Table of content

[Introduction 1](#_Toc367390416)

[Requirements 2](#_Toc367390417)

[Wholesaler 2](#_Toc367390418)

[Manufacturer 2](#_Toc367390419)

[Web-shop 3](#_Toc367390420)

[Web-shop customer 3](#_Toc367390421)

[Scenarios 3](#_Toc367390422)

[Purchasing from manufactorer 3](#_Toc367390423)

[Customer purchasing product 4](#_Toc367390424)

[Setting up a web-shop 4](#_Toc367390425)

[Design 4](#_Toc367390426)

[Choosing a database type 4](#_Toc367390427)

[Who is going to use it 5](#_Toc367390428)

[How is it going to be used 6](#_Toc367390429)

[Relational design / ER-diagram 6](#_Toc367390430)

[Overall ER-diagram 7](#_Toc367390431)

[Customer relationship diagram 8](#_Toc367390432)

[Transactional dependencies 8](#_Toc367390433)

[Performance and indexes 9](#_Toc367390434)

[Transactions 9](#_Toc367390435)

# Introduction

This report details the mini-project of Group 2. The report details the work which has been done to arrive at our final product. The focus of the report is on Database management System, and the formal semantics for a complete project has not been followed. There is no traceability for the requirements and design considerations, just like an acceptance test has not been devised.

The project takes its origin in a larger wholesaler who sells low end to mid-range cloth to different retailers. The wholesaler carries both their own brands (no-name) as well as brand names. The company is having a problem with falling sales as more and more clothing purchases are moving to the internet.

A third party person has contacted the company with a project proposal. If the wholesaler allows third party web-shops access to their distribution channels and product portfolio as well as makes it easy for the third-party web-shops to integrate with the wholesaler, then it is believed that it will be possible to boost sales considerably.

For this reason the wholesaler has contracted us to specify and design a database which allows multiple web-shops to integrate directly with the wholesaler’s product portfolio as well as allow for handling online payments and order tracking. At the same time the wholesaler would like their existing warehouse database replaced to integrate with the new system.

# Requirements

The database is used by a very easily defined group of people, each with their own requirements to the system:

* Wholesaler
* Manufacturer
* Web-shop
* Web-shop customer

## Wholesaler

1. The wholesaler must have unlimited access to the entire database
2. The wholesaler must be able to keep track of warehouse stock.
3. The wholesaler must be able to keep track of sales and extract assorted business statistics.
4. The wholesaler has only a single warehouse and the database should not support multiple warehouses.

## Manufacturer

1. A manufacturer will produce products to the wholesaler’s warehouse.
2. It must be possible for a wholesaler to agree on a price for a product in advance.
3. The price of a product may be dependent on the amount, delivery time, etc.
4. The wholesaler must be able to track the progress of a delivery from a manufacturer from order until it arrives at the warehouse.
5. It is not required to be able to trace a piece of clothing back to the manufacturer once it has been accepted at the warehouse.
6. Multiple manufacturers may product the same product.
7. It must be possible to have a product which exist in an assortment of variants (colour, pattern, ...)
8. It must be possible for a manufacturer to assign assorted attributes to the products like brand, colour, ...

## Web-shop

1. A third party must be able to build a web-shop around the interface provided by the wholesaler (the database).
2. The database must be able to handle online payments (DIBS).
3. It must be possible to track the progress of a delivery to a customer from order until it is shipped from the warehouse.
4. A web-shop may carry only a subset of the wholesaler’s portfolio.
5. It must be possible for the wholesaler to arrange different prices for different products with different web-shops
6. It must be possible for the wholesaler to set up different discount for the web-shop based on purchase amount, etc.
7. A web-shop may only see the products it carries.
8. A web-shop may only see the customers created inside the given web-shop.
9. A web-shop may only see order data relating to its own purchases and its own customers purchases.
10. It must be possible for the web-shop to define custom attributes which are stored with the customer.

## Web-shop customer

1. A web-shop must be able to create a customer with name, address, age, ...
2. It must be possible for the web-shop to arrange different prices for different products.
3. It must be possible for the web-shop to set up different discount for its customers based on purchase amount, etc.
4. It must be possible for a customer to use online payment (DIBS/PayPal/...).
5. It must be possible for a customer to pay via invoice.
6. It must be possible to handle customers returning purchased products.
7. It must be possible to handle customer complaints.
8. It must only be possible for a customer to see the order information relation to him- or herself.
9. The customer must never experience a delay of a database query of more than 1 second 99,98% of the time (from DB received the SQL until the response is ready to be sent).

## Scenarios

This section details different common scenarios of use.

### Purchasing from manufactorer

1. The wholesaler places an order with the manufactorer supplying a reference number and enters the order in ManufactorerOrders.
2. The manufactorer sends an order confirmation to the wholesaler who verifies it and enters it in the ManufactorerOrderConfirmations.
3. The manufacturer sends an invoice to the Wholesaler who verifies it and enters it in the ManufactorerInvoices.
4. When the order is received the shipping manifest is compared to the order and the products are added to the ManufactorerReceptions
5. When the invoice is to be paid the payment is made and the ManufactorerInvoices is updated to reflect that the payment is done.

