processing is done within Data Studio. The inbound data is raw and not optimized for analysis until it is imported to Data Studio. Currently, there are no steps in place to automate or optimize the data extraction process. In other words, every time a user wants to run an ad-hoc query on timely information, the data must be exported manually from the external sources and imported to Data Studio. Figure 1 shows the current data preparation process for PYX4's dashboard.

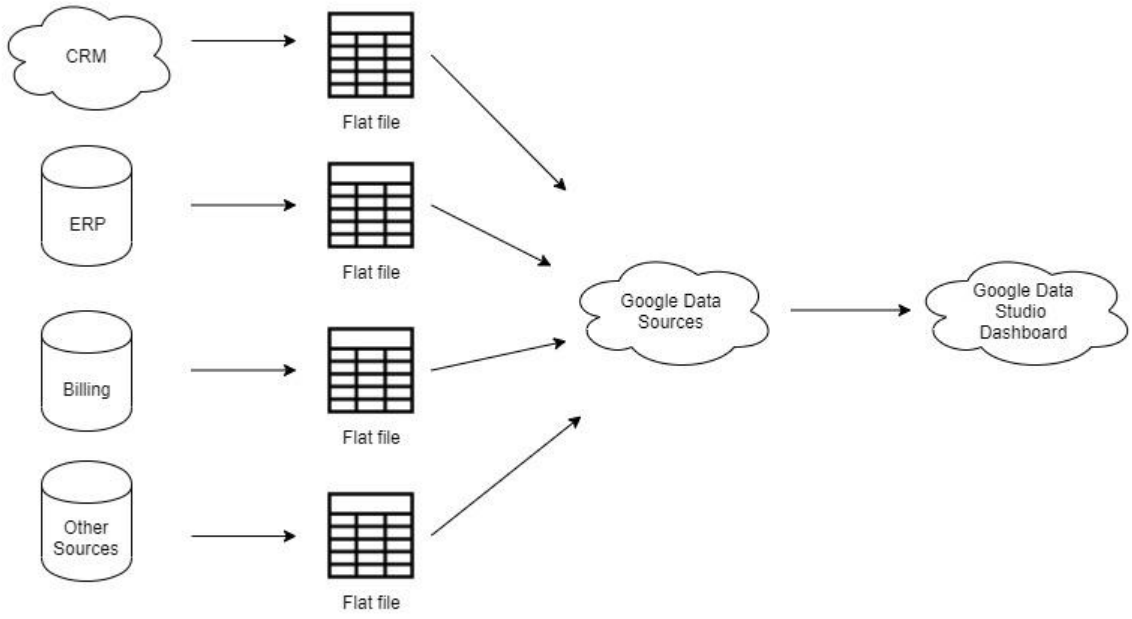


Figure 1 - Current data preparation process for PYX4 reporting

Most of the reports are built for strategic purposes. The reports are presented to upper management during periodic meetings to show the state of the company. These meetings are conducted on an ad-hoc basis and thus the reports must be refreshed when these meetings occur. The current reports show information about non-recurring and recurring income, sales opportunities and closed sales, receivables and billing. There is very little information that can be actionable on a day-to-day basis. At the time these reports were built, PYX4 needed to

understand and analyse its data promptly. The company quickly generated these dashboard reports without really considering their future use.

In order to properly understand PYX4's current situation, a brief and unstructured needs analysis was carried out. Email exchanges and in person interviews were conducted in order to better understand how to approach the company's current problems. The remainder of this section will discuss the main inquiries and conclusions relating to this needs analysis.

Is PYX4's current dashboard tool the right tool?

PYX4 is currently using Data Studio for its dashboard creation in addition to using it as a data repository for the data used in its reports. The data from all other sources is thus replicated in Data Studio. The use of Power BI and its free connectors would have enabled PYX4 to connect directly to its ERP database and CRM system. This in turn would have reduced data replication and given access to all the fields within these databases (or at least the necessary fields). For the time being, the cost and the ease of use of Data Studio is convenient for PYX4.

What are some of the security concerns associated with the current dashboard?

Currently, there is a manual process of exporting CSV files from its financial and CRM databases. These flat files are then transferred from an employee's local machine to Data Studio. It is not best practice to store these files on local machines as they can be transferred at will and can be accessed by other parties if that local machine is compromised. Using Power BI connectors could remove this step but would open other security concerns with regards to the connector itself. Another alternative would be to make web calls directly to these databases and create a script to manage the transfers.

What can be done to automate the flat file generation process?

It was clear that the labour-intensive process of asking for manual exports from different teams every time the dashboard report had to be updated was not desirable. Some automation would be beneficial.

As mentioned previously, Power BI connectors could be used to automate this process. The company was interested in using a connector to push data automatically. Unfortunately, Data Studio does not offer quality connectors. Most of the existing connectors are developed by third parties and have a cost associated to them. In addition, they would force PYX4 to lose control over some of their data. Understandably, PYX4 was not interested in considering this option, as it did not meet the security and functional criteria established by the company.

Another option explored was to use scripts to fetch the data from their systems. This method would require developing these scripts and ensuring the data is stored in a useable manner in Data Studio. While the data would still be replicated in Data Studio, only the required fields would be sent to Data Studio and would no longer require the labour-intensive process. Also, it would be a more secure method. The script option was thus favoured even if the company did not necessarily have the in-house knowledge to develop or maintain such scripts.

