A Multidimensional Database to Help Anticipate on Truck Arrival Time

Wouter Bol Raap
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands
w.bolraap@student.utwente.nl

ABSTRACT
Each distribution center has a different solution to prepare itself for incoming trucks. Due to different circumstances trucks can arrive earlier or later than anticipated. It is very hard, or maybe even impossible, for distribution centers to predict the actual truck arrival times, when they don’t know anything about the schedules from incoming trucks.

A solution for these distribution centers is a multidimensional database. This database gives information about the performances of the different haulers in different circumstances. This information can be used to predict a more exact arrival time of the incoming trucks, so the average time of the cargo in the holding area can be decreased and money will be saved. In this case study a schematic of a multidimensional database will be designed for a Dutch distribution center that doesn’t share any information with their haulers. During this case study a literature study, survey and interview will be done and business processes will be studied to determine relevant data to put in the multidimensional database. The scheme of the multidimensional database is applicable for all distribution centers which don’t have information about the incoming trucks and want a way to predict the arrival times of their incoming trucks more accurate.

Keywords
Data warehouse, multidimensional database, truck, external factors, distribution center, predicting, survey

1. INTRODUCTION
The distribution center, which name will be held unknown due to competition matters, makes daily schedules to handle all the incoming trucks. The schedules say when each truck is expected to arrive. This distribution center will be called Districenter. The schedules are used to put the cargo in the holding area in time, so arriving trucks can load their cargo in the holding area. This space cannot be used for cargo for the next incoming trucks. When a few trucks are too late and some trucks arrive on time or maybe earlier, a problem can arise because the cargo for this particular truck isn’t ready yet. A waiting line can develop which costs the distribution center a lot of money.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. 20th Twente Student Conference on IT, January 24th 2014, Enschede, The Netherlands.
Copyright 2014, University of Twente, Faculty of Electrical Engineering, Mathematics and Computer Science.

A good prediction of the arrival times can prevent this problem. The cargo will be put in holding area according to these predictions. Possible is that the cargo will be put earlier or later than two hours before the planned arrival time. At the end the cargo will spend an average less time in the holding area. This can save costs and creates flexibility to anticipate on unexpected situations. Other distribution centers have different solutions to predict actual arrival times. There are systems which can predict the arrival times of trucks. Two examples are sensors at access roads near a company or tracking the trucks with GPS data. Unfortunately these solutions are not applicable for Districenter which has no information about the truck at all.

1.1 Problem statement
Districenter experiences that a part of the arriving incoming trucks don’t arrive at the agreed time. In the contract with their partners is an agreement taken down which states that an incoming truck, which arrives within an hour before or after the planned arrival time, should be ready for departure within an hour. For each truck for which the distribution center cannot fulfill this agreement a fine must be paid. So it is important for Districenter to keep the flow of products in the holding area. Districenter wants to reduce the average time of the products in the holding area, so it is possible to better handle any difficulties or unexpected situations.

To reduce the average time, Districenter wants to predict the arrival time of the trucks. The problem for Districenter; they have no information about the truck. Other distribution centers have the same problem. These distribution centers don’t receive information from their partners either. For distribution centers which receive information from their partners, systems like a GPS system are applicable to determine the actual arrival time. In this research field a lot of research is done to predict the actual arrival time whereby the point of departure or the actual whereabouts is known. With this information the journey time can be determined through mathematical calculations. Unfortunately, solutions that need departure locations or actual GPS locations are not applicable for Districenter. There has been done some research in which the point of departure or the location of the truck is not taken along, but there is no research found that describes a method or creates a tool that helps determining the arrival time with no known information about the whereabouts of the truck and only with the information that is already inside the distribution center and which is available on the internet.

