Eveliina Nieminen

ANALYZING DATA WITH SAP BW/4HANA

Bike Sharing Application Project in SAP Next-Gen Lab

ABSTRACT

Centria University of Applied Sciences	Date December 2018	Author Eveliina Nieminen
Degree programme Business Management		
Name of thesis ANALYZING DATA WITH SAP BW/4HANA. Bike Sharing Application Project in SAP Next-Gen Lab		
Instructor Janne Peltoniemi	**Pages** 44	
Supervisor Janne Peltoniemi		

SAP Next-Gen Lab (NGL) in Centria started its activity on 8th of January, 2018. NGL is student-lead environment where students are able to do projects with SAP technologies. In January, the NGL activity started with overall eight students and the first project was given by University Competence Center Magdeburg.

This project was requiring SAP BW/4HANA- data warehousing system and analyzing tool SAP Lumira. The project was called "Bike Sharing Application- Project" and its aim was to use the data given by the bike sharing system NYC Bike Sharing Ltd, which is owned by the City of New York and Motivate. Motivate is the leading bike sharing system operator in the United States. The data from bike sharing system was combined with New York City's weather data in order to see how the weather impacts the bike trips made in New York City.

This thesis focused on the process of data gathering, bringing data to the SAP BW/4HANA system and data presentation in SAP Lumira. Besides that, this thesis described the difficulties and problematic situations of the process as well as was considering what could have been improved on this project process in future.

CONCEPT DEFINITIONS

aDSO   Data Store Object (advantaged)

BW   SAP Business Warehouse- system

BW/4HANA Business Warehouse for HANA

DTP   Data Transfer Process

ERP   Enterprise Resource Planning System

ETL   Extract, Transfer and Load -process

HPA   High-Performance Accelerator

NGL   Centria's SAP Next Gen Lab

OLAP   Online Analytical Process

OLTP   Online Transactional Process

SAP   Systems, Applications and Products in data Processing, multinational software corpora-
            tions

SAP Fiori   User Experience of SAP, designed especially for smart devices

SAP GUI   User Experience of SAP (old)

SAP HANA High performance ANalytical Appliance, database system offered by SAP

UCC   University Competence Center

ABSTRACT
CONCEPT DEFINITIONS
CONTENTS

FIGURES

GRAPHS

1. INTRODUCTION

Centria University of Applied Sciences announced with SAP SE to launch SAP Next-Gen Lab operations on Kokkola Campus on 12th of December, 2017. The idea of SAP Next-Gen Lab is to give students opportunity to work with SAP technologies and have a real-life project to study and research with the latest SAP technologies available. With six business students and two IT-students, the SAP Next-Gen Lab started its operations on 8th of January 2018.

This thesis is describing the operations of Bike Sharing Application project in Centria's SAP Next-Gen Lab, and focuses on work with business warehouse system, SAP BW/4HANA, and analyzing tool, SAP Lumira. The project was introduced by SAP UCC Magdeburg, and the idea of the project is to create SAP Fiori application. This application is predicting the bike rentals based on the weather forecast and history of the bike rentals. Therefore, the aim of the application is to help bike sharing systems to move the rental bikes to needed locations forehead before the weather changes. This is made based on the prediction.

Centria's SAP Next-Gen Lab was focusing more on customers' behavior of bike trips in the project. The most interesting aspects are to know where the trips were made the most, where the bikes are needed the most and how the weather is affecting it. In the project the case bike sharing system was New York City's NYC Bike Share, LCC (also known as Citi Bike). This bike sharing system was used as well in UCC Magdeburg's project. The weather forecast was gathered from APIXU. The project ended in August 2018.

In this thesis, the focus is on data gathering, data modeling and data presentation. The aim is as well to find which things in the project could have been done better and where the project had the most of the struggle. On other hand, this thesis is describing the process of bringing data to SAP BW/4HANA system until to make the data presentation to SAP Lumira, and its aim is to help the students and other people interested to work with SAP BW/4HANA system.

To be able to understand the process, it is important to understand the need of business warehouse, SAP HANA database and SAP BW/4HANA. In this thesis it is shortly descripted what SAP HANA database is, why business warehouse exists and why SAP BW/4HANA is different from previous business warehouse systems provided by SAP.

2. SAP SE

SAP SE is one of the leading multinational software corporations in the world. It was founded in 1972 in Germany and its headquarters are located in Walldorf, Germany. The term *SAP* stands for Systems, Applications and Products in Data Processing. SAP SE is well-known for its enterprise resource system (ERP) software and it is used by many largest companies globally. SAP SE has over 370 000 customers in 180 different countries and employs over 88 000 people. (Shalaby 2017, 1; SAP Integrated Report 2017; Walker 2012.)

2.1 About SAP Next-Gen Program

SAP Next-Gen program is an innovation community introduced and hosted by SAP. The innovation community's aim is to connect and innovate companies, partners and universities and work together with purpose of 17 UN Global Goals. These 17 UN Global Goals are following:

1) No Poverty

2) Zero Hunger

3) Good Health and Well-Being

4) Quality Education

5) Gender Equality

6) Clean Water and Sanitation

7) Affordable and Clean Energy

8) Decent Work and Economic Growth

9) Industry, Innovation and Infrastructure

10) Reduced Inequalities

11) Sustainable Cities and Communities

12) Responsible Consumption and Production

13) Climate Action

14) Life below Water

15) Life on Land

16) Peace, Justice and Strong Institution

17) Partnership for the Goals

SAP Next-Gen program offers different kind of services, which are for example innovation tours, boot camps and advisors. In other hand, SAP Next-Gen program has projects with academia. For universities and other educational institutions, SAP Next-Gen is an extension of SAP University Alliance (UA). SAP UA is program which is giving the latest SAP technologies to the educational institutions. These SAP technologies enables the educational institutions to use the technologies for practicing, teaching and researching purposes. It also gives possibility for the students from the educational institutes to practice with the latest SAP technologies available. (SAP Next Gen Innovation for Purpose 2018; The Global Goals for a Sustainable Development 2018.)

SAP UA is partnering with University Competence Centers (UCC) in order to be able to host the SAP software for educational institutions and to give support for teaching in areas of integrated business processes and digital innovation. There are globally six UCCs overall. Academic Competence Centers (ACC) are partnering with UCCs and SAP UA. ACCs are providing services with language translation and support for local curriculum and technical issues and they are targeted to educational institutions in certain regions. (SAP Next Gen Innovation for Purpose 2018.)

2.2 Description of SAP Next-Gen Lab's Activities in Centria UAS

On December 12th of 2017, Centria University of Applied Sciences and SAP SE made co-operation agreement of establishing the SAP Next Gen laboratory in Centria's Campus located in Kokkola. This is the first SAP Next Gen laboratory located in Northern Europe. Centria UAS is participating the SAP University of the Alliances and gets support and hosting from SAP UCC Magdeburg. In 10th of September 2018, it was announced by SAP SE that Centria UAS has got SAP Next-Gen Chapter- status. This means that Centria UAS will be supporting the other educational institutions in Finland with Next-Gen innovation. Centria UAS's expertise is focused on analytics and project management. (Centria-ammattikorkeakouluun Pohjois-Euroopan ensimmäinen Next-Gen Lab 2017; Centria is the first SAP Next-Gen Chapter in Northern Europe 2018)

Centria UAS is located in Central Ostrobothnia, in Finland. It has three campuses which are located in Kokkola, Pietarsaari and Ylivieska. Centria UAS educates around 3 000 students and offers to the students higher degree studies in technology, social- and healthcare and business. (Centria University of Applied Sciences – About Us 2018.)

Centria's SAP Next-Gen Lab (NGL) activities started in January 1st of 2018, including six business students and two IT-students and a teacher as an admin. The focus on the SAP software was on SAP's Business Warehouse solutions. Firstly, the team was using SAP Business Warehouse on HANA, but later on it was upgraded to SAP BW/4HANA. The NGL-team had support from SAP UCC Magdeburg by giving them study material for Business Warehouse and professional knowledge if needed. NGL team was using as well analyzing tools, such as SAP BusinessObjects, SAP Lumira or SAP Predictive Analysis. NGL is student-lead, which means that the students work towards the task given as a team, research and study for it. This can be done as an internship, optional courses or as a thesis work.

3. SAP HANA

As SAP's vision of in-memory computing, SAP HANA was launched in 2010 and is a database system which is designed to speed up database accesses. SAP HANA database offers more functionality compared to traditional databases and therefore there has been used also a term SAP HANA platform. SAP HANA has been developed into a platform to match the established databases in OLTP and OLAP and therefore SAP HANA platform is a logical platform for SAP's technology-based data warehouse implementations. (Walker 2012; Riesner & Sauer 2017.)

3.1 Background

SAP is one of the largest software companies and is having the largest companies in the world as customers. Most of the customers are using SAP's ERP systems which are designed to store all the information that businesses need for function normally. Therefore, ERP system is nowadays vital for companies.

Information, which is stored for businesses, is for example: companies' customer data, purchase- and sales orders, deliveries, financial accounting, warehouse management and so on. When incoming delivery is entered to ERP system, the system updates automatically the change of stock time. It means, that the data is in real-time and ERP system provides all the time current and fresh data. (Walker 2012.)

In 1997, SAP introduced SAP Business Information Warehouse (BIW) which was the first solution for data warehouse to the SAP ERP system provided by SAP. Business Warehouse (BW) system makes its user possible to report the data stored in ERP system and is capable to perform from simple analyses to complex simulations including multiple different data factors from the SAP ERP system. (Walker 2012.)

ERP systems stores data in the database, but instead of it, BW uses the data of the database to aggregate it for a presentation. This presentation is for reporting purposes and shows for example totals and trends to the user. BW systems have had traditionally performing difficulties to satisfy the BW users. These performing issues has been revealed for example when the data is modeled in BW it is needed to store initially to the staging area. From there the same data needs to be stored additionally in different layers. These layers differ depending on the way the data is stored and they are for example operational data

stores, stored data including business logics or aggregated data. Therefore, data must flow through different layers to reach the end-users for decision-making. The execution of reports takes lots of time depending on the structure and can be done from several minutes to hours. Therefore, is not very practical and the creation of repot might consume lots of time. (Walker 2012.)

There are other problems with traditional BW systems as well. One is that many times in BW system the data is stored repeatedly. Secondly, since the data loads are usually made once a day, the data is not up-to-date and the reports were presenting therefore the past rather than current situation. (Walker 2012.)

SAP HANA is SAP SE's answer to solve these problems mentioned above to satisfy the customers who were dissatisfied of the performance of SAP BW and SAP ERP systems together. For addition, the SAP software has been always relied on the third-party databases. When SAP HANA was introduced, there became an option to have a database offered by SAP. (Walker 2012; Mankala & Mahadevan 2013.)

3.2 Architecture of SAP HANA Database

Applications for company needs are having higher requirements all the time since companies are demanding more capabilities. Complex reports are needed and in the same time the transactional data, which is used in the reports, are updated and read by many users. SAP HANA's main goal is to provide a main-memory centric data management platform which is supporting traditional applications but as well is having a powerful integration model for SAP applications. Therefore, it is supporting a pure SQL (Structured Query Language) and it is designed to be common database for OLTP and OLAP -systems. SAP provides HANA Studio to the client side. SAP HANA Studio is a central development environment and an administration tool for SAP HANA database. (Dees, Färber, Grosse, Lehrer, May, Müller & Rauhe 2012; Mankala et al 2013; Walker 2012.)
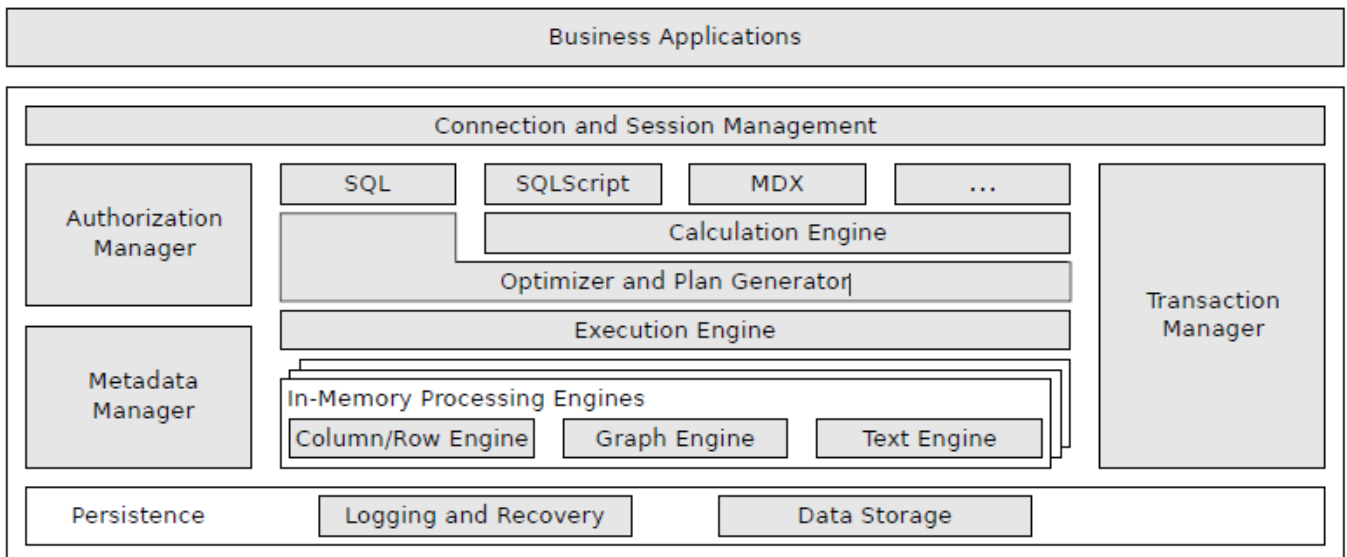
GRAPH 1. SAP HANA Architecture (Dees et al. 2012.)