Migration to newer technologies

PYX4 was moving to a new CRM tool called Hubspot and all the reports using CRM data from the legacy system would have to be remapped to the new system. They are also considering upgrading other older technologies. The migration to other new data sources could follow a similar process as the one used with this the new CRM.

What would the effort be to switch tools and is management interested in changing tools?

There would have been substantial work to be done if PYX4 were to migrate to either Power BI or another tool.

All the data sources and reports would have to be redone from scratch. In addition, upper management did not want to move away from Google solutions. It also made a lot of sense for PYX4 to keep its existing dashboard, as most of it was still functional.

2.4 Current Maturity Level

In order to move forward and properly understand where PYX4 stands with regards to its BI capabilities, it is important to understand what has already been done. Once this step is completed, it will be easier to assess which actions should be taken to improve its current BI solution. Evaluating the BI maturity level of PYX4 would greatly help understand where the company currently stands according to established frameworks and then guide the company regarding how to attain higher levels of maturity with regards to its BI initiatives.

The importance of evaluating a company's maturity is clear. First, it helps create structure around BI initiatives and determine where changes should begin. It also helps identify and define the organization's program goals and creates a process and vision that is communicable across the entire organization. A maturity model provides a methodology with which to measure and monitor the state of the program and the effort needed to advance past the current maturity stage. It serves as a kind of odometer to measure and manage progress and adoption within the company for an analytics or BI program (Halpe and Stodder n.d.)

There are several types of BI maturity models. Two different and relevant BI maturity models were selected for the analysis. The first is the HP BI maturity model and the second is the TDWI BI maturity model. These models are more practical than theoretical (Muller and Hart 2016). By choosing such models, we ensure the recommendations are more actionable.

While the HP BI model has great initiatives for a business and IT alignment strategy, its does not consider any technical considerations (Neves 2018). On the other hand, the TDWI model is helpful in guiding an organization towards its next technology advancements in terms of BI. While the model still investigates a company's business situation, it offers more insight into the next technological steps for the company.

The rest of this chapter discusses these models in more depth.

2.4.1 HP BI Maturity Model

The HP BI maturity model places a company on three different dimensions in order to assess maturity - business enablement, information management and strategy and program management. We can note that there are no technology considerations with this model (Neves 2018).

The five stages of the HP BI model are the following:

1. Operation: Running the business (pre-BI stage),
2. Improvement: Measuring and monitoring the business,
3. Alignment: Integrating performance management and intelligence,
4. Empowerment: Business innovation and people productivity,
5. Excellence: Strategic alignment and differentiation.

PYX4 fits within the second stage, that of Improvement.

In terms of business enablement, PYX4 is still focused on enhancing reporting and analysis capabilities, but it has also implemented basic dashboards and scorecards, which include some planning and forecasting applications (The HP Business Intelligence Maturity Model 2007).

In terms of information management, the company is still using multiple data sources that are not integrated. As discussed by the HP Information Management team (The HP Business Intelligence Maturity Model 2007), organizations that are in the second stage take advantage of web-based BI initiatives. Most importantly, data quality and cleaning activities are ad-hoc and involve a manual process. PYX4 uses the free Data Studio solution and does not have any automation measures in place.

In general, with respect to strategy and program management, C-level and executives are still hesitant to invest effort and generate tracking for BI initiatives mainly because it is hard to

justify their investment (Jaklič 2008). PYX4 is also hesitant to perform a deep dive analysis in terms of their BI needs and assessing how they could improve their current tools and BI system.

2.4.2 TDWI Maturity Model

The second maturity model considered is the TDWI model. The Data Warehousing Institute adapted its existing DW maturity model to tailor it to the BI environment. Similar to the HP BI model, it has five stages and the goal of this model is to guide companies to go from low value and low cost operation centers to a more strategic use of BI, which can provide high value and a competitive edge (Neves 2018).

The TDWI model places a company on five different dimensions in order to assess its BI maturity - Organization, Infrastructure, Data Management, Analytics and Governance. As opposed to the HP BI model there is more emphasis and technology and data.

The five stages of the TDWI model are the following:

1. Nascent stage: Pre-Analytics era, there is no infrastructure in place and only a handful of employees using spreadsheets use data to make decisions,
2. Pre-Adoption stage: Small investment or efforts are put in low-cost front-end BI or data discovery tools or data marts for managed reporting (Halpe and Stodder n.d.),
3. Early Adoption stage: Analytics tools and methodologies in BI are being established and put in place. The process goes a step forward into data management and the creation of dashboards,
4. Corporate Adoption stage: End-users begin getting involved and use the data to make better decisions. Business starts being transformed by these initiatives as the company starts using even more data sources,
5. Visionary stage: Executing analytics programs is smooth and a highly tuned infrastructure is in place. Projects and data governance strategies are well established.

Once again, the current situation at PYX4 fits within the second stage, that of Pre-Adoption.

