Adattárházak és üzleti intelligencia

Féléves feladat dokumentáció

Autó kereskedés adattárháza

Készítette: Gyurcsán Dávid Lóránt RP7XMC

Intézmény: Óbudai Egyetem – Neumann János Informatikai Kar

Szak: Mérnökinformatikus BSc

Tagozat: Nappali

Félév: 2023-2024 tavasz

Tartalomjegyzék

[Bevezetés 3](#_Toc166766517)

[A projekt témája 3](#_Toc166766518)

[A projekt alapját adó adatbázis 3](#_Toc166766519)

[Adatforrás 3](#_Toc166766520)

[Modellezés 4](#_Toc166766521)

[Use-Case meghatározása 4](#_Toc166766522)

[Use-Case tételesen 4](#_Toc166766523)

[DWH 5](#_Toc166766524)

[STAGE 5](#_Toc166766525)

[HST 6](#_Toc166766526)

[DM 7](#_Toc166766527)

[ETL 8](#_Toc166766528)

[EXTRACT 8](#_Toc166766529)

[TRANSFORM 9](#_Toc166766530)

[LOAD 10](#_Toc166766531)

[Power BI 12](#_Toc166766532)

[Riport 12](#_Toc166766533)

[Kódok 16](#_Toc166766534)

[STAGE 16](#_Toc166766535)

[HST 17](#_Toc166766536)

[DM 19](#_Toc166766537)

[Ábrajegyzék 21](#_Toc166766538)

# Bevezetés

## A projekt témája

A projekthez a Projektmunka I. tárgyon készített adatbázisom fogom felhasználni.

## A projekt alapját adó adatbázis

A feladat elkészítéséhez a Projektmunka 1-hez készített adatbázist fogom felhasználni, kiegészíteni a követelményeknek megfelelően. Az adatok „spawner.exe” segítségével generáltak.

## Adatforrás

Az adatforrásom egy kitalált autó kereskedés, márkaszervíz és roncstelep adatainak kezelését valósítja meg.

Az autókat megvásárolhatják a vevők az autó kereskedésben, probléma esetén az autót márkaszervízben javítják az autószerelők.

Nagyobb kár esetén, amelyik autót már nem lehet megjavítani az autót kivonják a forgalomból és a roncstelepre kerül. A képen szöveg, diagram, képernyőkép, sor látható

Automatikusan generált leírás

1. ábra adatbázis

# Modellezés

## Use-Case meghatározása

A rendszer számos felhasználói igényt kell kielégítsen. Ezek közé tartozik az eladások számának meghatározása, az üzletek számának meghatározása. Fontos még az ügyféladatok meghatározása. Emellett az autó modellek információinak megjelenítése, a modellek városonként való ábrázolása, az autókereskedések száma országonként és az ügyfelek száma országonként. A rendszernek képesnek kell lennie az értékesítések negyedévenkénti megjelenítésére.

## Use-Case tételesen

* Number of Sales
* Number of Carshops
* Costumer Details
* Model Details
* Models by City
* Number of Carshops by Coutry
* Number of Customers by Country
* Sale Number by Quarters

# DWH

## STAGE

A STAGE rétegben történik az egységesítés, ez azt jelenti, hogy közös adattípust kell meghatározni a táblák és mezők között, ezzel segítve a betöltés felépítését.

A STAGE réteg táblái megegyeznek a forrásadatbázis tábláival, viszont két különbség van köztük.

• A mezők mind szöveges adattípusúak.

• Elsődleges / másodlagos kulcsok és megszorítások nem szerepelnek.

A modelljét a következő ábra mutatja be:

A képen szöveg, képernyőkép, szám, Betűtípus látható

Automatikusan generált leírás



2. ábra stage táblák

## HST

A HST (History) rétegben történik az adatok historikus eltárolásra, transzformálása.   
Minden táblához bevezettem 2 új mezőt:

* start\_date
* end\_date

Ezek segítségével fogja elvégezni az ETL a historizálást, és minden mező konvertálásra kerül.