A multidimensional database is a solution which can categorize data as being facts with numerical measures [2]. One of the advantages in choosing a multidimensional database is that, it has cognitive advantages for the user. Its data presentation and navigation is easier than a normal relational database. A time dimension can easily be included to rolling-up and drilling-down across days, weeks, months and years. Disadvantages of a multidimensional database are it is less flexible and it requires a greater initial effort. This database will give facts about the
performances of the transportation companies and the freight owners by different circumstances, for example by different weather types. These facts can be used by schedulers to predict the actual arrival time of an incoming truck. Managers can use this database in their decision making process. This multidimensional database will be designed in this paper.

1.2 Research questions

From the problem statement the main research question can be formed:

Which scheme for a multidimensional database saves information which managers at distribution centers can use in their planning process?

To answer the main question it will be divided in several sub-questions. Before the multidimensional database can be designed, it needs to be clear which data are useful for managers and can be and will be put into the database.

First, the current process must be analyzed to find out how the process works and what its shortcomings are.

1. What is currently the planning process and what are its shortcomings?

The present decision and scheduling process needs to be researched to know which information is currently available for the managers. It is also important to find out which information the managers would like to have and which information they share with their partners.

2. What information do the managers currently have, need and share in their planning process?

Not only information from Districenter can be interesting. A lot of companies or governments organizations collect data and distribute this data. This data can be freely accessible or by payment and can be useful for Districenter. To know what kind of information is interesting, a literature study is necessary about what influences a truck its arrival time during its journey. With this information a specific search can be performed. In this case study only historical data are available, so the search is limited to historical data.

3. Which historical data are available, besides the current used information, which is be valuable for managers?

At this point there is a clear overview of the valuable and available data so the information found in the first two sub-questions will be used to define the representation and storage of data in a data warehouse. This should be designed in a way so that it is effectively stored in the data warehouse. Different dimensions must be created for the multidimensional database by structuring the found data. This will be done according to the theory of Kimball [11][12].

4. What dimensions are needed in the data warehouse using the data found in the first two sub-questions?

With the four sub-questions answered it is clear which information is valuable for managers to support their decisions and how a multidimensional database can be designed to help making those decisions.

1.3 Method

This case study consists several steps. First, we started with one interview with two managers and one interview with a scheduler of Districenter. These interviews were held to get to know what the current process is at Districenter, what kind of problems they experience and what kind of solution they want. The documents sent by the managers will be analyzed to this process and problem.

To design a multidimensional database it must be known which information is important. The process to model must be clear before starting with searching for information. When this is clear, the managers and schedulers are asked what kind of information they would like to have in the database. The schedules, which are made available by Districenter, are analyzed to find the data which are connected with the chosen process.

Also a literature study is performed. This must reveal what affects the truck’s arrival time during his journey towards Districenter. During this literature study all available papers which discuss the fluctuation of arrival or journey time of a truck are searched for possible causes for delay. Only the causes which can be applied to the particular problem of Districenter are taken along. The number of appearances in the literature will be added and so the most interesting causes of delay, according to the literature, will pop out. The causes which stick out the most will be tested by a survey. This survey will be held among all the incoming truck drivers. Every incoming truck driver will be asked to fill in this questionnaire. The results must give insights in the acceptance of the found causes. The availability of the causes, which have influence on the arrival time according the truck drivers will be searched on the internet. If the information is available, the cause will be taken along.

When all the information which is available is found, the database will be designed according to the theory of multidimensional modeling process from Kimball [11][12], explained in the book of Jensen et al. [10]. This process is divided in four sub-processes. The first step is choosing the business process to model. Not all the business processes are equally important for Districenter, so the process with the largest potential to solve the problem must be chosen. The second step is choosing the granularity of the business process. The data granularity that best matches the analysis must be chosen to prevent the database growing unnecessary large. Step three is designing the dimensions. The granularity will be used to refine the scheme into a complete dimension with levels and attributes. According to Jensen et al. [10] querying is easier with a star scheme, therefore a star scheme will be designed instead of a snowflake scheme. The last step is choosing the measures for the attributes.