At the organization level, internal use of business intelligence is not sought out by many individuals. If reporting exists, it is usually done with a free tool. PYX4 has a group of people using reports. These individuals use Google Data Studio to drive their own initiatives. A discussion around the importance of using data to make better decisions has been going on for some time now.

Multiple data sources are being leveraged but they come from many different sources so managing this is a hassle. Company leadership is beginning to think holistically about data provided from the disparate sources around the company (Halpe and Stodder n.d.). There has been some talk about centralising data in a data warehouse for reporting purposes, but no projects have been formalised.

Within the company the use of business intelligence is still rather primitive. As mentioned earlier, efforts have been made to try and understand the data. There has also been some coordination between different departments to streamline the different business intelligence initiatives discussed and have been implemented with variable success.

PYX4 has put some efforts in BI initiatives and this maturity assessment clearly demonstrates this. However, more work needs to be done to improve dashboard reporting and BI maturity at PYX4. This BI maturity analysis will help guide the improvements that will be discussed and implemented in the next chapters.

CHAPTER 3

Methodology and Applied Case Study

3.1 Context

Now that we comprehend the current state and understand what PYX4 can do to improve its business intelligence maturity level, two initiatives have been put forth in order improve the current situation in terms of its reporting capabilities.

The first part of this chapter discusses the technologies used for this case study. The second section will provide details of the two improvement initiatives that were implemented.

3.2 Technologies Used

Several technologies and libraries have been used as part of this case study and are discussed below.

HubSpot

HubSpot is a CRM application. Many companies use a CRM application to manage all relationships and interactions with current and potential customers. The goal of a CRM tool is to streamline customer relationships, which can in turn improve profitability for the company (SFDC 2019). It is often implemented when a company starts growing and begins to have multiple business relationships.

HubSpot is free *software as a service* tool. In order to maximise its benefits, various modules can be purchased to facilitate customer relationship management. PYX4 currently uses several of these purchased modules.

For the purpose of this case study, we will be using PYX4's development environment, which has a certain amount of data. The development environment allows users to create applications. An application was created to enable the connection to Google Sheets and then push data fields there. The development environment also contains a subset of data from the live production environment.

Google Data Studio

This is the visualisation tool that has already been referred to on many occasions in this paper. It allows companies to use their own data and display it in a way that is adequate for their reporting purposes. As with many Google tools, it is user-friendly and made available through Google Drive as a free *software as a service* solution. It allows users to unlock the power of their data with interactive dashboards and beautiful reports that inspire smarter business decisions (Google Data Studio 2019). It can also be used to create ad-hoc reports to compare filters and organise the right data on the fly, creating unique reports. Data Studio uses its Data Source functionality to connect to various data points which can then be used for reporting purposes.

Google Data Studio Data Sources

Data Sources is the functionality within Google Data Studio that centralizes all the data points used for a Data Studio report. As mentioned previously, there are many in-house and third-party connectors that are available to connect to a wide range of databases and systems in order to extract data. For the purposes of this project, the in-house Google Sheets connector was used to retrieve the data for HubSpot.

Google App Scripts

For this case study, Google App scripts were used to extend the functionalities of Google Sheets. Essentially, custom JavaScript functions can be developed to interact with

spreadsheets. These scripts are written and executed directly within the browser. In addition, one can connect to different libraries to facilitate this process. The OAuth2 (version 35) library was used for authentication and to get access to the HubSpot data (Koleda and Timmerman 2019). With the combined power of spreadsheets and scripts, it was possible to fetch the required data fields from HubSpot.

3.3 Improvements

The remainder of this chapter will discuss the two improvement implementations. The first initiative is the automation of a data extract from a new HubSpot CRM integration. The data from this source has never been exported yet. As discussed in previous chapters, the old dashboard required manual intervention in order to retrieve CRM data. This initiative will automate this process so that the relevant data is automatically sent from the CRM to a Google Sheet that can then be used for reporting purposes.

The second initiative is to reconstruct part of the current dashboard so that it is no longer mapped to old data sources that are no longer refreshed. The new reports that have been constructed are highly inspired by the existing system. These reports are connected to the newly created HubSpot data source that refreshes regularly.

3.3.1 Automation

One of the main concerns with the previous reporting process was that it was labour-intensive, requiring a person to get the exports from several data sources manually. To help PYX4 grow in terms of its BI maturity level, it was necessary that an automated process be put in place for data imports. Eventually, the implemented automation process could be mimicked for other data sources to increase the usability of the reports. Additionally, automation would allow the reporting to at least be considered near real-time, thus increasing its daily usage.

As mentioned, initially the use of a native connector was sought. The trial version of this connector was tested. In addition to the fact that the connector was not free and lose of control over data, it was discovered that the list of fields being transferred over this connector was incomplete and thus inadequate for PYX4's needs.

The script development option was favoured. It would provide the ability to retrieve any relevant data field directly from HubSpot. While this option would require more effort, it was certainly more adaptable to the specific needs of the company. Figure 2 shows the to-be data preparation process that would be possible with the help of scripts.

