Haaga-Helia
University of Applied Sciences

Improving solutions for analytics services in a mid-sized insurance company

Isabel Viljanen

Bachelor's Thesis
Degree Programme in
Business Information
Technology
2020

Haaga-Helia
University of Applied Sciences

Abstract

May 2020

Author(s)
Isabel Viljanen

Degree programme
Business Information Technology

Report/thesis title
Improving solutions for analytics services in a mid-sized insurance company

Number of pages and appendix pages
60

The purpose of this research-oriented thesis is to provide an understanding of the process of improving solutions for analytics services in a mid-sized insurance company. As there is not one solution that would fit for all, Finnish P&C Insurance Ltd will be used as a case company to demonstrate the process and to find a possible solution that could respond to their current situation better. Finding a suitable analysis tool is important in order to improve the quality of data analysis, to respond to the increasing amount of data and case specific problems in data analysis.

In the study, the focus is on reporting and visualization software due to the limited timeframe. Other data analysis solutions such as ETL tools, OLAP processing or database engines are excluded from the possible solutions. The business needs, type of data, user groups and infrastructural requirements and preferences were considered when choosing a suitable solution for the company. The purpose is to compare the current software in use to the alternative software.

The theoretical background will be based on literature review to understand analytics as a phenomenon and to understand the type of data, analytics use cases, end user groups and infrastructural requirements that are specific to the insurance industry. The empirical part uses both qualitative and quantitative analysis.

The empirical part has been divided into two iterations. On the first one, the alternative software will be evaluated based on the requirement list of desired features and capabilities. On the second iteration, the software fulfilling most of the criteria will be tested. A data analysis scenario will be designed and built using the alternative and current software in order to measure query times of the software, and to evaluate restrictively the user interface.

Results and conclusions of the study are presented in the final chapter. According to the analysis made, three different solutions with software were suggested that might benefit the case company. Also, a suggestion of an improvement of the study is presented considering the things that did not fit for the timeframe of this study. These additional points would be to test the suggested software with larger datasets and more complex analysis use cases. In addition, end user group analysis could be carried out and the usability and adaptability for the software usage considered to be done.

Keywords
Business Intelligence, descriptive analytics, reporting and visualization.

Table of contents

1 Introduction

As the amount of data expands enormously now, and on the upcoming years, and the technologies of collecting, processing and analyzing it improve, the need for proper analytics software increases among the companies. This same need can be seen in an insurance field where the data will play an important part of decision making and offering good quality customer service. Therefore, the purpose of this thesis project is to understand the analytics as a phenomenon, the needs of data analysis in an insurance field and the related aspects of it such as end user groups and usability and infrastructural limitations or possibilities and current and future trends of analytics. The purpose of this thesis project is to also offer a way to revise the current analytics software solutions in the company and to help to decide whether they could be improved in order to improve the data analysis capabilities in the company and through it to help decision making.

1.1 Presenting the case company

The case company in the project is Finnish P&C Insurance Ltd., a middle-sized B2C general insurance company, which started operating in 2012. It is 100% owned by POP Holding Oy and provides insurances together with POP Bank Group and Finnish Savings Banks Group under the brand names POP Vakuutus and Säästöpankin vakuutukset. Finnish P&C Insurance currently offers personal insurances for over 143 000 customers and they have over 100 employees. As a company, Finnish P&C Insurance markets itself as a modern digital insurance company. (POP Vakuutus; Suomen Vahinkovakuutus Oy a; Suomen Vahinkovakuutus Oy b.)

The company offers personal insurances such as motor insurance for different vehicles (passenger car, van, motorbike, moped, snowmobile, tractor, trailer), boat insurance, home insurance, travel and luggage insurance and accident insurance. The company does not offer life insurance nor insurances for companies. The company also provides insurances for persons having poor credit scores. These insurances include, along with the obligatory traffic insurance which companies are entitled to grant for their customers, extra cover for traffic insurance and home insurance. (POP Vakuutus; Säästöpankin vakuutukset.)

In 2019, the gross premiums earned totaled 42,5 million euros, claims expenses were 31,6 million euros and operating expenses were 11,2 million of euros. The total of the whole fiscal year profit was profitable by 1,5 million of euros. The biggest insurance group is motor insurance that will bring most of the gross revenue, obligatory traffic insurances form 51% of the income and motor insurance altogether 82%. In 2019, the number of new

customers per month was approximately 3800. In 2019, the company underwrote approximately 130 500 new insurance contracts and the premium income was 45 787,5 euros resulting 17,8 percentage growth compared to previous year. The main functions of the company include insurance underwriting, marketing, risk management, actuarial function, finance, investment management, outwards reinsurance, claims management, ICT-function, internal audit and compliance function. The CEO of the company is Harri Mattinen who has been working in the position since August 2017. The organizational structure consists of HR and legal services, risk management, insurance investigation, actuary and finance, ICT, marketing and communications, underwriting, customer and product management and claims management. (Suomen Vahinkovakuutus Oy a; Suomen Vahinkovakuutus Oy b.)

I have been working myself in the company nearly two years now, first as an analyst, and currently I am working as a data integration developer taking care of the company's data warehouse, ETL and analysis tools among other tasks. This project is, therefore, also offering me new perspectives to review and improve the existing analytics solutions in the company and to recognize the needs for them better. As my work is closely related to the topic of this project, parts of the current problems and needs of the case company are based on my own observations and discussions I have had with my supervisor and team members during the time I have been working in the company.

1.2 Thesis objectives

The objective of the thesis project is to improve an insurance company's solutions for analytics services based on their business needs, user groups and their current infrastructure and technology. The objectives are to review the current analysis software used in the case company and to search new candidates that will meet the criteria set for the software. The aim of the search will be to find good candidates to either replace the existing software or to use on the side of the current ones in order to improve the data processing and analyzing capabilities. The thesis project may also help other companies in the insurance field, and other fields, to review their analytics software solutions and to find suitable solutions to improve the quality of the data analysis and to respond to the existing problems.

1.3 Delimitations

The thesis project has been delimited to review the Business Intelligence software meant for data visualization and reporting due to the limited time frame and, therefore, solutions

used mainly for data modelling and integration i.e. ETL tools, OLAP processing or database engines will not be considered as a solution. However, possibility to use data modelling software and engines such as BigQuery or Microsoft Analysis Services cube technology will be taken into a consideration when reviewing the suitable software. The purpose is not to find a solution that would fit for all insurance companies because there is not a one solution that will fit all but instead to give an example to support their own process of reviewing the analysis software needs and usage. Also, budget review for the software will be excluded from the project since the software licensing prices are around the same range in all products reviewed here, and, because none of them exceeds the case company's current budget significantly.

1.4 Research problem and task setting

A research problem in this thesis project is trying to find a proposal of a suitable analytics software solution for a mid-sized insurance company based on the industry specific use cases, understanding user groups and considering the infrastructure and other requirements for the analytics software using Finnish P&C Insurance Company Ltd. as a model case company. This will be conducted by revising current technologies in use and searching possible candidates by checking the needed features. In the end, the software that seem to best respond to the requirements, will be taken into a closer test that will be conducted by using built data analysis scenario. Things tested will include query speed and the usability of the software interface.

1.5 Structure of the thesis

There will be an overview of analytics as a phenomenon and how analytics is used in an insurance field. Different terms and concepts of analytics will be introduced as well as typical use cases of data analytics in an insurance field and data type and quality. Theoretical framework includes also explanation of the typical end user groups of analytics software and infrastructure and requirements for it. In the end, current and future trends in analytics will be considered.

In the empirical part, the reasons for the project will be introduced in more detail and based on the requirements for the software, some commercial and open source solutions will be chosen for the closer analysis. From those, the software that has most of the searched features will be taken into a testing face. Also, other opinions about software will be introduced. In the end, the most suitable solution for the company will be introduced.

In the discussion results will be discussed, their reliability and significance explained, and conclusions drawn. The discussion will provide an overview of the study, results and what do they mean for the company.

2 Theoretical framework

In this chapter, there will be an overview of analytics in an insurance field as a phenomenon, what can be understood as analytics and what does it mean specifically in an insurance field. What kind of needs analytics has from the perspective of business processes, decision making and planning? What kind of requirements or restrictions the type of data and infrastructure will set for the analytics software solution, what kind of solutions are popular among users and companies right now and in the past, and in which direction analytics is moving in the future in the insurance field as the goal is to improve the accuracy of the analysis and to improve the business performance of the companies. One interesting question is also, how the end users will affect for the decision making when choosing adequate analytics solution for a company.

2.1 Analytics as phenomenon and analytics in an insurance field

Analytics, in general, means understanding data and finding useful information and patterns out of data through different processes, scientific techniques and with different tools and software. Techniques include, for example, statistics and mathematics or machine learning such as predictive modelling. Analytics can be divided into further categories. The most common division is between descriptive, predictive and prescriptive analytics. Descriptive analytics focuses of what happened and why, and it uses usually historical data for analysis. Predictive analysis tells what is probably going to happen in the future and uses statistical modelling and machine learning's supervised learning. Prescriptive analytics instead suggests what should be done in order to change things happening now or in the future so basically it will be offering new operational models instead of analysing solely data with focus on the past events and, for example, successfulness of the past marketing strategy. The research, European Vertical Market Survey conducted by International Data Corporation (IDC) in 2016, suggested that almost half of the insurance companies are using analytics and Big Data in their business. Other wider perspectives of analytics are Business Intelligence and Artificial Intelligence. Figure 1 in the end of the chapter is compiled based on the chapter 2.1 to explain better the relationship of different forms of analytics. (Cowdrey 2018; Hirst 2017; Knight 2018a.)

2.1.1 Descriptive, predictive, prescriptive and operational analytics

Descriptive analytics can be described to be the very first form of analytics when starting to process data further. For example, as trying to understand Big Data, forming chunks of information out of data that will be useful, would be called as descriptive analysis. Data mining and preparing for use will be the main goals as especially Big Data is not usually

understandable for humans as it is but requires further processing. Information retrieved summarizes the past evets to describe what happened in the past. (Bertolucci, 2013.)

Predictive analytics can be developed when descriptive analytics has taken its place. The meaning of it is to process current and past data further so that it can be used to predict probabilities in future trends. Predictive analytics uses data mining/text mining, machine learning, statistical and predictive modelling and processes the data further from the first understandable form of it. It tries to fill in the gaps of which are not know yet by the data the company has. (Bertolucci 2013; Saporito 2014.)

Prescriptive analytics is a more advanced form of predictive analytics, also focusing on the future outcomes and predictions. However, it differs from predictive analytics by not only predicting future trends but also predicting possible outcomes from a course of action by predicting its outcome and consequences. Basically, it can predict the best possible course of action to take from all the forecasted possibilities. (Bertolucci 2013.)

Operational (embedded analytics) and traditional analytics in an insurance field rely heavily on predictive analysis. Operational analytics is often divided between different business operations such as marketing, sales or claim and compensation services and to the applications they use to find solutions for the current problems often real-time. As an example, to review a credit worthiness of a customer before the purchase. The traditional analytics include more wider view of the company's operations and outcomes and suggests future trends based on past events. An example would be to follow a successfulness of a new marketing campaign on the long run, has it been increasing the sales of a certain product compared to the past. Is it useful to continue the campaign or to stop it if it does not benefit the company or causes possibly even financial loss over time? (Saporito 2014.)

Some of the key terms explained: Analytics also include, is part of, or is closely related to Data Science, Business Intelligence (more commonly BI), Big Data and Machine Learning and Artificial Intelligence (AI). As mentioned above, predictive modelling, which is a subarea of machine learning, can be used in prescriptive analytics. Machine learning itself will be recognizing and creating rules itself without human interaction in other than defining the desired result. Machine learning is based on programming and mathematics. Analytics itself is a part of Data Science in general. The basics of data science is to understand data and build different processes and models from it, but it contains a larger scale of understanding of different topics. Machine Learning, for example, can work as tool inside Data Science. The difference in machine learning compared to other rule-based analysis is that

the logic which computer has built cannot be usually converted into readable form for humans and, therefore, will be difficult to, for example, document or verify. (Scherbak 2019.)

2.1.2 Business Intelligence and other concepts of analytics

Business Intelligence, generally referred as BI, is not only a form of data analysis but is a concept containing the methodology, processes and the technological tools needed for processing and analysing data. It includes, for example, processing the data into usable form before storing it into a data warehouse, data lake or data marts, storage processing such as ETL (extract, transform and load) or ELT (extract, load, transform) and access to the data such as BI portals that offer graphical user interface for the data manipulation and processing, analysis and querying and also reporting and other end results like dashboards and data visualizations that will follow the analysis made. BI tools' most important tasks will be data extraction, architecture and modelling, querying and visualization capabilities and they will be possible to achieve with variety of different tools and software. Generally, this stack will be the area of a BI analyst in the company to take care of the data ETL processes and/or reporting and ad-hoc queries. The results received in BI will support process making, market research and other business areas of the company. Especially on the insurance field data-analysis has been considered important due to very risky business environment. (Knight 2018b; Peterson 2013.)

Business Intelligence (BI) is using descriptive analytics and, therefore, will tell the company of what happened in the past or what is happening now and helping the company with its decision-making and to gain competitive benefit. It describes the state of company and its business processes and operations or productivity. Another term what is used is Business Analytics (BA) which uses predictive and/or prescriptive analytics instead and therefore will focus on predicting the future by analysing the data further. Business Intelligence and Business Analytics are so far their own fields of analytics. In BI different tools and software will be used to gain information useful to business decisions. The meaning of BI tools is to both store, access and analyse data and to prepare graphs, charts or other type of reporting from the outcomes of analysis produced. Data storages include data warehouses, data lakes, data marts and, for example, cloud storages. The usage of BI will help the company to automate the process of analysing data and fuse/combine it from the different sources. There is new division in the field of traditional BI, self-service BI, which makes end users of the company to participate more of the analysing of data. Of course, even in self-service BI, IT-professionals will make most of the data processing in the background to help to build the insights but when the software are evolving they can dig the data themselves up to a certain point without need to have everything completely ready-processed. In the future, BI might be turning into augmented analytics basically

containing both business analytics and business intelligence in a form of machine learning combined with the traditional BI software. (Knight 2018a; Knight 2018b; Pratt 2017.)

Analytics itself is a broader view of different processes, tools and techniques for interpreting, reading or predicting things from data and making it into a meaningful form. Business Intelligence could be referred as a subset of analytics, or Business Analytics (BA), as its concept is more narrowed down and easier to define. Business intelligence supports the decision-making and business processes, optimizes them and evaluates their efficiency by backing it up with analysis of past and current data using different technologies, methods and solutions. The goal of BI is to increase the productivity and make the processes and decision making reasonable so that it supports the business growth and prunes the inefficient processes and decisions or finds possibilities for improvement. As BI focuses heavily on current and past processes and to interpret their successfulness, analytics as a broader sight will bring the points of analysing the data further for the root causes or consequences. Analytics will be used to widen the angle and to bring more meaning the decision making and data results. (Adair; Caldwell 2017.)

Big Data can be defined to be large volumes of data that will not make sense itself without processing, storing and analysing it but which cannot be done with regular database software and tools. There is no specified limit for the Big Data but instead the definition of it can change over time as the systems processing it will evolve and be able to handle more information. On different fields also, the size that is counted as Big Data can vary. As the amount of information is constantly increasing, big data and technologies using it will make more and more importance both for the consumers and companies. When talking about big data, the amount of it to be captured and processed is another question of the technological point of view but another thing is how to make use of it and can it be made useful or meaningful. Data can be collected from mobile phone users, from websites, smart devices, sensors in cars, organizations and much more. The daily use of different multimedia data will create lot of trails and content what we understand as Big Data. Companies are storing a lot of data about their customers that will be created as a side-product for example when visiting on the company's website. The study conducted by McKinsey Global Institute form 2011 suggests that the usage of big data will bring important value for the world economy by improving productivity and competitiveness both on public sector and companies. One example has been given from U.S. health care sector where the estimated benefit from using big data could be even 300 billion dollars yearly which would include the saving from health care expenses around 8%. Therefore, usage of big data is crucial on every sector and on insurance field where it could be used for example detecting frauds and understanding or even predicting customer behaviour. (Brown & al. 2011.)

Artificial Intelligence is another form of analytics where the data will be used in order to create processes that learn and automate functions that humans have traditionally made. Examples of these are, for example, facial and image recognition and self-driving cars. On an insurance field the typical use cases for AI and machine learning algorithms are using chatbots when offering customer service online, automating claim handling processes or parts of it, for example, by settling a claim or predicting repair costs or using Big Data to recognize or evaluate damages from photos. (National Association of Insurance Commissioners 2020.)

As discussed about different forms of analytics that are appearing in an insurance field, on this project we will be focusing on the Business Intelligence, Business Analytics and descriptive analytics as it is the foundation of other forms of analytics and which still plays the major and most critical role in most of the business functions in an insurance field. When the preparation and data architecture have been organized wisely and the right tools to make most out of BI have been chosen, after that also the use of predictive or prescriptive analysis with machine learning, AI and/or Big Data will come into a question. Therefore, the main target of this project would be middle-sized insurance companies that will be benefitting most of this project but also larger companies that have been on the field decades might find benefits from this study. Basically, using data on an insurance field gives nearly endless opportunities to develop the efficiency of business processes, lessen the workload of human impersonator by automating the processes, to make better decisions based on more accurate numbers and results of the analysis and to predict usefulness of the ideas based on future predictions such as how the profit will vary depending of different factors and past variables.

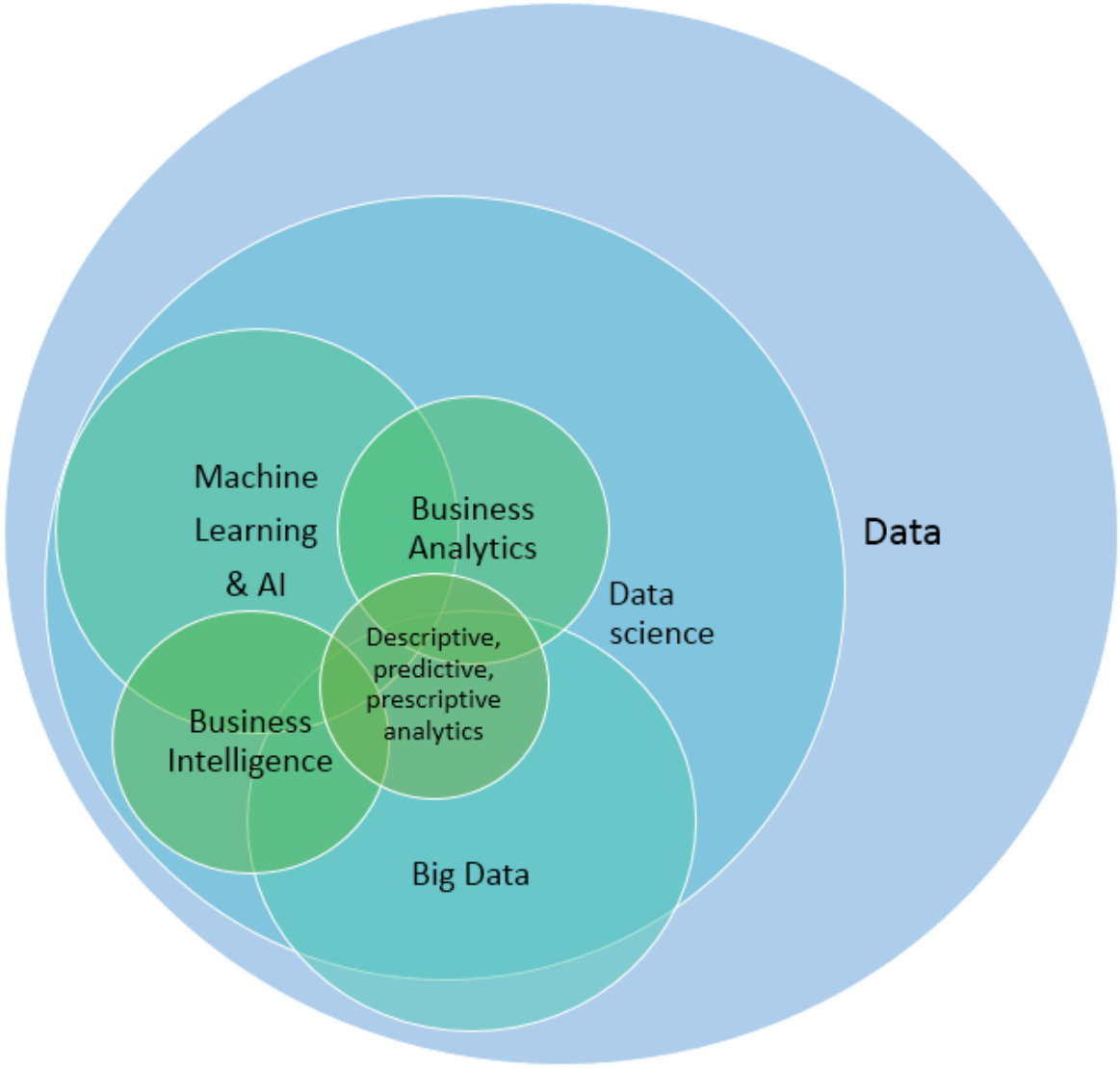


Figure 1. Graph explaining the relations of different terms of analytics

2.2 Use cases of analytics and data

In the insurance industry, Business Intelligence and descriptive analytics have many use cases on different business processes, decision making and planning. To conduct a reliable analysis and to be able to use the needed software for it, the data quality and governance must be ensured, and its storage solutions thought carefully.

2.2.1 Business processes and decision making

The need and possibility for data analysis and the use of adequate software will arouse from the companies needs such as business processes, decision-making and planning. In an insurance field the major areas are risk assessment and pricing, fraud detection, customer experience management, marketing, clientele and sales insights, process automation and financial decision making some to mention. Collection of the analytics use cases can be seen in Figure 2.

Sales and clientele base analysis is important for the company to understand, for example, how the sales are developing, what kind of customers are the major groups that will bring profit, which customers will profile to be the most risky ones for the business, how different campaigns are turning out or should they be stopped or maybe not to be initiated at all based on previous experience. Risk assessment will lessen the risk from different factors in the company and fraud detection will catch the claims which should not be compensated due to a fraud. Different automated processes such as segmentation or pricing will require also analysis made based on the data available from the policies and customers. Marketing requires analysis to define the groups to target with, for example, different campaigns with discounts. Analysis will be used also for claim process automation where big data might be in use, customer interaction with chatbot technology and image detection. However, as the concentration of this project is merely on Business Intelligence, these areas will not be covered deeper. Data-analysis is also required on customer and sales insight development. Understanding the type of customer base of the company helps making targeted and personalized marketing measures to selected groups. Understanding and analyzing the sales data makes the financial decisions and decision of the successfulness of a project easier and it results in more reliable decision making. More specialized marketing based on the customer data and the data analysis about the information related to service improvement makes the customer experience better and more personalized and, therefore, also benefits the company and its reputation. Fraud detection will be made easier when patterns can be found from the customer behavior based on the earlier fraud detection cases. Data can be processed and be more easily available to claim handlers or automation system who can detect possible risks beforehand and make them to be handled, for example, manually. Financial decision making, for example, centralizes on financial solvency in the company. All the processes and decision making that supports it will make the insurance company more successful and reliable on the long run or in the eyes of the customer and investors. Solvency is supervised in Finland by Financial Supervisory Authority (FIN-FSA). (Financial Supervisory Authority 2018; Mallon 2018.)