1.4 Structure

The paper will follow the next design. Each chapter will discuss a sub-question named in section 1.2. In chapter 2 the current process will be examined. In chapter 3 the information that Districenter has within the company will be presented and the interesting information will be provided. Chapter 4 will provide the information that is available on the internet and by other organizations which also can be interesting. Chapter 5 will explain the process towards the database schematic. In chapter 6 follows a discussion and in chapter 7 the conclusions of the work and possible future extensions are presented.

2. CURRENT PROCESS

As stated in the problem statement Districenter wants to reduce the average time that the cargo spends in the holding area. Before the multidimensional database can be designed, the current scheduling process must be identified and examined. An interview with two managers is held and an internal document with all the schematic processes is examined. To get a better view of what the managers and the schedulers want to improve in their current system, they are also asked what information they would like to have.
2.1 Daily scheduling process

Currently Districenter depends on schedules. These schedules, built in Microsoft Excel, will give them insight of how, when and where the cargo must be ready and when they can expect the trucks. Good schedules will not only deliver satisfaction at the customer side, but it will also deliver more flexibility within the company.

The process of making a daily schedule involves the distribution center and the different haulers. Each day, before 5 pm the two largest haulers can submit their orders. These two haulers have fixed ports which they can use every day. Managers of the distribution center will look for difficulties, and when there are none, the order will be accepted. When it’s 5 pm the gatekeeper will add the variable orders in to the planning. The variable orders are orders of haulers who are not having fixed ports at the distribution center. He will assign the variable orders to the ports. The planner will check the schedule and send it to the haulers. Haulers can call or mail within the hour for forgotten orders and add these in the schedule or can move an order. After 6 pm, the schedule will be final.

2.2 Interview

At the start of this case study one interview was held with two managers from Districenter. Each day around 250 trucks arrive at Districenter. For each truck the cargo must be put into the holding area so that, when a truck arrives, it can be loaded fast. Now, two hours before the planned arrival time the cargo is put in the holding area. The process from the planning to the arrival of the truck is given in Figure 1.

The distribution center has agreements with the companies who store their products in the warehouses. These agreements state that a truck, when it arrives on time (within the agreed time window), should be loaded and departed within an hour. They notice that not every truck arrives on time. This can cause delay, but it also can cost a lot of money when agreements are not fulfilled.

The managers would like to decide if the cargo can be put out later than the two hours before the planned arrival. They want to have a system that helps them in this decision making process. This system must show who, when and why it is too late. For this problem a data warehouse is an appropriate solution. “A data warehouse can categorize data as being either facts with associated numerical measures or as being dimensions that characterize the facts and are mostly textual.”[2, page 1].

Also a scheduler is interviewed to know what kind of additional information is needed to use in their scheduling process. Firstly, they would like to know when multiple orders are shipped in the same truck. This happens regularly and brings the distribution in trouble because they need to move the shipment all across the building. Secondly, planners would like to standardize the planning document. It takes a lot of extra acts to add the orders from the different haulers in the Districenter’s document.

The information the schedulers named was information they can use in their current scheduling process. That the retrieved answers weren’t as expected was probably due to not clear formulated questions. It wasn’t clear for the schedulers that the questions were about the scheduling process with the dimensional database. The received answer was information of their current process. Therefore their wishes are not taken into account.

In this chapter is examined how Districenter is currently working. In the interview the managers explained why a database can help them in the daily processes. The managers have a clear view of what they can expect and to do with the system while the schedulers haven’t seen the potential of a database yet. The schedulers don’t know what a multidimensional database can do and that is why they cannot explain what kind of information they need.

3. SCHEDULING INFORMATION

As described in the chapter ‘Current Process’ each day a schedule with the planning of that day is made. In this chapter these schedules are examined for its content.

3.1 Information in the schedule

Districenter makes schedules every day to plan when the truck should arrive and when the cargo must be ready for this truck. The data in the scheduling document is given in Appendix A.