In the graph 1 is shown the architecture of SAP HANA database. When the graph is looked more closely, Business Applications are presenting any application business is using, such as SAP ERP or SAP BW/4 HANA. The heart of the database is in in-memory processing engines. They are storing the relational data in tables, which can be either column- or row layout. These are combined in column- and row engines which can convert the tables to row- or column layout if necessary. The data which contains graph data is in graph engine and the data containing the text data is in text engine. Because of the extensible architecture, it is possible to add more engines if needed. These engines keep all the data coming from applications in-memory as long as there is enough space. By using compression schemes, the data can be compressed. If the space is limited, with control, the data is unloaded from main memory for example to secondary database. When the data is needed again, the in-memory processing engines does load the data back automatically. The persistent layer is storing the same data as the main memory, and its' duty is to back-up and recover the data if it is lost in main memory. The persistent layer updates all the logs and last committed states as well. (Dees et al. 2012.)

SAP HANA database provides multiple interfaces, such as standard SQL, SQL Script and MDX. SQL queries, which includes either the incoming or out coming data, are translated to the execution plan in Plan Generator. After the execution, the queries go via Optimizer. From there the queries are executed by Execution Engine to the in-memory processing engines. When the queries are coming from other interfaces, the queries go through complex and abstract data flows in Calculation engine before they are translated to the execution plan in Plan Generator. Otherwise, the queries go same path as SQL queries.

The role of the Session Manager is to manage the queries executed and as well to control the individual connections between database and applications. Authorization Manager is controlling the user permissions and Transaction Manager implements the isolation snapshots or weaker isolation levels. Unlike Transaction Manager, the Metadata manager consists of local and global parts of distribution. It is repository of data and descripts the tables and other data structures. (Dees et al. 2012.)

3.2.1 Column- versus Row-based Formatting

One of the improvements of SAP HANA database is the capability of storing the data in column-based format. In traditional databases, the data is only stored in row-based format. As in the graph shown (GRAPH 2), in the row-based format, every row is stored as a unique, own row. The engine reads the data from every row which makes it slow process and the reporting is naturally slower then as well. (Walker 2012.)

```
1,Joe,Smith,35000,EUR
2,Emma,Thomson,40000,USD
3,Sam,Wiggins,42500,USD
```

GRAPH 2. Row-based Formatted Data (Walker 2012.)

When column-based format is used to store data in the database, the database is able to read quicker the data than in the situation when it is reading the data stored with row-based format. This is a very important element when the data is reported (for example by using SAP BW system) since the results are given to the user faster than when the row-based formatting is used. Example of column-based formatted data is seen in graph 3. (Walker 2012.)

```
1,2,3
Joe,Emma,Sam
Smith,Thomson,Wiggins
35000,40000,42500
EUR,USD,USD
```

GRAPH 3. Column-based Formatted Data (Walker 2012.)

3.2.2 Data Compression

Another new feature to SAP HANA database is the data compression. Instead of storing the data multiple times by repeating values, the idea of data compression is to mention the value only once alongside the occurrence of the value. For example, instead of repeating the currency in the last row in graph 3 (EUR, USD, USD), it can be expressed as *EUR, 2:USD* and therefore the data is compressed.

With large companies, lots of new data is needed to be stored and memory is consumed. When the amount of data stored is huge, storing data with data compression saves lots of space of the memory and money as well. According to SAP, the data compression makes the tables stored between 10 per cent and 25 per cent smaller than its' original size. Therefore, it is possible to keep data in memory at once from 4 to 10 times more than with traditional way of storing data. (Walker 2012.)

3.3 SAP HANA Studio

SAP HANA Studio is designed to be a central development environment and an administration tool for SAP HANA database. The HANA Studio is developed in Java programming language and is running on Eclipse Platform. The HANA Studio is interacting with the HANA database by using SQL. Therefore, HANA Studio is designed to work on development activities of the HANA database and is designed to be used by developers, modelers and technical users. (Mankala et al. 2013.)
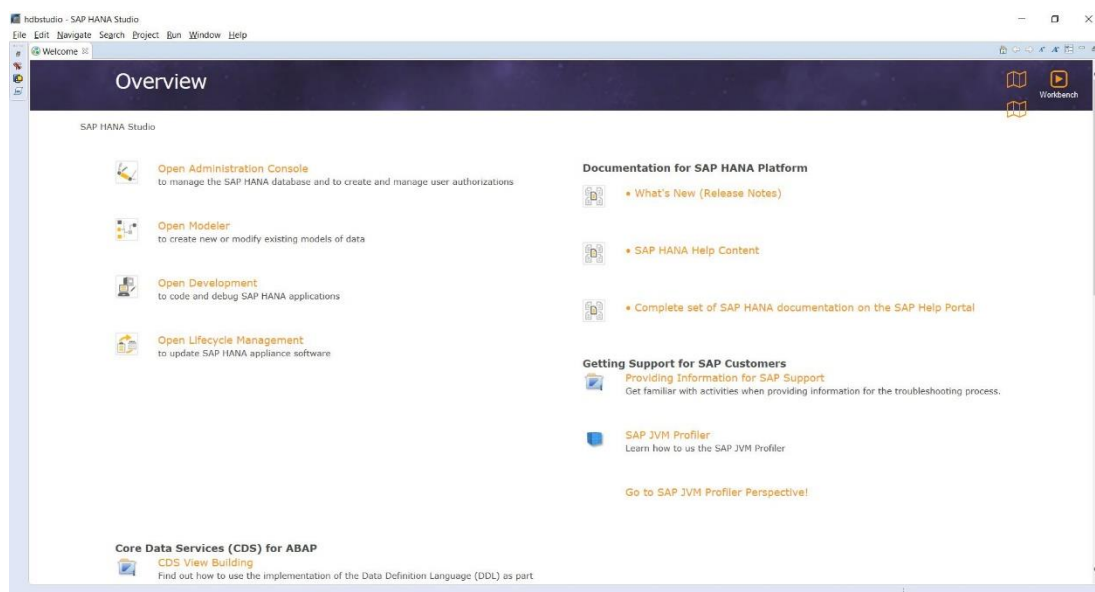
FIGURE 1. Overview of SAP HANA Studio (Centria University of Applied Sciences 2018.)

SAP HANA Studio has different *perspectives* for different tasks. For administrators, it is a tool to manage administrative activities, users and authorizations but as well to monitor systems and to configure the settings of the systems. Developers are using SAP HANA Studio to create contents, which are for example information views or stored procedures. These contents are stored in SAP HANA repository. (Mankala et al. 2013.)
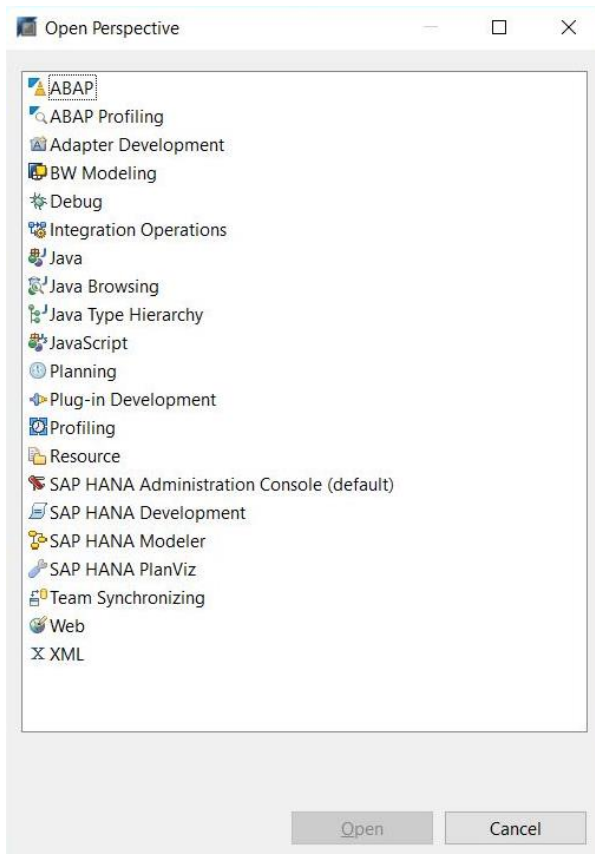
FIGURE 2. Perspectives of HANA Studio (Centria University of Applied Sciences 2018.)

For SAP HANA appliance software, HANA Studio is a collection of applications. It can be interface between HANA database and reporting layer or it can be interface between the database and presentation layer. SAP BW/4 HANA is running fully inside of HANA Studio instead of SAP GUI. The previous business warehouse applications were running on SAP GUI or partly in SAP HANA Studio. HANA Studio is not depending on the location of the database, the database can be in the same environment, but it can be located remotely as well. (Mankala et al. 2013.)

4. SAP BUSINESS WAREHOUSE

In the 1990s, SAP's ERP system R/3-system was having troubles. The information systems, which had the duty of storing the information gathered from the transactions of ERP systems, reached their limits. This led to the separation of the OLTP and OLAP systems and to the creation of the first Business Information Warehouse (BIW) in 1997. In the same year, BIW was published as well. (Riesner et al. 2017.)

4.1 OLTP versus OLAP system

To understand the need of Business Warehouse, it is important to understand Online Transactional Process (OLTP) and Online Analytical Process (OLAP) systems. OLTP system, which is for example ERP system, need to be fast and running lean, when OLAP system has as much data as possible to make as accurate analytics as possible. For example, Business Warehouse system is OLAP system. OLAP systems as well can have different data sources, since usually in enterprise the data is spread to many places located in different platforms. In analytical point of view, it is useful to have external data sources and therefore it is naturally one capability of OLAP systems.

A big difference to OLTP systems is that OLAP systems need to have more computation the data, and single value is not so important than the volume of the data. Therefore, it is understandable, that OLAP systems are needing complex queries when OLTP systems are using very simple queries of data to deliver the data fast to the user.

OLTP systems are available usually to lots of the users, which are for example, customers, suppliers or employees, but OLAP systems are used by small number of skilled users. It is important for the OLTP systems to be current and valid, and the data is changing as for example, new sales orders are entered to the system, but when OLAP systems are used, traditionally, the data is stored first and then, without having changes meanwhile, the data is modified to have an analysis from it. (Celko 2006.)

4.2 Timeline of SAP Business Warehouse

All the basic concepts of business warehouse were born when the first Business Information Warehouse was introduced in 1997. These basic concepts are still visible on the newest business warehouse systems. For example, Info Objects with their characteristics and key figures are still in BW/4HANA and are the base pillars of the system. Info Cubes were having a role of managing transaction data with extended star schema. In the first business warehouse system, data was loaded either from flat files or from SAP systems. (Riesner et al. 2017.)

In 2001, the Business Information Warehouse was shortened to Business Warehouse (BW). In 2004, the Business Warehouse was integrated with NetWeaver, and the name was changed to NetWeaver 2004 Business Intelligence (BI). Later on, SAP released High-Performance Accelerator (HPA) which was designed to load the data by in-memory structures. It enabled to run the reports faster than before and solved problem of the previous releases which was reports running long. (Riesner et al. 2017.)

The huge step forward, after long time no new improvements since HPA, was in 2011, November. SAP released Business Warehouse 7.3 on HANA which meant simply that it was possible to have SAP HANA as a database for Business Warehouse. This database speeds up the data load to the report faster than traditional database could have do. Therefore, it was more flexible than previous releases. Smaller innovations were released as well in data modeling, data staging, administration and monitoring. (Riesner et al. 2017.)
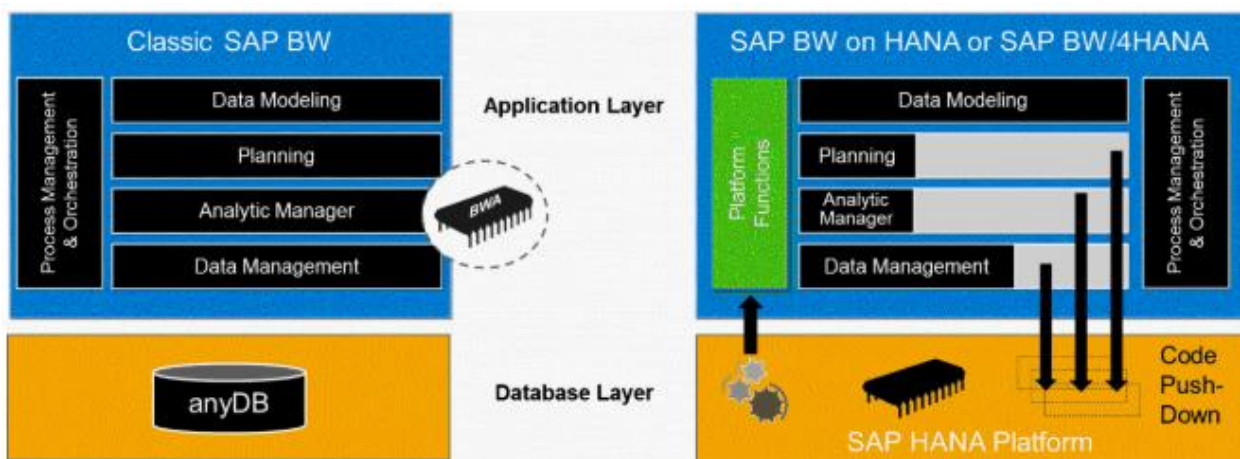

GRAPH 4. Classic Business Warehouse vs SAP Business Warehouse on HANA or SAP BW/4 HANA (Riesner et al. 2017.)