Figure 2. Use cases of analytics in an insurance field

As there are many business cases producing data, the amount of data has been growing in the insurance industry as in many other businesses. Especially the amount of real-time data has been increasing enormously during past decade. The increase of data in itself has given both new opportunities and developmental challenges in an insurance field. Data analysis has become something that all the companies use as an advantage, but the techniques and solutions are individual for everyone and based on their exact needs. The type and amount of data defines which methods would be most efficient in use. Since lots of data comes from variation of different sources, also the solutions need to be adjusted according to it. Part of this is, for example, the data that comes from sales applications such as through websites, mobile apps or other software used to register customer and his/her information. On the other hand, there is data that is coming from third parties and which will be used for different purposes such as service or marketing development. This could be, for example, data from Google Analytics which can help companies to understand how the customers are interacting on the company website some to mention. Example of the data sources can be seen on Figure 3. From the data side, the important thing to understand is the quality of data which is one of the foundations of successful data analysis and the data to be used through any software. (Google c.)

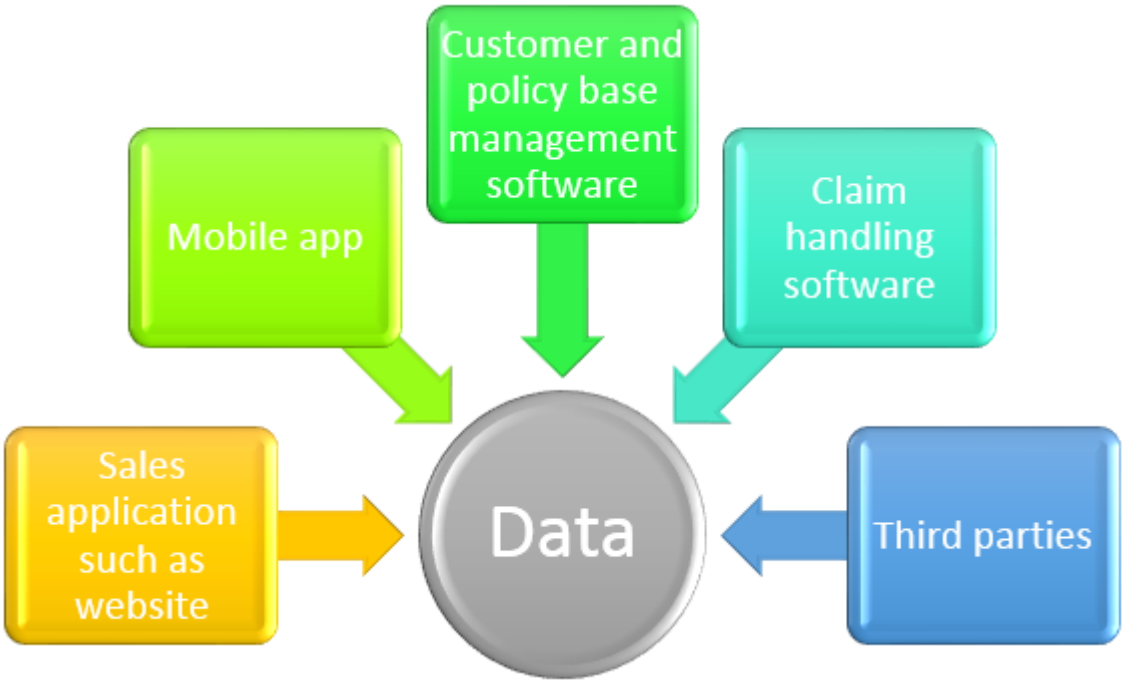


Figure 3. Example of data sources in insurance field

2.2.2 Examples of data analytics uses cases

One example of the use cases where data analysis has been used as an advantage in order to improve pricing by evaluating the risk of the policy holder to the insurance company, is by If P&C Insurance Ltd. In 2015, If P&C Insurance Ltd. started to use new factors to price their passenger car insurances. The new factors were the age of the driver and the frequency of usage of the car. This decision was based on the analysis which showed that, for example, the young driver with less experience is more vulnerable to damages than a more experienced driver. More damages seem to happen also to drivers that drive a lot than for those that drive very little per year. For that reason, If P&C Insurances Ltd. decided to add question about driver's age (under 24 years old) and the estimate of kilometres driven per year. These two factors are used for pricing of the Casco insurance and for the experienced driver who drives shorter distances, the insurance will be cheaper than for a young driver with long distances. (Moottori 2015.)

Another example is Pay how you Drive risk-based pricing method that at least have been considered by Fennia Mutual Insurance Company, If P&C Insurance Ltd., Pohjola Insurance Ltd. and Turva Mutual Insurance Company. In addition, Pohjantähti Mutual Insurance Company has already tested the Pay how you drive pricing method in the past but declared that the technology to use it were not enough for it back then. Pay how you Drive method is based on the idea that a customer gives permission to the insurance company to follow their driving and the insurance company will collect data, analyse it and use it for

the pricing of the customer's insurance. In this method, the customer would have possibility to affect for the cost of the insurance by driving with as little risks as possible. The data collected would consist of, for example, type of driving (acceleration, braking and motion control), hours of driving (rush hours vs. silent hours), place (capital area or low-density residential area) and speed. This data can be then analysed and used for the pricing. In Germany, many insurance companies have already taken this pricing method into use and it lessens the cost of insurance especially for young drivers whose insurance costs typically are higher than older drivers. (Ziemann 2016.)

2.2.3 Type and quality of data

The challenge is how to maintain data, how to store it, how to ensure the data integrity and quality checking and how to find the suitable software, tools and methods to make most out of it with data analytics. As there is such a huge amount of data available from different sources, also combining and analysis of it becomes more complex than ever before. This causes challenges for the companies to decide which would be the most suitable architecture for the data and which would be the most suitable tools to handle and analyse it. In this project we will be focusing on the problem of analysing data that comes from multiple sources and is, therefore, difficult to combine with other data in an efficient way to perform analysis. The data collected can be both structured, semi-structured and unstructured data but the amount of semi-structured and unstructured are increasing currently. When talking about data itself, there are different concepts regarding data governance such as data veracity, consistency, quality assurance and cleansing of data, and on the other side data integration and data access which will be discussed more on the infrastructure section.

Structured data is the type of data that is in the most understandable form and easiest to use in analysis without the need of processing it too much. It is created either by machine or human and it can be stored on database tables. Example of the data from the insurance field could be, for example, all the customers policies and product types with their prices. The data entities can form relations and they are easily combined with other data such as customers' purchased insurance types could be combined with the ages of customers to see what kind of insurances people by at certain age. Structured data is the most typical type of data to have been used on data analysis. Typically, analysis is made using structured query language (SQL) to retrieve and combine data from the relational databases. (Marr 2019.)

Unstructured data cannot be generally used without pre-processing it first. It is bigger junks of information such as text files, audio or pictures. It usually cannot be stored to database column either and it does not have a data model where it could be combined to some other data as on the structured data example. As processing the useful information out, for example, from the text file is more difficult, this type of data usage has become popular only after machine learning algorithms and artificial intelligence has evolved. Using that technology, it is possible to notice patterns of useful information from unstructured data. If structured data is stored on relational database, unstructured can be stored on data lake solutions, data warehouses and NoSQL databases. (Marr 2019.)

Semi-structured data has elements from both structured and unstructured data. It might contain blocks of structured data but contains also structured data that can help to link the data with other information on the database. As an example, claim report might contain the explanation of the accident happened which is unstructured data, but the document also contains possibly identifier for the claim and customer and date when the document is made and/or when the accident happened containing therefore structured data which can be used to combine the claim data to other customer data. (Marr 2019.)

Data governance is used to make sure that the data is consistent, reliable and integrated. It includes the data's quality assurance and cleansing of the data, recovery and access regulations as well as security and privacy. If the data is not managed properly, problems might occur, for example, with data accuracy or with misuse of the data. In data governance, the data will be combined from smaller blocks, data silos, that contain data from one business unit or area such as data from marketing, customer data or claim data. The data can be combined to one centralized system such as data warehouse or data lake where it can be modelled and combined accordingly. Data governance also includes data management by monitoring and making sure that there is no erroneous or duplicate data and that the data will be handled and accessed by the regulations. These regulations try to guarantee data consistency, veracity and right usage of it and an example of it is General Data Protection Regulation (GDPR) in European Union. The company can make rules, policies and procedures of how to take care of data quality and usage according to the law. When the data quality is high and the right persons have sufficient access rights on the data, the

more reliable decisions can be made also based on the analysis and the risk, for example, for data leakage decreases. The main principles of data governance are presented in Figure 4 below. (Rouse & Vaughan 2019.)

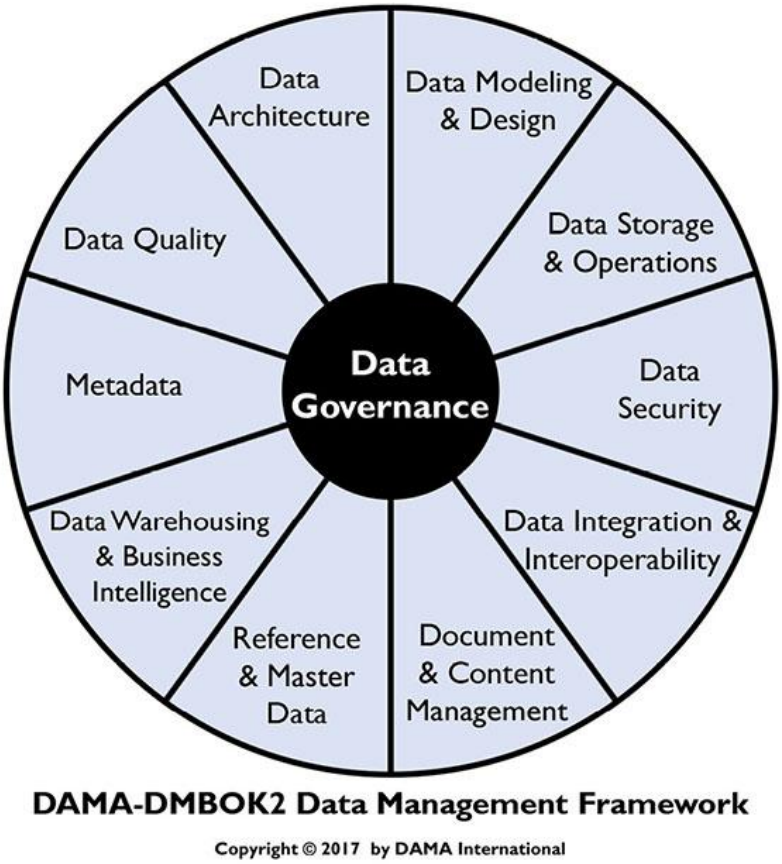


Figure 4. Main principles of data governance. (DAMA International 2017.)

2.3 End user groups and usability

The end user groups on this case will be the analytics team in the company and other employees from different departments whose responsibilities will include to monitor and analyze the produced data, for example, in risk management. End user group types will be divided into different categories based on their skill level regarding of current analytics software and tools in use. End users can be generally divided into beginner, intermediate and expert users. Of course, when the use of the software becomes more familiar, also the skill level will change, therefore as it is not static state, the definition will be for the user's current level in analytics software usage. (Mitchell 2017.)

For the beginners, the complexity of the software can be challenging as they do not have enough acquired skills to fully understand the use of the software and its features. They often also do not learn by reading the manual or instructions but through trial and error. They have some interest usually to progress and move forward for the intermediate level

and the learning process can be rather fast but there is also risk that if learning to use the software feels too demanding, they will not continue to use it. Prospective buyers will have usually similar mindset than beginner users to evaluate and test the new software. However, prospective buyers will not be the end user group in this project. (Mitchell 2017.)

Intermediate users will be the most common group among the end users. They have the basic knowledge of the general features and they will learn fast those ones that they need on their work most frequently. Very often even though they do not have interest to continue exploring the so-called extra features when they are not needed and will therefore advance slowly into a next user level. Also, from intermediate group the progression of level can turn to lower level when the software is not used often, and they might not remember after a while where to find or how to use all the features. (Mitchell 2017.)

Expert users are the users that have the most advanced skill level among the end users of the software. They typically have a very good understanding of how to use the different parts and features of the software and will remember where to find everything as they usually use the software professionally. They can also advice others to use it. Maintaining the skill level requires frequent use of the software and new updates and features are often appreciated also. Normally, the users can be experts in some parts of the software they are using but it is also possible to be intermediate or end user on other levels. (Mitchell 2017.)

When comparing the users in this project to the defined user groups above, we could state that by the experience in using different analytics software and tools, the users from analytics team would fall into the category of either intermediate or expert user. Analysts typically will be expert users on the different analytics features of the software and experts on the administration features if it belongs to their job maintaining the correct settings and to solve possible problems. Other users that will not be using it daily could be either intermediate or beginner users. Overview of the different user groups is presented in Figure 5 below.

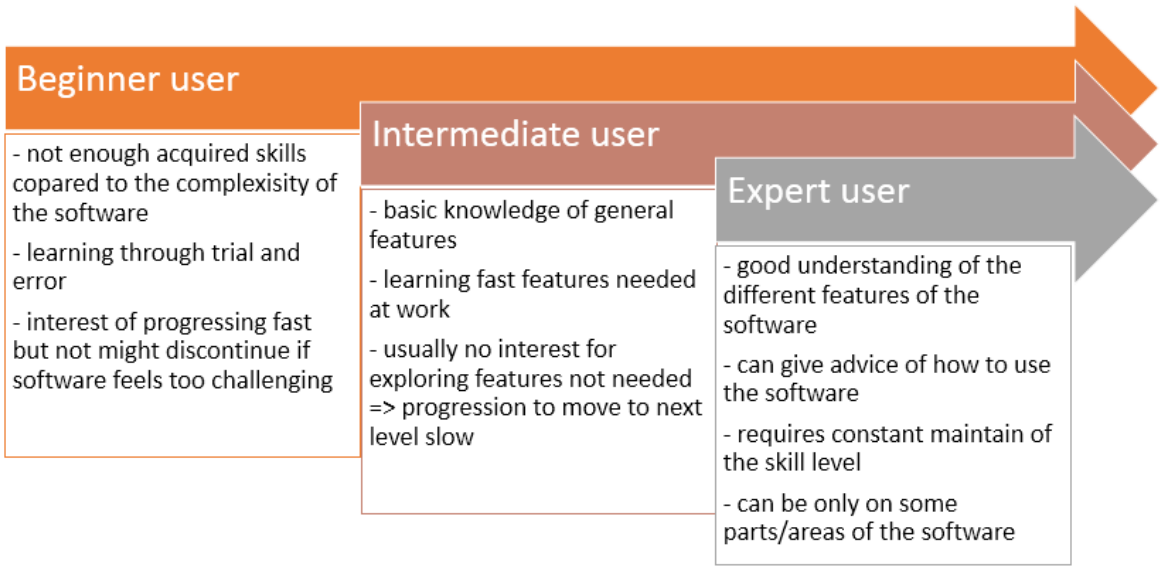


Figure 5. Overview of different user groups.

2.4 Infrastructure and requirements for analytics software

Business Intelligence solutions are built on top of the IT architecture in the company. It is beneficial that the architectural solution is not built only for one software but that the requirements work for multiple software as the requirements and need may change over time. Therefore, the system should be flexible for the possible future changes. One of the first things to consider when building business intelligence architecture, is what kind of deployment environment there is for data storage and computing; if it is data warehouse or data mart in physical server, a cloud-based solution or a hybrid which is the combination of both. If there is a physical server, it is important to consider of the resources and how powerful it is, accessibility of it with other software, if the servers are clustered and what kind of system and data recovery there is in order to guarantee fast processing of data and failover functions that guarantee that the system does not go down when it is needed. The powerfulness of the hardware is also dependant of the amount of data that needs to be handled and stored including processed data, metadata and the data moved and processed through ETL's. This depends of the accepted requirement for the speed compared to the amount of data being processed. Also, how the data will be processed, is important regarding the choices made in analysis such as online analytical processing or online transaction processing. (Cook 2013.)

When considering a new analysis software for the company, the technology that already exists needs to be considered. This includes the things mentioned above such as speed, access to data or migration with other software, and availability of the data from current technological stack. As the amounts of data that will be processed grows larger, to be able

to use it efficiently, the data processing needs to be relatively fast. When the data processing happens with faster speed, the data can be used to decision making, for example, by ad-hoc queries on a short notice. Also, the analysis software should be considered through the speed of data processing. The technological architecture can be supporting either data updates on real-time or in batches. What is the requirement in this for the analysis purposes? How fast it will be to use either on-premise or cloud solution for the data processing. Another fact to consider is how the chosen analytics platform can access the data. Can it access the current data warehouse, cloud storages and other separate data sources, excel files etc. An example of the IT architecture enabling data analysis is presented in Figure 6. Also, the analysis done needs to be able to be delivered and available for the persons needing it. If the system is unavailable, or the data processing takes too long, or all the needed data sources cannot be integrated with the analytics software, then the data probably will not be available to the personnel either. From the infrastructural side, this means good management, system flexibility and disaster recovery, some to mention. As a conclusion, when choosing the right analytics software, it is crucial to consider where the data is stored, how it is stored i.e. batch runs or real-time updates and what system integration will be needed. (Perret 2014.)

As for the deployment environment, both the server solutions and analytics software will be offered as on-premise and cloud-based solutions. On-premise is a physical server that will need to be maintained either in-house or by a third party and cloud-based server is a virtual server managed through cloud computing. Main differences lie in costs, reliability, security and scalability. Setting up an on-premise server is usually more expensive than setting up cloud-based solutions as it requires installations and hardware, security is the responsibility of the company and maintenance will be required after installation. Cloud-based solution will be quicker to set up, the maintenance costs are not as high, and the solution provider takes care of the security. Of course, if somebody else takes care of the security then there will not be the same control over it by your company. On-premise servers work without internet connection but cloud-based will need one to work, also in an on-premise server scaling requires manual work and hardware and software resources when in cloud-based solution adding, for example, more storage is typically a simple process. Besides, there are also hybrid solutions available where the company can make use of the both types of solution and their features based on the needs of the company. (Munk 2019.)

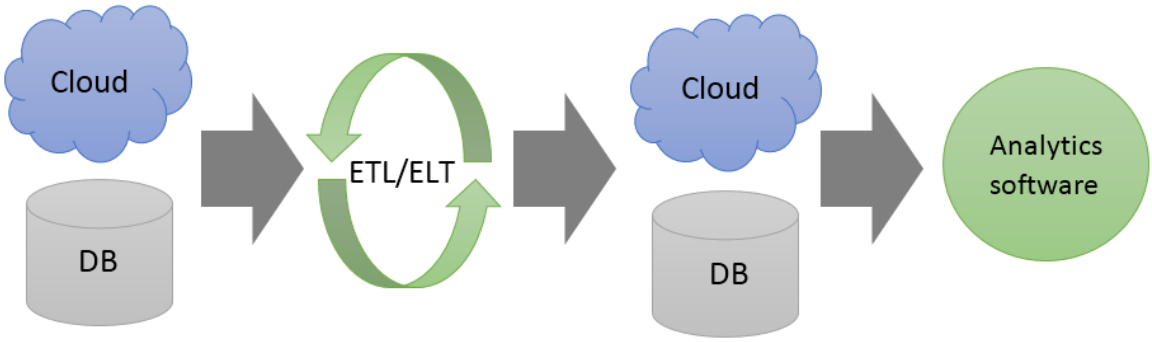


Figure 6. Common architecture behind data analytics software solution

2.5 Current trends on analytics software and future development

In this section, there will be some thoughts collected from multiple sources regarding what are the current trends in analytics and what are the future development areas where the data analytics is heading to. What type of software solutions are popular, in which direction they are evolving, for example, as there are new cloud-based products on the market such as Google BigQuery and on the other hand, are older technologies disappearing or do they still have their space in the markets in the future.

According to Datapine's article, analytics has been evolving fast on the past decade and past year. The focus has been changed from the general analytics made by data analysts more to the direction of self-service analytics that means that people who need the results can produce them also by themselves helped by analysts on the background. The cloud services have changed the focus partly from on-premise services to the cloud and hybrid services that will offer easily scalable and easily set up environments. Outcomes will be more of interactive dashboards and visualizations that can benefit larger scale of people than static spreadsheets did before. The focus will be on the question of how the data should be analysed and presented instead of the question whether it is needed or not in the company. The amount of different solutions and platforms and the multiple forms of analytics are also growing and offering more choices to choose from based on the individual needs of the company. (Durcevic 2018.)

According to the latest study conducted by BARC (Business Application Research Centre) the most important BI trends for the year 2020 will be master data and data quality management, data discovery and visualization, establishing a data-driven culture in the company, data governance, self-service BI and data preparation by business users, which all got reviews more than 6 points out of then by importance. BARC has made a comparation

between most important Business Intelligence trends in 2020 to the equivalent study conducted in 2016 from where it is possible to see that the trends in rise are MD/DQ management (which holds the first place this year), data governance, data preparation by business users, data storytelling, using external/open data, cloud for data and analytics and analytics teams and data labs. The biggest growth from these areas was with analytics teams and data labs by 1 full point. The trends that dropped their points, even though some of them are still in the top of the list, were data discovery, self-service BI, agile BI development, advanced analytics and machine learning, mobile BI, integrated platforms for BI and PM, real-time analytics and big data analytics. The biggest drop was with mobile BI that dropped 0.7 points down from year 2016. According to BARC, it is possible to see the importance of data management, quality and processing to increase when companies want to get the most out of the data and also regarding that, the improvement of the data warehouse solutions and technologies inside the companies. On the other hand, the importance of the self-service BI has decreased even though it is still considered as an important asset. On the other hand, based on BARC survey, Mobile BI has been dropping maybe due to its availability or because it is not serving the purpose of analytics anymore. Another surprising result was to see the machine learning, advanced analytics, and AI to drop their importance in the eyes of BI professionals. One of the stable trends for years have been integrated BI platforms and performance management which are important for the successfulness of analysing huge amounts of data efficiently. (BARC.)

Another research conducted by Dresner Advisory Associates, Wisdom of Crowds® Business Intelligence Market Study, suggests that Bi adoption will be biggest in departments such as operations, sales, finances, and executive management. Also, the most vital BI initiatives in companies include reporting and data visualizations such as dashboards, data integration and end-user self-service trying to improve the data storytelling, governance, and importance in decision-making. According to the study, BI will be used more and more to increasing revenue instead of traditional risk and compliance management and the usage of many different BI tools is most common in Marketing, Sales and Operations departments. Also, the companies with biggest budgets for BI and analytics invested more heavily on reporting, dashboards, and data integration. The respondents of the study are the research community of the company conducting the study including organizations and customers and other interviewed with qualifications found through social media. (Columbus 2019.)

3 Implementation of software testing

This project will be taking into a consideration the problems and current and future needs of the company in the field of descriptive analytics. The current software will be revised based on the requirements and compared to the other options available at the market. The end goals and desired outcome will be defined along as the suggestions of the software solution type, whether it should be in-house product or cloud based. (Stencil 2018.)