Not all of the data can be put into the database. Some content is not interesting to put in the database. Some data catch the eye immediately like ‘Planned arrival time’ and ‘Actual arrival time’.

The managers said in the interview that they would like to see which, when and why trucks are not on time. Considering this, interesting data in the schedule are:

- Planned arrival time
- Actual arrival time
- Transportation company

The Transportation Company tells who are too late while the Planned and Actual arrival time explain when the trucks are too late. The content in the scheduling document does not show why trucks are too late.

Concluding, the schedule document tells which and when trucks arrive too late, but they don’t give information why they arrive too late. The fields “Planned Arrival Time”, “Actual Arrival Time” and “Transportation Company” are interesting to be included in the database.

4. EXTERNAL INFORMATION

Not only data from Districenter is interesting, also information available on the internet or collected by companies can be

![Figure 1: Process of putting the cargo ready](image)
valuable. In this chapter this valuable data, other than the known data at the distribution center, will be found. Firstly, a literature study is done to find several causes of delay. Secondly, a survey was executed to check if the results of the survey match with the literature study. Last, a research is performed to check the availability of the valuable found data types.

4.1 What causes delay?
To create insights of what can cause delay a literature study was executed. Literature is searched for any causes what leads to delay without knowing where the truck is. For finding the literature, the citations and references of the already known literature is valued. This is also done for every new found paper later in this literature study. For finding the papers the databases for scientific literature Scopus and Google Scholar is used. In both databases a combination of several words is used. The words that are used in the search term are: travel time, journey time, reliability, delay, variability, model, simulation, congestion, survey, estimation, arrival time, truck and traffic. The abstract of papers with an interesting title are read and based on the abstract the papers are scanned for causes for delay. In total, 14 available papers are found [2-9, 14-17, 19, 20]. Some interesting papers are not available, for example the papers from the database of Transportation Research Board. I had no rights to read these papers online. For every found cause the number of appearances in the papers is counted. By counting the number of appearances the most interesting causes of delay named in the literature will pop out.

In Appendix B A diagram is given which shows the different causes and the number of appearances in the found literature. Only the causes are showed that are named more than once. The different causes are categorized by Low, Medium and High what tells how much influence it has on the journey time. For example, day- or night movement has a high influence on the journey time. During the day, there is more traffic on the road than during the night so the chance of being in a traffic jam is much higher during the day than at night. The diagram shows which six different causes strikes the most and appears to be, according to the found literature, the most interesting causes of delay. Some of the found causes don’t have to be taken into account because they consist of several other causes or there aren’t any records of the cause. For example, road constructions and accidents lead to traffic congestion. Therefore only traffic congestion is taken into account. That leaves three interesting causes of delay: day or night time movement, traffic congestion and the weather.

4.2 Survey
To test if the results are applicable for this case study, we created a questionnaire for the incoming truck drivers at Districenter. The base of this questionnaire is the found causes of the literature study and the interesting information found in the schedules. For several weeks, every incoming truck driver was asked to fill in this questionnaire. The truck drivers were asked what kind of causes made them arrive on time or too late.

In Appendix C an example of the questionnaire is given.

The results of this questionnaire are given in Appendix D. In the results an overview is given of the different causes of delay and the relative frequency of when the truck driver arrived. Later or Earlier is still within the agreed time window and Much Later and Much Earlier is not within the agreed time window. The results show that traffic congestion, the weather and the time of movement have great influences on the arrival time. These results match with the findings of the literature study.

4.3 Available data types
Historic weather information of the Netherlands is available via the website of the Dutch KNMI Weather Research [13]. Traffic jam information is unavailable on the internet. The Rijkswaterstaat measures at many points in Holland the amount of traffic. The delay between day or night time movements is also available and can be found in the daily schemes of the distribution center.

The literature study and the survey gave interesting data types that will be taken into account. The literature mentioned three important causes for delay: day or night time movement, traffic congestion and the weather. The survey agreed that these three causes are indeed three important causes for delay. It can be that some data types are missing because of the non-availability of some literature, but this will be covered by choosing the only three causes which stick out.