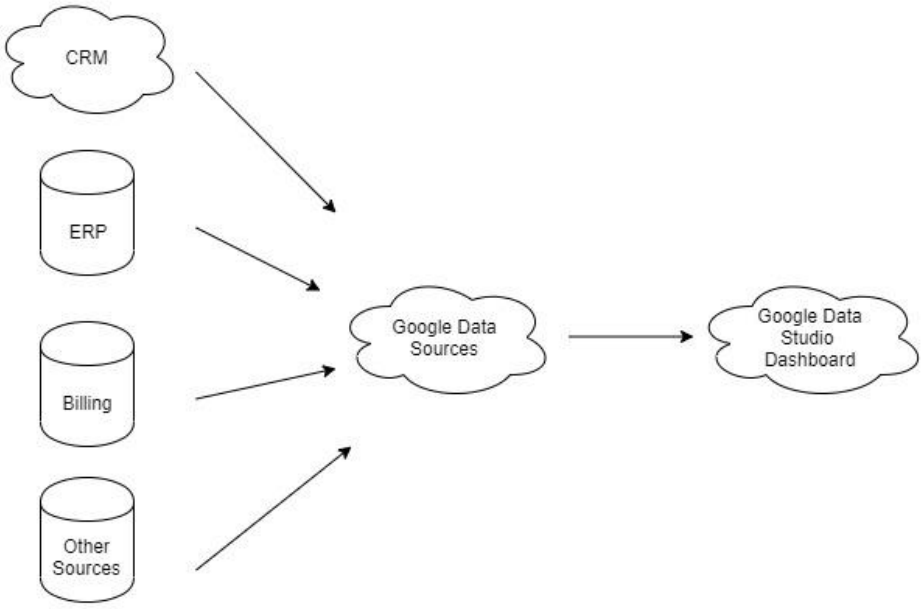


Figure 2 - To-be data preparation process for PYX4 reporting

3.3.2 Script Creation Process

The HubSpot solution has a development platform enabling companies to further benefit from the out of the box features. API documentation is also available on their website which provides the necessary guidance on how to retrieve, write or edit data from external sources.

A web application had to be created within the HubSpot developer environment. Once this was completed, OAuth ID and secret key was provided. This information would eventually be used in the Google App script as parameters for the OAuth2 function libraries.

With the help of HubSpot's API documentation (HubSpot n.d.), and the HubSpot community (Bedoret 2019), the script was developed in Google App Scripts.

The objective of this automation initiative was to be able to get the data from HubSpot and to write it to Google Sheets. Once the data was written within the Google Sheets spreadsheet, it would be easy to add this new data source within Data Studio for it to be used for the dashboard. Data pertaining to HubSpot's Stage, Deals and Contacts table were processed in this script and stored in three different sheets respectively.

Parameters, which use the internal name found in HubSpot, are passed in a URL and the script parses through all the pages in HubSpot to get records with the relevant information. The developed script has seven functions including a "get" function that stores the data in an array and a "write" function that processes and writes the information in Google Sheets. These 2 functions are used for the stage, deals and contacts tables. A "run" function was also built to execute the other six functions. Table 1 summarizes these functions and explains their purpose.

Table 1 - Script functions and their purpose

	Function name	Explanation
1	getStages()	Gets the Stages data found in the URL parameters and stores in an array
2	getDeals()	Gets the Deals data found in the URL parameters and stores in an array
3	getContacts()	Gets the Contacts data found in the URL parameters and stores them in an array
4	writeStages()	Writes the Stages array to Google sheets
5	writeDeals()	Writes the Deals array to Google sheets
6	writeContacts()	Writes the Contacts array to Google sheets
7	run()	Executes the authentication functions and then function 1 to 6

In addition, this script is executed at specific time intervals making it a data source that could be used in real-time. The trigger function of Google App Scripts allows us to run specific functions of the script at any given time interval. Currently, the script is set to run once a day in the middle of the night. The frequency of this trigger can be increased in order to achieve real-time reporting.

Several fields were pushed into the Google spreadsheet. That being said, no analysis was done on which fields should be sent. Best judgment was used to automate the fields believed to be optimal and valuable for reporting purposes. Of course, an analysis at this level would be beneficial in order to determine which specific fields are needed.

Some pre-processing activities were accomplished directly in the Google Sheets. For example, the HubSpot API pushes date field types in Unix timestamps. An additional column was created in Google Sheets to convert the Close and Start Dates of Deals to a human readable format. The columns were easily added into Google Data Studio to perform time related analysis. A similar issue was seen with the ten possible values within StageID that display in a numerical format without any meaning. A calculated field was created, within Google Studio's Data Sources, mapping these values to strings and making them available as part of a picklist. The values of this picklist then became meaningful for the user.

Another issue was that the fields obtained via HubSpot are extracted from a development environment and might not be representative of the fields currently used in the PYX4's live environment. Minor modifications to the script and HubSpot environment would be needed to extract the data from a live environment in the future.

The data source was added in Google Data Studio and can be used for reporting and dashboard creation. Unfortunately, there is no unique key to identify the customers from the CRM data so it was not possible to combine it with other data sources such as the billing or subscription

data provided by PYX4. The reports that will be discussed in the next section do not share data points from multiple sources. The reports use data from the HubSpot Deals and Stage tables.