Compared the Classic BW, BW on HANA and BW/4HANA are having functions more in database layer than application layer (GRAPH 4). This is called code push-down. It means, that instead of going through Data management, Analytical Manager or Planning in application layer, which is taking lots of time to prepare, it is already done in database, and therefore is saving time of making actual reports out of the data. This makes the data processing efficient. With the Business Warehouse 7.4 on SAP HANA, the work with Business Warehouse came simpler and more flexible. The Modeling tools were shifted from SAP GUI to SAP HANA Studio as well. (Riesner et al. 2017.)

4.3 SAP BW/4HANA

SAP BW/4HANA was presented publicly on September 7th of 2016 in San Francisco as a new product of SAP. To simplify the application, BW/4 HANA is focusing on one database technology which was unlike the previous BW products. To achieve one database technology, lots of code from previous BW products was left behind. That is why the old BW objects and Business Explorer- tools have been disappeared in BW/4 HANA.

Therefore, SAP BW/4HANA runs only with SAP HANA database and is now based on a lean ABAP application server. It has modeling interface fully in HANA Studio, as it was partly already in Business Warehouse 7.4 on HANA. In the beginning, SAP BW/4 HANA was technically very similar to BW 7.4 but later on, new functionality has been delivered from SAP with the Support Packages. SAP BW/4 HANA is built for the Cloud as well. In the graph 5, the overview of the SAP BW/4 HANA is shown. The interaction between other applications and the database is presented in the graph as well. (SAP Blogs 2016; Riesner et al. 2017.)
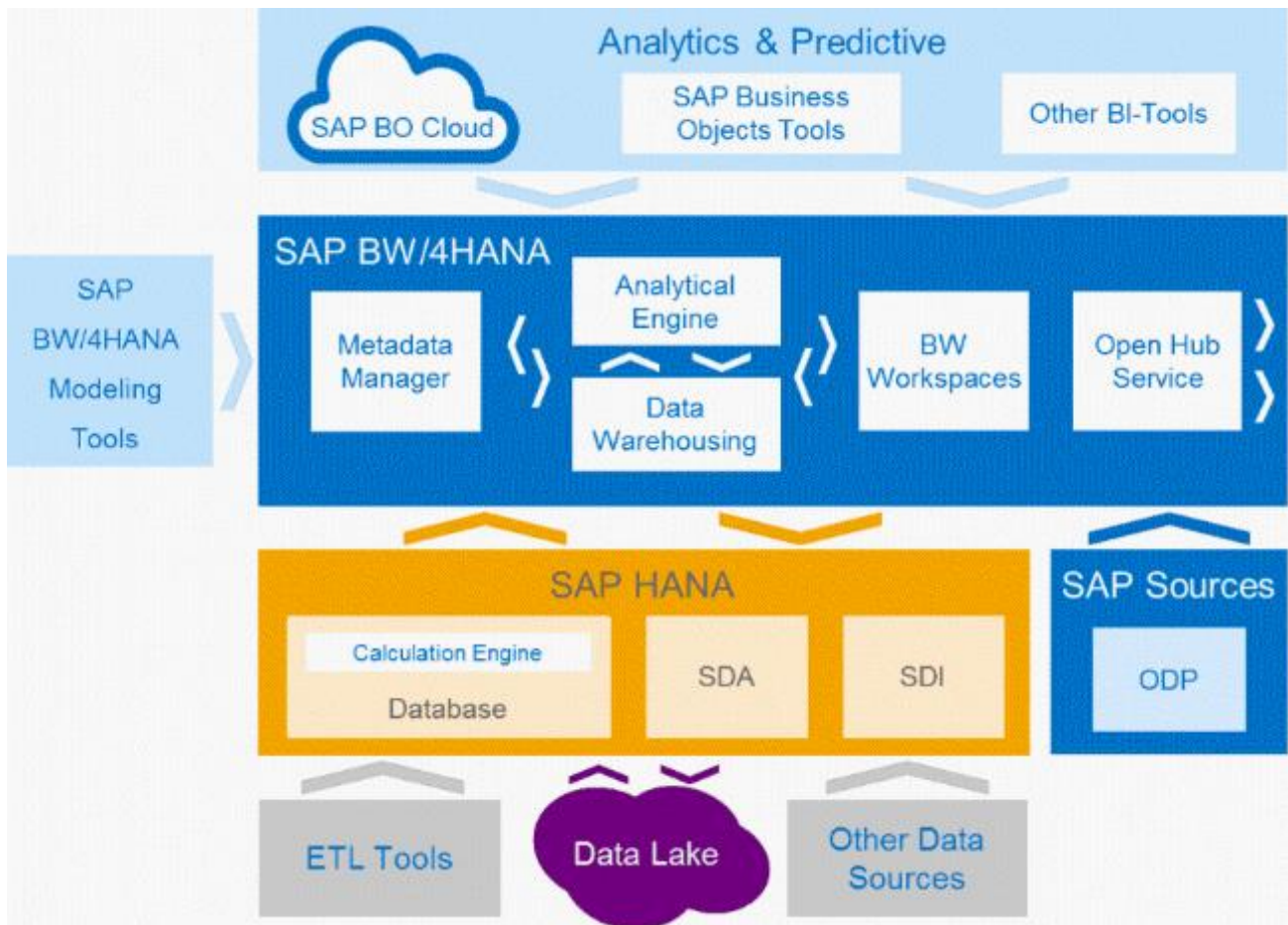
GRAPH 5. Overview of SAP BW/4 HANA (Riesner et al. 2017.)

4.4 SAP HANA-Specific BW Objects

Since lots of codes were left behind from previous releases in SAP BW/4HANA, the new business warehouse objects were only available in SAP BW/4HANA. These BW objects are halved from original objects, in previous releases they were eight but in SAP BW/4HANA there is only four BW objects left. They are also grouped to two categories: virtual objects and persistent objects. The change of the BW objects is presented in graph 6. (Riesner et al. 2017.)
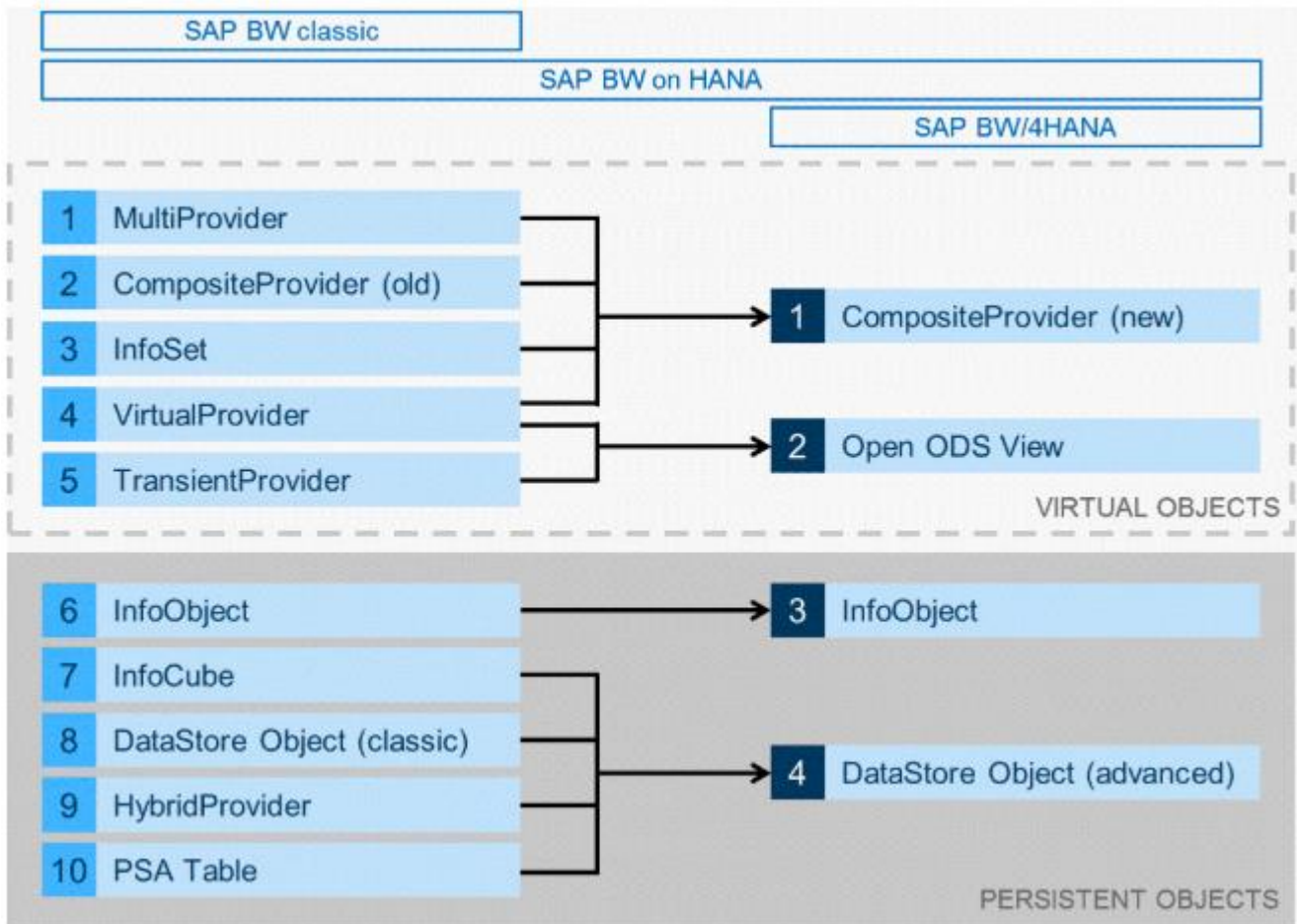
GRAPH 6. SAP BW/4HANA Objects (Riesner et al. 2017.)

Data Store Object (advantaged) (aDSO) is central data model which has duty of storing and managing transaction data. It was already introduced in BW 7.4 in Support Package 9, and from SAP BW/4HANA, it is replacing Info Cube, classis Data Store Object, Hybrid Provider and PSA Table. aDSO is only possible to define in SAP HANA Studio. The aDSO consist of InfoObjects or fields or the combination of InfoObjects and the fields. (Riesner et al. 2017.)

With aDSO, it is possible to create change logs and manage data via load request in a write-optimized way. It has also interfaces for direct writing. An aDSO is a large fact table which cannot overwrite the data but add more needed. The new data records will be updated when the selection of "Active Data" in Modeling Properties is ticked. It also gives more modeling possibilities following:

- **Write Change Log**: If there is new or changed data records, then they will be entered to change log. This makes the delta extraction possible, which means that only the updated data will be loaded.

- **Keep Inbound Data**, Extract from Inbound Table: Data extraction is possible to do fully or in delta option from inbound table, but changes are not written in change log.

- **Unique Data Records**: If the data records will be only new but there is no changes in existed data surely, then this option can be used.

- **Snapshot Support**: If the Data Source, where the data is coming from, is non-delta-compatible, the Snapshot Support is able to upload the transactions which are already deleted. When this is activated, it recognizes the records existing already and recognizes, when the records are missing new load request and handles them accordingly.

In aDSO it is possible to define planning applications as well and it is possible to implement a logical partitioning of large data quantities when semantic group is included. The aDSO consist always three tables which are inbound table, table of active data and change log table. Even though these tables would not be used actively, they are always created in aDSO when it is created. (Riesner et al. 2017.)

Info Object is a persistent object, which means that the data are loaded and stored via staging. They are used as a business evaluation objects and are the smallest units in Business Warehouse. Info Objects are divided into characteristics, key figures, units, time characteristics and technical characteristics. (SAP Help Portal – SAP BW/4HANA 2018.)

Characteristics are keys for sorting purposes, such as company code, product, region and customer group. They are meant to be to specify the classification of datasets and are the reference objects to the key figures. For example, a characteristic can be identifying the product name, and key figure is showing amount of the products sold. Time characteristics are representing for example date or fiscal year. Technical characteristics are only for administrative purposes in business warehouse, and it can be for example a request number and is for helping to locate the request if it is needed later.