5. DESIGN
In this section the actual schematic for the multidimensional database will be designed. The multidimensional modeling process from Kimball [11][12] named in the book of Jensen et al. [10] will be followed. This process is divided into four subprocesses:

1. Choose the business process(es) to model.
2. Choose the granularity of the business process.
3. Design the dimensions.
4. Choose the measures.

In the following sections the four steps are executed.

5.1 Modeling business process
The first sub-process in the multidimensional modeling process of Kimball is choosing which business process must be modeled [11][12]. Not every business process is equally important, so the business process with the largest potential of reaching the target must be chosen.

In the interview of section 2.2 the managers indicated that they would like to shorten the average time of the cargo in the holding area. Figure 1 shows that employees put the cargo in the holding area according to the arrival time in the daily schedule. A more precise arrival time can be predicted by the schedulers when they got the right tools to do it. They can use a multidimensional database as a tool. Central in this process must be the historical facts of each incoming truck with the external conditions around that time. Each incoming truck has an agreed arrival time, actual arrival time and a difference between the actual and agreed arrival time. Therefore the process of the incoming trucks will be chosen as business process.

5.2 The granularity
The second sub-process is determining the data granularity. It is important to use the granularity that best matches with the needs of the analysis. The granularity of the data refers to the size in which data fields are sub-divided.

The multidimensional database will not contain real time information. Therefore for all the days in the past the planned agreed arrival time and the actual arrival time of each truck are known. Jensen et al. [10] say that in general the finest aggregation of the data should be used. This gives the largest flexibility for the analysis. Within aggregation peaks of the
difference of agreed and actual arrival time cannot be found again. Therefore each arrival time has to be taken into account individually.

Jensen et al. [10] also state that a key difference in the multidimensional modeling is that neither all the available data nor all the existing relationships in the data in the model should be included. Only the aspects that matter or are essential should be included. In chapter 3 and 4 of this paper is the essential information found.

As discussed earlier the incoming truck is the business process to model. Each incoming truck has an agreed arrival time and an actual arrival time. Earlier is found what information is interesting to put in the database. These are the different parts of the multidimensional database. Combining this selection of data the granularity for the database can be formulated: “Incoming truck per Transportation Company per weather type per traffic situation per time”. Each incoming truck has to be taken into account individually to get the best results.

The managers of Districenter were asked to validate the formulated granularity. They missed the data of the freight owner. At the Districenter the cargo can be picked up by the owner of the cargo, but this company can also hire a third-party hauler. To complete this granularity, they asked to add the information about the owners of the freight. The granularity will be after validating: “Incoming truck per Transportation Company per Freight Owner per weather type per traffic situation per time”

5.3 Dimensions

The third sub-process is designing the dimensions. In this sub-process the scheme of each part of the granularity must be refined into a complete dimension with levels and attributes.

From the granularity five dimension candidates can be found, with the incoming truck as fact table: Transportation Company, Freight Owner, Weather Type, Traffic Situation and Time. Each of the dimensions needs several attributes which describe the dimension and choose the level of each dimension.

Because traffic situations are very local, some traffic situations are more relevant than other traffic situations. Splitting the traffic situations in different zones is helpful. This way the traffic situation for the truck will be measured for only the zone where the truck will be during that time. Other roads, from outside the zone, where the truck will not be are not interesting to take into account. The border of each zone will be an one hour drive, except the last one. Only the traffic situations in the Netherlands are taken because we have only data from the Netherlands. The zones are interesting to look at. For example, the influences of the traffic situation at one hour, two hours or three hours before the arrival time can be analyzed. There will be three zones: 0-80 km, 81-160 km and 161+ km.