3.3.3 Report Creation

As mention previously, the company had an existing dashboard. It comprised of five reports that used legacy data sources. The dashboard had been operational for some time but had not been updated with current data. The main problem was getting exports from external sources regularly as these reports were not automated. Every time the reports were to be updated, someone had to manually go into these systems and request exports and update the dashboard in Data Studio. Additionally, the reports used data from the payroll system and the CRM system, which were siloed. The data structure from the new HubSpot CRM is quite different and had to be remapped and remodelled for new reports in the dashboard.

While the old structure for these reports is still present, a lot of work must be done to make these dashboards operational. The ETL process, the columns available and the data types are also different. This can be seen within the different flat files provided and the existing raw data seen in Google Data Studio. While the new data provided is not sufficient to replace all of the existing reports, efforts have been deployed to replicate a portion of these reports, so that actionable information is available for better decision making. The new reports can be used as a complement to the old ones.

Two new reports were developed that use the new automated data source. The old reports were used as inspiration. It was concluded that using this new data source and linking it to older data was not desirable – it would not be possible to use these new sales dashboards in real-time. Not all fields from the old data sources were available in the new data source, which is why the report creation efforts were focused on only two reports. While some visual elements were removed in the process, other visual aids and levels of data granularity were added.

From a practical sense, it would be a waste of effort to re-create these reports from scratch as most of them had been used in the past with the old and depreciated data sources. While there is still a lot of work to be done with regards to data cleansing and data extraction activities, the raw data extracted could easily be used for some reasonable reporting functions.

3.3.4 Report Implementation Process

The new reports were inspired by the existing legacy Data Studio solution. In order for the reports to be as useful as possible for its users, they have been created as general as possible, allowing the user to perform ad-hoc queries based on several data filtering options. The reason for implementing the reports in such a way was to allow the user to query the data as he saw fit depending on his needs at any given time. The legacy reports were not interactive and were intended for presentation to upper management executives. In addition, the legacy system provided very little insight of the future.

Compared with the legacy reports, the new reports were much more dynamic thus giving more power and flexibility to the user. Contrary to the legacy reports, the intended users are managers that are involved in the sales process at a more operational level. These sales managers can thus query and filter the report based on a variety of dimensions in order to understand the past, current and future states of their sales.

The rest of this chapter discusses the two reports that were created in the dashboard.

General Deals Ad-Hoc report

This report provides the user with a very high level, yet detailed report on the current and future deals in the sales process. It includes data filtering options and visual aids to interpret the data.

The filtering dimensions available for this report are:

- Date range - Based on the Close Date of the deal,
- Deal ID,
- Sales Channel,
- Deal Owner,
- Consultant,
- Stage Name.

All these elements are enclosed by the red lines shown in Figure 3. They can be modified at the convenience of the user. Changing the filtering dimensions will modify what is displayed in the visual aids. Figure 3 can be viewed at the following address: https://datastudio.google.com/u/0/reporting/af88aadd-3023-4667-8659-d248a3137775/page/30kCB

The visual aids are enclosed with the green lines shown in Figure 3. The top left table displays the sum in euros for the deal owner based on the filtered deals. The deal owner is the person in charge of the deal throughout the sales process. The second table displays the same sum but with the associated company. The associated company is the client or potential client, depending on where the company is situated in terms of the stage of the deal.

The bottom of the report shows a time series analysis. As with the tables above it, it computes a sum in euros of the filtered deals and shows its progression over time. The deals have been separated by stage name and this is displayed in the different lines of the graph.