Attributes, texts and hierarchies are known as data-bearing characteristics and can be included in characteristic. If Master Data, which contains data unchanged over long period of time, is needed, it can be referred to all Info Providers in BW. Attributes and hierarchies gives structure if the values of characteristics. For example, the location of company is an attribute of customer. It is possible to make a hierarchy for the customer characteristic to make its structure clearer. Hierarchies are always made for the characteristics. (SAP Help Portal – SAP BW/4HANA 2018.)

Key figures are the values which are reported in the query. These can be for example amount, quantity or number of items. Key figures are assigned to additional properties which have an effect on the data

loading and to the display of the query. These are currencies, unit of measures, setting aggregations and specifying number of decimals. For example, with quantity -key figure it is needed to have unit of measure, to specify are the quantities in pieces, kilos, liters and so on. The Info Objects have the capability to be Info Providers or they can be used in Info Provider. (SAP Help Portal – SAP BW/4HANA 2018.)

To integrate data in to the BW system flexible, the Open ODS View is needed. To have Open ODS View, it is required to have SAP HANA as a database. In Open ODS View, the data is not needed to convert to Info Objects, but it can be remained in the original format which is a field-based format. Open ODS View gives virtual access to BW and this makes data possible to be integrated without physical replication. Anyway, if there is a need to store the data physically, aDSO can be made. It is possible as well to model the data extendedly with Composite Provider and the data from Open ODS View is directly available to the reporting purposes. (Riesner et al. 2017.)

Open ODS View consist of field structures and the data types of the fields. It is a view of the data source and this view is enriched with analytical metadata. It is possible to map with Open ODS View persistent tables and database views of BW system's HANA database, data sources from other databases by using virtual SAP HANA tables. These tables are possible to have via SAP HANA Smart Data Access (SDA) and SAP HANA Smart Data Integration (SDI). If the BW data sources has property "Direct access", it can be mapped in Open ODS View as well. Open ODS View does not provide information about database connections or relevant database schemas but this information is defined in the properties of SAP HANA source systems.

It is possible to define the data to transactional data, master data or texts. To activate the data model, it is needed to define characteristics and key figures of the data and to define the needed unit of measures, which are such as currencies and units. If needed, the data can be associated to an Info Object. In BW/4HANA and BW 7.5 Open ODS View is only possible to do in SAP HANA Studio, when previous releases allowed to use it via SAP GUI (with the transaction code RSODSVIEW). Open ODS View is a good option to have compared to Composite Provider when the data is in raw format, data is master data or texts, the data is integrated via SDI or SDA from external sources or it is foreseeable. (Riesner et al. 2017.)

Composite Provider in BW/4 HANA and BW 7.4 SP5 is current central, virtual data model. Its duty is to merge data from Info Providers (for example aDSO, Info Object) by using SQL operations: SQL UNION or SQL JOIN. Composite Provider makes the data, which is combined from multiple sources,

available to the reporting. Possibilities of Composite Provider are to merge with BW Info Providers and data from SAP HANA models or make virtual data models which are using only BW data.

If Composite Provider has set as "This Composite Provider can be added to another Composite Provider" it is possible to nest the Composite Providers by one level. This means that the Composite Provider serves as Part Provider to the Composite Provider it is added on. (Riesner et al.2017.)

5. BIKE SHARING APPLICATION PROJECT

Centria's Next-Gen Lab team got its first project which was introduced by SAP UCC Magdeburg in January 2018. The goal of the project was to build a Fiori application for bike sharing system. The idea of the bike sharing system is to give opportunity to customers to take a bicycle from a bike station and return in certain time passed to another bike station. The bike stations are located in different areas of the city where the bike station is used. This data in this project was collected and modeled by using BW 4/HANA and analyzed by using SAP Lumira.

In this project, the case company was NYC Bike Share LLC and the location of the bike share system is in New York City, United States. NYC Bike Share LCC provides its data and it is consisting of how long trips are, from where they are started and to where the bikes are returned in New York City. (Citi Bike – System Data 2018.)

5.1 Citi Bike

Citi Bike, launched in May 2013, is bike share system located in New York City and offered by NYC Bike Share LCC. It is the largest bike share system in United States and has giving an alternative option for traditional public transportation in New York City. It is recognized as healthier and more sustainable option for the environment as well. (Citi Bike – About 2018.)

NYC Bike Share, LCC is a subsidiary of Motivate and has a public-private partnership between Motivate and City of New York (NYC DOT). Motivate is a private company which owns the bike share systems and operates them in several locations in US. Motivate is leading company globally in bike sharing and it manages currently all the bike sharing systems located in big cities in United States. Citi Bike has sponsors with Citi Bank (a title sponsorship) and MasterCard as a preferred Payment Partner. (New York Bike Share 2018; Citi Bike – About 2018; Bloomberg 2018.)

Citi Bike offers either one-time use ride or annual membership. There are over 750 bike stations across Manhattan, Brooklyn, Queens and Jersey City including over 12 000 bikes overall. For one-time user, there is possibility to ride 30 minutes without extra fee and for the annual member, the ride without extra fee is 45 minutes. It means, that the Citi Bikes are targeted especially for short rides, such as for tourists

to go around the city but as well for daily needs to go to work for the citizens. The bike sharing system operates in New York City every hour of the day and every day of the year. (Bike Rental NYC – Citi Bike 2018.)

5.2 Aim of the Project

The aim of the project is to educate students to learn to use SAP BW4/HANA and various analyzing tools offered by SAP. The project gives experience for the students how to work in the real project environment. As well, the aim of the project is to offer NYC Bike Share valuable information about the Citi Bike such as:

1. In which time the bikes are used the most?
2. In which weather the bikes are used the most?
3. In which locations and stations bikes are used the most or the least?
4. Is there more demand for bike stations in certain areas?
5. Is there less demand for bike stations in certain areas?

There is possibility to see the customer behavior of the rentals, such as gender, age, if the customer is a subscriber or a one-time user. Also, there is possibility to see the differences of customer behavior in different locations in the city if there is any. Based on the results, the project's aim is to give valuable information to improve the efficiency of the bike sharing system in New York City, supply the demand of the bikes in certain areas, and reduce the number of bikes available in locations which are not used so frequently.

6. PROCESS OF ANALYZING DATA

Centria's SAP Next Gen Lab started The Bike Sharing Application -project on 8th of January 2018 with six business students. The project was introduced by UCC Magdeburg, and it was introduced with 10 modules overall. Later on, next month, there were two IT students taking part in the project. With assistance of admin teacher, the project team started working with SAP Business Warehouse on HANA and on April 2018, the team got chance to use SAP BW/4 HANA in their project. Project continued with three students after May 2018 and they were able to accomplish the project in end of August 2018.

Centria's SAP Next-Gen lab is continuing the project still in autumn 2018, when there will be full capabilities of SAP BW/4 HANA, such as administration capabilities. Naturally, without these capabilities the project was limited in certain areas. This thesis is focusing on the project made from beginning, until end of August 2018, covering shortly about the data acquisition, data modelling and data presentation.

6.1 Target of the Process

UCC Magdeburg introduced Centria's NGL the Bike Sharing Application- project which was including 10 modules. Those modules are possible to see in the graph 7. In the graph, the data is flowing to SAP HANA database through GBFS (General Bikes Feed Specification) API and Weather API. Because of the possibilities of SAP HANA database, the data could be updated to the system in real-time. In SAP BW/4HANA side, the data is modeled and aggregated (in graph 7: Station Data, Current Weather and Weather Forecast). These aggregated data tables can be combined. These tables, containing bike station data from NYC Bike Share LCC and weather data from the source wanted (for example APIXU), can show together which kind of effects there is in usage of the bikes when the weather changes. From weather forecast and historical data of current weather and bike stations' status, the machine learning module can be done. This machine learning module is making predictive analysis, and therefore it is possible to see how the usage of bikes is predicted when the weather forecast is already known.

This kind of prediction is beneficial for the bike sharing system companies to react faster if the weather changes on different parts of city. The problem of nowadays is, that there might be in one location having lots of bikes, even if the usage is little. The other parts of the city might need the bikes more in the same

time. When the changes are finally made, it is usually already too late since packing the bikes to vehicle and transport them takes lots of time.

The presentation of the reports can be done on the presentation tool wanted, such as data analyzing tool, SAP Lumira. When it is wanted, the presentation can be shown as well in SAP Fiori Application (Fiori App), which is a new user experience of SAP software. Fiori App can be used in computers, but as well in tables and phones, which makes the experience more user-friendly when compared the traditional user experience SAP GUI from SAP. (SAP – Products / SAP Fiori 2018; Wegener & Böttcher 2018.)
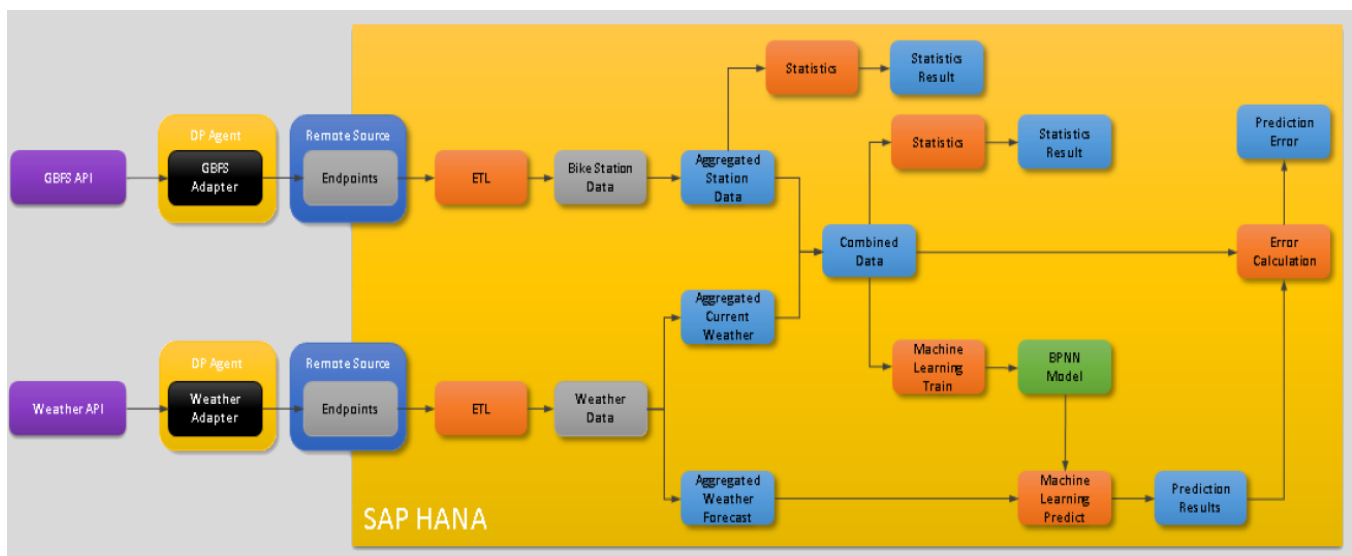
GRAPH 7. Overview of Bike Sharing Project by SAP UCC Magdeburg (Wegener et al. 2018.)

On 31th of August 2018, the Bike Sharing Project in Centria UAS was done with the limitations of the current system. Instead of the bike station data, the project is containing trip data of bikes. In graph 8 it is possible to see what fields the trip data is consisting. The data is historical data and is updated once in the month which telling the behavior of trips monthly. Trip specific data fields are following: trip duration, start time, start date, end time and end date of the bicycle trip the customer has made. The bike which has travelled can be located with Bike ID. There is customer specific data such as customer type (subscriber or one-time user), gender and birth year. Top of all, there is information about which station the bike has been taken and to where it has been returned. (Citi Bike – System Data 2018.)
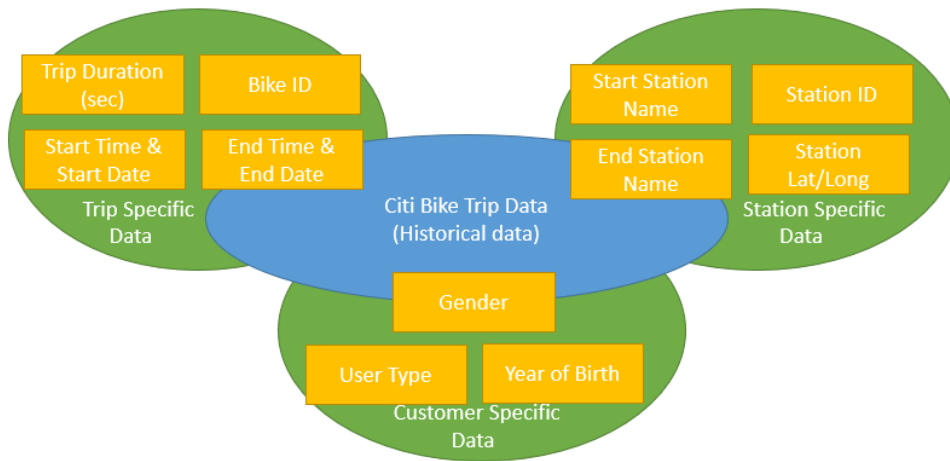
GRAPH 8. Trip data

In graph 9 is presented the project overview from August 2018. Before loading the data to the system, some data preparation is needed. With help of the data engineer of the team, the data can be prepared. The raw data feed from GBFS-feed (General Bikes Feed Specification) is in json -programming language format. By using Python -programming language, the data engineer of the team has been able to make the data into csv-files. The csv-files are loaded to SAP BW/4 HANA system as flat files. In there, the extract, transfer and loading (ETL) process is done, as well as modeling to aggregate the data in the form which can satisfy the needs of reporting. These data sets are combined, and data presentation is made with SAP Lumira. If wanted, the presentation can be presented in Fiori App. These steps are done by data scientists of the SAP Next-Gen Lab team. (Wegener et al. 2018.)