Empirical part of the project will be conducted as qualitative research as the main research method. In the project, different BI solutions will be compared to each other and to the BI solutions already in use in the company in order to find the most suitable one based on the theoretical framework and the requirements defined beforehand. In addition, the project also contains empirical approach and results by testing the most suitable software solutions based on the qualitative research. This project will be made through a case company which in this case will be a middle-sized insurance company situated in Finland.

On the following chapter, the different aspects of the causes of the project, software candidate search and comparation, testing chosen solutions with demo and reviewing the options will be gone through. In the end of the process, the best applicable solution will be chosen as the suggestion for the future BI tool for the company. The meaning of this project is not to implement the software in place but to give a suggestion for the useful tool to be used by the BI professionals in the company in order to improve the efficiency and possibilities of analytics in the company.

3.1 Research methodology

In the theory part, qualitative research will be used. Collecting data about software currently in use in the company, and the issues faced in data analysis when using the software, have been collected by using qualitative analysis methodology and also through my own observations as I am working currently in the company and with the software frequently. Based on the recognition of requirements and problems faced, and discussion with my supervisor and team members during the time I have been working in the company, the information for this project has been collected.

The software search and testing will be conducted in two iterations. The first is to define the qualities and specifications of the software and choose few of them which will have most hits with the specs specified. The second iteration includes building a test query to analyse the data and to compare the running-time of queries between different software and their usability. Test comparison for the query will be also conducted straight against

the database and through current software the company is currently using. The results will not provide generalization that could be used to decide software for every company's analysis purposes nor it will give absolute certainty about the effectiveness of the analytical capabilities and query processing of the software, but it can be used to give an indication about the usefulness of the software for this specific study. The methodology on software specification comparation and testing, therefore, is using qualitative research such as unstructured observations by using company websites to gather information about software, and quantitative research in the second iteration to measure query times.

3.2 Base for the project

Currently, the case company provides analysis and reporting for the different departments through automated reporting by email and embedded dashboards. For this purpose, the company is using IBM Cognos Analytics and IBM Cognos Framework Manager where most of the data has been collected from the company's data warehouse and, recently, the company has adopted a new technology stack by implementing SSAS tabular cubes from Microsoft and building reporting on top of it with Power BI and Excel. In addition, R and Python are used for advanced analytics and Google Analytics and Google Data Studio, among others, for online analytics.

With the current technology there exists certain restrictions that have been leading for the fact that new software has been considered and looked for. IBM Cognos Analytics have proved to be very efficient and useful for reporting purposes as long as the queries and logic built are not too complicated. That said, the problem is exactly with those business needs that will require more complicated processing of the data. Currently, the longest reports can run approximately four to five hours and if there are any mistakes or errors why the report run fails, it is not fast to get the report for the departments needing it.
SSAS tabular cubes and Power BI dashboard reporting was implemented to help with the issues faced in Cognos Analytics reporting but it has brought its own issues along with it. The major problem with SSAS tabular cubes that Finnish P&C Insurance are implementing, is the memory limit the current stack sets. As the company is using Microsoft SQL Server Standard edition, the Microsoft Analysis Services tabular instance has the limit of 16GB to use for all usage related to it i.e. processing data. This limit is not a problem in enterprise edition but IBM Cognos Analytics, which is located on the same server, does not support other editions than Microsoft SQL Server Standard, Datacentre and Essentials. Also, the enterprise edition is a core-based licence which will be much more expensive than the server-based licence currently in use, so it is not considered as a first option. (IBM 2020; Microsoft 2020e.)

The 16GB memory limit in SQL Server Standard Edition is a problem for the following reason: when processing the cube or refreshing all of its data, Analysis Service keeps a copy of the cube's data structure in memory which will be deleted when processing is ready. Therefore, the memory consumption of a tabular cube is approximately two or two and half more than the actual size of the cube. As an example, if the tabular cube size is 4GB, the memory consumption of it will be around eight to ten GBs. If there are multiple cubes of the same size that needs to be processed concurrently, the memory will not be enough. This problem, of course, does not exist only when doing a full refresh for the cube but also when changing the structure of it, because in tabular model, there is the so-called shadow copy or workspace database created of the cube which will be removed from the memory and deleted (or stored on the disk) when the project is closed if this is assigned in the setup. Then the workspace database does not consume memory after closing the project. (Microsoft 2020e; Microsoft 2020f; Microsoft 2020l.)

The Finnish P&C Insurance has tried to overcome the problem by trying to find a workaround to the 16GB limit of memory in the tabular instance. As for now, the company has SQL Server Developer edition licence installed for developing the projects because it does not have any maximum limit, and the SQL Server Standard Edition is used for the production environment. The developer edition license is meant only for testing and developing environments and not for production use, therefore, the company has kept the standard edition for production use. For the standard edition, it is possible to install two additional instances for production use, so when the 16GB is not anymore enough, 3x16GB of memory can be achieved by installing two instances more. However, this solution has still an issue because the Analysis Services process will not release the memory usage very effectively after the process has run, and therefore, the daily SQL Server Agent processes refreshing data requires reboot of the Analysis Services in order to achieve successful completion without memory error. (Microsoft 2020e.)

Another solution would be to use the cubes with Microsoft Azure Analysis Services on the cloud, which is also more costly solution, or to work with multidimensional cubes instead of the tabular ones. The difference of the multidimensional is that it is not using memory the same way as tabular models are as it is using a resource management layer instead of VertiPaq in-memory analytics engine. Multidimensional solutions are also harder to master than tabular solutions. As a technology, multidimensional cubes are not the future technology the same way as tabular cubes are. Microsoft recommends using tabular solutions on the new projects instead of multidimensional cubes mainly because it is easier to design, test and deploy than multidimensional solutions and because the support with

self-service BI applications and cloud service solutions is better compared to the multidimensional solution. (Microsoft 2020b; Microsoft 2020j.)

Based on the current needs and restrictions, and possible future scope of cloud services, the main requirements with reporting software are:

1. Possibility for complicated logic behind reporting i.e. building framework models and make calculation and processing of data (preferably on visual form without needing to write SQL, R or Python).
2. To be able to produce reports and dashboards and share them efficiently with other departments and personnel i.e. sending them with email or sharing in Microsoft Teams and automated reporting plus bursting options.
3. To support MS SQL Server/Database, MSAS (Microsoft SQL Server Analysis Services, tabular), MySQL and other data sources such as excel/csv sheets, XML or JSON.
4. Possibility to work with large datasets fast enough (i.e. hundreds of millions of rows).
5. Open source solutions are a plus but not obligatory.
6. As for a solution, it could be both in-house or cloud based, depending of the costs and functionalities of the solution.
7. Connectors for Google data sources such as Google Analytics.
8. Easy introduction and usability for end users (analysts).
9. Simple/affordable pricing plan.
10. Data centre's location where the data is being processed should preferably be within Europe.

It is not necessary for the company to change all the software in use for something else as lot of reporting is already built on top of IBM Cognos for example. But certain analysis is hard to do with the current solutions so the desired result and end goal would be to find an alternative for the advanced analysis purposes; a solution for processing data more efficiently meeting all the mentioned criteria above so that it supports the already existing architecture in the company and facilitates the data usage. As the data comes from multiple sources, in the future, it would be also more practical to be able to combine the sources within a software easily. At this point, the new software solution should offer connectors to the existing data sources. In the future, cloud services could be a solution also and are therefore considered in the requirements.

In general, in the insurance field, the need for data collection and storing is relatively big in order to have all the necessary information, for example, for claim handling or to provide the insurances for the customers. Therefore, the necessity of handling large amount of data and complex queries can be seen as an industry specific problem. In Finland, the amount of insurance companies is limited, and it results also to a challenge where the number of customers per company is large. This results for larger organization, many branches etc. that will rely their decision-making also on reporting. That is why it is crucial to have efficient report and dashboard delivery to the employees needing it.

3.3 Searching for candidates and short list

There are two types of solutions that might be considered: cloud-based reporting and visualization software with enough capabilities with framework modelling and another option would be to choose technology for modelling the data better i.e. Google Big Query or Microsoft Analysis Services databases and to choose the reporting software on top of it. In this project, due to the limited scope, we will focus on going through the Business Intelligence software meant for data visualization for building reports and dashboards.

Based on the needs of the insurance company, some of the available and most used/recommended software on the market have been chosen for the closer review of the features and functionality. The chosen software includes both commercial and open source software. Also, the current solutions will be reviewed and their new functionalities or possibilities that are not yet utilized will be taken into a consideration. After the preliminary check the most promising software based on their features will be taken into testing phase. The following software will be checked for further analysis:

1. Current software
    a. IBM Cognos Analytics
    b. Microsoft Power BI
2. Commercial software
    a. Clic Data
    b. Chartio
    c. Databox
    d. Dundas BI
    e. Google Data Studio
    f. Qlik Sense
    g. Tableau
3. Open source software
    a. BIRT Report Designer
    b. KNIME Analytics Platform & Report Designer
    c. Seal Report

3.3.1 Current software

IBM Cognos Analytics supports visual SQL and is offered as on-premise and cloud-based solutions. It can be used to produce both traditional excel reports and more visual dashboards for the users and it is possible to either send scheduled reports as email in various formats including excel, csv, pdf or to share a dashboard on Cognos Analytics for other Cognos users to see. Bursting options are available for automated reports. Cognos Analytics can access multiple data sources such as Microsoft SQL Server, Microsoft Analysis Services, MySQL, Google BigQuery and Google Cloud SQL. Also, file exports are supported for XSLX, CSV and XML formats. Data source package and data modelling can be

managed on IBM Cognos Framework Manager on on-premise version and all the automated reports, run history and database connections can be managed through Administration Console. IBM has data centers located in Europe currently. As for the current usage, the main problem seen with Cognos Analytics has been slow processing of complex queries. (IBM a; IBM b; IBM c; IBM d; IBM e.)

Microsoft Power BI is currently a fast-evolving reporting and visualization software which is focused on dashboards. Dashboards can be shared in Power BI service, shared as links, publishing them in an application, embedding them in secure portals such as Microsoft Teams or SharePoint Online, or sharing via Power BI Mobile. Sharing a dashboard privately with other employees requires licence for the viewer except on Premium licence. On embedded dashboards report bursting can be achieved with row-level security settings. Dashboards can be also printed or saved as PDF. Excel reports are not supported nor sending reports attached to email. Available data sources include, for example, Excel, CSV, XML, JSON, PDF, Azure SQL Database, Azure Analysis Services Database, SQL Server Analysis Services, SQL Server, MySQL, Google BigQuery, Google Analytics and Snowflake. Data can be queried by R and Python also. Power BI's national cloud is situated in Germany and is meant for European customers.

Power BI has two suitable licences, Power BI Pro and Power BI Premium. Cloud-based Pro licence costs 8,40€/user/month and Premium 4212,30€/month and it is available both on-premise and cloud. Power BI has some limitations regarding data processing. If direct query is used for database connection, there is a limitation of 1 million rows per dataset. Larger datasets can be connected to Power BI, for example, by using Microsoft SQL Analysis Services tabular models. Other than Premium licence, Power BI has also a 1 GB dataset limit and 10GB maximum storage limit. On Premium licence the limits are 10 GB and 100TB. There exists also maximum number of data refresh per day which is 8 for Pro licence and 48 for Premium. (Microsoft 2019a; Microsoft 2019b; Microsoft 2020 a; Microsoft 2020c; Microsoft 2020h; Microsoft 2020i; Microsoft 2020k; Sharabi 2019.)

Table 1. Collection of the features on current software

1. Current software	IBM Cognos Analytics	Microsoft Power BI Service
Visual SQL	Yes	No
Sending scheduled reports by email	Yes	No
Embedded/shared dashboards	Yes	Yes
Bursting options	Yes	Yes
Microsoft SQL Server & Analysis Services	Yes	Yes
Microsoft Azure	Yes	Yes
MySQL/MariaDB	Yes	Yes
XSLX, CSV, JSON, XML	Yes, except no JSON	Yes
R & Python	No without Jupyter Notebook	Yes
Google Ads, Analytics, BigQuery, Cloud	Yes, Google Cloud & BigQuery	Yes, Google Analytics & BigQuery
Large datasets (> 1 million rows)	Yes, but depends from data source	Yes, except in direct query
On-premise or cloud based	Both	Cloud, Power BI Report Server both on-prem & hosted cloud
Data center in Europe	Yes	Yes

3.3.2 Commercial software

Clic Data is a cloud-based reporting and visualization software which has multiple data centers located in Europe. Clic Data has four different license types: Premium, Team, Enterprise and Dedicated. The number of available users and storage size varies depending of the license. Premium license includes 10 users which can be upgraded up to 50 and 5 GB of storage which can be scaled up to 10 GB. Premium license costs 62€/month. Team licence includes 25 users, can be scaled up to 200, and 10 GB of storage which can be scaled up to 15 GB and costs 200€/month. Enterprise license instead includes 50 users, scaled up to 500, and 10 GB of storage which can be maximum of 50 GB. Enterprise license costs 380€ per month. Dedicated license starts from 780€/month and includes 100 users and 100 GB – 8TB of storage. The number of users can be scaled up unlimitedly. All licenses include unlimited number of dashboards. (Clic Data a; Clic Data b; Clic Data c; Clic Data d; Clic Data e; Clic Data f; Clic Data g; Clic Data h; Clic Data i.)

For all licenses, Clic Data has connectors such as Google Ads, Google Analytics, Google Sheets and files and spreadsheets. From Team license and up, MySQL, MariaDB, ODBC

(Microsoft Azure, Microsoft Analysis Services), Microsoft SQL Server and Google Analytics Query are available. For Enterprise and Dedicated licenses also, connector for Google BigQuery is included. R & Python integration is available in Dedicated license. Clic data offers both sending scheduled reports and dashboards by email and sharing them with the users. Dashboards can be sent as PDF, image, Excel, CSV, Power Point or as Live Link. Live Link usage requires viewer's license. The data refreshes can be also automated depending of license from 5000 data refresh up to unlimited. From Team license and up, the charts and tables are editable for others when sharing. Dashboards can be personalized for the viewers by using parameters and filtering data (row-level security). This option is available from Team license and up. (Clic Data a; Clic Data b; Clic Data c; Clic Data d; Clic Data e; Clic Data f; Clic Data g; Clic Data h; Clic Data i.)

Chartio is software that offers visual interface for building SQL queries. It is cloud-based and even though its data centers are not located in Europe, it is GDPR compliant. Chartio has three licenses: Startup 40$/user/month, Growth 60$/user/month and Premium which has custom pricing. All licenses include unlimited number of charts and dashboards and they include scheduled emails for sending reports (PDF or CSV). Email size is limited to 25MB and they can be sent to non-Chartio users as well. Embedding dashboards is also possible, for example, via Slack or they can be shared to other Chartio users. Reports and dashboards can be personalized by user using row-level permissions to manage the content. Chartio offers data connectors such as to CSV, Google Analytics, Google Cloud SQL, Google Sheets, MySQL, Maria DB, Microsoft Azure and Microsoft SQL Server. Google AdWords connector available through partner. Growth and Premium licenses offer also alerts for changing data and connections to BigQuery and Snowflake. (Chartio 2017; Chartio 2018; Chartio 2020a; Chartio 2020b; Chartio 2020c; Chartio 2020d; Chartio 2020f; Chartio 2020g; Chartio 2020h; Chartio 2020i.)

Databox has its data centers in the US but it is GDPR compliant like Chartio. Databox has four license types Free, Basic, Plus, Business and on top of that a custom plan. Those cost 0$, 49$, 99$ and 248$ per month. Custom plan depends the needs of the company. Business plan, for example, contains 50 different data sources and dashboards and 20 users. Data connections include REST API, Google Analytics, HubSpot Marketing, Google AdWords, Google BigQuery, Snowflake, Google Sheets, MySQL, Microsoft SQL Server and Microsoft SQL Azure. Through REST API it is possible to integrate Python also. It is possible to send snapshots by scheduled emails, share dashboards with links and set up alerts, for example, by email, Slack or Push notifications. (Databox 2020; Databox, Inc. 2020a; Databox, Inc. 2020b; Databox, Inc. 2020c; Databox, Inc. 2020d; Vidovic 2016.)

Dundas BI provides data connections to data sources such as Microsoft SQL Server, MySQL, Google BigQuery, Google Sheets, Microsoft SQL Azure, Google Analytics, Snowflake, Microsoft SQL Analysis Services, Azure Analysis Services, Microsoft Power-Pivot. Also, Dundas BI supports ODBC and JDBC drivers, XML, Excel, CSV, dBase, R and Python. In Dundas BI, it is possible to send reports by scheduled emails, share URL links, share the dashboards in Slack or export the file to a certain location on the disk regularly. For the data changes, notification alerts can be created which reminds viewer when new data is available. Dundas BI supported export formats are XSLX, CSV, PDF, PNG and PowerPoint. Bursting options are also available for personated reports. Also, there is graphical drag and drop user interface to facilitate data usage. Dundas BI has data center in Europe and offers both cloud and on-premises options. (Dundas Data Visualization, Inc. 1999-2020a; Dundas Data Visualization, Inc. 1999-2020b; Dundas Data Visualization, Inc. 1999-2020c; García 2015; Zenko.)

Google Data Studio has a bit different pricing plan from the other software in the review as it is based data usage such as data import, refresh and queries. Google's data center is located in Finland, Hamina. Google native data source connections include Google services such as BigQuery, Ads, Analytics, Sheets, Cloud Storage. Other connections are, for example, CSV files (dataset limit 100MB) and MySQL. Google does not currently offer native Microsoft SQL Server connector, but the data transfer is available through paid third-party data pipeline. Reports can be delivered through scheduled emails in PDF format. Sharing a link to the dashboard or embedding it to any application or site supporting HTML iframe tag or embedding to Google sites is also possible. Dashboards can be shared through social media that supports Open Graph protocol, for example, via Google Hangouts or Slack. If dashboard is shared for people without Google account, editing of the dashboard is not possible. Row-level security exists by filtering the data with email. In general, viewer cannot refresh the data itself on an embedded report or dashboard. Some other limitations regarding the sharing options apply as well. (Google a; Google b; Google 2020a; Google 2020b; Google 2020c; Google 2020d; Google 2020e; Google2020f; Google 2020g; Google 2020h.)

Qlik Sense is a reporting and visualization software provided by Qlik. Businesses may decide themselves the location of used data center of Qlik's offering i.e. Ireland in Europe. Licenses include Qlik Sense Business 30$/month/user, Enterprise Analyzer 40$/month/user and Professional 70$/month/user. For Enterprise Analyzer and Professional license both cloud and on-premise solutions are available and included to the price. The licenses include limitations to the amount of cloud storage, scheduled data reload

concurrency and scheduled data reloads. The cloud storage size limitations are 250GB for Business and 500GB and up for Enterprise. The allowed amount of data reloads for Business license is 50 and for Enterprise 100 and up. Qlik Sense data connectors include Google Ads, Google Analytics, Google Drive, spreadsheets, ODBC connector (Azure SQL 2005-2014, MariaDB, Snowflake, Google BigQuery, Microsoft SQL Server 2012-2017, MySQL Enterprise), REST (JSON, CSV, XML), Microsoft Analysis Services and XSLX files. R and Python can be integrated in Enterprise version. Custom reports can be delivered by scheduled email, by saving to disk, sharing on web or through NPrinting NewsStand portal or Qlik Sense hub. Supported report formats are Excel, Power Point, Word, HTML and PDF. Qlik Sense offers also bursting options for email delivery. (QlikTech International AB 2018; QlikTech International AB 1993-2019; QlikTech International AB 1993-2020a; QlikTech International AB 1993-2020b; QlikTech International AB 1993-2020c; QlikTech International AB 1993-2020d; QlikTech International AB 1993-2020e; QlikTech International AB 1993-2020f; QlikTech International AB 2020a; QlikTech International AB 2020b.)

Tableau is a software offered by Salesforce company and it has both cloud and on-premise versions. Tableau has data center located in EU, Ireland. License options with Tableau Server (either public cloud or on-premises) are Tableau Creator 70$/user/month, Tableau Explorer 35$/user/month and Tableau Viewer 12$/user/month. The same licenses with Tableau Online (hosted by Tableau) are Creator 70$/user/month, Explorer 42$/user/month and Viewer 15$/user/month. Tableau has different data connectors to Google services such as Ads, Analytics, BigQuery and Cloud. Also, Microsoft SQL Server, Microsoft Analysis Services, Microsoft Azure, MySQL, XSLX, CSV and JSON connectors are available. R & Python integration is possible too. Tableau offers graphical SQL builder VizQL™ which makes building queries easier for users not familiar with SQL. Complex queries and large datasets are supported by Tableau's data engine. Currently Tableau does not offer scheduled reports sent by email, but dashboards can be published and shared with Tableau Viewers and other Tableau users and the links of the dashboards can be sent also. Every viewer that receives a dashboard needs an own license or the business must have Enterprise/server license. Dashboards can be personalized for different viewers with row-level permissions. (Tableau Software, LLC 2003-2020a; Tableau Software, LLC 2003-2020b; Tableau Software, LLC 2003-2020c; Tableau Software, LLC 2003-2020d; Tableau Software, LLC 2003-2020e; Tableau Software, LLC 2003-2020f; Tableau Software, LLC. & A Salesforce Company 2003-2020a; Tableau Software, LLC. & A Salesforce Company 2003-2020b; Tableau Software, LLC. & A Salesforce Company 2016; Tableau Software, LLC. & A Salesforce Company 2019; Woodall 2017.)