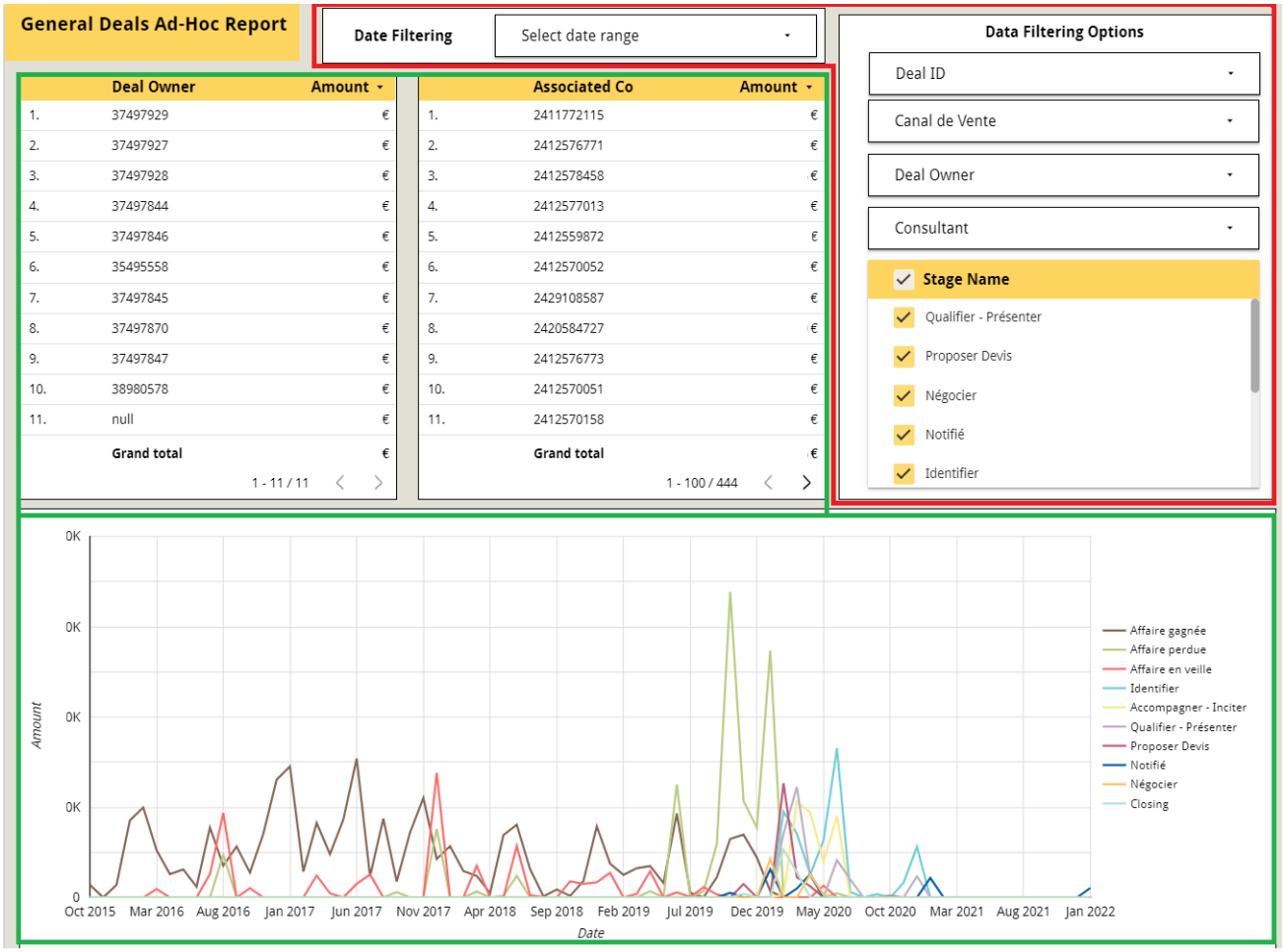


Figure 3 - New implemented report: General ad-hoc deals report

Sales follow-up planning report

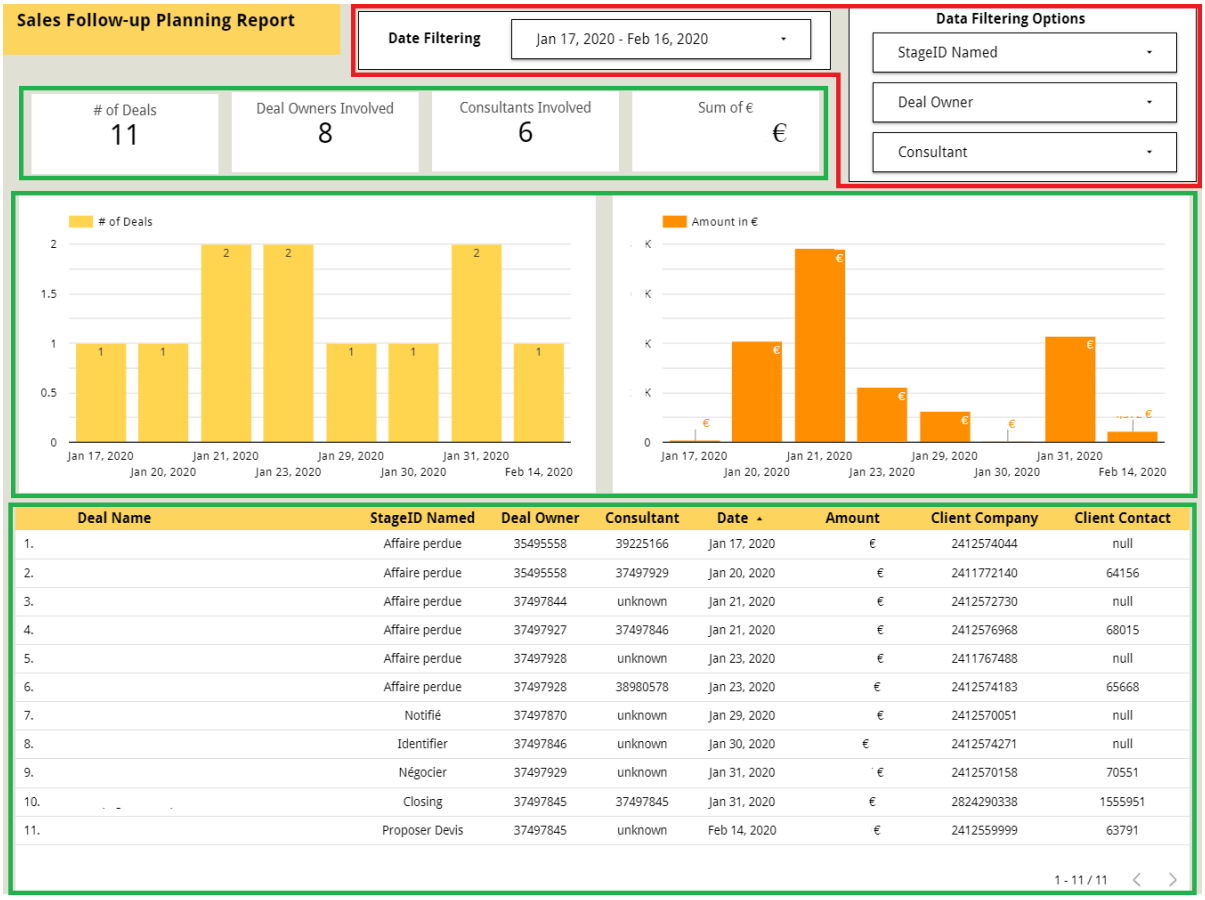
This report provides the user with information about upcoming deals. It is intended to be used by deal owners and managers to track the future actions to be taken by indicating when to plan a follow-up for a potential deal. It includes data filtering options and visual aids to interpret the data.