GRAPH 9. Overview of the Bike Sharing Project by Centria

6.2 Data Acquisition

To get the reports from the data, the data needs to be gathered, prepared and then to be loaded to the system. In this project the data preparation is made by the data engineer of the team, but to be able to do it, the student needs to listen the requirements of the report carefully. It requires thigh communication with the team and understanding the goal of the project. In the project, the data was gathered from Citi Bike's System Data and the weather data from APIXU.

Since the original source is in json -format (FIGURE 3), to get the data gathered to csv-files, the data engineer must have skills required to do that. In the team, the data engineer used Python -programming language to solve this issue. The data engineer was able to do code, which was calling the source to give information, and with control, it was possible to gather and organize the data in wanted format in CSV (FIGURE 4).

```
{"last_updated":1538333901,"ttl":10,"data":{"stations":
[{"station_id":"281","num_bikes_available":22,"num_ebikes_available":0,"num_bikes_disabled":2,"num_docks_available
:33,"num_docks_disabled":0,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1538333897,"eightd_has
available_keys":true,"eightd_active_station_services":[{"id":"32461582-cd1e-4ecf-a5ea-563593fa7009"}]},
{"station_id":"72","num_bikes_available":2,"num_ebikes_available":0,"num_bikes_disabled":1,"num_docks_available":3
,"num_docks_disabled":1,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1538333864,"eightd_has_av
ilable_keys":false},
{"station_id":"79","num_bikes_available":13,"num_ebikes_available":0,"num_bikes_disabled":2,"num_docks_available":
8,"num_docks_disabled":0,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1538333231,"eightd_has_a
ailable_keys":false},
{"station_id":"82","num_bikes_available":8,"num_ebikes_available":0,"num_bikes_disabled":0,"num_docks_available":1
,"num_docks_disabled":0,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1538333731,"eightd_has_av
ilable_keys":false},
{"station_id":"83","num_bikes_available":41,"num_ebikes_available":0,"num_bikes_disabled":3,"num_docks_available":
8,"num_docks_disabled":0,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1538333879,"eightd_has_a
ailable_keys":false},
{"station_id":"119","num_bikes_available":12,"num_ebikes_available":0,"num_bikes_disabled":1,"num_docks_available"
6,"num_docks_disabled":0,"is_installed":1,"is_renting":1,"is_returning":1,"last_reported":1538321412,"eightd_has_a
```

FIGURE 3. Data from the data source in json format (Citi Bike – System Data 2018.)

```
ROWID;TRIPDURATION;STARTDATE;STARTTIME;STARTHOUR;STARTSTATIONID;ENDSTATIONID;BIKEID;USERTYPE;BIRTHYEAR;GENDER;COMB_WITH_WEATHER
000000000000004CDC82;693;2013-06-01;00:00:08;00:00:00;444;434;16649;Subscriber;1984;Male;E28
000000000000004CDC83;2059;2013-06-01;00:00:44;00:00:00;406;406;19599;Customer;NULL;Unknown;E28
000000000000004CDC84;123;2013-06-01;00:01:04;00:00:00;475;262;16352;Subscriber;1960;Male;E28
000000000000004CDC85;1521;2013-06-01;00:01:22;00:00:00;2008;310;15567;Subscriber;1983;Male;E28
000000000000004CDC86;2028;2013-06-01;00:01:47;00:00:00;485;406;18445;Customer;NULL;Unknown;E28
000000000000004CDC87;2057;2013-06-01;00:02:33;00:00:00;285;532;15693;Subscriber;1991;Male;E28
000000000000004CDC88;369;2013-06-01;00:03:29;00:00:00;509;521;16100;Subscriber;1981;Male;E28
000000000000004CDC89;1829;2013-06-01;00:03:47;00:00:00;265;436;15234;Subscriber;1984;Male;E28
000000000000004CDC8A;829;2013-06-01;00:04:22;00:00:00;404;303;16400;Subscriber;1987;Male;E28
000000000000004CDC8B;1316;2013-06-01;00:04:28;00:00:00;423;314;19781;Subscriber;1960;Male;E28
000000000000004CDC8C;1456;2013-06-01;00:04:41;00:00:00;502;532;18886;Customer;NULL;Unknown;E28
000000000000004CDC8D;386;2013-06-01;00:05:13;00:00:00;241;365;19039;Subscriber;1981;Male;E28
000000000000004CDC8E;924;2013-06-01;00:05:21;00:00:00;486;521;16608;Customer;NULL;Unknown;E28
000000000000004CDC8F;1233;2013-06-01;00:06:44;00:00:00;527;296;14761;Subscriber;1987;Male;E28
000000000000004CDC90;512;2013-06-01;00:03:36;00:00:00;309;300;19080;Subscriber;1979;Female;E28
000000000000004CDC91;505;2013-06-01;00:03:45;00:00:00;309;347;16798;Subscriber;1984;Male;E28
000000000000004CDC92;833;2013-06-01;00:07:29;00:00:00;503;503;19072;Customer;NULL;Unknown;E28
000000000000004CDC93;1818;2013-06-01;00:08:10;00:00:00;257;500;20349;Customer;NULL;Unknown;E28
```

FIGURE 4. Data gathered and organized to CSV-file (Centria University of Applied Sciences 2018.)

6.2.1 Creating Data Source

When Data Source is created, with transactional data is used "Data Source Type: Transactional", in the case is used only flat files and therefore, as seen in figure 5, when fields in "Extraction" -tab is filled, direct access is not allowed and the settings in adapter is set as "Load Text-Type File from Local Work-station"

Since in the project there is used CSV-files, in Data Format-field the settings for CSV- file is needed to set correct. The right file is chosen in "Extractor-Specific Properties" and set up the rows need to be ignored from header. If it is not set, the system reads the header row which are trip date, trip time, trip duration for example as a value rather than descriptive information.

After the Extraction- field is set correct, the preview is possible to see in "Fields" -tab. The Data Source is first level how to model the incoming data to the system. To be able to model the data correctly, the system makes a copy of the data which can be modeled. The data will flow through data models to its target when all the modeling properties and target has been created and set correctly. If there is new data from the source, it can be loaded with choosing "Delta Process". In this case data is loaded fully only.

FIGURE 5. Properties of Extraction when Creating a Data Source (Centria University of Applied Sciences 2018.)

6.2.2 Data Fields

To make the extract, transform and load process successful, it is important to understand the data fields of the source data. These data fields need to be studied in order to be able to make corresponding settings to the corresponding Info Providers. If there is any mistake done in data fields, more likely an error will occur.

In figure 6 it is seen, that when the flat file is loaded to the system, it automatically categorizes the types of data. When data types are set, it is important to understand to where the data is used. For example, in the project the TRIPDURATION is a key figure, but its' data type is INT8 (range of values are -9,223,372,036,854,775,807 to 9,223,372,036,854,775,807) since the trip duration is in seconds and the amount can be quite high, especially if counted together.

On other hand, the length of the data type is important as well and should be thought closely which length as its' maximum value should be allowed. When figure 6 is viewed, GENDER is known that is maximum 10 values long. If the length of any data on that field is longer than specified, the error of overlapping data can be caused when the data is loaded to the system.

Field	Key	T...	Data Type	Len...	Field Type
[ROW_ID] ROW_ID	🔑	☑	CHAR	20	
[TRIPDURATION] TRIPDURATIC		☑	INT8	19	
[STARTDATE] STARTDATE		☑	DATS	8	
[STARTTIME] STARTTIME		☑	TIMS	6	
[STARTHOUR] STARTHOUR		☑	TIMS	6	
[STARTSTATIONID] STARTSTATI		☑	CHAR	10	
[ENDSTATIONID] ENDSTATION		☑	CHAR	10	
[BIKEID] BIKEID		☑	CHAR	10	
[USERTYPE] USERTYPE		☑	CHAR	10	
[BIRTHYEAR] BIRTHYEAR		☑	CHAR	10	
[GENDER] GENDER		☑	CHAR	10	
[COMB_WITH_WEATHER] COM		☑	CHAR	10	

FIGURE 6. Data Fields of Trip Data (Centria University of Applied Sciences 2018.)

6.2.3 Master Data Source

All the data which is available are not always changing all the time. This kind of data can be categorized as a Master Data. For reporting purposes, this Master Data is usually needed to load only once to the target system. In this project, Station information (bike station related data) was decided to categorize as a Master Data. Since Info Object can be Info Provider, in this project it was decided to make an Info Object which was providing the data from Station Information. To be able to provide that data, the Info Object needs to contain attributes. (SAP Help Portal – SAP BW/4HANA 2018)
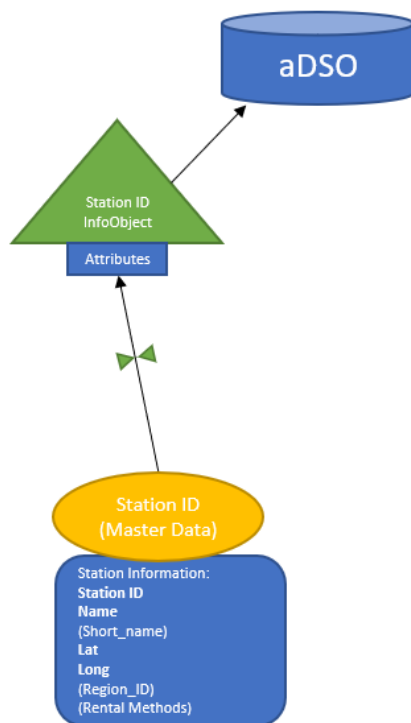
GRAPH 10. Station ID – Info Object shown as an Info Provider

In graph 10 is shown, how the master data has been planned to transfer to BW system. Firstly, the Station ID (RE1STAD in figure 7) is made as an Info Object. This Info Object is defined to be Info Provider, but as well it is necessary to include master data and texts –properties correctly to the Info Object. In the Attributes-tab, it is defined what are the attributes of the Info Object. These attributes are going to be providing the information from master data and for these attributes is necessary to make own corresponding Info Object. In the same tab it is possible to definite the Navigation Attribute.

FIGURE 7. Definition of Attributes for the Info Object (Centria University of Applied Sciences 2018.)

After the following has been done, Station ID Data Source is created with Data Source data type "Master Data – Attributes". When the data has been previewed, the transformation can be done. In transformations, it is needed to define the attributes but as well the texts which are in the Master Data. In the project the key attribute is station ID (STATION_ID). Since station ID and station name are characteristics, they are needing to have text transformations as well. (See figures 8 and 9)

FIGURE 8. Transformation rules made for Attributes (Centria University of Applied Sciences 2018.)

FIGURE 9. Transformation rules made for Texts (Centria University of Applied Sciences 2018.)

The Info Object, which is created with the attributes mentioned, will be linked to aDSO. Therefore, the Master Data which is needed for the reporting can be included in the report. In this case, it means that the data has also additional Master Data features such as the names of the stations and the latitude and longitude of the stations. With location information, latitude and longitude, it is possible to map the stations locations geographically as well.