Table 2. Collection of the features on commercial software

2. Commercial Software	Clic Data	Chartio	Databox
Visual SQL	No	Yes	No
Sending scheduled reports by email	Yes	Yes	Yes
Embedded/shared dashboards	Yes	Yes	Yes
Bursting options/row-level permissions	Yes, through parameters	Yes, row-level permissions	No
Microsoft SQL Server & Analysis Services	Yes	Yes, SQL Server	Yes, SQL Server
Microsoft Azure	Yes	Yes	Yes
MySQL/MariaDB	Yes	Yes	Yes
XSLX, CSV, JSON, XML	Yes	Yes, only CSV	No
R & Python	Yes, only on Dedicated license	No	Yes, Python
Google Ads, Analytics, BigQuery, Cloud	Yes, Ads, Analytics, BigQuery	Yes, all except Ads	Yes, except Cloud
Large datasets (> 1 million rows)	--	--	--
On-premise or cloud based	Cloud	Cloud	--
Data center in Europe	Yes	No, GDPR compliant	No, GDPR compliant

Table 3. Collection of the features on commercial software

2. Commercial Software	Dundas BI	Google Data Studio
Visual SQL	Yes	--
Sending scheduled reports by email	Yes	Yes
Embedded/shared dashboards	Yes	Yes
Bursting options/row-level security	Yes	Yes
Microsoft SQL Server & Analysis Services	Yes	Not provided by Google
Microsoft Azure	Yes	No
MySQL/MariaDB	Yes	Yes
XSLX, CSV, JSON, XML	Yes, except JSON	Yes, only CSV
R & Python	Yes	No
Google Ads, Analytics, BigQuery, Cloud	Yes, Analytics & BigQuery	Yes
Large datasets (> 1 million rows)	--	--
On-premise or cloud based	Both	Cloud
Data center in Europe	Yes	Yes

Table 4. Collection of the features on commercial software

2. Commercial Software	Qlik Sense	Tableau
Visual SQL	No	Yes
Sending scheduled reports by email	Yes	No
Embedded/shared dashboards	Yes	Yes
Bursting options/row-level security	Yes	Yes, row-level permissions
Microsoft SQL Server & Analysis Services	Yes	Yes
Microsoft Azure	Yes	Yes
MySQL/MariaDB	Yes	Yes
XSLX, CSV, JSON, XML	Yes	Yes, except XML
R & Python	Yes, depends on license	Yes
Google Ads, Analytics, BigQuery, Cloud	Yes, except Cloud	Yes
Large datasets (> 1 million rows)	--	Supported
On-premise or cloud based	Both, depends on license	Both
Data center in Europe	Yes	Yes

3.3.3 Open source software

BIRT Report Designer is an Eclipse-based open-source solution for building reports. It supports data sources such as CSV, XML, XSLX and JDBC driver connections such as Microsoft SQL Server, Microsoft Azure and MySQL databases. As a community contributions Google Analytics, Google Docs and REST services can be connected. Supported report output formats are paginated HTML, PDF, Microsoft Office formats and Open Document. Reports can be saved to disk and the supported formats for report outputs are HTML, Word, XLS, PS, ODT, ODS, ODP or PDF. It is also possible to export the report data into CSV. There seems to be no documentation about sharing reports or sending them by email regularly. (Actuate Corporation 2013; Dodson 2013; Freeman 2015; The Eclipse Edition 2014a; The Eclipse Edition 2014b; The Eclipse Foundation 2014.)

KNIME Analytics Platform & KNIME Report Designer are an open-source ETL and reporting and visualization tool. It is possible to build visual workflows in KNIME and they can be scheduled also. As data sources, KNIME supports for example CSV, JSON, XML, XSLX, PDF, Microsoft SQL Server, Microsoft Azure, Google Sheets, MySQL, SOAP and REST services, JDBC drivers, Snowflake, Google BigQuery and other Google APIs such as Google Analytics. R and Python integration is supported also. (KNIME AG a; KNIME AG b; KNIME AG c; KNIME AG 2014; KNIME AG 2018.)

The third open-source option for reporting software would be Seal Report. It is built as Microsoft .NET Framework tool and it supports OLEDB and ODBC drivers which can be user to access, for example, MySQL, Microsoft SQL Server, Microsoft Analysis Services

etc. Reports can be created as HTML document and automated reporting by email is possible as well as saving the report to disk or to publish them in web through Web Report Server. Scheduled report execution is possible through Windows Task Scheduler integration. Different alerts can be created for reports as well and the report output can be changed depending of the viewer by using Razor Script. PDF and Excel converters for reports are offered as commercial components. (Seal Report a; Seal Report b; Seal Report c; Seal Report d.)

Some other open-source products are Knowage, Metabase and Jaspersoft Studio and they have a paid version of the software as well. Knowage, for example, does not offer scheduled reporting in its free open-source version and Metabase also has quite limited features in it so they could be more suitable for personal use than for companies. However, Metabase open-source solution offers graphic SQL which can be used to build SQL queries. For comparison Metabase's Enterprise version starts from 10 000$ per year. (Engineering 2017a; Engineering 2017b; Metabase 2020; TIBCO Software Inc. 2020; Wolski 2001-2020.)

Table 5. Collection of the features on open-source software

3. Open-source software	BIRT Report Designer	KNIME Analytics Platform & Report Designer	Seal Report
Visual SQL	No	No	Yes
Sending scheduled reports by email	--	Yes	Yes
Embedded/shared dashboards	--	--	Yes
Bursting options/row-level security	--	--	Yes
Microsoft SQL Server & Analysis Services	Yes	Yes, SQL Server	Yes
Microsoft Azure	Yes	Yes	Yes
MySQL/MariaDB	Yes	Yes	Yes
XSLX, CSV, JSON, XML	Yes, except JSON	Yes	--
R & Python	--	Yes	--
Google Ads, Analytics, BigQuery, Cloud	Yes, Analytics	Yes, BigQuery & Analytics	--
Large datasets (> 1 million rows)	--	--	--
On-premise or cloud based	On-premise	On-premise	On-premise
Data center in Europe	--	--	--

KNIME, BIRT and Seal Report as open source platforms will require more work from the end user than other solutions, especially connecting to the data sources, building the

workflow and to get automated reports. Also, KNIME seems to be more of an ETL tool than reporting and visualization tool even though it offers reporting capabilities. For these three, Seal Report will be the lightest option to use. In general, many companies offer open source community editions or free-to-use versions of their products but with limited capabilities (users, amount of dashboards, size of the reports or amount of queries, sharing options) and they are mainly meant for single users such as students or researchers and for smaller projects. For companies many software providers have their own solutions and licensing also from open source software. Examples of such products include Knowage, Metabase and JasperSoft Studio.

3.4 Testing the software chosen from preliminary analysis

In the test, Microsoft SQL Server 2016 Express will be used with SQL Server Management Studio. As an example, Microsoft's sample data warehouse AdventureWorksDW2016_EXT will be used for producing the test. The public dataset has been chosen due to confidentiality issues, as the purpose is to present the type of data analysis case used in the testing, and because the type of data used in the test is not as relevant as explaining how the test scenario is built and conducted. Also, to prevent confidential information appearing on the screen captures taken to demonstrate the results and test scenario. (Kess 2017; Microsoft 2020d; Microsoft 2020g.)

A simple query is designed to retrieve data from data warehouse. The query will be run through SQL Server Management Studio's Query Editor and the run time for the query will be measured. Then the same query will be built using the chosen software either graphically or by inserting the created T-SQL to the software as custom table. Then the run time of query will be measured again and compared to the other results. The query times will be measured with phone's chronometer if the software does not have a counter for query run time. Every query's run time will be measured three times and the results will be used to measure average query time for each software.

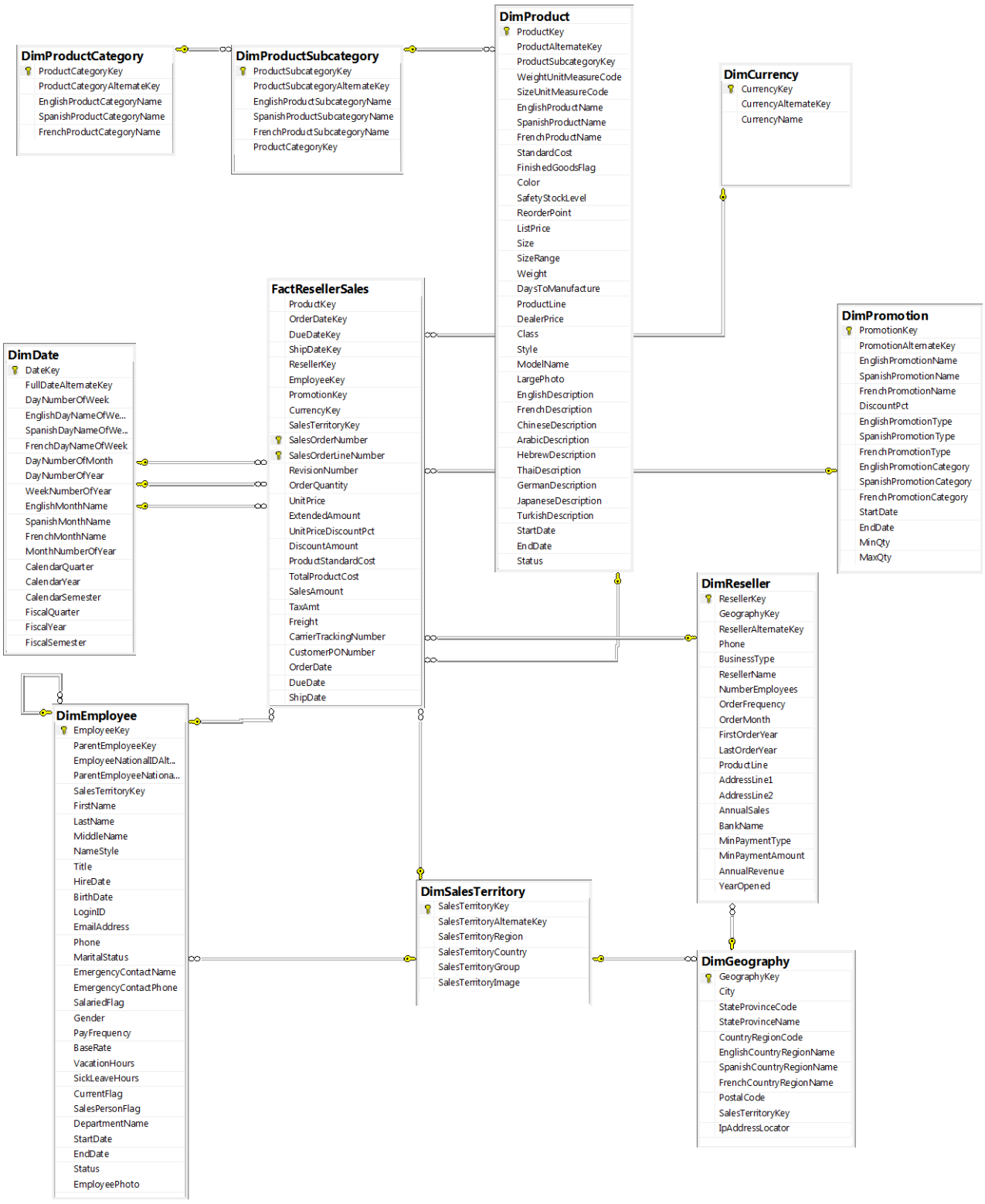


Figure 7. Database diagram model for AdventureWorksDW2016_EXT

Based on the above database diagram (Figure 7), the analysis case will be to define how many orders for reseller there have been every year by sales territory group (North America, Europe, Pacific, N/A), what products they sold and how many orders and products were sold to reseller. The meaning of it will be to follow the amount of orders and the

quantity of products that have been sold by country each year. The T-SQL query producing desired results can be seen in Figure 8 below. Run times for three attempts with the query (Figure 8) in Microsoft SQL Server Management Studio were 00:00:51, 00:00:44 and 00:00:41. Results are presented in Figure 9.

```
select
a5.OrderYear, a5.SalesTerritoryGroup, a5.EnglishProductCategoryName,
count(distinct a5.SalesOrderNumber) as CountOfSalesOrderNumber, sum(a5.OrderQuantity) as SumOfOrderQuantity
from
(
    select a4.*, b4.SalesTerritoryGroup from
    (
        select a2.ProductKey, a2.EmployeeKey, a2.SalesTerritoryKey, a2.SalesOrderNumber, a2.OrderQuantity,
        year(a2.OrderDate) as OrderYear, b2.EnglishProductCategoryName
        from AdventureWorksDW2016_EXT.dbo.FactResellerSalesXL_CCI a2
        left join
        (
            select a1.ProductKey, a1.ProductSubCategoryKey, b1.EnglishProductCategoryName
            from AdventureWorksDW2016_EXT.dbo.DimProduct a1
            left join
            (
                select a.ProductSubcategoryKey, b.ProductCategoryKey, b.EnglishProductCategoryName
                from AdventureWorksDW2016_EXT.dbo.DimProductSubcategory a
                inner join AdventureWorksDW2016_EXT.dbo.DimProductCategory b
                on a.ProductCategoryKey = b.ProductCategoryKey
            ) b1
            on a1.ProductSubcategoryKey = b1.ProductSubcategoryKey
        ) b2
        on a2.ProductKey = b2.ProductKey
    ) a4
    left join
    (
        select a3.EmployeeKey, a3.SalesTerritoryKey, b3.SalesTerritoryGroup
        from AdventureWorksDW2016_EXT.dbo.DimEmployee a3
        inner join AdventureWorksDW2016_EXT.dbo.DimSalesTerritory b3
        on a3.SalesTerritoryKey = b3.SalesTerritoryKey
    ) b4
    on a4.EmployeeKey = b4.EmployeeKey
) a5
group by a5.OrderYear, a5.SalesTerritoryGroup, a5.EnglishProductCategoryName
order by a5.OrderYear desc, a5.SalesTerritoryGroup asc, a5.EnglishProductCategoryName asc;
```

Figure 8. T-SQL of the desired outcome of data set

Results | Messages

	OrderYear	SalesTerritoryGroup	EnglishProductCategoryName	CountOfSalesOrderNumber	SumOfOrderQuantity
1	2014	Europe	Accessories	7834	37689
2	2014	Europe	Bikes	14716	136095
3	2014	Europe	Clothing	9646	52225
4	2014	Europe	Components	16031	202596
5	2014	North America	Accessories	63459	309064
6	2014	North America	Bikes	118935	1102220
7	2014	North America	Clothing	77848	421951
8	2014	North America	Components	130223	1648544
9	2014	Pacific	Accessories	2576	12261
10	2014	Pacific	Bikes	4940	46062
11	2014	Pacific	Clothing	3210	17382
12	2014	Pacific	Components	5433	69525

Figure 9. Result set from the query

The software chosen for the testing are the ones that met most of the criteria would be from the search of software requirements (part 3.3). From commercial side, the chosen

software are Dundas BI and Tableau, and from open-source software Seal Report. It might be beneficial to include Chartio also to testing because it has advanced visual SQL feature which is not very common among reporting and visualization software. The comparison will be conducted with current software the case company is using (Microsoft Power BI and IBM Cognos Analytics). The query times will be measured with phone's chronometer if there is no time provided in the software to count query time. Every query time will be measured three times and the results will be used to measure average query time for each software.

3.4.1 Dundas BI

The first software to be tested was Dundas BI. They offered 25 days free trial of their product either as online or on-premises version. The online version supports only excel as data source and it has few ready-made data source samples such as AdventureWorks OLAP database. Based on the documentation, the on-premises installation requires several hours of work and since the scope and timeframe of the project are limited, the testing will be conducted with the online version's AdventureWorks database which is pre-installed there as an example. Since it is different version from what has been used in other tests, the query results are not equal. Testing of the query times is not either possible as planned but limited opinion of the user interface and usability can be created. Since the data model for AdventureWorks is ready and all the measures also calculated, the items can be just dragged and dropped to the wanted visualization as can be seen in Figure 10. Final results of the data set are visible below on Figure 11.

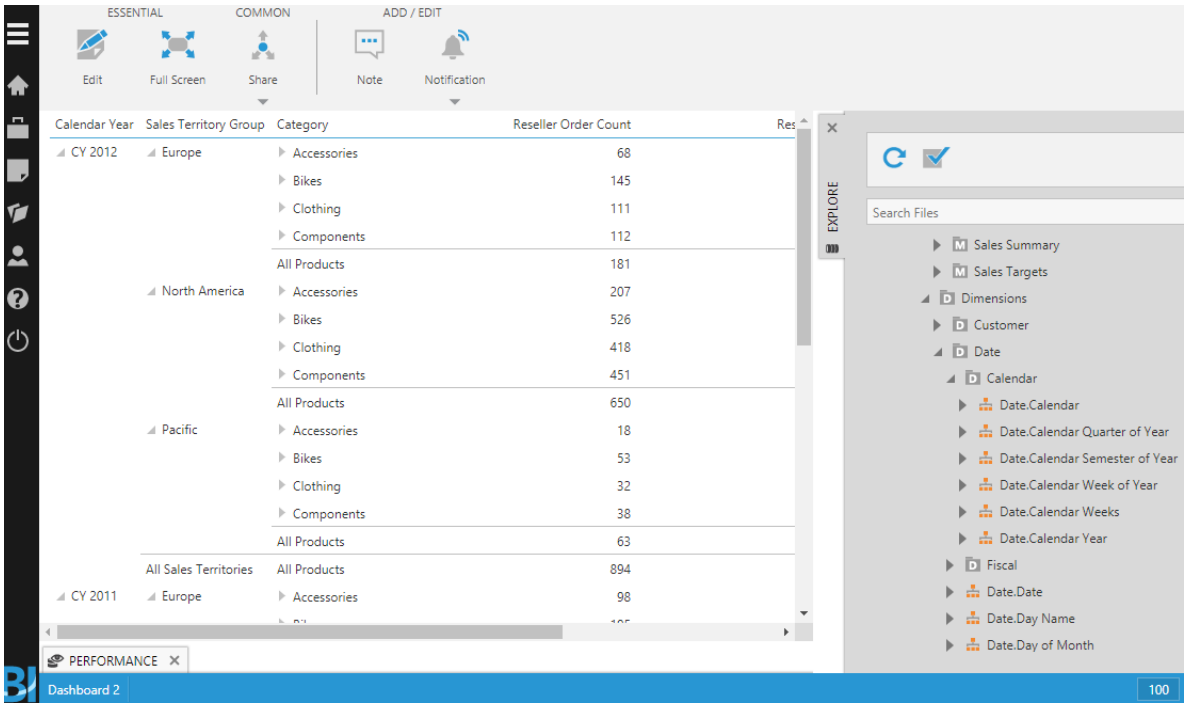
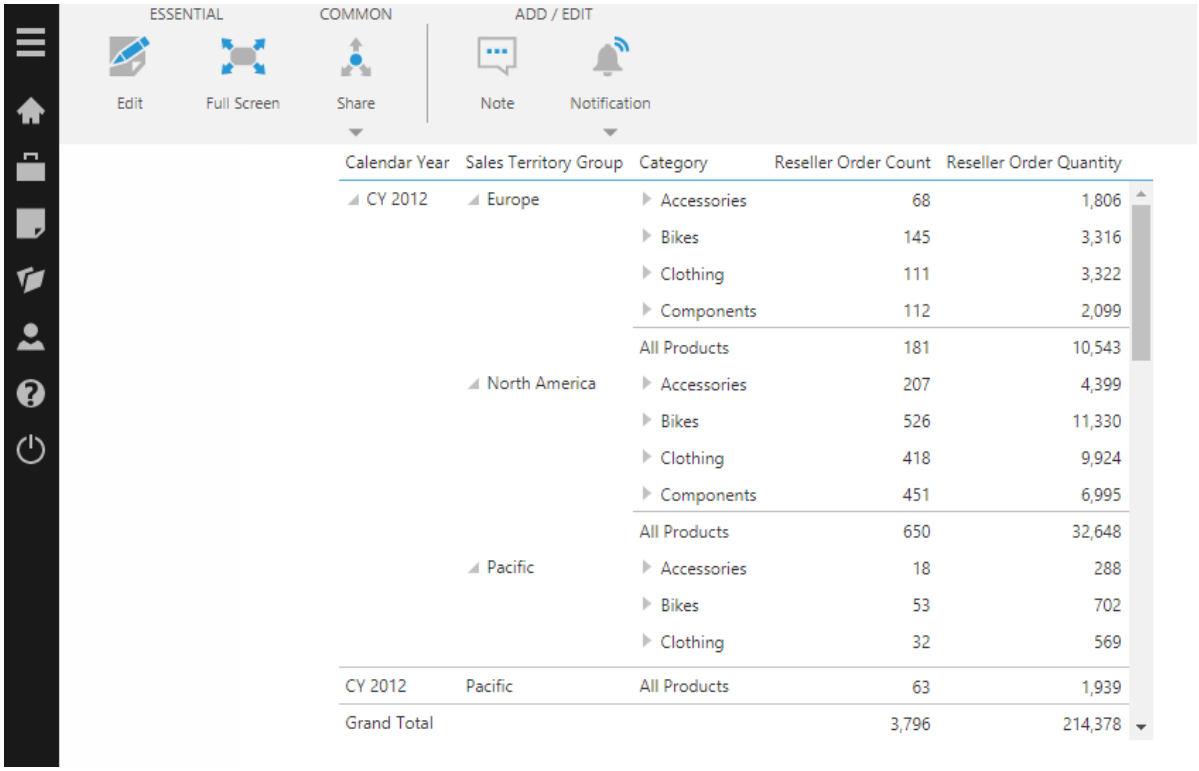


Figure 10. Dundas BI drag and drop user interface



Calendar Year	Sales Territory Group	Category	Reseller Order Count	Reseller Order Quantity
CY 2012	Europe	Accessories	68	1,806
		Bikes	145	3,316
		Clothing	111	3,322
		Components	112	2,099
		All Products	181	10,543
	North America	Accessories	207	4,399
		Bikes	526	11,330
		Clothing	418	9,924
		Components	451	6,995
		All Products	650	32,648
	Pacific	Accessories	18	288
		Bikes	53	702
		Clothing	32	569
CY 2012	Pacific	All Products	63	1,939
Grand Total			3,796	214,378

Figure 11. Dundas BI, final result set

3.4.2 Tableau

The second software used for testing is Tableau which offers 14 days trial version of the product. It's a desktop version with easy installation and will be ready in few minutes. The desktop version seems to contain full features of Tableau software. Connecting to the data source is easy with native connector.

First, a data model will be created with tables needed in our scenario and the relationships are formed. The tables can be dragged and dropped to the modification area and when dropped next to other table, Tableau suggests the relationships automatically. Multiple column relationships are supported as well. It is also possible to add new tables to the data model by using custom SQL. The custom SQL in this case is the script constructed for the test scenario. Creating table with custom SQL was fast and took only few seconds. Example of the data model builder in Tableau can be seen in Figure 12. After the data model is ready, it can be saved, and the user may proceed to worksheet.

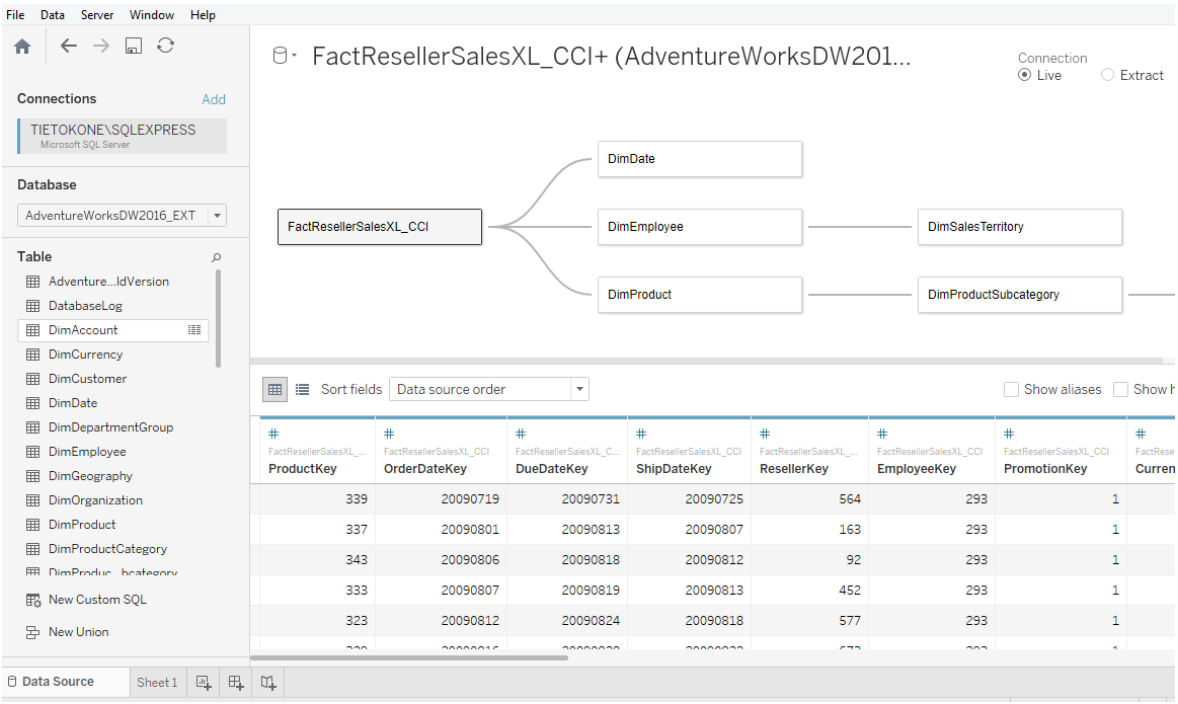


Figure 12. Tableau, creating data model

On the worksheet, it is possible to create, for example, calculated columns (Figure 13). When adding data to the worksheet, the query to retrieve data will be executed. This might take longer than running the SQL script built in Microsoft SQL Server Management Studio. In some cases, adding new fields to the visualization could take up to one minute. Adding data to the visualization from custom built SQL took around 34 seconds.