The Incoming Truck dimension, what will be used as fact table, needs some attributes. Each incoming truck is unique. As discussed in section 3.1 the Planned Arrival Time and Actual Arrival Time are interesting to save in the database. This information will be saved here because each incoming truck has its own Planned and Actual Arrival Time. Also the difference between the different times will be saved to make it easier to compare the performances of the trucks. Incoming Truck:

- Planned Arrival Time
- Actual Arrival Time
- Difference

For each of the dimensions the chosen attributes will be given. Chosen is to include only the necessary attributes to prevent that the database becomes too large and takes too much space.

For Districenter is it interesting to know who arrived too late. Therefore the Transportation Company needs its own dimension. For the Transportation Company it is necessary to know the name of the company so they know who isn’t on time. Transportation Company:

- TransComID (Surrogate key)
- Name

Districenter would like to get information about the Freight Owner. Information they desire is the name of the company. Therefore the dimension Freight Owner will also have the attribute Name. Freight Owner:

- FreightOwnerID (Surrogate key)
- Name

The Weather Type is taken apart from the Traffic Situation so it can be used separately from the Traffic Situation. The availability of the chosen attributes is checked on the website of KNMI [13]. Agarwal et al. [1] says that that severe rain, snow, and low visibility cause the most significant reductions in capacities and operating speeds. Rain (more than 0.25 inch/hour), snow (more than 0.5 inch/hour), and low visibility (less than 0.25 mile) showed capacity reductions of 10%-17%, 19%-27%, and 12% and speed reductions of 4%-7%, 11%-15%, and 10%-12%, respectively. In their paper they have taken five different variables into account. These are Rain, Snow, Temperature, Wind Speed and Visibility. Because Snow data isn’t available at the KNMI, the other four are taken along. The KNMI has data available about the least and the most sight of the day. Because the least sight at a day is the most interesting, this will be taken along. Weather Type:

- WeatherTypeID (Surrogate key)
- Date
- Temperature
- Visibility
- Rain
- Wind Speed

As earlier explained, the Traffic Situation is divided into three zones. These three together is the whole traffic situation of the Netherlands. Therefore the whole traffic situation doesn’t have to be a dimension on its own. The data found gives only the number of passed cars per hour, therefore only the number of passing cars is given together with the zone, date and time. Traffic Situation:

- TrafficSituationID (Surrogate Key)
- Zone
- Date
- Time
- Number Of Passing Cars

A guideline tip of Rizzi [18] tells that at least one of the dimensions should represent time, at any granularity. It is important to search at dates and times so a Time dimension is needed. This time dimension is needed as a predefined hierarchy for rolling-up and drilling-down across days, weeks, months, years and certain time periods. It also eliminates the effort required to build a hierarchy every time a database is set up. Time:

- TimeKey (Surrogate Key)
- Day
- Month
7. CONCLUSIONS

In this paper the goal was to answer the following main research question: Which scheme for a multidimensional database saves information which managers at distribution centers can use in their planning process?

An interview with the managers from Districenter revealed the problems and their view for a solution. They experience that a group of trucks doesn’t arrive on time and their cargo fills expensive space in the holding area. A problem is that they don’t know anything of the incoming truck except for whom the cargo is and the agreed arrival time. The schedulers couldn’t give information what was helpful for them to put in the database.

A lot of information is saved in the schedules, but most of information is not useful to put in a database. The time-related information (the agreed and actual arrival time) and the information about the transportation company are interesting to put in the database. This information can answer who and when is on time or too late.

A literature study is conducted to find the most common causes of delay. The found literature is studied for causes of delay. The most interesting found causes are day or night movement, weather and the traffic situation. To validate the found causes a survey was conducted for several weeks. The results are matching the found causes. On the internet data for the weather and traffic situations are available. The day or night movement can be checked with the arrival times of the trucks.