### Customer purchasing product

1. The customer places an order with the web-shop and an entry is added to the CustomerOrders.
   1. If the customer pays with a credit card a NETS interface is used to validate the card and an ID is generated. This ID is used later when the money is to be transferred (may also be used in case the order is returned.
2. An order confirmation is generated and sent to the customer.
3. The order is packaged and shipped and the shipping manifest is added to CustomerDeliveries
4. An invoice is sent to the customer and it is entered in the CustomerInvoices
5. If the customer paid via credit card the money is withdrawn from the customer’s accont and the CustomerInvoices is updated accordingly.
6. If the customer pays the invoice manually the CustomerInvoices is updated when the money is received.

### Setting up a web-shop

TBD

# Design

When designing the database it is important to not only look at the requirements, but also to consider the domain in which the database is expected to be used.

Before delving into the database design some of the more overall considerations will be considered.

1. Who is going to use it?
2. How is it going to be used?
3. What guarantees can we have about the input?
4. What performance requirements are there?
5. What other layers are there apart from the database?

## Choosing a database type

Most traditional databases are normalized. This means that data preferably only exist once and are linked via keys. This is referred to as normalized data. The advantage of this format is that it has a low footprint as duplicate data only exist in form of keys. To retrieve the data the different tables must be joined. Unfortunately the joining operations are costly, and if the throughput is very large, normalized data is actually not preferable.

Another important consideration is consistency. Traditional databases use ACID (Atomicity, Consistency, Isolation, Durability), which ensures that when data is written to the database subsequent requests is guaranteed to retrieve the new update. It also ensures that simultaneous updates are protected from each other and that transactions are supported. Unfortunately this form of access is also expensive, and sometimes not needed. An alternative is BASE (Basically Available, Soft state, Eventual consistency). Sometimes it is OK that we simply know that an update will eventually be completed, and that inconsistencies are OK for a limited time. This form of database is also used for very large project, and often combined with a distributed database. It often relies on NoSQL (Not Only SQL) as opposed to SQL.

As we are designing a database for a larger cloth wholesaler which includes payments, it is beneficial to have ACID ensured, however it would be quite possible to implement it on BASE. Also, since the database is to be used by a limited number of simultaneous users (< 10000000) an ACID DB should be sufficient, and normalized data will also be acceptable, and the databases indexing and performance optimization (keeping some tables joined in RAM to improve performance) should be sufficient.

## Who is going to use it

If the database is only used by fully trusted personnel then there is no need to place limitations on the access, however if the database is to be accessed by people we do not trust, then it is important to ensure that the users are not able to access more than he or she is allowed to.

Access to the database is not directly part of the SQL standard, but most database implementation works with Users, permissions and possibly Roles. The normal implementation allows for limiting a user’s permissions to one or more specific tables or views as well as whether the user is allowed to insert, update or select. Unfortunately this is insufficient when data from multiple individuals exist in the same table and it is therefore necessary to add another layer of protection. This could be on the form of a REST service with its own layer of authentication and exposing a limited interface to the user.

An alternative is the database principle of Row Level Security (RLS), which basically allows a database to limit access to a given row of a database based on the content of an attribute.

RLS is a very interesting, but though there is an implementation in a patch to postgresql, it has been decided to move the limitation on the row level to a layer above the database. The design of the database will however be in such a way that it make it simple to create a layer on top for row level authentication.

In the tables were RLS is required a GUID attribute is added to each row. This GUID is linked to a user, and the higher level authentication ensures that all quires includes a “WHERE GUID=”...”. If e.g. a 128bit GUID is used then it is infeasible that anyone can guess the correct GUID, and the limitation on the upper level authentication is simply to ensure that the where-clause is included, as that is not possible to enforce using standard SQL users, roles and privileges.

There is a special work-around which involves having all requests to the database go through functions, as functions may enforce a WHERE-clause, but this use of transactional SQL one every query has a high performance penalty and is not very “pretty” from a design perspective.

As a minor note, if the users are allowed to insert arbitrary data it is important to encode the text to prevent them from inserting SQL-statements inside the text.

## How is it going to be used

As mentioned before if the database is used through a predefined interface,e it is simple to limit the access based on authentication. We are going to attempt to create the database so it may be access through the normal postgresql access and authentication, so the database may simply be exposed “online”, yet with the simple pre-processing of validating that the WHERE-clause with the GUID is include for the shared tables.

The actual implementation of the layer above the database which enforces this WHERE-clause will not be a part of the project, and it will simply be assumes that it is included where appropriate.

## Relational design / ER-diagram

The relational design may be done from the requirements and the overall design considerations. A way of representing the design at an overall level is using an ER-diagram.

The ER-diagram suffers from the same problem as UML, that the diagrams very quickly become extremely large and detailed.

For this reason we have decided to split the ER diagrams up into a collection of diagrams focussing on a specific subsection of the diagrams, and increasing in detail.

* Overall ER-diagram
* Manufacturer relationship diagram
* Web-shop relationship diagram
* Customer relationship diagram
* Product attribute diagram
* Manufacturer attribute diagram
* Web-shop attribute diagram
* Customer attribute diagram

### Overall ER-diagram



In the above may be found the following entities

### Customer relationship diagram



## Transactional dependencies

# Queries and views

# Performance and indexes

# Transactions