In figure 10, it is possible to see the total Data Flow of the Master Data. As it is seen, the Master Data is also loaded to another Info Object, RE1ENDST. This Info Object is representing the end station, where the bike has been left after the trip has been made. For reporting purposes, it is important to identify from which station the trip has started and to which station the trip has ended.
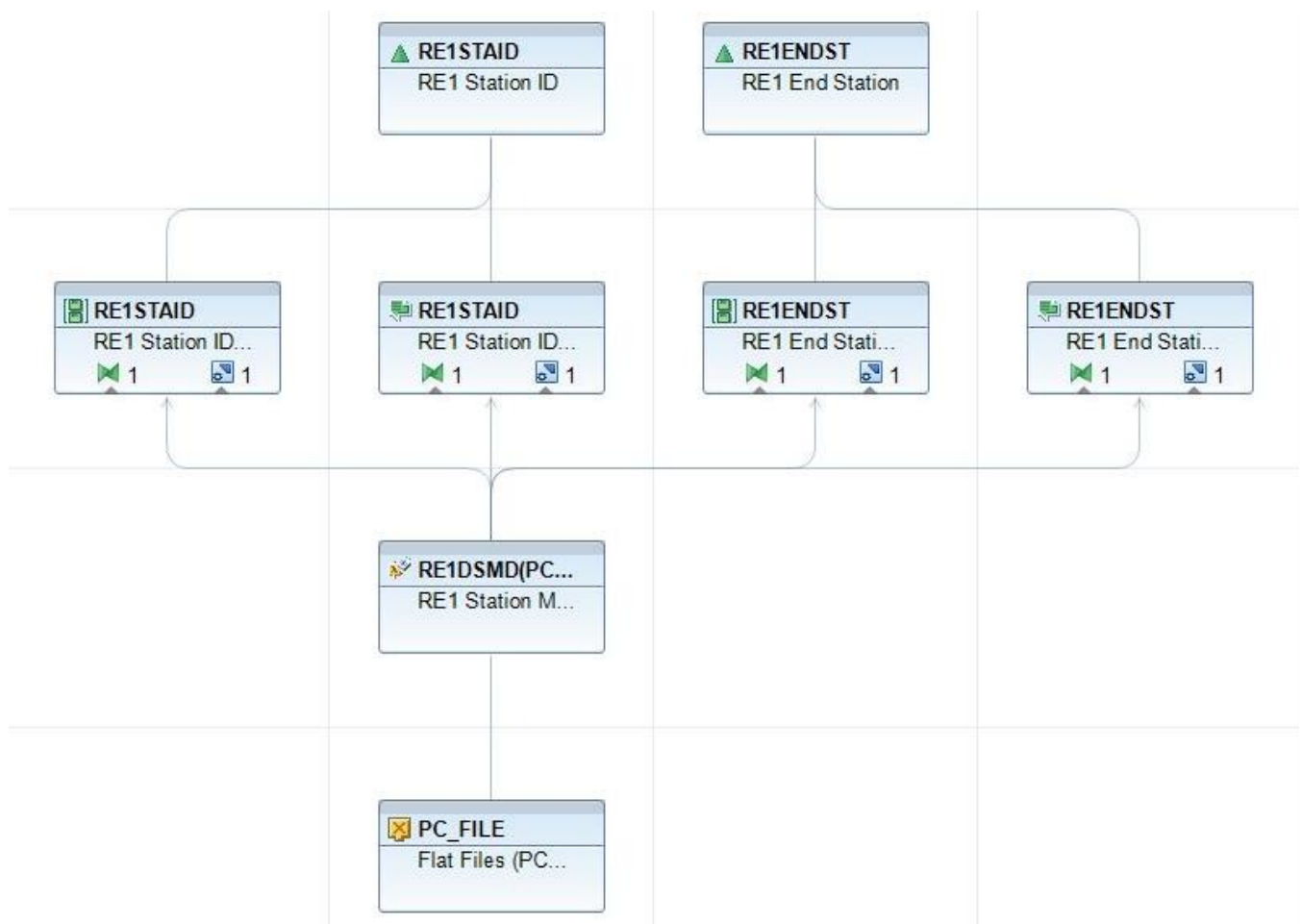
FIGURE 10. Data Flow of the Info Object having Attributes (Centria University of Applied Sciences 2018.)

6.3 Data Modelling

The first step of data modelling is made when the Data Source is made. After the creation of the Data Source, it is needed to make corresponding Data Store Object (advantaged) (aDSO) where the data will be stored physically since it is persistent BW object. Transformations are made from Data Source to aDSO. The rules of the transformations are defined when the transformations are created as well. If there are needed any formulas, they can be added (such as calculating gross margin). Also, the transformation rules of time characteristics are defined there. When the transformations are successfully set, it is possible to do Data Transfer Processes, DTPs, to load the data physically to aDSO. In figure 11, the data flow of the whole project process is presented. The Data Sources are in the bottom of the flow and in top is Composite Provider consisting combination of trip and weather data. (SAP Help Portal – SAP BW/4HANA 2018)

As from figure 11 can be seen, the data flow is consisting of multiple Data Sources and aDSOs with Trip Data. Each aDSO is consisting of trip data from one year. As it can be seen, each trip data aDSOs have multiple DTPs. It means that once the data is modeled to system, there can be multiple data transfers made via this data model. After the data is flown to aDSO, the Composite Provider is made to aggregate the data. From each aDSOs which are including trip data or weather data is made own Composite Providers in order to see is the data flowing correctly to the target. Another reason is to find the problems if they are occurred in Composite Provider level easier, since the data scope is narrow. In the end, these Composite Providers are combined together with SQL UNION in order to be able to use all the fields in trip data and weather data in reporting.
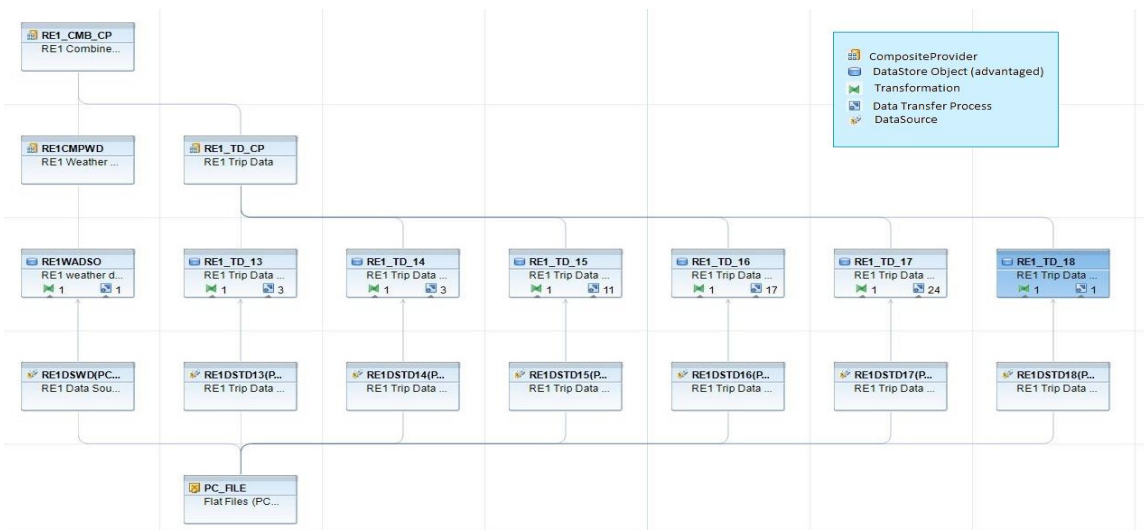
FIGURE 11. Finalized Data Flow (Centria University of Applied Sciences 2018.)

6.3.1 Creation of Info Object

Info Objects, the tiniest entities of BW objects, are the pillars of business warehouse. They define are the incoming data amounts, what color the products have and in which warehouse they locate for example. When Info Object is going to be created, it is important to remember for what purpose the data is used for and which are the minimum requirements they need. These requirements come based on the requirements of data fields.

When Info Object is created it can be defined as an Info Provider but as well it is possible to define how the data is behaving when it is received. With key figures, the data can be shown as a sum or average of the records. Data can be defined as a decimal number for example in key figure level. With Characteristics, the length of the characteristic is important to define but as well time-depended aspects or language-depended aspects can be added there. With characteristics it is possible to define a hierarchy or to have attributes in order to organize the incoming data more deeply. In figure 12 is presented General view of Info Object when it is Characteristics.

FIGURE 12. General view of Info Object (Characteristic) (Centria University of Applied Sciences 2018.)

6.3.2 Creation of Data Store Object (advantaged)

To create the aDSO, it is needed to know what kind of data is coming to the aDSO. Therefore, the Data Source which has been created is used as a model to create the aDSO correctly. Then it is less likely to make mistakes in creation level of aDSO since the Data Source is defined correctly in earlier step. The aDSO has lots of duties in SAP BW/4HANA and there are lots of important settings as well. There are different options of Model Templates to serve the different needs of the companies, and aDSO has characteristics of Classic BW Objects, such as "Planning on Info Cube-like".

In the project the model template used is "Data warehouse layer – Data Mart". As a new thing in SAP BW/4HANA, the data can be tiered. In the Data Tiering Properties, the data can be selected to be Hot (standard), Warm or Cold. When hot data is widely used and needed, the cold data is rarely used, and because of it, is stored in cold store. With this option, it is possible to save storage of the data in the database, since the hot data is needed to be looked up more often and therefore is needed to request from database many times more than cold data. The cold data is needed less often and therefore can be stored to secondary options. The General properties of aDSO is seen in figure 13. (Riesner et al. 2017.)
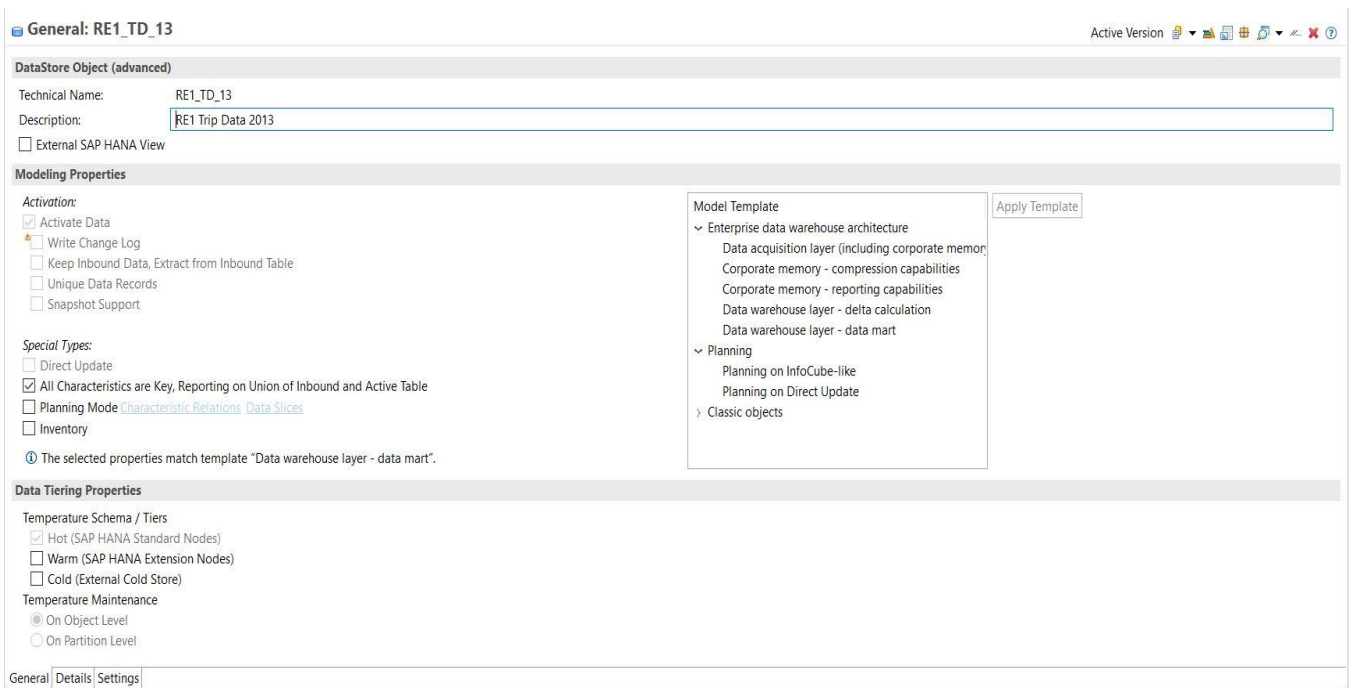

FIGURE 13. General Properties of Advantaged Data Store Object (Centria University of Applied Sciences 2018.)

In Details- tab the Info Objects of the aDSO need to be defined. Because of the copy of the Data Source, it is easy for the user to find correct Info Objects when the data fields of the Data Source are seen. Before this, all the Info Objects which are needed should be already created and ready to be used. It is important to set for all the Info Objects correct settings according to the types of data to be received. If the data which is coming does not have corresponding Info Object, the data from that field won't be seen. The system warns it when the aDSO is activated. For example, in figure 14, the warning would be caused because "RE1_ROW_ID" is not assigned to any Info Object. On other hand, if the data is seen unnecessary on this step, it can be skipped and add later. The data can be grouped on this stage in order to organize the data easier. In figure 14, data is grouped in four groups: TRIP, START, END and USER. The decision of the groups has been done based on characteristics of the data. For example, data fields, which are in TRIP-group, are presenting all the data fields which aree relevant to actual bike trips. USER- group is presenting more about the type of the bike user: how old the user is and which gender the user is representing. The groups are seen until the Composite Provider and Query -level. Therefore, if there is lots of data fields, it is important to group them in order to make the data analyzing process easier in further steps.

When the aDSO is activated, the system will be notifying if the aDSO is working properly. Warnings can occur when some data field is missing Info Object. Errors can occur for example when aDSO is activated too early, but as well if the data fields and types of Info Objects are not matching at all.

FIGURE 14. Details of the Advantaged Data Store Object (Centria University of Applied Sciences 2018.)