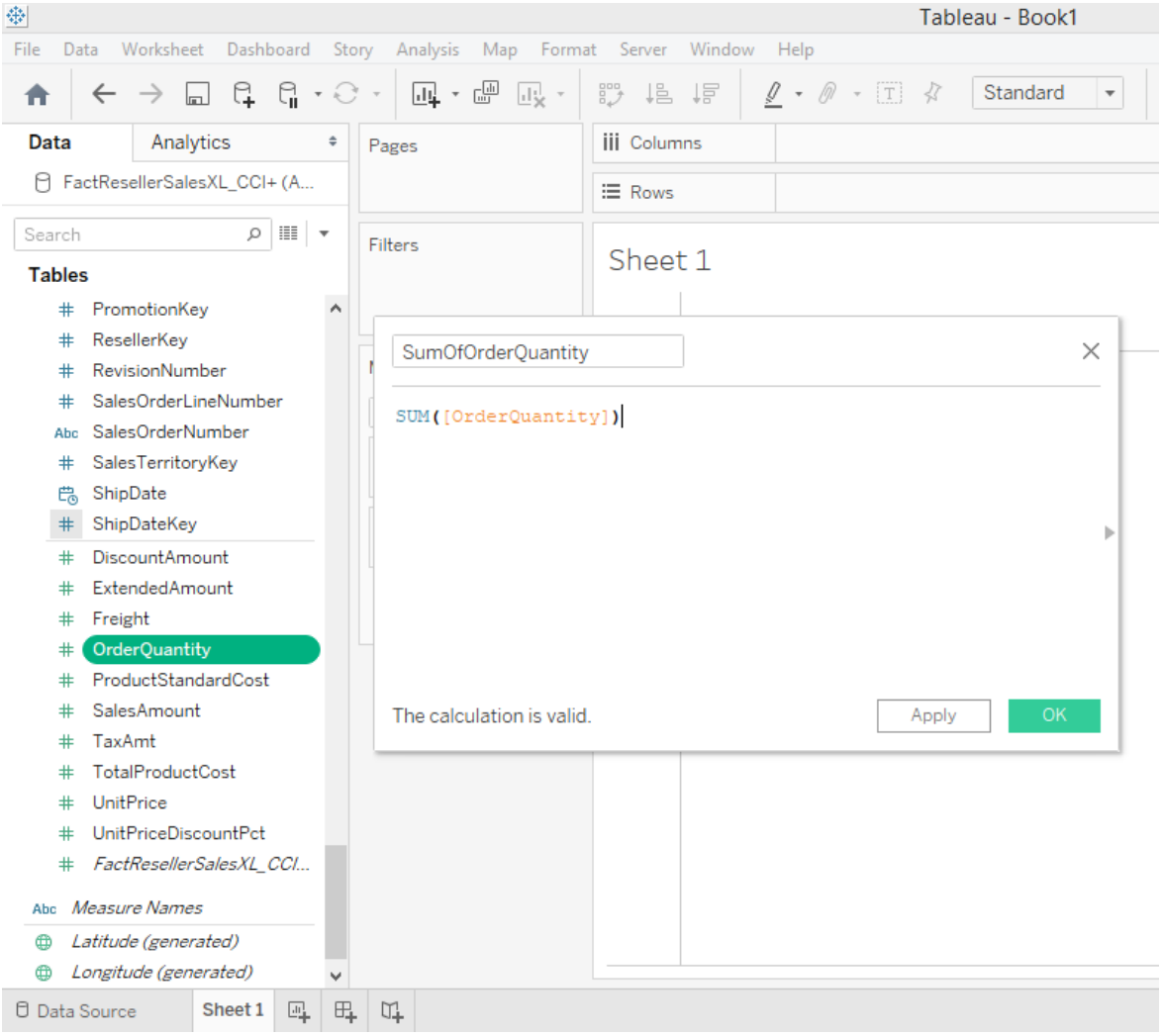


Figure 13. Tableau worksheet and creation of a calculated column

Count of orders and order quantities by sales territory , product and year

OrderDate (..	SalesTerritoryGroup	EnglishProductCategoryName	CountOfSalesOrderNumber	SumOfOrderQuantity
2014	Europe	Accessories	7 834	37 689
		Bikes	14 716	136 095
		Clothing	9 646	52 225
		Components	16 031	202 596
	North America	Accessories	63 459	309 064
		Bikes	118 935	1 102 220
		Clothing	77 848	421 951
		Components	130 223	1 648 544
	Pacific	Accessories	2 576	12 261
		Bikes	4 940	46 062
		Clothing	3 210	17 382
		Components	5 433	69 525
2013	Europe	Accessories	8 193	39 589
		Bikes	15 052	140 365
		Clothing	9 920	53 649
		Components	16 448	209 283
	North America	Accessories	64 736	315 279
		Bikes	121 065	1 123 011
		Clothing	79 430	432 021
		Components	132 477	1 677 834
	Pacific	Accessories	2 638	12 810
		Bikes	5 002	46 944

Figure 14. Tableau, final results of the query

As a conclusion, it was possible to build the same results (Figure 14) without using direct SQL query but it's a slightly slower option, and more complex queries might be more reasonable to build as SQL query instead the drag and drop visualization builder. In Tableau, there is no possibility to take out the converted SQL from visualizations. Opening the ready-made dashboard took 44, 54 and 38 seconds with three attempts. Opening the custom SQL dashboard took around 24 seconds instead and was, therefore, a bit faster.

3.4.3 Chartio

Chartio offers also a 14-day trial completely cloud-based, so no installation is required. The access is allowed when new user account is created. Connecting to Microsoft SQL data sources is not as simple as in Tableau. There are two options of how to connect to the database. The first is through direct connection and the second using SSH tunnel. On the direct database connection, the user must define the external IP address of the server and the port database is using. Also, the configuration for Windows Firewall needs to be made in order to get the connection working between Chartio and MSSQL Server. For both options, separate database reader user role must be created in MSSQL Server. In this case SSH tunneling were used through autossh program. Setting up the connection required some time but was not too hard for someone with experience of SSH tunnels.

After the database connection was established, the needed tables will be selected from AdventureWorksDW2016_EXT database by using database alias (Figure 15). It is also possible to add custom queries to Chartio. Adding the SQL script took around 45 seconds. In the query editor, it is also possible to format the query for more readable form (Figure 16).

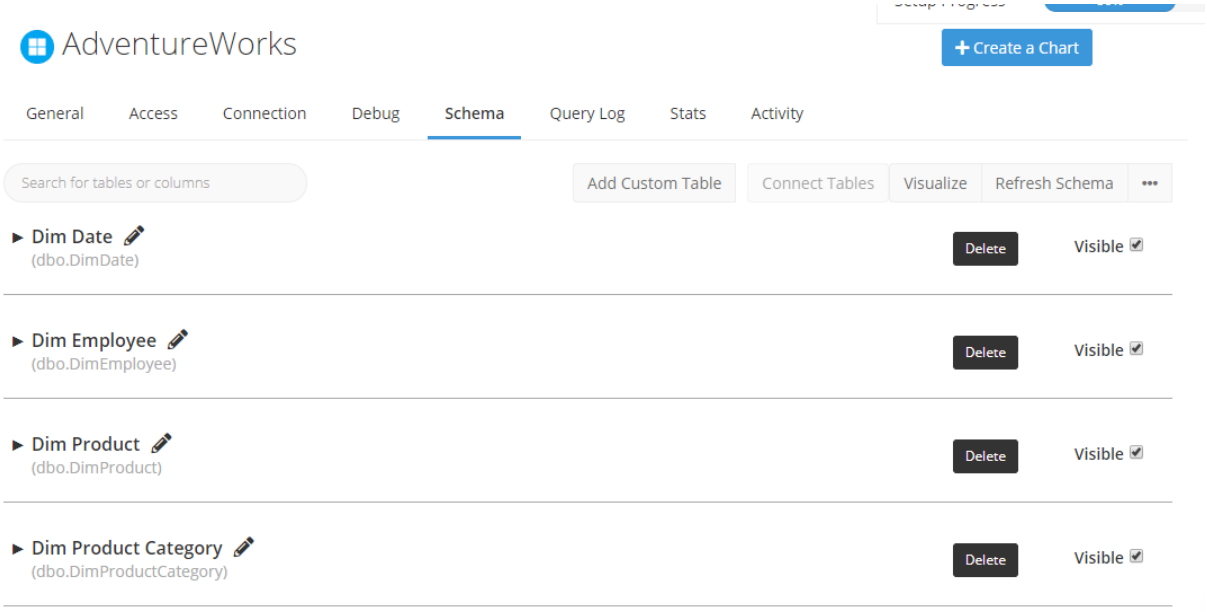


Figure 15. Chartio, choosing tables

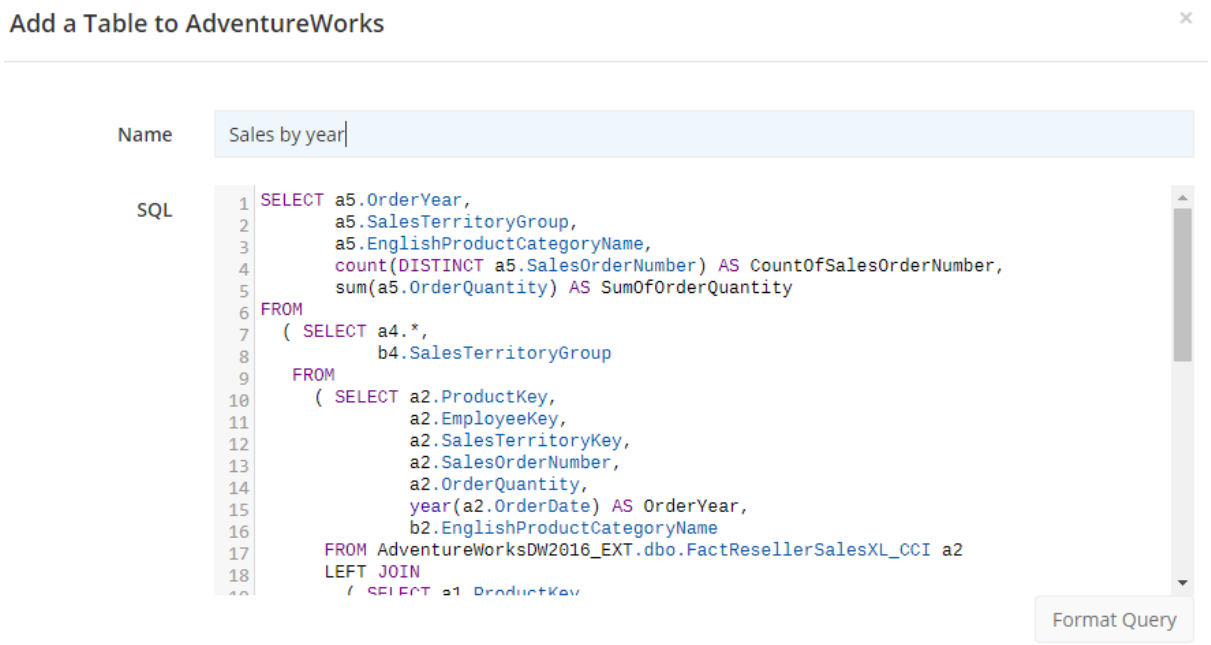


Figure 16. Adding custom queries from MSSQL Server in Chartio

It is possible to use the visual query builder of Chartio to build and merge the queries without writing any SQL (Figure 17). Also, the visual SQL is turned to SQL, so it can be taken out from the report in any phase of the query building if needed (Figure 18). In one query it is possible to combine all items from tables that are joined with primary and foreign keys. If the table does not have defined relationship with the other tables, it is possible to add other queries and join them (Figure 19). The query results are limited to 100 000 rows per query and the limit needs to be changed manually. Final query and result set are presented on Figure 20 and Figure 21. (Chartio 2020e; Chartio 2020j.)

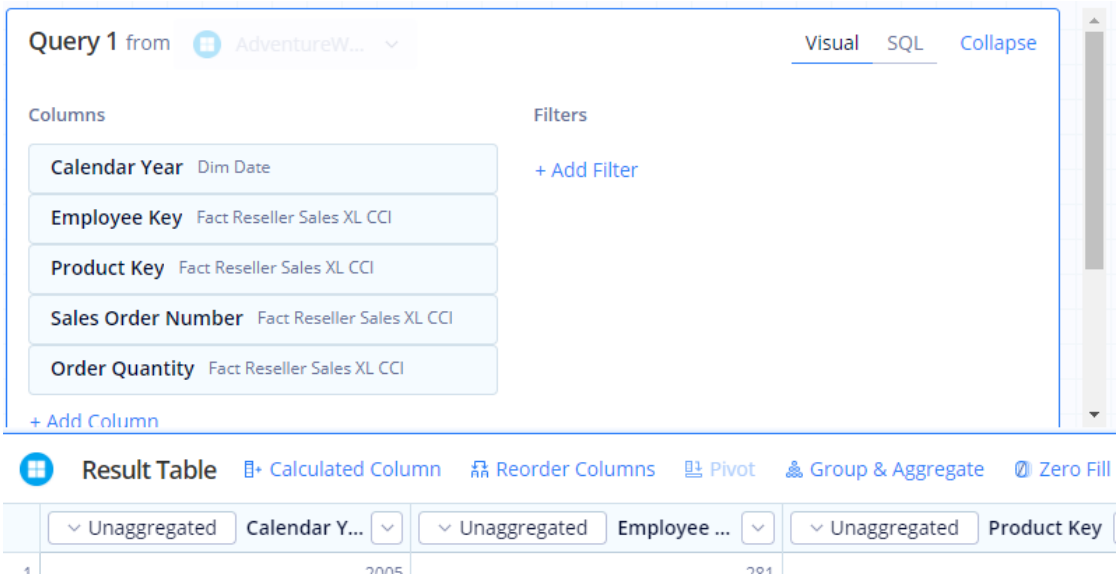


Figure 17. Building query in Chartio using visual SQL

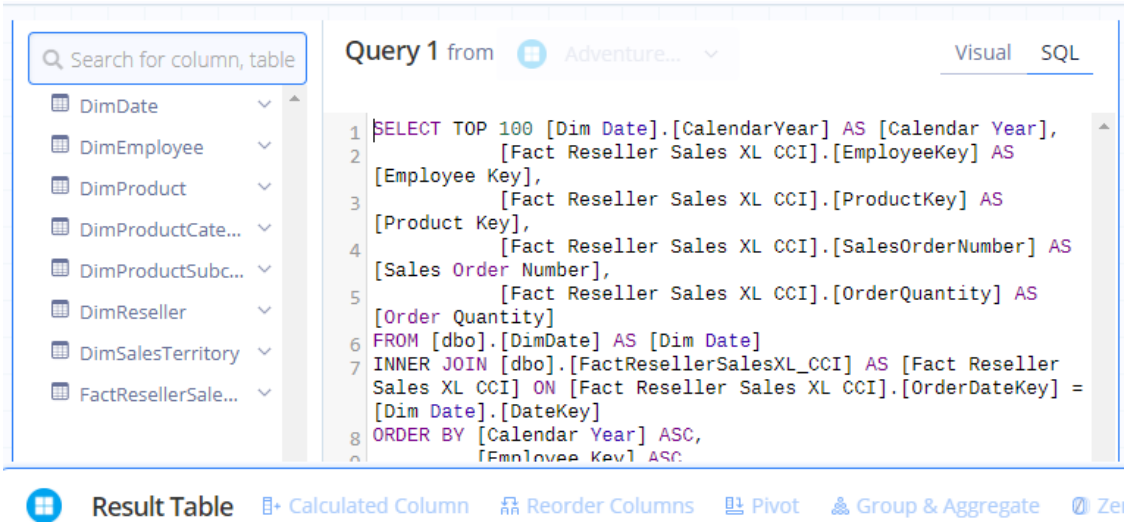


Figure 18. Taking SQL query out from visual SQL in Chartio

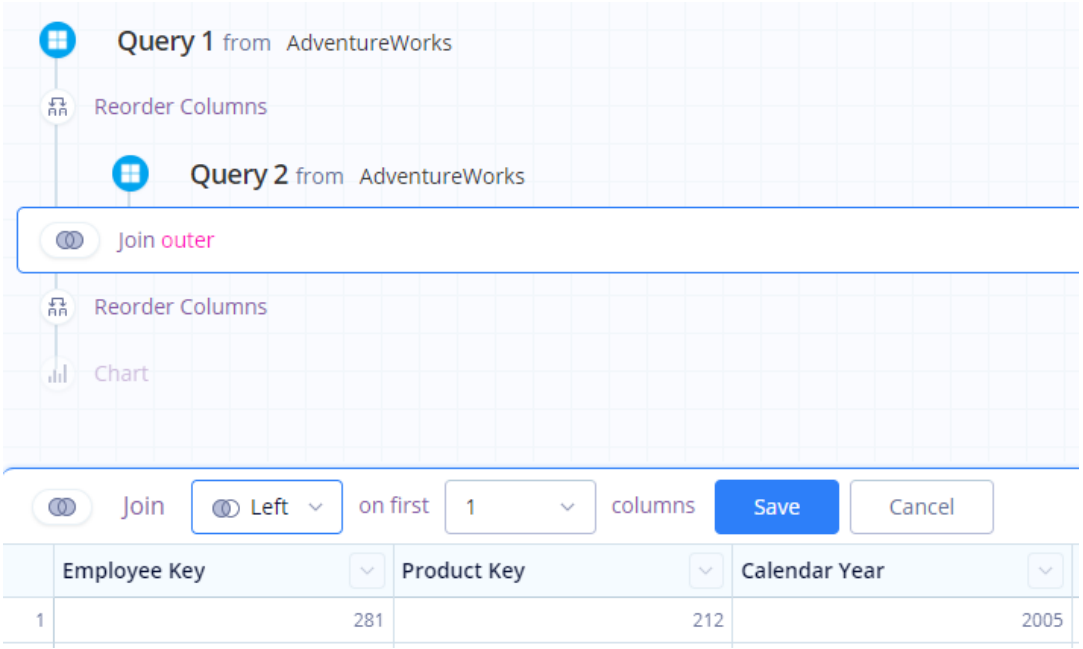


Figure 19. Merging queries through visual SQL in Chartio

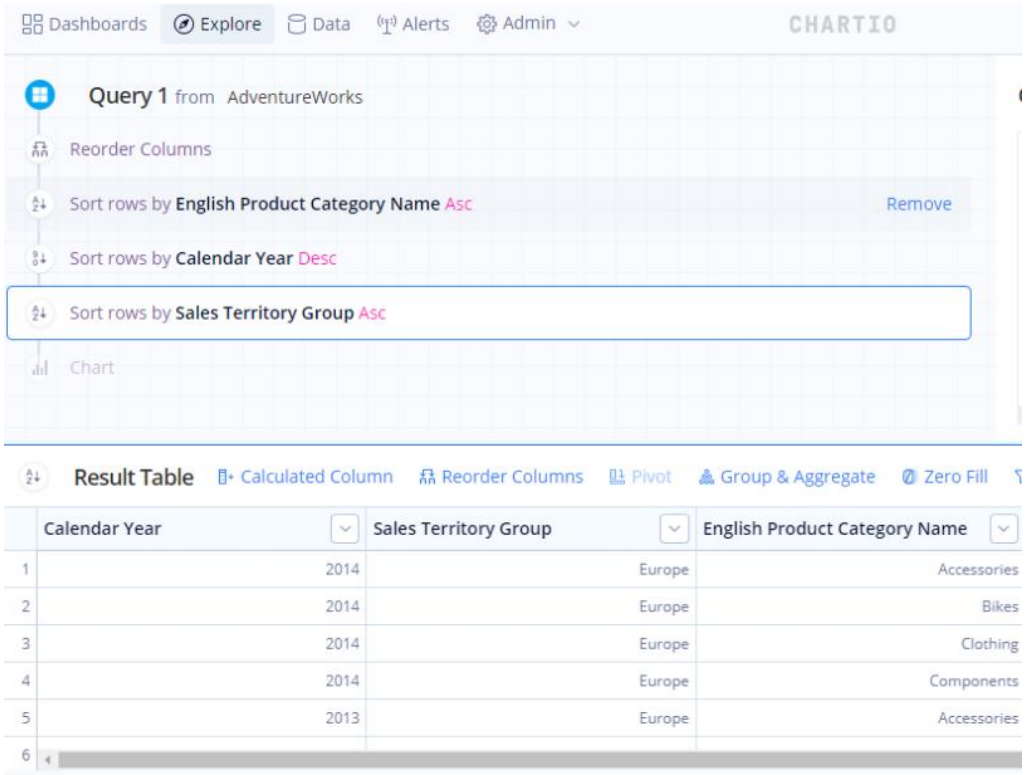


Figure 20.Final query built in Chartio

Calendar Year	Sales Territory Group	English Product Category Name	Sales Order Number	Order Quantity
2,014	Europe	Accessories	7,834	37,689
2,014	Europe	Bikes	14,716	136,095
2,014	Europe	Clothing	9,646	52,225
2,014	Europe	Components	16,031	202,596
2,013	Europe	Accessories	8,193	39,589
2,013	Europe	Bikes	15,052	140,365
2,013	Europe	Clothing	9,920	53,649
2,013	Europe	Components	16,448	209,283

View all 120 rows

Figure 21. Result set in Chartio

The query made from Chartio Visual SQL runs in Microsoft SQL Server Management Studio around 35 seconds on first run, 57 on second and 44 on third one. When updating the data in Chartio dashboard (running query on the background), the query lasts around 45, 44 and 42 seconds on three attempts.

3.4.4 Seal Report

The next software to be tested will be open-source product Seal Report. After downloading and installing the software, there will be two parts installed: Seal Server Manager and Seal Report Designer. First, the new data source connection to MSSQL Server will be created using Seal Server Manager. Setting up the connection was relatively easy and took less than 15 minutes, ODCB DSN driver needed to be configured if it has not been done beforehand to create the connection though ODBC driver. After the connection had been

established, it is possible to bring tables either separately by defining the SQL or through table catalogue as seen on Figure 22. Joins between tables were not created automatically so they needed to be added manually instead (Figure 23).