The filtering dimensions available for this report are:

- Date range - Based on the Close Date of the deal,
- Stage Name,

- Deal Owner,
- Consultant.

The filtering elements are enclosed in the red lines as shown in Figure 4. The date filtering option is defaulted to the previous and upcoming 15 days from the current date, spanning a 30-day period. This will help the user track any deals he might have failed to follow-up on and the upcoming deals that require attention. The date range can be changed by the user as needed. Figure 4 can be viewed at the following address: https://datastudio.google.com/u/0/reporting/af88aadd-3023-4667-8659-d248a3137775/page/zwkCB



Deal Name	StageID Named	Deal Owner	Consultant	Date	Amount	Client Company	Client Contact
1.	Affaire perdue	35495558	39225166	Jan 17, 2020	€	2412574044	null
2.	Affaire perdue	35495558	37497929	Jan 20, 2020	€	2411772140	64156
3.	Affaire perdue	37497844	unknown	Jan 21, 2020	€	2412572730	null
4.	Affaire perdue	37497927	37497846	Jan 21, 2020	€	2412576968	68015
5.	Affaire perdue	37497928	unknown	Jan 23, 2020	€	2411767488	null
6.	Affaire perdue	37497928	38980578	Jan 23, 2020	€	2412574183	65668
7.	Notifié	37497870	unknown	Jan 29, 2020	€	2412570051	null
8.	Identifier	37497846	unknown	Jan 30, 2020	€	2412574271	null
9.	Négocier	37497929	unknown	Jan 31, 2020	€	2412570158	70551
10.	Closing	37497845	37497845	Jan 31, 2020	€	2824290338	1555951
11.	Proposer Devis	37497845	unknown	Feb 14, 2020	€	2412559999	63791

Figure 4 - New implemented report: Sales follow-up planning report

The visual aids are enclosed by the green lines shown in Figure 4. The top four boxes display the sum and counts that summarize the findings of the other visual aids. The two middle bar graphs show a count of deals and a sum in euros of the filtered deals respectively. The bottom table presents detailed information regarding the specific deals that might require follow-up actions.

With these filtering dimensions and the help of visual aids, it is simple for the sales manager to interpret the data in these two reports, thus increasing his ability to make decisions based on accurate and up to date data.

CHAPTER 4

Moving Forward

4.1 Additional Recommendations

The limited scope and time for this project only permitted two improvements to be put forth for PYX4 as part of this case study. Throughout the project other opportunities were identified.

First, the next step for PYX4 in order to climb in maturity level and to enhance its automation efforts is to integrate more data sources automatically. A similar process to the one used in this case study can be used to automate other data sources used by PYX4. Billing and subscription data could potentially be the next sources to be automated. Most *software as a service* solutions provide documentation on how to extract, edit and write data via web calls.

Additional reports can be constructed for other operations with more data sources providing even more insight. An in-depth analysis of the concrete needs of PYX4 would also be beneficial for these new reports.

At a higher level, PYX4 could invest in a data warehouse in order to centralise its data. Currently, the data is spread everywhere. Each department is the owner of its own data. This can complicate any BI efforts that hope to integrate multiple sources. It would also help reduce the redundancy of data in places like Google Data Sources. Google Data Studio or any other BI tool could be connected directly to this warehouse allowing for full automation.

Finally, PYX4 would certainly benefit from increased organization of BI initiatives, with dedicated resources for these types of projects. This would also help build structure and a clear process for BI initiatives.

4.2 Conclusion

This project aimed to improve PYX4's current BI environment as part of a case study. After a brief needs analysis, two improvements were identified and implemented.

The first improvement was to automate the extraction of a data source. The chosen source was PYX4's new CRM system, HubSpot. A script was developed to extract the required fields and make them available for reports in the dashboard. The main advantage of this improvement is that it eliminated the need for a manual data preparation process. The data from this source is now accessible in real-time and users do not need to manually refresh the data every time they access the reports, saving them time. It also set the path for other automatic data extractions as other sources could use a similar process.

The second improvement was creating two additional reports for the dashboard using the automated data source. The new reports are highly customizable and help users make decisions in real-time. The main advantage of this improvement is that the reports that were created are much more dynamic and provide the user with the flexibility to do any ad-hoc query desired. Additionally, the reports can be used by operational managers as well as upper management, which is not the case with the older static reports.