6.3.3 Transformation Rules

When the ETL-process is made, one important step is to make a Transformation from the source location to the target location. This step makes the model which will be applied when the data is transferred to the system and therefore it is important to think closely how the data transfer should be modelled. Once the model is made, multiple data transfer processes, even real-time loading, can be done via the transformation. Transformation is needed to make only once to the target. In graph 11 is shown, how in the project the transformation from Data Source to aDSO is made only once, but multiple data transfer processes (DTP) can be made.
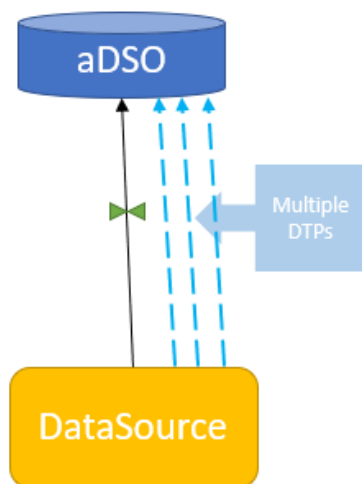
GRAPH 11. Data Transfer Process

When transformations are created, the rules of the transformations need to be defined as well. In figure 15, can be seen how the transformation rules are made. System defines automatically the transformation rules, but these need to be checked are they correct. It is important to make sure, that the data field is added to corresponding Info Object because the Info Object is also an Info Provider and therefore it will provide data to the reports made. The overall Trips have been calculated by using ROWID in the project and this can be seen from figure 15. To get the total trips, the BW system is told that one data unit from ROWID is identified as one trip. After the transformation rules are defined, the system can be requested to check are the rules defined correctly. When the transformation is activated, it is possible to process DTP.
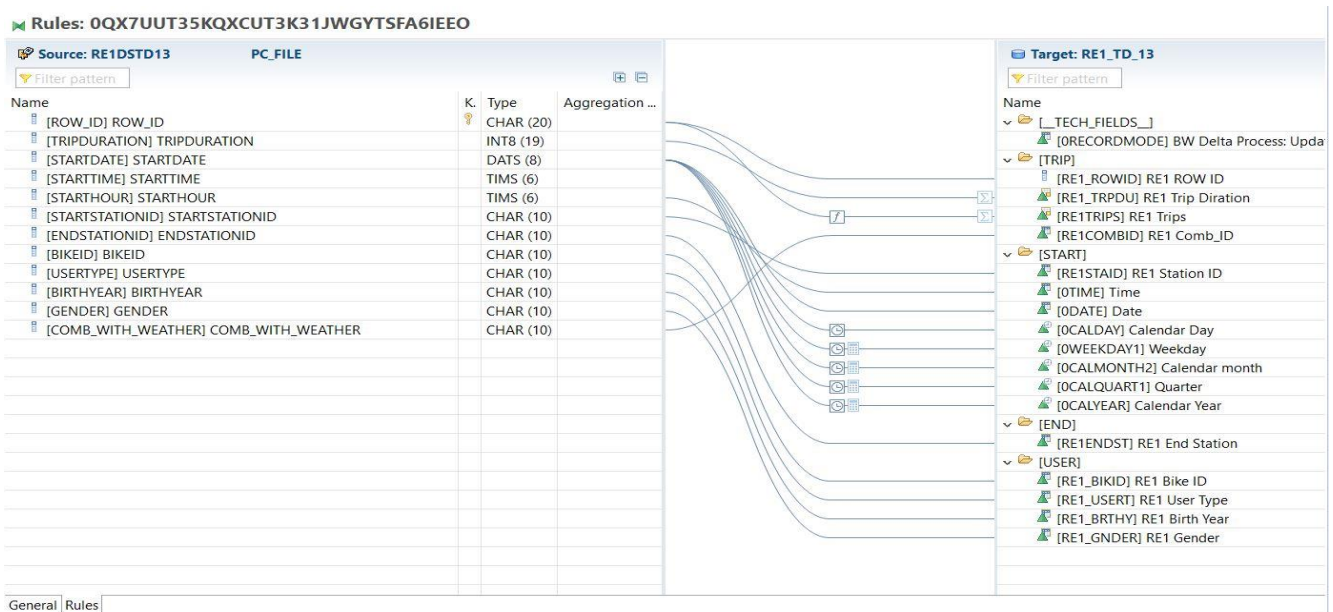
FIGURE 15. Transformation rules (Centria University of Applied Sciences 2018.)

6.3.4 Successful Data Transfer Process

In DTP, the data is loaded to system. If any mistakes have been done during the data modeling, in this stage the mistakes are shown the latest more likely. When the DTP is going to be processed, there is options of full data transfer process or delta transfer process. After verifying the source information which are copied from Data Source, the data transfer process can be processed. After the DTP, it is important to checkup has the data transferred successfully by selecting "Monitor Data". The errors faced during the project has been for example mistakes of modelling data in earlier processes, overloading data (data fields lengths incorrect), corresponding Info Object has not been matching with data field or the source field has had special letters such at @. (Korkatti, Bitarafan & Kujala 2018.)

6.3.5 Composite Providers

When Composite Provider is created, the data fields are copied on source wanted. In the project, the data fields were copied from existing and active aDSO. Since aDSOs were created based on trip data from each year, in the project the Composite Provider was created to add all these year-based aDSOs. With weather data, only one aDSO was created and therefore when making Composite Provider for weather

data, only one aDSO is needed to be add to this Composite Provider. Based on the analytical need, when Composite Provider is created, the aDSOs, or other Info Providers, can be added as an SQL UNION or SQL JOIN. The difference between SQL UNION and SQL JOIN is that in SQL UNION all the data fields are taken to the Composite Provider, meaning that all the fields are possible to use in further analyzing. SQL JOIN therefore joins some of the data fields of the source to the target object. In the project, SQL UNION was used in all the Composite Providers. SQL JOIN is useful, when lots of data sources exists, and only certain data fields are needed to reporting purposes. In figure 16, it is possible to see how Composite Provider is made from trip data aDSOs.
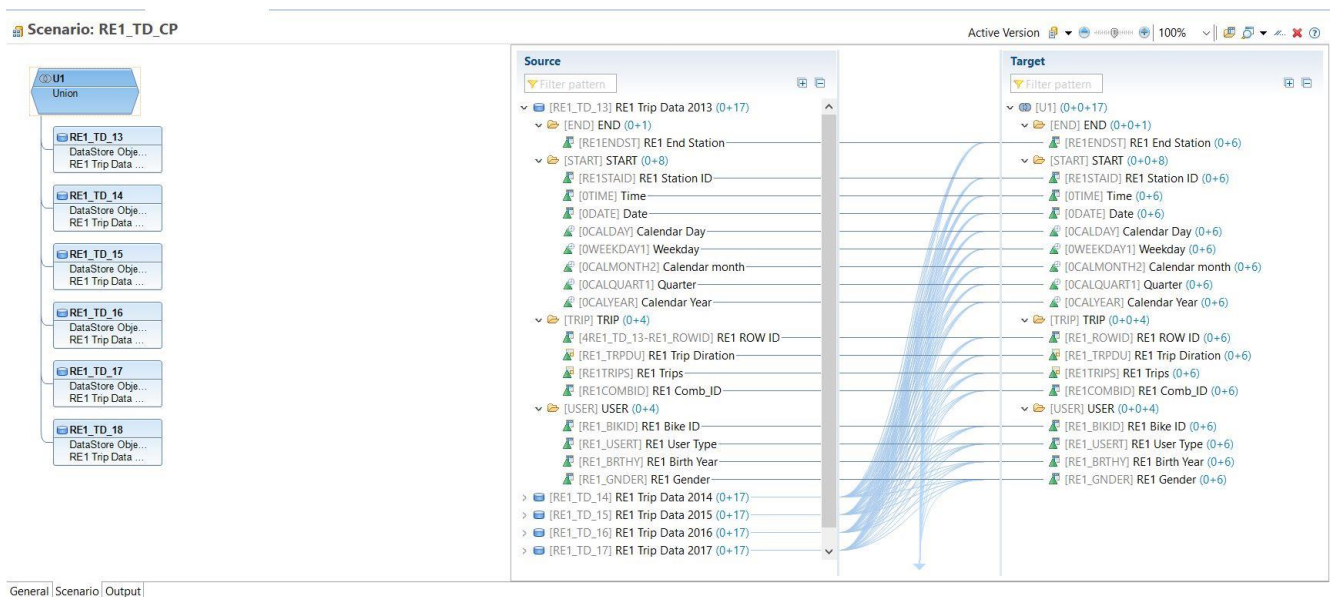

FIGURE 16. Composite Provider in Scenario view (Centria University of Applied Sciences 2018.)

When Composite Provider for trip- and weather data is made, it is possible to combine these data sets with another Composite Provider. Simply this can be used, when source Composite Providers' properties are set to "This Composite Provider can be added to another Composite Provider". If some data fields or properties of the Info Object have been not set right, it can cause that the Composite Provider is not getting data correctly from aDSO.

After the combined data is in target Composite Provider, the Queries can be done based on the analyzing questions. These were for example:

1. Where are most of the trips made?
2. What is the average duration of the trips?
3. Which gender does the most trips?

4. In which age the customer is more likely to use Bike Sharing system's bikes?

In the project each question was answered based on one query. In order to be able to do so, some formulas are needed to add to the queries. For example, when the question "what is the average duration of the trips" is answered, the formulas are needed to add to the query. Since the original data is presented in seconds, it does not make no sense to show what the average seconds of the trips are. Instead of that, it is possible to calculate via formula the trip duration in minutes when there is available the information how many seconds each trip took. Then to calculate the average trip duration in minutes, it is needed to divide the overall trip duration in minutes with number of trips done. (Korkatti et al. 2018.)

6.4 Data Presentation

For data presentation, there was used SAP Lumira Designer in the project, which is data visualizing application and is provided by SAP. Firstly, it is important to make a plan of layout how the report should be looked like. This can be done by placing SAP Lumira Designer's objects by dragging and dropping or by specifying the place in object's properties.

To make charts in SAP Lumira Designer, it is needed to bring data source from BW-system. When adding the data source, it is possible to add the exact queries made in SAP BW/4HANA. Then the filtering and specific formulas are already done and not so much modification is needed in SAP Lumira Designer. If it is wanted, it is possible to add as well a whole Composite Provider to SAP Lumira Designer. When the data sources are added, the charts can be created. The default charts are shown, but by clicking "Allow Data Source Modification" in chart properties, it is possible to modify the charts with Chart Configuration. This allows the user to modify the chart in many features, which are for example used measures and dimensions, secondary axis, types of charts and so on. In figure 17, it is possible to see the example report created in the project with SAP Lumira design.

Since there was latitude and longitude in station's Master Data, it is possible to have a geographical map of the data gathered by having a source which is including this Master Data. In the figure 18 it is seen, that the locations of the stations can be seen by blue bubbles, and by the size the popularity of the station. This is made by choosing the data point type as a bubble and choosing the right data source to the geographical map. Also, geo information type should be Latitude & Longitude. The information from source will come from corresponding latitude and longitude Info Objects.
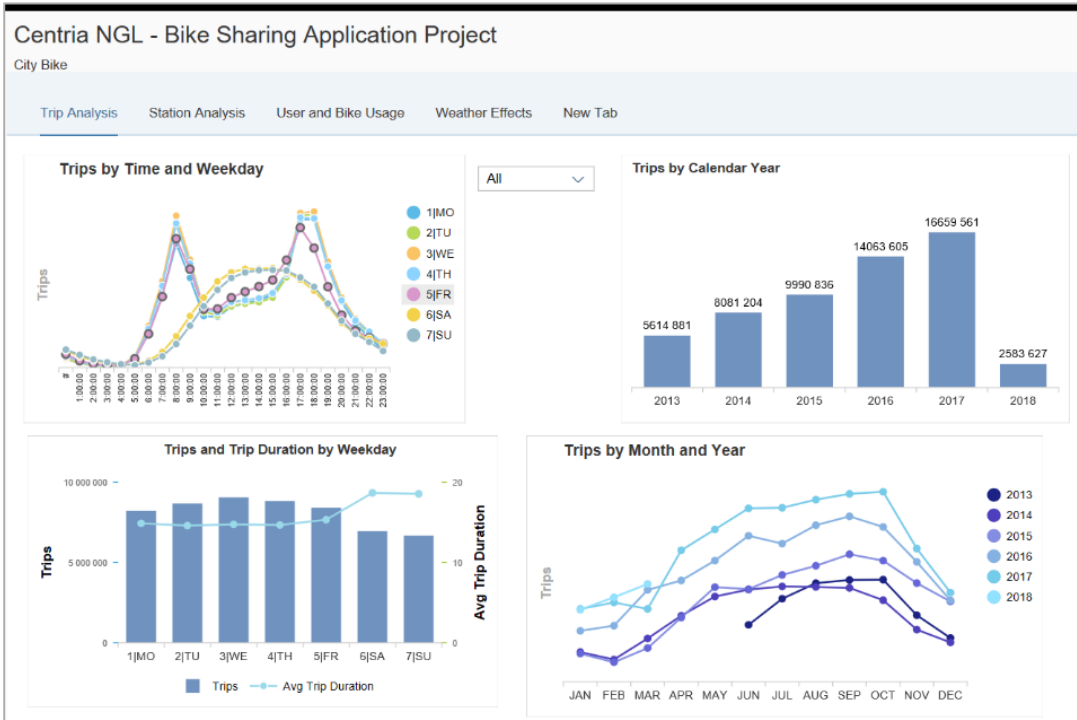
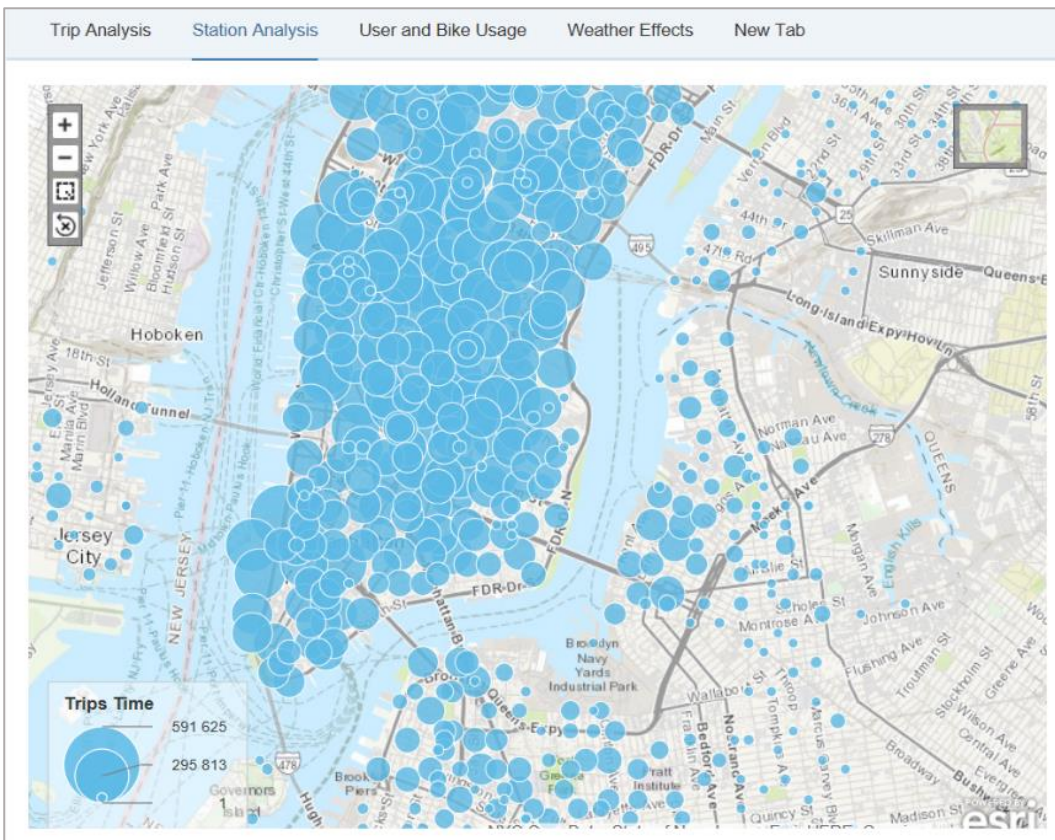FIGURE 17. Charts in SAP Lumira Designer (Korkatti et al. 2018.)