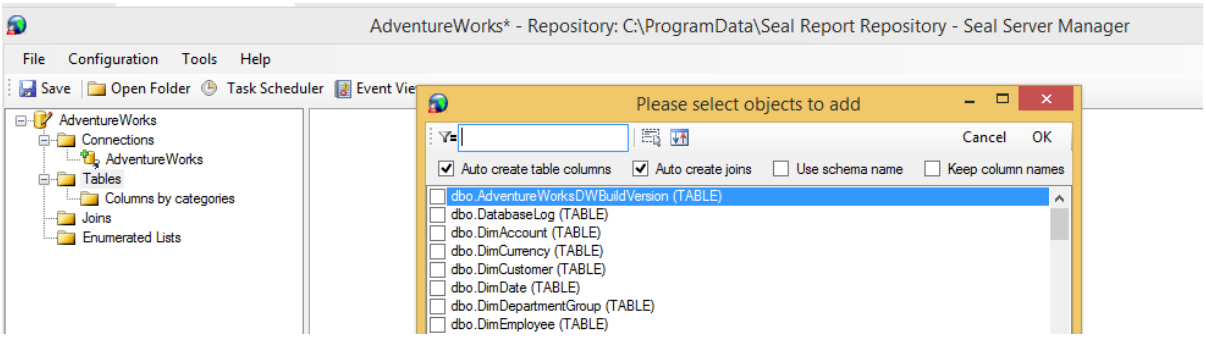


Figure 22. Selecting tables in Seal Server Manager.

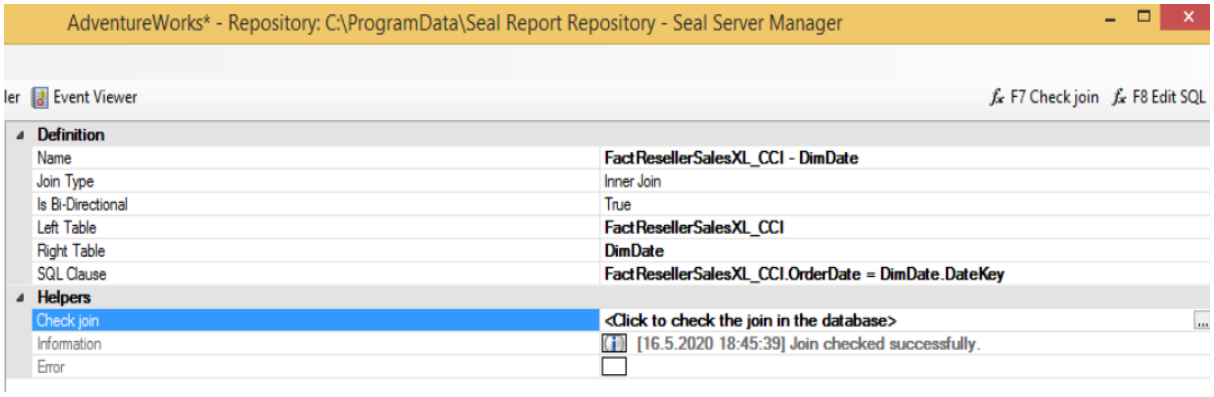


Figure 23. Creating relationships between tables in Seal Server Manager

When the data source, tables and relationships have been defined, the Seal Report Designer can be opened to create a new report. The report view is presented in Figure 24. Executing the report opens Seal Report Viewer where the final results are shown (Figure 25). In Seal Report, it is also possible to take out the visually build solution as SQL script as seen in Figure 26.

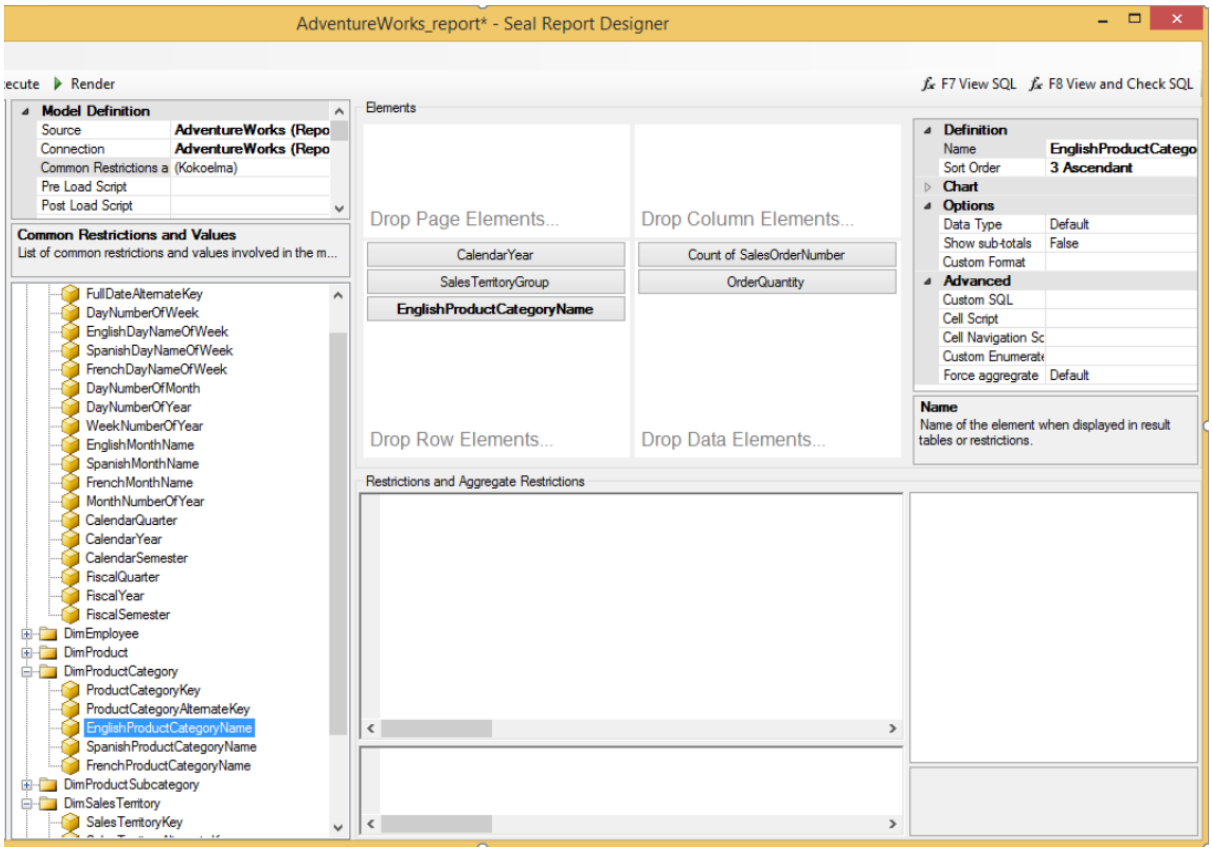
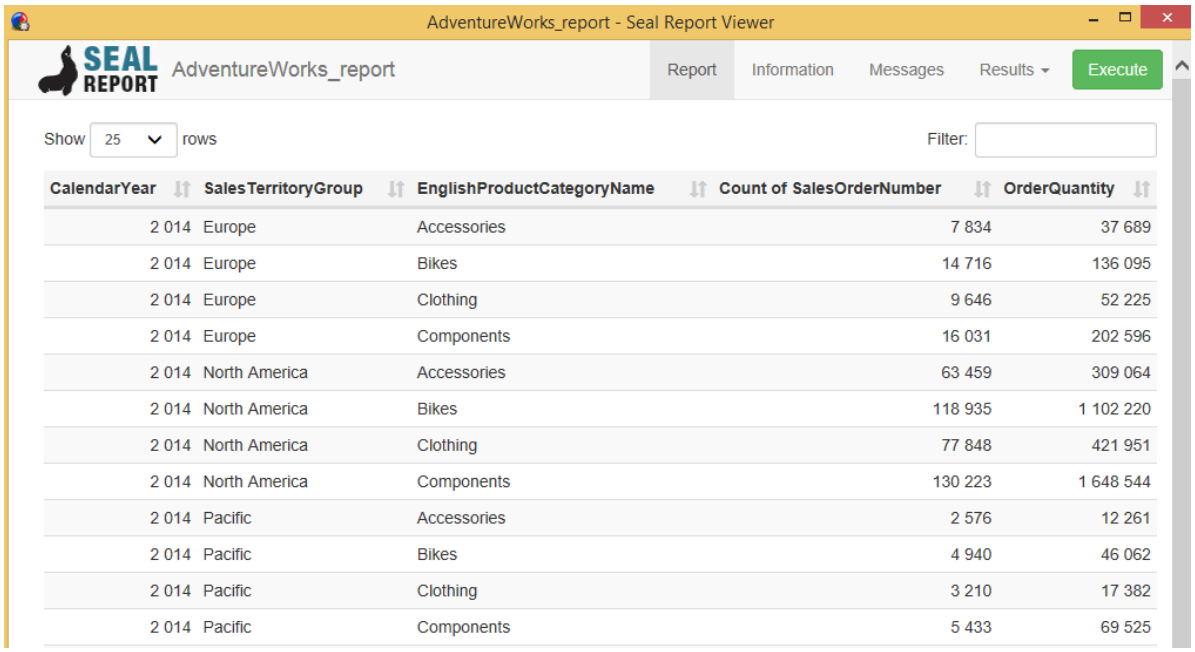


Figure 24. Building report in Seal Report Designer



CalendarYear	SalesTerritoryGroup	EnglishProductCategoryName	Count of SalesOrderNumber	OrderQuantity
2 014	Europe	Accessories	7 834	37 689
2 014	Europe	Bikes	14 716	136 095
2 014	Europe	Clothing	9 646	52 225
2 014	Europe	Components	16 031	202 596
2 014	North America	Accessories	63 459	309 064
2 014	North America	Bikes	118 935	1 102 220
2 014	North America	Clothing	77 848	421 951
2 014	North America	Components	130 223	1 648 544
2 014	Pacific	Accessories	2 576	12 261
2 014	Pacific	Bikes	4 940	46 062
2 014	Pacific	Clothing	3 210	17 382
2 014	Pacific	Components	5 433	69 525

Figure 25. Executing results of report in Seal Report Viewer

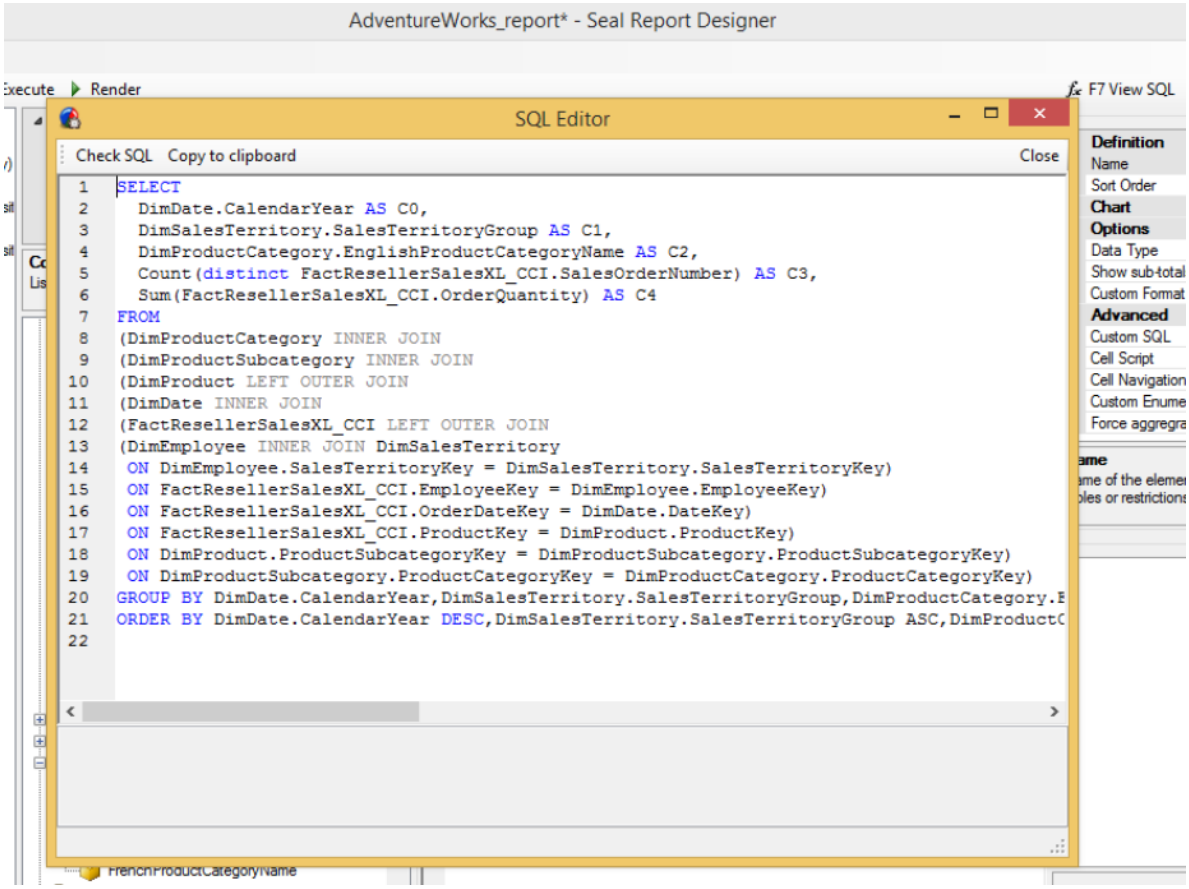


Figure 26. Seal Report Designer's SQL query produced based on the graphical visualization interface

In Microsoft SQL Management Studio, the SQL created by Seal Report software, runs on first attempt 35 seconds, second 38 seconds and third 41 seconds. When running the report, the program opens the Seal Report Viewer and the execution of the report takes around 42, 49 and 42 seconds on three attempts.

3.4.5 Microsoft Power BI

The current software tested includes first Microsoft Power BI. From Power BI desktop, the database connection will be created first. From navigator the wanted database tables will be selected (Figure 27). When loading the tables, Power BI creates the data model automatically based on the primary and foreign key relations of tables.

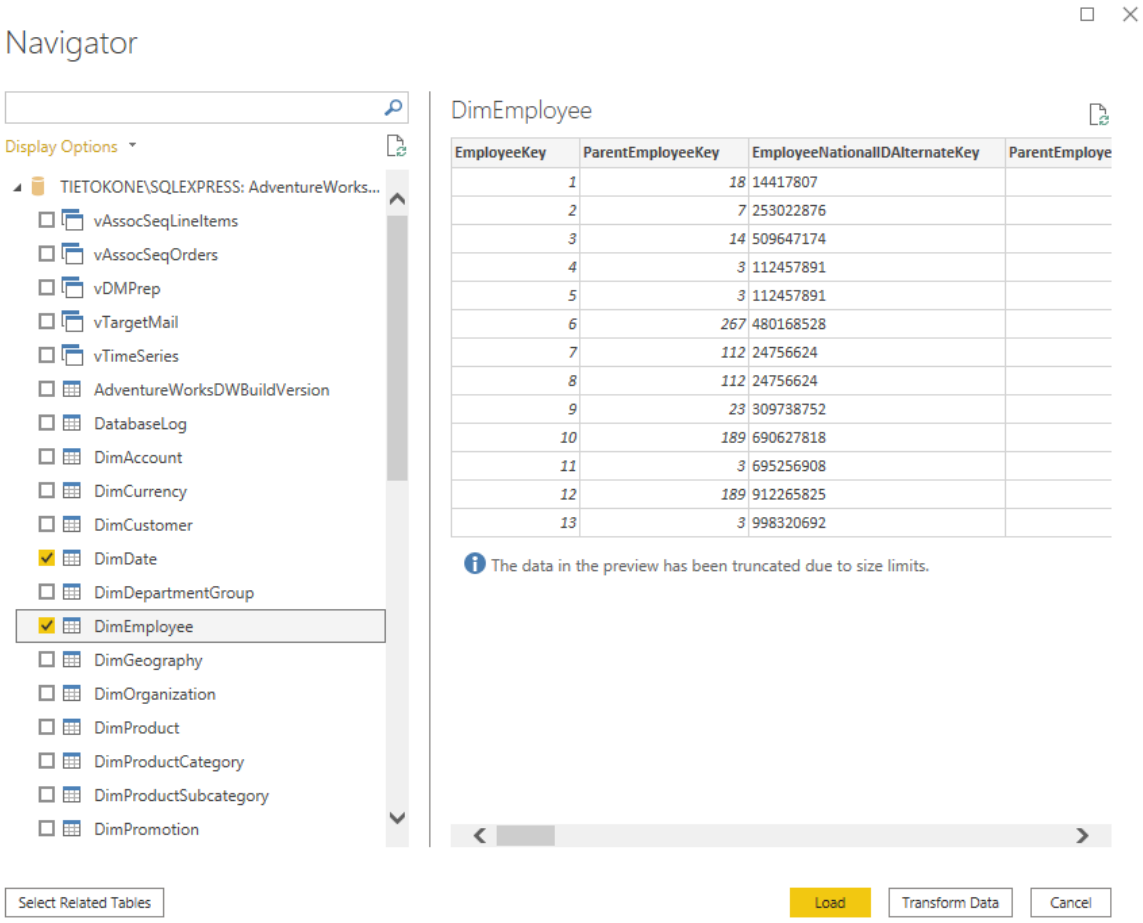


Figure 27. Selecting tables in Power BI

Since the data model has been created, the fields can be dragged and dropped to the chosen visualization to create the wanted data collection. The chosen fields be modified with calculations and formatting when added to the table as seen in Figure 28. Results are presented in Figure 29. When opening the report, loading data took 57, 60 and 62 seconds on three attempts. It seems that despite of the graphical interface, Power BI cannot take out SQL query for the created table, therefore, the load times will be measured only from building the table view.

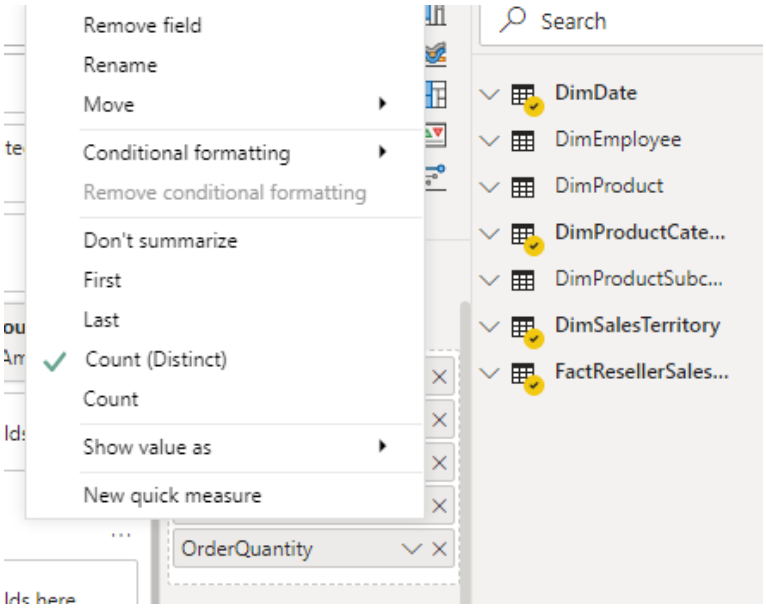
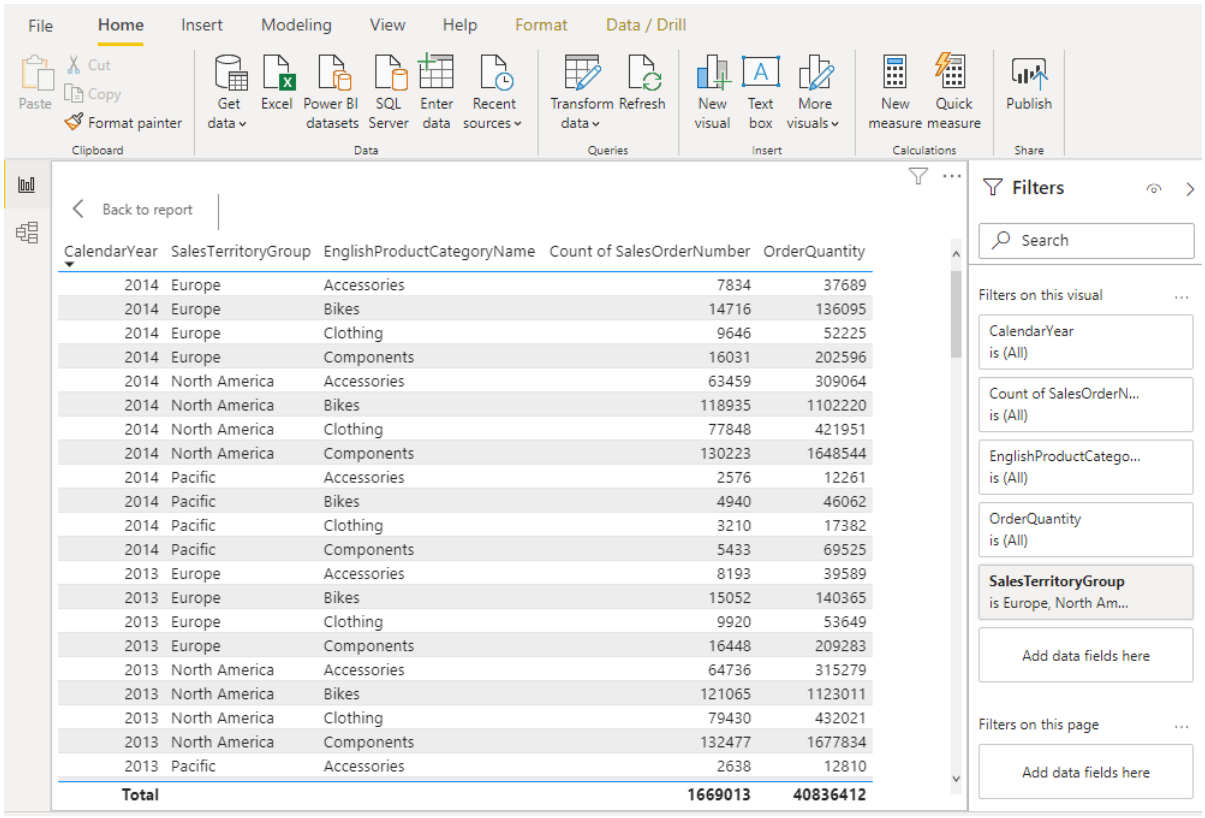


Figure 28. Creating calculated columns in Power BI



CalendarYear	SalesTerritoryGroup	EnglishProductCategoryName	Count of SalesOrderNumber	OrderQuantity
2014	Europe	Accessories	7834	37689
2014	Europe	Bikes	14716	136095
2014	Europe	Clothing	9646	52225
2014	Europe	Components	16031	202596
2014	North America	Accessories	63459	309064
2014	North America	Bikes	118935	1102220
2014	North America	Clothing	77848	421951
2014	North America	Components	130223	1648544
2014	Pacific	Accessories	2576	12261
2014	Pacific	Bikes	4940	46062
2014	Pacific	Clothing	3210	17382
2014	Pacific	Components	5433	69525
2013	Europe	Accessories	8193	39589
2013	Europe	Bikes	15052	140365
2013	Europe	Clothing	9920	53649
2013	Europe	Components	16448	209283
2013	North America	Accessories	64736	315279
2013	North America	Bikes	121065	1123011
2013	North America	Clothing	79430	432021
2013	North America	Components	132477	1677834
2013	Pacific	Accessories	2638	12810
Total			**1669013**	**40836412**

Figure 29. Final query results in Power BI

3.4.6 IBM Cognos Analytics

Comparation of the query building and run-times will be also done with IBM Cognos Analytics. First, the Microsoft SQL Server connection will be created. For database connection, IBM secure gateway connection and JDBC driver connections need to be configured

successfully. When the connection is ready, a data model will be created to choose the necessary tables and to create the relationships between tables as seen in Figure 30.