Designing the schematics of the multidimensional database is done by the model of Kimball [11][12]. The incoming trucks are chosen as the model to process because of the incoming trucks is central in the process of predicting the arrival times. By determining the granularity of the process the dimensions are designed. The following dimensions are used for the database: Incoming Truck as fact-table and Time, Weather Information, Traffic Situation, Freight Owner and Transportation Company as dimensions. To answer the main research question, Figure 4 shows a scheme of the multidimensional database which managers can use to better predict the arrival time of a truck but also understand the performance of the transportation companies. The scheme is not a tool what calculates the actual arrival time of the truck but this scheme will help, with the forecasts of the weather and traffic situation of the next day, giving a better estimated arrival time. It gives facts about the performances of different companies under different circumstances what schedulers can use when making the schedules.

7.1 Further Work

The purpose of this case study was not to find the best solution for distribution centers that have the same problems as Districenter. We were interested to find a solution how a multidimensional database must be designed to solve the problems of the distribution centers.

This scheme for the multidimensional database for Districenter is not tested. To know if the suggested scheme is working for distribution centers who don’t know anything about the incoming trucks the scheme must be implemented and tested for several months. The interview with the schedulers at Districenter revealed that they are not common with these types.
of systems and don’t know the advantages of a multidimensional database.

This scheme of a multidimensional database can serve as base for an actual prediction tool. Now, with this tool, the schedulers need to do its own prediction of the actual arrival time. Further research is necessary to find a mathematical formula and build a tool which uses this information and this formula to predict an actual arrival time.

8. REFERENCES


APPENDIX

A. Data fields in the schedules

Table 1: Information in the schedules

<table>
<thead>
<tr>
<th>Data on the schedules</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ok</td>
</tr>
<tr>
<td>Planned Arrival Time</td>
</tr>
<tr>
<td>Package Number Time</td>
</tr>
<tr>
<td>Actual Arrival Time</td>
</tr>
<tr>
<td>Destination</td>
</tr>
<tr>
<td>Cargo ID</td>
</tr>
<tr>
<td>UITSB</td>
</tr>
<tr>
<td>Dock</td>
</tr>
<tr>
<td>Instructions</td>
</tr>
<tr>
<td>Seal Check</td>
</tr>
<tr>
<td>Remarks</td>
</tr>
<tr>
<td>Transportation Company</td>
</tr>
<tr>
<td>By</td>
</tr>
<tr>
<td>Acts</td>
</tr>
</tbody>
</table>

B. Findings literature study

Figure 2: Different causes of delay found in the literature with their number of appearances
C. Example survey

Survey: Factors influencing arrival time

For scientific research, we are interested in which factors you think influence your arrival time at [insert location]. We will use your answers to better predict arrival times. All your answers will be treated confidentially and anonymously. Thanks in advance for filling out the survey.

1. Date: ____________________________

2. What time did you arrive?

__________ : __________
hour minute

3. What was your planned arrival time?

__________ : __________
hour minute

4. What was your previous stop? _________________________________

5. What is your next stop? _________________________________

6. When do you plan to deliver the goods you picked up at [insert location] today?

__________ : __________
hour minute

7. Where will you deliver the goods you picked up at [insert location] today?

___________________________

Please answer how each factor below influenced your arrival time.

<table>
<thead>
<tr>
<th>Made me arrive much earlier</th>
<th>Made me arrive much later</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weather</td>
<td></td>
</tr>
<tr>
<td>Traffic/congestion</td>
<td></td>
</tr>
<tr>
<td>My driving style</td>
<td></td>
</tr>
<tr>
<td>Cargo</td>
<td></td>
</tr>
<tr>
<td>My company's planning</td>
<td></td>
</tr>
<tr>
<td>Distance I had to travel</td>
<td></td>
</tr>
<tr>
<td>Type of truck/trailer</td>
<td></td>
</tr>
<tr>
<td>Time of day</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Example of the English survey
D. Findings survey

Figure 4: The results of the survey done at Districenter

E. Hierarchies of the dimensions

Figure 5: The hierarchies of the dimensions chosen in section 5.3.

F. Multidimensional Database Scheme

Figure 6: Multidimensional Database Scheme