FIGURE 18. Map of the Stations in SAP Lumira Designer (Korkatti et al. 2018.)

7. CONCLUSION

The process from data acquisition to data presentation is a long journey for the students who are doing this kind of project for the first time in their lives. The SAP Next-Gen Lab environment for the students is unique, since it is possible to use SAP technologies and real-life projects in the environment. By being first students in SAP Next-Gen Lab in Centria University of Applied Sciences, it is understandable lots of things need to be improved. Many things in this half a year were working well already. Everyone in the team were able to catch the project fast and were excited to work on this project and it is always very important in order to be successful in such a project.

One improvement to the operations of SAP Next-Gen Lab needed is to structure the team in very beginning. The reason for that is that everyone in the team is knowing what to do and which areas they are responsible for. In the first SAP Next-Gen Lab this was not decided clearly firstly, and developed during the time, but processes would have been faster if everyone had known in the beginning. Unnecessary hassle would have been avoided as well if everyone had been clarified who should work on one specific area. These responsibility areas for student could be decided based on students' interest areas and strengths. It is recommended to look up the roles together closely at first, and if team cannot decide which areas are whose responsibility, this can be done for example by interviewing each person of the team and decide based on the profile of the student.

Another improvement to add is that research work in the project was limited and in order to make the process faster, more research hours should be added. This thesis is written afterwards, and therefore many things done in the project have been clarified during thesis writing process. Basics of business warehouse were studied in the project period, but for instance SAP HANA –database knowledge was limited. These research hours could be applied for example in beginning in the project, by giving every student certain area to research and when the research is done, it could be presented shortly to the team. Certain areas, such as data types, are important to everyone in the team to know, since this knowledge applies from data acquisition steps until the presentation steps.

In the project one huge problem that came across was that the limitations of the current system, firstly SAP BW on HANA hosted by UCC Magdeburg and later SAP BW/4HANA hosted by UCC Magdeburg were not studied. The problems that came across quickly were simply because of the system limitations. Those were in the project for example that the team did not have administration rights and therefore

some steps (for example having real-time data) were impossible to do. Therefore, when any new system is used, the limitations of the system used should be researched beforehand in order to not get stuck on them.

In project process the first step of all is the planning and making clear with all the team members that the target of the project is clear to everyone. Planning is important in every area of the project: from data gathering, transferring them to CSV-files until to the report of the analyses. Another thing to remember is to keep the scope of data small at first. At the beginning of the project, the team tried to load lots of data records to the system, but very fast it was realized that when the error occurs, there is too much data to look up where the mistake is. In order to make the process of the project more efficient, it is needed to proof that the project is possible to do with tiny set of data. Once tiny set of data are brought successfully to the system until reporting, the scope can be expanded. It is also important to remember to plan separately the bike trip data and weather data, and how they should be combined. To combine the data there should be one data field which is the same in both of data sets from different sources. In the project it was decided to be ROW_ID –field as well as date and time.

In data acquisition, the data engineers need to have good communication with the data analysts who are working on data presentation, since the requirements of data preparation come from the report requirements. Also in this project, there should be removed every kind of special letters (such as @, #, ") from the source data. These might cause problems further on the process. Data engineer might need to be able to add a new column for the data table. In the project, there was added "ROW ID" to identify each row from the data. In other hand, deletion of some columns and *data clean-up* might be needed. In the project the information which were seen irrelevant (such as phone number) were deleted. It also reduces errors when only the relevant data is chosen to load in the system. For each student it is important to understand how system behaves when data is loaded to the system. Sometimes the error can occur from tiniest mistake from data gathering but also when the tiniest mistake is in data modelling.

When data is modelled, lots of attention is needed in the structure. These structures, which tells to the system how data should be flowing to the system, should be planned beforehand. Every plan which has been done, even when it has been failed, should be documented in order to know which plans did not work. In data modelling, it is important as well to be knowing what kind of data is going to be loaded to system. All the data types need to be defined in order to be able to make corresponding Info Objects which are the pillars of the SAP BW/4HANA. If the Info Objects are not defined correctly, errors occur for sure. Each data field of the source should be studied and categorized. Data should be defined either

Master Data or Transactional data, and by the behavior of the data it is needed to define which data is characteristics or key figures and what kind of characteristics and key figures are they. The successful data flow flows when it is able to be seen until the reporting. Sometimes the error occurs on this stage. In SAP BW/4HANA and other releases, there are many ways to check-up if the data flows are correct. They are offered by SAP itself so in most of the cases, they are the most trustable way to verify if the data sets are correctly flowing. In SAP BW/4HANA, the transformation rules are mainly possible to define by system, except when formulas are made. These features make tinier chance of human errors and therefore, should be used. In data presentation, the biggest difficulties are to find how to answer questions to satisfy reporting and therefore there is a need to research the tool (SAP Lumira) beforehand in order to be able to make the reporting needed.

As a conclusion, SAP Next-Gen Lab is a great opportunity to students to learn real-life projects. With SAP technologies it is possible to gain experience within short time and as well as interaction between business and information technology students is important. Business students are able to experience different technologies and to learn the limitations of these technologies. Therefore, the business students are able to communicate better with IT students. On other hand, IT students are gaining by learning how business students think and how to communicate with them better in order to understand what is needed in business side. Both sides, IT and business students learn what tools are needed and how to use these tools in order to finish the project. In our project, the biggest limitations were that the project was not structured well and the system limitations. Top of all, the time, six months, was limited. In order to be able to do the project in six months, no bigger problems should have been occurred.

When experience of the NGL environment is more gained and improved, the students are able to be more efficient than how it was with the first project in SAP Next-Gen Lab. Students should be encouraged to study further how to develop SAP Next-Gen Lab environment in Centria, and those studies should be applied by university if they are seen as good improvement ideas.

Working with SAP Business Warehouse solutions in Centria is with its baby steps but even in this half a year period lots of knowledge has been gained, not only with students but with teachers as well. To make the study environment even greater experience, the learning of students and teachers should be supported as it has been so far. To be able to make greater projects with SAP BW/4HANA, in this thesis is recommended to give deeper ground studies of data warehousing and SAP HANA as a database.

In the perspective of SAP BW/4HANA tool, SAP Next Gen Lab is in the future recommended to study

further the different ways of bringing data in the system. For example, Open ODS View is not practiced in the first project. Open ODS View is enabling to skip physical data replication by integrating data virtually in the system. Therefore, the data is modelled in Calculation views and the source of data can contain lots of interesting data without making limitations to the system. Smart Data Integration or Smart Data Access can be applied with Open ODS View as well. But with Open ODS View, it is important to understand the interaction between BW/4HANA system and SAP HANA database and it requires students' interested in databases. Another study area is the different features of aDSO, since that has lots of characteristics of traditional BW objects. Surely, there are many interesting bigger or tinier projects which can be applied to the SAP BW/4HANA, since it is a tool widely used with companies with different purposes.

REFERENCES

Bike Rental NYC. Citi Bike. Available at: https://www.citibikenyc.com/how-it-works/bike-rental-nyc. Accessed 2.8.2018.

Bloomberg. Company Overview of NYC Bike Share, LCC. Available at https://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=285318548. Accessed 3.8.2018.

Celko, J. 2006. Joe Celko's Analytics and OLAP in SQL. San Francisco: Elsevier Science & Technology.

Centria-ammattikorkeakouluun Pohjois-Euroopan ensimmäinen Next-Gen Lab. Centria University of Applied Sciences. Published on 11.12.2017. Available at: https://web.centria.fi/uutinen/centria-ammattikorkeakouluun-pohjois-euroopan-ensimmainen-sa/2626.Accessed 27.7.2018.

Centria is the first SAP Next-Gen Chapter in Northern Europe. Centria University of Applied Sciences. Published on 13.9.2018. Available at: https://web.centria.fi/uutinen/centria-is-the-first-sap-next-gen-chapter-in-northern-europe/2839. Accessed 13.11.2018.

Centria University of Applied Sciences – About Us. Available at: https://web.centria.fi/about-us. Accessed 27.7.2018.

Centria University of Applied Sciences. Screenshots of HANA Studio and SAP BW/4HANA. 2018. Accessed 31.9.2018.

Citi Bike – About. 2018. Available at: https://www.citibikenyc.com/about. Accessed 2.8.2018.

Citi Bike – System Data. Available at: https://www.citibikenyc.com/system-data. Accessed 5.10.2018.

Dees, J., Färber, F., Grosse, P., Lehrer, W., May, N., Müller, I. & Rauhe, H. 2012. The SAP HANA Database – An Architecture Overview. Bulletin of the IEEE Computer Society Technical Committee on Data Engineering.

Global Goals for Sustainable Development. Available at: https://www.globalgoals.org/. Accessed 27.7.2018.

Korkatti, L., Bitarafan, A. & Kujala, J. 2018. Centria UAS SAP Next-Gen Lab Report 2018. Centria University of Applied Sciences.

Mankala, C. & Mahadevan, G. 2013. SAP HANA Cookbook. Birmingham: Packt Publishing Ltd.

New York Bike Share. New York City DOT. Available at: https://nycdotbikeshare.info/home. Accessed 2.8.2018.

Riesner, F. & Sauer, K-P. 2017. SAP BW/4HANA and BW on HANA. 2nd Extended Edition. Gleichen: Espresso Tutorials.

SAP Blogs. BW4/HANA in a Nutshell. 2016. Available at:
https://blogs.sap.com/2016/11/02/bw4hana-in-a-nutshell/. Accessed 11.9.2018.

SAP Help Portal – SAP BW/4HANA. 2018. Available at:
https://help.sap.com/viewer/107a6e8a38b74ede94c833ca3b7b6f51/1.0.9/en-US/d9adb4c23ccb49fdaaaa0da5acfb9a4e.html. Accessed 11.9.2018.

SAP Integrated Report 2017. SAP SE. Available at: https://www.sap.com/integrated-reports/2017/en.html.Accessed 27.7.2018.

SAP Next-Gen Innovation with Purpose. SAP SE. Available at: https://www.sap.com/corporate/en/documents/2018/04/f80714c1-fd7c-0010-87a3-c30de2ffd8ff.html.Accessed 27.7.2018.

SAP – Products / SAP Fiori. SAP SE. Available at: https://www.sap.com/products/fiori.html.Accessed 2.10.2018.

Shalaby, M. 2017. Development and Deployment of Enterprise Applications on SAP HANA Cloud Platform. Helsinki Metropolia Univeristy of Applied Sciences. Bachelor of Engineering Thesis.

Walker, M. 2012. SAP HANA Starter. Birmingham: Packt Publishing Ltd.

Wegener, B. & Böttcher, T. 2018. SAP Next Gen – Bike Sharing Project. SAP UCC Madgeburg.