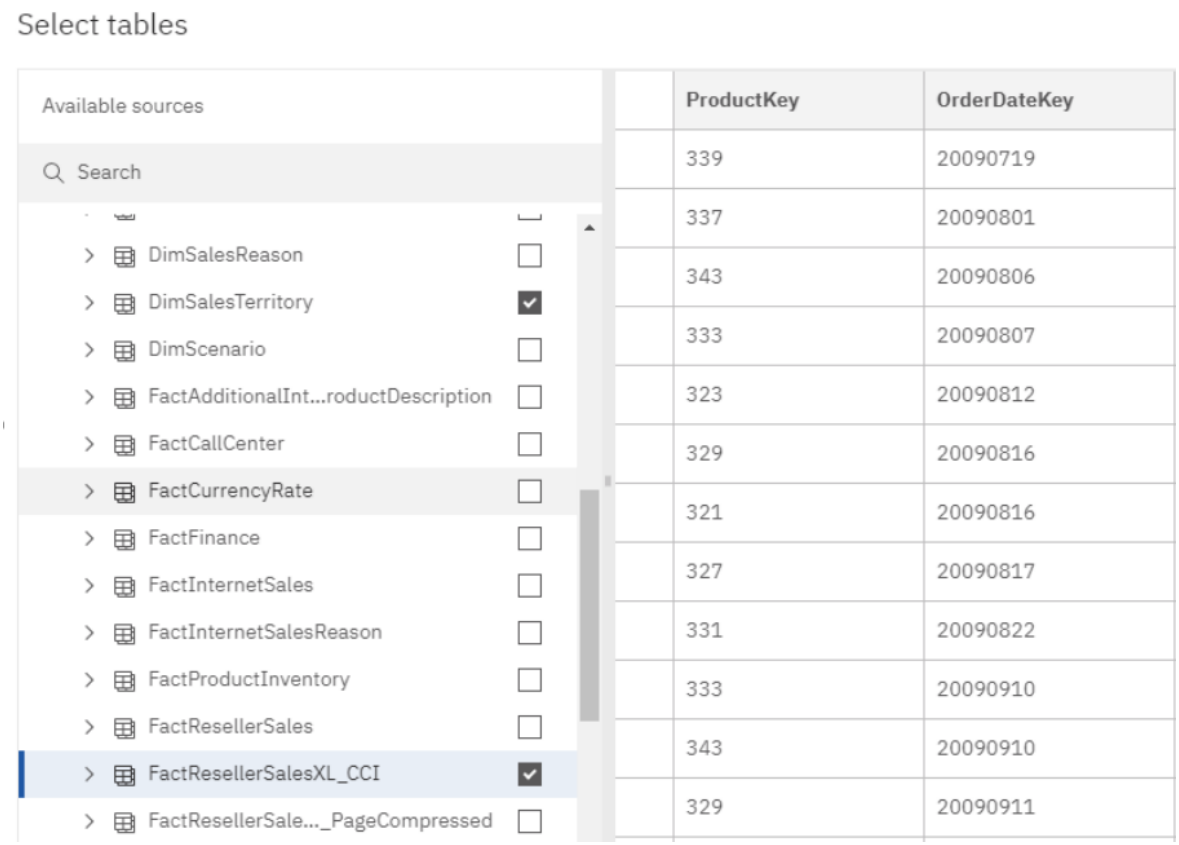


Figure 30. Creating data model in Cognos Analytics

When creating a report based on the data model, the fields will be selected in a query editor window in the report view (Figure 31) and results can be viewed through tabular query window as in Figure 32. In Cognos Analytics, it is also possible to retrieve the visually built queries as native SQL (Figure 33).

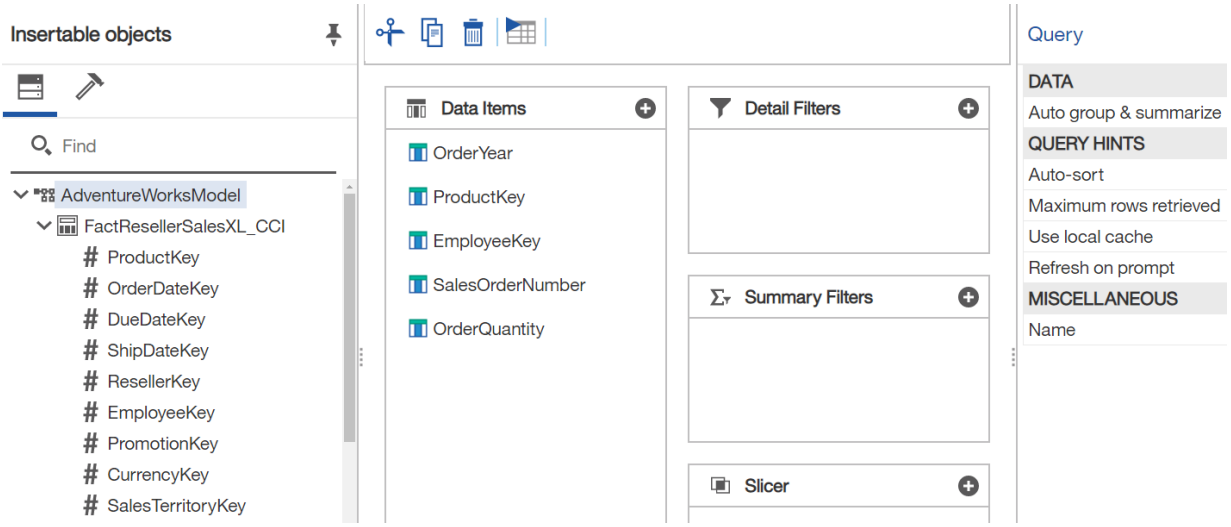
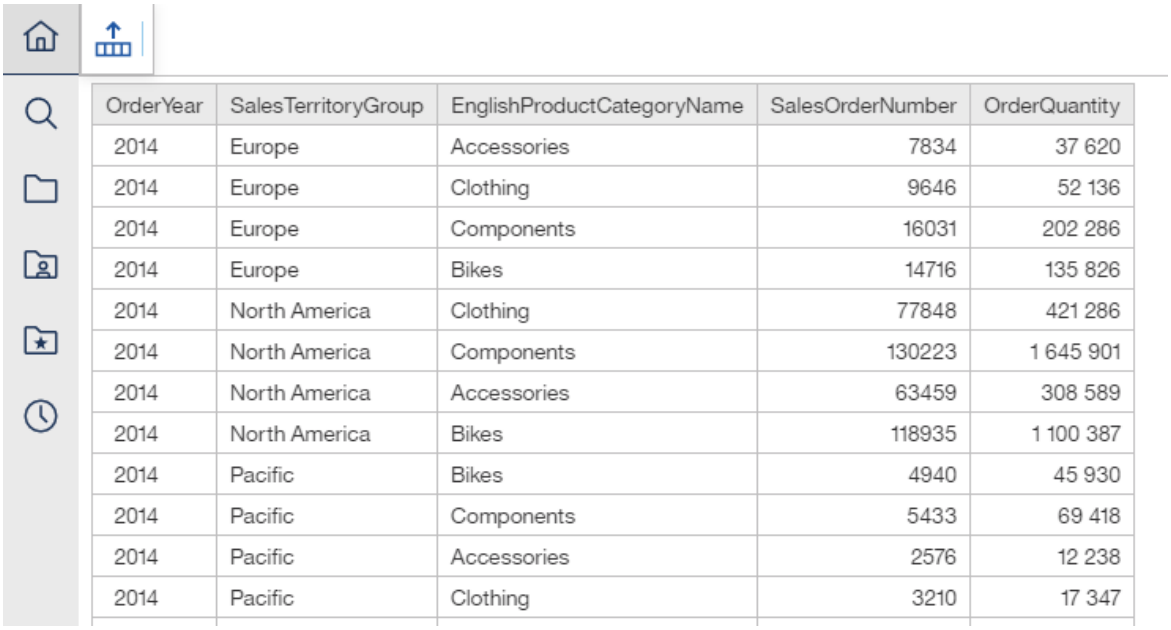


Figure 31. Building queries in Cognos Analytics



OrderYear	SalesTerritoryGroup	EnglishProductCategoryName	SalesOrderNumber	OrderQuantity
2014	Europe	Accessories	7834	37 620
2014	Europe	Clothing	9646	52 136
2014	Europe	Components	16031	202 286
2014	Europe	Bikes	14716	135 826
2014	North America	Clothing	77848	421 286
2014	North America	Components	130223	1 645 901
2014	North America	Accessories	63459	308 589
2014	North America	Bikes	118935	1 100 387
2014	Pacific	Bikes	4940	45 930
2014	Pacific	Components	5433	69 418
2014	Pacific	Accessories	2576	12 238
2014	Pacific	Clothing	3210	17 347

Figure 32. Final result set run in tabular view in Cognos Analytics

Figure 33. Graphical SQL query converted to SQL query in Cognos Analytics

Running the SQL query in MSSQL Server Management Studio, what Cognos Analytics produced, took 00:01:26, 00:01:39 and 00:01:19 minutes on three attempts. In Cognos Analytics, the result-set built runs on the first try 1 minute 30 seconds, second time 1 minute 13 seconds and third time 1 minute 7 seconds.

4 Discussion

Based on the theoretical and empirical part, the results will be presented and discussed in this chapter, and the conclusions made based on them. The trustworthiness of the study will be reviewed, and possible suggestions and development ideas taken into a consideration. In the end of the chapter the thesis process evaluation and learning from writer's point of view will be discussed.

4.1 Results and trustworthiness of the study

In the first iteration of empirical part, chapter 3.3, where different capabilities of software were compared to each other, the requirements were separated into 13 different sections in the tables from one to five. Based on those requirements, points were given for each software in order to measure which alternative would fill most of the criteria specified for the reporting and visualization software. The results are presented in Table 6 where also the points for the current software have been counted for comparison.

Some solutions showed considerably less desired features than others and were, therefore, easy to draw out from the testing phase. Databox and Google Data Studio were the excluded software of the commercial software, and BIRT Report Designer and KNIME Analytics Platform & Report Designer of the open source software. The rest of the software from commercial side were relatively even.

Tableau and Dundas BI covered most of the requirements and were chosen for the testing phase. Qlik Sense had a little bit better review than Clic Data or Chartio, but as visual SQL is an important consideration for more complex queries, Chartio was taken as a third option to be tested. From open source software, the only fitting software seemed to be Seal Report, so it was taken into a testing phase also to be compared to the commercial software. As previously mentioned, the tests were conducted to the current software as well.

From my opinion, it was sometimes hard to find the information about the features of software. It seemed that all the documentation was not absolute regarding, for example, data source connectors or sharing reports. Often the information needed to be gathered from various pages which might have resulted to the fact that some information was not found even though the software would have the feature. Therefore, I would say the listing of the features is more directional for choosing the software for the next step, also, some features will probably have more importance for the company than the others.

Table 6. Requirement points of the software

Software	Points from filling the requirements (max 13 points)	Software type
BIRT Report Designer	6	Open source
Chartio	10	Commercial
Clic Data	10	Commercial
Cognos Analytics	12	Current
Databox	7	Commercial
Dundas BI	12	Commercial
Google Data Studio	8	Commercial
KNIME Analytics Platform & Report Designer	8	Open source
Power BI	11	Current
Qlik Sense	11	Commercial
Seal Report	8	Open source
Tableau	12	Commercial

In the second iteration (software testing phase), the query built was not too complicated by its logic as in the scope of the project more complicated query would have taken too much time to plan and build with all the software. Also, as the Microsoft's sample database was used, there was not larger tables than 11 million rows. However, that amount is already considerable and can affect to the performance of the product used so it was used to build the test scenario. The tests in the second iteration did not show notable difference in the query results. The slowest results were shown in IBM Cognos Analytics but as the difference was relatively small compared to other query times, the difference might be also due to network traffic or other test environment related issues. For Dundas BI the testing of query times could not be conducted due to the different dataset. Also, as some software did not offer the possibility to take out the SQL query from built dataset, those queries are not tested in Microsoft SQL Server Management Studio. The comparation of query results can be seen from Table 7 below. The test T-SQL that I had written by myself for comparison, took 45 seconds (the average of three attempts).

Table 7. Average query times for the software

Average query times (seconds)	Software	Microsoft Management Studio
Dundas BI	--	--
Tableau	45	--
Chartio	44	45
Seal Report	44	38
Power BI	60	--
Cognos Analytics	77	88

Overall, none of the software tested showed clear advantage over each other in query times. When the query is run multiple times, the trustworthiness of the result increases leaving less possibility for the environmental factors, such as network traffic, to affect to the query result. Therefore, in this project the query times were measured generally three times. Despite of that, as some of the measurements have been taken by phone's chronometer, there might be less than five seconds difference to the actual query time due to human error in measuring the time (pressing the button in exactly the same second as the query starts running might not have succeeded and the error is also hard to detect).

The usability of the software was not tested with different users, but the review is based on the own observations of the writer. Some of the software were very fast and easy to set up. These programs were Tableau, Power BI and Seal Report. Dundas BI had long instructions so based on those, the installation of the full program would have taken maybe one day or more, therefore, the cloud-based version with limited connectors were used. Chartio and Cognos Analytics were more complicated mostly due to being on the cloud and the database connections to local server took more work than desktop version installations (i.e. SSH tunnel, DSN, TCP ports and secure gateways needed to be adjusted according the software specific instructions). However, for experienced user using database connection properties, they can still be set up relatively fast.

The data modelling and relationships between tables and creating reports itself was not too complicated in any of the options. As Cognos Analytics and Power BI are already familiar to me, it is hard to evaluate the usage of those user interfaces. The most time I spent learning Chartio visual SQL query because it was clearly very different from the other user interfaces. The main idea was similar than in Cognos Analytics, which is another software offering graphical SQL, but the implementation was quite different as it worked as a pipeline where each of the actions made will create a mark on the data flow.

Even though there was no installation required for Dundas BI cloud version, I felt that the user interface was not very clear to use and with this software I spent some time figuring out how to create the desired report. Seal Report and Tableau were both easiest options to use and did not require any documentation or other reference to figure out how the reports are built there. From my own observations and user experience, I created a table (Table 8) to present my opinion of the ease of use of the software based on the time I spent building reports.

Table 8. Ease of use of the software

Ease of use (writer's own observation)	Hard-Low (5-1)
Chartio	5
Dundas BI	4
Seal Report	2
Tableau	2

Most of the programs created data models and relationships automatically based on the tables chosen and the primary key and foreign key relations. Seal Report was the only one where all the relationships between tables needed to be created manually. For the test case designed, it was possible to create the report in all of the tested software but in most of the programs, if the queries needed to be merged, it should have been done in the data model. For example, once the user enters the dashboard with drag and drop functionalities, only smaller modifications to the measurements, hierarchies and calculated columns can be made. For more complex queries, merging and modification of data will be easiest to accomplish in Chartio and Cognos Analytics.

Overall, all of the tested software had their pros and cons. Seal Report was easy to use and had open source licensing, in addition, SQL script could be taken out from visually formed query. However, it lacks quite many of the required features such as using external files as data source, not being able to connect Google services and information was not found about using python or R for advanced analytics purposes. Besides, the software is only offered as on-premise version.

Chartio and Cognos Analytics have clear advantage with visual SQL builder that enables more complicated query logics than other software. Also, they cover most of the connectivity requirements and the visual SQL can be transferred into T-SQL. The minus points for Chartio comes from limiting the query results to 100 000 rows which might make larger datasets hard to use. Also, Python and R are not supported, and they do not have data

center in Europe even though they advertise themselves as GDPR compliant for customer's in Europe.

Dundas BI and Tableau cover nearly all of the requirements and they have relatively advanced graphical user interface for building reports. Minus point for Tableau comes from not being able to send scheduled reports by email. Also, the use of these software will probably require knowledge of building SQL queries.

From the commercial software, Chartio was the only one to offer only cloud-service. Dundas BI and Tableau had also possibility to on-premise version of the software. For the current infrastructure both options suit well but, in the future, if the data warehouse will be wanted to move to cloud, cloud-based reporting and visualization software might serve the purpose better.

As a conclusion, there is few different options of how to carry on with the visualization and reporting software in the company. As comparing all the information gathered and all the points considered in this discussion part, Chartio might be possible option instead of two software (Cognos Analytics and Power BI) since it has visual SQL query builder, data modelling option and possibility to build dashboards and more traditional tables. However, the limitations need to be considered carefully and more testing would be required to see how well it works with larger datasets.

Another solution would be to continue with Cognos Analytics and Power BI as they are fulfilling many of the requirements together and based on the study it is relatively hard to find alternatives that would fit the case company's needs. Microsoft's Power BI has already been implemented alongside of SSAS tabular cubes in order to respond better for more complex queries and slower processing times. However, if the current software solutions will be kept, some changes will be required to do to fix the memory issues i.e. by moving the analysis services into Microsoft Azure. Another option could be possibly to use cube technology in Cognos Analytics. Microsoft Power BI as a solution itself would not probably work because there is a limit of how much data it is possible to use and bring through connectors. The limit itself is not big enough for the company's usage and excel reports are not supported.

A third option would be to convert to use either Tableau or Dundas BI instead of Cognos Analytics and Power BI. As both have quite good advanced graphical interfaces and lot of native connectors available, the lack of visual SQL could be bypassed by using another software, such as Metabase, to create the SQL queries visually. As mentioned earlier,

with Metabase it is possible to build visual SQL and convert it to actual SQL script. The software is also open source solution. Options are presented in Table 9.

Table 9. Options to improve analytics services solution

Option A	Option B	Option C
To move from Cognos Analytics and Power BI to Chartio.	To continue with Cognos Analytics and Power BI and consider making improvement with them i.e. using Analyses Services through Azure	To switch using Tableau or Dundas BI and open source software for visualization such as Metabase.

4.2 Suggestions and development ideas

The research could be improved by testing the query times with larger datasets and with more complicated query logics. The more complicated the query, and larger the dataset, the bigger are the differences in performance generally. Also, to fully test the features, more precise and profound use would be required, and more capabilities should be tested.

Also, if considering switching the software in use, it might be beneficial to test the usage of the software on a larger scale among the other end users in the case company to see how complicated or easy it is to use the software from their opinion. The primary group of the users (using the software daily or weekly) in the case company would probably fall into the category of intermediate or expert user so based on writer's observations, none of the tested software would be too demanding to be used. But as stated above, more analysis and testing would be required to decide whether the end users would be willing to adapt for the usage of the new software.

These additional tests could be used to narrow down more of the three different solutions this study suggested. In the first and third option might be beneficial also to consider what is the gained benefit from changing the software compared to the work with moving all the existing reports to the new software.

4.3 Evaluation of thesis process and professional development

From my point of view, doing this thesis project was somewhat difficult as I was working full-time the whole time and was also learning lot of new things at my workplace due to changing my job and responsibilities inside the company. Mostly, this was shown in the schedule of the project and I had to improve my time management skills in order to be

able to stay in the project schedule and to find a balanced way to combine time between work, thesis project and personal life. However, I think this has been beneficial not only for time management skills but also to possibly support my idea of making further studies alongside of work in the future in order to develop my expertise.

I ended up choosing this thesis project topic because as I have been working in the company, I have recognized together with my supervisor and other team members issues in the data visualization and reporting that could be possibly accomplished more efficiently and wanted to try to find an alternative solution. Also, as working in the insurance field, I was interested to learn how the process of choosing software for this specific field would require and how implementing it would work in the case company. As I have been working closely with thesis related things such as creating reports and visualizations and later by supporting the infrastructure behind the software and being responsible of improvement of the technological solutions related to them, I wanted to understand the whole scope behind software solutions in use. Overall, I think this project has helped me to develop my expertise regarding the technological solutions chosen and to understand better the context and requirements for them. I also think that considering my current position in the company, it will bring benefit also to my employer in the future development projects.

Besides of the work-related matters, I had first difficulties of defining the scope of the project and to realize what things will be possible to conduct in the given timeframe for the project. The thesis process has helped me to understand better the importance of delimitation of the project scope in early enough phase and to consider the relevance of the different tasks and plans in the project.

References

Actuate Corporation 2013. About BIRT data sources. URL: https://help.eclipse.org/2019-09/index.jsp?topic=%2Forg.eclipse.birt.doc%2Fbirt%2Fcon-AboutBIRTDataSources.html. Accessed: 18.4.2020.

Adair, B. The Difference Between Business Intelligence vs Business Analytics Solutions. URL: https://selecthub.com/business-intelligence/business-intelligence-vs-business-analytics/. Accessed: 25.08.2019.

BARC. Top Business Intelligence Trends 2020: What 2,865 BI Professionals Really Think. URL: https://bi-survey.com/top-business-intelligence-trends. Accessed: 30.12.2019.

Bertolucci, J. 2013. Big Data Analytics: Descriptive Vs. Predictive Vs. Prescriptive. URL: https://www.informationweek.com/big-data/big-data-analytics/big-data-analytics-descriptive-vs-predictive-vs-prescriptive/d/d-id/1113279. Accessed: 07.09.2019.

Caldwell, C. 2017. Business Intelligence vs. Analytics: What's the Difference? URL: https://www.logianalytics.com/bi-trends/business-intelligence-vs-analytics-whats-the-difference/. Accessed: 25.08.2019.

Chartio 2017. GDPR Statement. URL: https://chartio.com/blog/gdpr-statement/. Accessed: 26.4.2020.

Chartio 2018. Privacy Shield Statement. URL: https://chartio.com/about/legal/privacyshield/. Accessed: 26.4.2020.

Chartio 2020a. Build interactive dashboards from your company's data with Chartio. URL: https://chartio.com/product/dashboards/. Accessed: 18.4.2020.

Chartio 2020b. Email Reports. URL: https://chartio.com/docs/sharing/email-reports/. Accessed: 18.4.2020.

Chartio 2020c. Embedding. URL: https://chartio.com/docs/embedding/#embedding-security. Accessed: 26.4.2020.

Chartio 2020d. Finally, a visual version of SQL. URL: https://chartio.com/product/visual-sql/. Accessed: 18.4.2020.

Chartio 2020e. Merge Queries. URL: https://chartio.com/docs/visual-sql/merge-queries/. Accessed: 16.5.2020.