LIST OF BIBLIOGRAPHICAL REFERENCES

Albusaidi, H.S., Udupi, P.K. and Dattana, V. (2016). "Integrated Data Analytic Tourism Dashboard (IDATD)." In *5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Noida, India, pp:497–500.

Bedoret, A. (2018). "Create an Automated Hubspot Custom Dashboard with Google Spreadsheet and Data Studio." *Medium*. https://medium.com/how-to-lean-startup/create-a-hubspot-custom-dashboard-with-google-spreadsheet-and-data-studio-27f9c08ade8d (December 16, 2019).

Carr, N.G. (2003). "IT Doesn't Matter." *Harvard Business Review*, May 2003, https://hbr.org/2003/05/it-doesnt-matter (December 12, 2019).

Google. (2019). "Extending Google Sheets | Apps Script." *Google Developers*. https://developers.google.com/apps-script/guides/sheets?hl=fr (December 28, 2019).

Google Data Studio. (2019). "Outils de création de tableaux de bord et de visualisation des données – Google Data Studio." *Google Marketing Platform*. https://marketingplatform.google.com/intl/fr/about/data-studio/ (December 28, 2019).

Halpe, F. and Stodder, D. (2020). "TDWI Analytics Maturity Model Guide." *Transforming Data with Intelligence*, Renton, WA, https://tdwi.org/whitepapers/2014/10/tdwi-analytics-maturity-model-guide/asset.aspx (December 16, 2019).

Hočevar, B. and Jaklič, J. (2010). "Assessing Benefits of Business Intelligence Systems – A Case Study." *Management: Journal of Contemporary Management Issues*, 15(1): 87–119, hrcak.srce.hr/53609 (March 9, 2020).

HubSpot. (2019). "API Overview | HubSpot API." https://developers.hubspot.com/docs/overview (December 16, 2019).

Öykü, I., Jones, M.C. and Sidorova, A. (2012). "Business Intelligence Success: The Roles of BI Capabilities and Decision Environments." *Information & Management* 50(1):13–23, DOI: 10.1016/j.im.2012.12.001 (March 9, 2020).

Kumar, S.M. and Belwal, M. (2017). "Performance Dashboard: Cutting-Edge Business Intelligence and Data Visualization." In *International Conference On Smart Technologies For Smart Nation (SmartTechCon)*, Bangalore, India, pp:1201–1207.

Nuno, L., Pedrosa, I. and Bernardino, J. (2019). "Open Source Business Intelligence on a SME: A Case Study Using Pentaho." In *14th Iberian Conference on Information Systems and Technologies (CISTI)*, Coimbra, Portugal, pp:1–7.

Muller, L. and Hart, M. (2016). "Updating Business Intelligence and Analytics Maturity Models for New Developments." In : Liu S., Delibašić B., Oderanti F. (eds) Decision Support Systems VI - Addressing Sustainability and Societal Challenges. ICDSST 2016. Lecture Notes in Business Information Processing, vol 250. Springer, Cham, pp:137–151.

Neves, J.M.R.M (2018). "Analysis of Lisbon Startups' Business Intelligence Capabilities.", master thesis, NOVA – School of Business and Economics, Lisbon, superv,: Prof. M. Neto, 25p. https://pdfs.semanticscholar.org/f007/71085a381cf8d9d7fbc957ee5d825d914278.pdf (March 9, 2020).

Ng, C.Y. and Cheng, E.A. (2018). "Using Dashboard and Automated Scripts to Improve End-to-End Processes, From Ordering to Cataloguing." In *12th International Conference on Performance Measurement in Libraries*, https://scholarbank.nus.edu.sg/handle/10635/136459 (December 12, 2019).

Nofal, M.I. and Yusof, Z.M. (2013). "Integration of Business Intelligence and Enterprise Resource Planning within Organizations." In *4th International Conference on Electrical Engineering and Informatics (ICEEI), Procedia - Technology*, 11:658–665.

PYX4. 2020. "Discover the Values Driving the PYX4 Team." *PYX4*. https://www.pyx4.com/en/about/ (February 16, 2020).

SFDC. 2019. "What Is CRM? - Salesforce.Com." https://www.salesforce.com/crm/what-is-crm/ (December 28, 2019).

Silahtaroğlu, G. and Alayoglu, N. (2016). "Using or Not Using Business Intelligence and Big Data for Strategic Management: An Empirical Study Based on Interviews with Executives in Various Sectors." *Procedia - Social and Behavioral Sciences*, 235:208–215.

HP Information Management. 2007. "The HP Business Intelligence Maturity Model.", HP invent series, 4AA1-5467ENW, September 2007, 7 p. https://pdfs.semanticscholar.org/926f/fd57893d1df28eb4b564282379204ef814b5.pdf (March 9, 2020)

Thusoo, A., Shao, Z. et al. (2010). "Data Warehousing and Analytics Infrastructure at Facebook." In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (SIGMOD '10), Indianapolis, Indiana, USA, June 6-10, pp:1013–1020, http://doi.acm.org/10.1145/1807167.1807278 (December 12, 2019).

Watson, H. (2009). "Tutorial: Business Intelligence: Past, Present and Future." *Communications of the Association for Information Systems*, 25(39):487–510.