Chartio 2020f. Priced to scale. URL: https://chartio.com/product/pricing/. Accessed: 18.4.2020.

Chartio 2020g. Quick set-up, countless integrations. URL: https://chartio.com/product/data-sources/. Accessed: 18.4.2020.

Chartio 2020h. Set row-level permissions and filter data based on the viewer. URL: https://chartio.com/docs/data-sources/faqs/row-level-permissions/. Accessed: 26.4.2020.

Chartio 2020i. Share Dashboards. URL: https://chartio.com/docs/sharing/dashboards/. Accessed: 18.4.2020.

Chartio 2020j. Start a Query. URL: https://chartio.com/docs/visual-sql/start-a-query/. Accessed: 16.5.2020.

Clic Data a. Create New Schedule. URL: https://app.clicdata.com/help/docs/341. Accessed: 12.4.2020.

Clic Data b. Local File. URL: https://app.clicdata.com/help/docs/local-file. Accessed: 25.4.2020.

Clic Data c. Our plans & connectors. URL: https://www.clicdata.com/pricing/data-connectors/. Accessed: 12.4.2020.

Clic Data d. ODBC. URL: https://app.clicdata.com/help/docs/connectionodbc. Accessed: 25.4.2020.

Clic Data e. Plan Details. URL: https://www.clicdata.com/pricing/comparison/. Accessed: 12.4.2020.

Clic Data f. Pricing. URL: https://www.clicdata.com/pricing/. Accessed: 12.4.2020.

Clic Data g. Publish, share & collaborate. URL: https://www.clicdata.com/product/automation/publish-share/. Accessed: 12.4.2020.

Clic Data h. Security & trust center. URL: https://www.clicdata.com/company/security/. Accessed: 25.4.2020.

Clic Data i. Share a dashboard with parameters. URL: https://app.clicdata.com/help/docs/share-a-dashboard-with-parameters. Accessed: 12.4.2020.

Columbus, L. 2019. What Matters Most In Business Intelligence, 2019. URL: https://www.forbes.com/sites/louiscolumbus/2019/06/09/what-matters-most-in-business-intelligence-2019/#20249b71702d. Accessed: 30.12.2019.

Cook, R. 2013. Building a BI Infrastructure. URL: https://it.toolbox.com/blogs/erpdesk/building-a-bi-infrastructure-120913. Accessed: 17.11.2019.

Cowdrey, B. 2018. What exactly is analytics? URL: https://www.quora.com/What-exactly-is-analytics. Accessed: 24.08.2019.

DAMA International 2017. DAMA-DMBOK: Data Management Body of Knowledge (2nd Edition). URL: https://www.oreilly.com/library/view/dama-dmbok-data-management/9781634622479/Chapters-5.xhtml. Accessed: 08.12.2019.

Databox 2020. Overview: Scheduled Snapshots. URL: https://help.databox.com/article/138-overview-scheduled-snapshots. Accessed: 26.4.2020.

Databox, Inc. 2020a. Get all of your data into Databox. URL: https://databox.com/product/any-data. Accessed: 20.4.2020.

Databox, Inc. 2020b. Pricing that grows with your Business. URL: https://databox.com/pricing#. Accessed: 20.4.2020.

Databox, Inc. 2020c. Receive alerts & recommendations when they matter most. URL: https://databox.com/product/alerts-insights. Accessed: 20.4.2020.

Databox, Inc. 2020d. Security Policy. URL: https://databox.com/security-policy. Accessed: 20.4.2020.

Dodson, V. 2013. Visualizing Big Data with Eclipse BIRT. URL: https://wiki.eclipse.org/images/a/a0/Visualizing_Big_Data_with_BIRT_EclipseDay_12182013.pdf. Accessed: 18.4.2020.

Dundas Data Visualization, Inc. 1999-2020a. Dashboards. URL: https://www.dundas.com/dundas-bi/features#reporting. Accessed: 19.4.2020.

Dundas Data Visualization, Inc. 1999-2020b. Notifications (alerts). URL: https://www.dundas.com/support/learning/documentation/share-collaborate/notifications-alerts. Accessed: 19.4.2020.

Dundas Data Visualization, Inc. 1999-2020c. Share or export your work. URL: https://www.dundas.com/support/learning/documentation/share-collaborate/share-or-export-your-work. Accessed: 19.4.2020.

Durcevic, S. 2018. Top 10 Analytics And Business Intelligence Trends for 2019. URL: https://www.datapine.com/blog/business-intelligence-trends/. Accessed: 31.10.2019.

Engineering 2017a. Community Edition. URL: https://www.knowage-suite.com/site/licensing/community-edition/. Accessed: 21.4.2020.

Engineering 2017b. Knowage editions. URL: https://www.knowage-suite.com/site/licensing/knowage-editions/. Accessed: 21.4.2020.

Financial Supervisory Authority 2018. Solvency II. URL: https://www.finanssivalvonta.fi/en/regulation/regulatory-framework/solvency-ii/. Accessed: 3.5.2020.

Freeman, J. 2015. Beginning BIRT: A Practical Introduction. URL: https://www.eclipse.org/community/eclipse_newsletter/2015/september/article3.php. Accessed: 18.4.2020.

García, J. 2015. BI Buyer's Guide Data Discovery and Visualization. URL: https://www.dundas.com/resource/getwhitepaper?whitePaperName=07-10-2015-TEC-Buyer%27s-Guide-2015%2FTEC-2015-BI-Data-Discovery-and-Visualization-Buyers-guide-Dundas.pdf. Accessed: 19.4.2020.

Google a. Connect to Data. URL: https://datastudio.google.com/data. Accessed: 19.4.2020.

Google b. Hamina, Finland. URL: https://www.google.com/about/datacenters/locations/hamina/. Accessed: 19.4.2020.

Google c. Transform your business through insights. URL: https://marketingplatform.google.com/about/resources/analytics-product-overview/. Accessed: 2.5.2020.

Google 2020a. About file upload. URL: https://support.google.com/datastudio/answer/7333350?hl=en-GB&ref_topic=7333349. Accessed: 19.4.2020.

Google 2020b. About sharing. URL: https://support.google.com/datastudio/answer/6287179. Accessed: 19.4.2020.

Google 2020c. Embed a report. URL: https://support.google.com/datastudio/answer/7450249. Accessed: 19.4.2020.

Google 2020d. Filter by email address. URL: https://support.google.com/datastudio/answer/9713766?hl=en. Accessed: 19.5.2020.

Google 2020e. Manage data freshness. URL: https://support.google.com/datastudio/answer/7020039. Accessed: 19.4.2020.

Google 2020f. Share reports on social media. URL: https://support.google.com/datastudio/answer/9069190. Accessed: 19.4.2020.

Google 2020g. Schedule email delivery. URL: https://support.google.com/datastudio/answer/9263641?hl=en. Accessed: 27.4.2020.

Google 2020h. Visualize spend over time with Google Data Studio. URL: https://cloud.google.com/billing/docs/how-to/visualize-data. Accessed: 19.4.2020.

Hirst, A. 2017. Digital Transformation In the Insurance Industry: A Snapshot. URL: https://www.digitalistmag.com/customer-experience/2017/10/09/digital-transformation-in-insurance-industry-snapshot-05410435. Accessed: 15.09.2019.

IBM a. IBM Cognos Analytics. URL: https://www.ibm.com/fi-en/products/cognos-analytics. Accessed: 11.4.2020.

IBM b. Creating a data source connection. URL: https://www.ibm.com/support/knowledge-center/SSEP7J_11.1.0/com.ibm.swg.ba.cognos.ug_cra.doc/t_asg_createdatasource.html. Accessed: 24.4.2020.

IBM c. Global locations for your global business. URL: https://www.ibm.com/cloud/data-centers/. Accessed: 11.4.2020.

IBM d. Release 11.0.6 – March 2017. URL: https://www.ibm.com/support/knowledge-center/SSEP7J_11.0.0/com.ibm.swg.ba.cognos.ca_new.doc/c_ca_nf_11_0_6.html. Accessed: 24.4.2020.

IBM e. Release 11.0.8 – November 2017. URL: https://www.ibm.com/support/knowledge-center/SSEP7J_11.0.0/com.ibm.swg.ba.cognos.ca_new.doc/c_ca_nf_11_0_8.html. Accessed: 24.4.2020.

International Business Machines Corporation (IBM). 2020. Software Product Compatibility Reports. URL: https://www.ibm.com/software/reports/compatibility/clarity-reports/report/html/softwareReqsForProduct?deliverableId=BE691D30C50C11E9B0F2505999809CBE&osPlatform=Windows. Accessed: 01.03.2019.

Kess, B. 2017. AdventureWorks sample databases. URL: https://github.com/Microsoft/sql-server-samples/releases/tag/adventureworks?fbclid=IwAR0uPxyQgyWxtDo8VKJ-ItsufwngvCAIbW_5n9YHwSUVPbTzEuJ04yHNkA. Accessed 4.5.2020.

Knight, M. 2018a. What is Analytics? URL: https://www.dataversity.net/what-is-analytics/#. Accessed: 24.08.2019.

Knight, M. 2018b. What is Business Intelligence? URL: https://www.dataversity.net/what-is-business-intelligence/. Accessed: 22.09.2019.

KNIME AG a. Chapter 2. KNIME Server: Automation and Deployment. URL: https://www.knime.com/knime-server-course/chapter2. Accessed: 19.4.2020.

KNIME AG b. KNIME Analytics Platform. URL: https://www.knime.com/knime-analytics-platform. Accessed: 19.4.2020.

KNIME AG c. KNIME Report Designer. URL: https://www.knime.com/knime-report-designer. Accessed: 27.4.2020.

KNIME AG. 2014. Querying Google Analytics in KNIME. URL: https://www.knime.com/blog/querying-google-analytics-in-knime. Accessed: 19.4.2020.

KNIME AG. 2018. KNIME Database Extension Guide. URL: https://docs.knime.com/2019-06/db_extension_guide/index.html. Accessed: 19.4.2020.

Mallon, S. 2018. Predictive Modeling and Big Data Are Insurance Industry Powerhouses. URL: https://www.smartdatacollective.com/predictive-modeling-and-big-data-are-insurance-industry-powerhouses/. Accessed: 25.3.2019.

Brown, B., Bughin, J., Byers, A.H., Chui, M., Dobbs, R., Manyika, J. & Roxburgh, C. 2011. Big Data: The next frontier for innovation, competition and productivity. URL: https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.ashx. Accessed: 09.10.2019.

Metabase 2020. Pricing. URL: https://www.metabase.com/enterprise/pricing.html. Accessed: 21.4.2020.

Microsoft 2019a. Data sources for the Power BI service. URL: https://docs.microsoft.com/en-us/power-bi/service-get-data. Accessed: 18.4.2020.

Microsoft 2019b. Share Power BI dashboards and reports with coworkers and others. URL: https://docs.microsoft.com/en-us/power-bi/service-share-dashboards. Accessed: 18.4.2020.

Microsoft 2020a. Business analytics in a cloud you can trust. URL: https://powerbi.microsoft.com/en-us/clouds/. Accessed: 18.4.2020.

Microsoft 2020b. Comparing tabular and multidimensional solutions. URL: https://docs.microsoft.com/en-us/analysis-services/comparing-tabular-and-multidimensional-solutions-ssas?view=asallproducts-allversions. Accessed: 01.03.2020.

Microsoft 2020c. Data Sources in Power BI Desktop. URL: https://docs.microsoft.com/en-us/power-bi/desktop-data-sources. Accessed: 18.4.2020.

Microsoft 2020d. Download SQL Server Management Studio (SSMS). URL: https://docs.microsoft.com/en-us/sql/ssms/download-sql-server-management-studio-ssms?view=sql-server-ver15. Accessed: 3.5.2020.

Microsoft 2020e. Editions and supported features of SQL Server 2016. URL: https://docs.microsoft.com/en-us/sql/sql-server/editions-and-components-of-sql-server-2016?view=sql-server-ver15. Accessed: 01.03.2020.

Microsoft 2020f. High Availability and Scalability in Analysis Services. URL: https://docs.microsoft.com/en-us/analysis-services/instances/high-availability-and-scalability-in-analysis-services?view=asallproducts-allversions. Accessed: 19.5.2020.

Microsoft 2020g. Microsoft SQL Server 2016 Service Pack 2 Express. URL: https://www.microsoft.com/en-US/download/details.aspx?id=56840. Accessed: 3.5.2020.

Microsoft 2020h. Power BI pricing. URL: https://powerbi.microsoft.com/en-us/pricing/. Accessed: 18.4.2020.

Microsoft 2020i. Publish an app in Power BI. URL: https://docs.microsoft.com/en-us/power-bi/service-create-distribute-apps. Accessed: 18.4.2020.

Microsoft 2020j. SQL Server Analysis Services server management. URL: https://docs.microsoft.com/en-us/analysis-services/instances/analysis-services-instance-management?view=asallproducts-allversions. Accessed: 01.03.2020.

Microsoft 2020k. Ways to share your work in Power BI. URL: https://docs.microsoft.com/en-us/power-bi/service-how-to-collaborate-distribute-dashboards-reports . Accessed: 18.4.2020.

Microsoft 2020l. Workspace database. URL: https://docs.microsoft.com/en-us/analysis-services/tabular-models/workspace-database-ssas-tabular?view=asallproducts-allversions. Accessed: 01.03.2020.

Mitchell, S. N. 2017. How To Attract Users by Targeting Their Actual Abilities. URL: https://uxplanet.org/how-to-attract-users-by-targeting-their-actual-abilities-3c4ae1121644. Accessed: 27.10.2019.

Marr, B. 2019. What's The Difference Between Structured, Semi-Structured and Unstructured Data? URL: https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/#f1f876e2b4d3. Accessed: 17.11.2019.

Moottori 2015. Vakuutusyhtiö: Jatkossa ajokilometrit ja kuljettajan ikä määräävät hinnan. URL: https://moottori.fi/liikenne/jutut/vakuutusyhtio-jatkossa-ajokilometrit-ja-kuljettajan-ika-maaravat-hinnan/. Accessed: 08.12.2019.

Munk, D. 2019. Cloud-Based Vs. On-Premise Servers. URL: https://www.forbes.com/sites/forbestechcouncil/2019/03/22/cloud-based-vs-on-premise-servers/#70e4c88879e2. Accessed: 31.10.2019.

National Association of Insurance Commissioners 2020. Artificial Intelligence. URL: https://content.naic.org/cipr_topics/topic_artificial_intelligence.htm. Accessed: 29.4.2020.

Perret, R. 2014. 3 key IT infrastructure requirements for big data and analytics. URL: https://www.ibm.com/blogs/systems/3-key-it-infrastructure-requirements-for-big-data-and-analytics/. Accessed: 16.11.2019.

Peterson, E. 2013. What is Business Intelligence? URL: https://www.businessnews-daily.com/4692-business-intelligence.html. Accessed: 07.10.2019.

POP Vakuutus. POP Vakuutus on suomalainen vakuutusyhtiö. URL: https://www.popvakuutus.fi/yritys. Accessed: 21.1.2020.

Pratt, M. K. 2017. What is business intelligence? Transforming data into business insights. URL: https://www.cio.com/article/2439504/business-intelligence-definition-and-solutions.html. Accessed: 22.09.2019.

QlikTech International AB 2018. General Data Protection Regulation (GDPR) and Qlik Sense Cloud. URL: https://www.qlik.com/us/-/media/files/resource-library/global-us/direct/datasheets/ds-gdpr-qlik-sense-cloud-products-faq-en.pdf . Accessed: 20.4.2020.

QlikTech International AB 1993-2019. Qlik Sense connection to MariaDB. URL: https://support.qlik.com/articles/000050451. Accessed: 20.4.2020.

QlikTech International AB 1993-2020a. About Qlik Connectors. URL: https://help.qlik.com/en-US/connectors/Content/Connectors_Home/Home.htm. Accessed: 20.4.2020.

QlikTech International AB 1993-2020b. Built-in Qlik Web Connectors. URL: https://help.qlik.com/en-US/connectors/Subsystems/Integrated_Web_Connectors_help/Content/Connectors_QWC_BuiltIn/Introduction/IntegratedWebConnectors.htm. Accessed: 20.4.2020.

QlikTech International AB 1993-2020c. ODBC Connector Package. URL: https://help.qlik.com/en-US/connectors/Subsystems/ODBC_connector_help/Content/Connectors_ODBC/Introduction/ODBC-connector.htm. Accessed: 20.4.2020.

QlikTech International AB 1993-2020d. Qlik Connectors. URL: https://www.qlik.com/us/products/qlik-connectors. Accessed: 27.4.2020.

QlikTech International AB 1993-2020e. Qlik NPrinting. URL: https://www.qlik.com/us/products/nprinting. Accessed: 20.4.2020.

QlikTech International AB 1993-2020f. Qlik Pricing. URL: https://www.qlik.com/us/pricing. Accessed: 20.4.2020.

QlikTech International AB 2020a. Advanced Reporting and Distribution. URL: https://www.qlik.com/us/-/media/files/resource-library/global-us/direct/datasheets/ds-qlik-nprinting-en.pdf?la=en&hash=5B9997A0BC471C36B24574EDBBF78D8CCDD32F3D. Accessed: 20.4.2020.

QlikTech International AB 2020b. R Integration with Qlik Sense. URL: https://support.qlik.com/articles/000039436. Accessed: 27.4.2020.

Rouse, M. & Vaughan, J. 2019. data governance (DG). URL: https://searchdatamanagement.techtarget.com/definition/data-governance. Accessed: 17.11.2019.

Saporito P, L. 2014. Applied Insurance Analytics: A Framework for Driving More Value from Data Assets, Technologies and Tools. URL: https://learning.oreilly.com/library/view/applied-insurance-analytics/9780133760729/ch01.html. Accessed: 07.08.2019.

Scherbak, M. 2019. When Business Analytics Meet Machine Learning. URL: https://towardsdatascience.com/when-business-analytics-meets-machine-learning-10ecaada9d8. Accessed: 25.08.2019.

Seal Report a. Data Sources. URL: https://sealreport.org/Sources. Accessed: 21.4.2020.

Seal Report b. Recipes. URL: https://sealreport.org/Recipes. Accessed: 28.4.2020.

Seal Report c. Reports. URL: https://sealreport.org/Reports. Accessed: 21.4.2020.

Seal Report d. Welcome to Seal Report. URL: https://sealreport.org/. Accessed: 21.4.2020.

Sharabi, K. 2019. Row-level security with Power BI Embedded. URL: https://docs.microsoft.com/en-us/power-bi/developer/embedded/embedded-row-level-security. Accessed: 24.4.2020.

Some, K. 2018. Top 7 Big Data Analytics Trends For 2019. URL: https://www.analyticsinsight.net/top-7-big-data-analytics-trends-for-2019/. Accessed: 31.10.2019.

Stencil, S. 2018. Choosing the Right Business Software. URL: https://search-proquest-com.ezproxy.haaga-helia.fi/docview/2048234085/fulltext/573089F71C104174PQ/1?accountid=27436. Accessed: 21.1.2020.

Suomen Vahinkovakuutus Oy a. URL: https://www.suomenvahinkovakuutus.fi/. Accessed: 21.1.2020.

Suomen Vahinkovakuutus Oy b. Vakavaraisuutta ja taloudellista tilaa koskeva kertomus 2019. URL: https://www.suomenvahinkovakuutus.fi/wp-content/uploads/2020/03/SVV-Vakavaraisuutta_ja_taloudellista_tilaa_koskeva_kertomus_2019.pdf. Accessed: 12.4.2020.

Säästöpankin vakuutukset. Tutustu säästöpankin vakuutuksiin. URL: https://www.saastopankinvakuutukset.fi/fi-fi. Accessed: 25.1.2020.

Tableau Software, LLC. 2003-2020a. Restrict Access at the Data Row Level. URL: https://help.tableau.com/current/pro/desktop/en-us/publish_userfilters.htm. Accessed: 26.4.2020.

Tableau Software, LLC. 2003-2020b. Best Practices for Published Data Sources. URL: https://help.tableau.com/current/pro/desktop/en-us/publish_datasources_about.htm. Accessed: 11.4.2020.

Tableau Software LLC. 2003-2020c. Microsoft Analysis Services. URL: https://help.tableau.com/current/pro/desktop/en-us/examples_msas.htm. Accessed: 27.4.2020.

Tableau Software, LLC. 2003-2020d. Publish a Data Source. URL: https://help.tableau.com/current/pro/desktop/en-us/publish_datasources.htm. Accessed: 11.4.2020.

Tableau Software, LLC. 2003-2020e. Supported Connectors. URL: https://help.tableau.com/current/pro/desktop/en-us/exampleconnections_overview.htm. Accessed: 11.4.2020.

Tableau Software, LLC. 2003-2020f. Tableau Technology. URL: https://www.tableau.com/products/technology. Accessed: 27.4.2020.

Tableau Software, LLC. & A Salesforce Company. 2003-2020a. Pricing for data people. URL: https://www.tableau.com/pricing/teams-orgs. Accessed: 11.4.2020.

Tableau Software, LLC. & A Salesforce Company. 2003-2020b. R for Statistical Computing & Analysis. URL: https://www.tableau.com/solutions/r. Accessed: 27.4.2020.

Tableau Software, LLC. & A Salesforce Company. 2016. Tableau Expands Cloud Capability with Data Centre in Europe. URL: https://www.tableau.com/about/press-releases/2015/tableau-expands-cloud-capability-data-centre-europe. Accessed: 27.4.2020.

Tableau Software, LLC. & A Salesforce Company. 2019. Tabcmd for Generating a PDF of a Tableau View and Emailing. URL: https://kb.tableau.com/articles/howto/tabcmd-for-generating-a-pdf-and-emailing. Accessed: 11.4.2020.

TIBCO Software Inc. 2020. Business Intelligence Software Editions. URL: https://www.jaspersoft.com/editions. Accessed: 11.4.2020.

The Eclipse Edition 2014a. About. URL: https://www.eclipse.org/birt/about/. Accessed: 18.4.2020.

The Eclipse Edition 2014b. Demos. URL: https://www.eclipse.org/birt/demos/. Accessed: 18.4.2020.

The Eclipse Foundation 2014. Integrating BIRT. URL: https://www.eclipse.org/birt/documentation/integrating/reapi.php. Accessed: 28.4.2020.

Vidovic, G. 2016. Databox for Developers. URL: https://databox.com/databox-for-developers. Accessed: 26.4.2020.

Wolski, C. 2001-2020. How to create a SQL dashboard that pulls in data from your database. URL: https://www.klipfolio.com/blog/create-sql-dashboard. Accessed: 27.4.2020.

Woodall, Z. 2017. Save time with data-driven alerts in Tableau 10.3. URL: https://www.tableau.com/about/blog/2017/4/save-time-data-driven-alerts-tableau-103-67888. Accessed: 11.4.2020.

Zenko, J. Innovation Never Rests: Delivery API. URL: https://www.dundas.com/resources/blogs/innovation-never-rests-delivery-api. Accessed: 19.4.2020.

Ziemann, M. 2016. Vakuutusyhtiö haluaa tarkkailla ajamistasi – kiltti saisi kunnon alennuksen vakuutusmaksuista. URL: https://yle.fi/uutiset/3-9278541. Accessed: 09.12.2019.