A historizáció ábrája:

A képen szöveg, képernyőkép, szám, Párhuzamos látható

Automatikusan generált leírás



3. ábra hst táblák

## DM

A DM réteg tartalmazza csillagsémát, ami HST réteg tábláiból képzett dimenzió és tény táblákat tartalmazza. A Dátum dimenzió táblát sql-ben készítettem.

Az alábbi dimenzió és tény táblák találhatóak meg a DM rétegben:

* fact\_car\_sales: Autó eladások
* dim\_car\_models: Autók adatai.
* dim\_car\_customers: Vásárlók adatai.
* dim\_car\_shops: Autókereskedések adatai.
* dim\_date: Dátum adatok.

A teljes csillagsémát a következő ábra mutatja be.

A képen szöveg, képernyőkép, diagram, szám látható

Automatikusan generált leírás

4. ábra DM táblák

# ETL

## EXTRACT

Az Extract folyamatban történik a STAGE rétegben található táblák adatokkal való feltöltése.

Az ábra az Extract folyamatot mutatja be.

A képen szöveg, képernyőkép, Betűtípus, tervezés látható

Automatikusan generált leírás

5. ábra extract folyamat

A képen látható a táblák feltöltése a forrásadatokkal, ezt 10db „OLE DB Souce”-al, 10db „Data Conversion”-al és 10db „OLE DB Destination”-al oldottam meg.  
Minden tábáblát külön-külön, az adatkonvertálásnál, minden adat szövegként kerül be a mezőkbe.

## TRANSFORM

A Transform folyamat felelős az adatok vissza konvertálásáért és a historizálásért.

A képen szöveg, képernyőkép, Betűtípus látható

Automatikusan generált leírás

6. ábra transform folyamat

A képen szöveg, képernyőkép, diagram, Betűtípus látható

Automatikusan generált leírás

7. ábra transform folyamat

A táblákat külön DataFlow-kba kezeltem. A 7-es számú ábrán a customer-ek adatait historizáló folyamata látható. SCD szempontjából a historikus (SCD 3) tárolást választottam.

## LOAD

A Load folyamat felelős a csillagsémában található dimenzió és tény táblák feltöltéséért. Megoldásom során a „LookUp”, „Sort” és „Merge Join” komponenseket használtam, ezek segítségével készítettem a dimenzió táblákat és a fact táblát.  
A dim\_Date dimenziót sql procedure-al csináltam.

A képen szöveg, képernyőkép, Betűtípus, sor látható

Automatikusan generált leírás

8. ábra load folyamat

A képen szöveg, képernyőkép, Betűtípus, diagram látható

Automatikusan generált leírás

9. ábra dim tábla

A képen szöveg, képernyőkép, Multimédiás szoftver, szoftver látható

Automatikusan generált leírás

10. ábra dim tábla

A képen szöveg, képernyőkép, diagram, Betűtípus látható

Automatikusan generált leírás

11. ábra fact tábla

# Power BI

## Riport

A riportban a Use-Case fejezetben meghatározott kérdésekre választ adó vizualizációk találhatóak meg.

* Number of Sales

A képen szöveg, képernyőkép, Betűtípus, szám látható

Automatikusan generált leírás

12. ábra 1.use\_case

* Number of Carshops

A képen szöveg, képernyőkép, Betűtípus, szám látható

Automatikusan generált leírás

13. ábra 2.use\_case

* Costumer Details

A képen szöveg, képernyőkép, szám, Betűtípus látható

Automatikusan generált leírás

14. ábra 3.use\_case

* Model Details

A képen szöveg, képernyőkép, Betűtípus, szám látható

Automatikusan generált leírás

15. ábra 4.use\_case

* Models by City

A képen szöveg, térkép, képernyőkép, atlasz látható

Automatikusan generált leírás

16. ábra 5.use\_case

* Number of Carshops by Coutry

A képen szöveg, képernyőkép, Betűtípus, Párhuzamos látható

Automatikusan generált leírás

17. ábra 6.use\_case

* Number of Customers by Country

A képen szöveg, képernyőkép, Betűtípus, szám látható

Automatikusan generált leírás

18. ábra 7.use\_case

* Sale Number by Quarters

A képen szöveg, képernyőkép, Diagram, sor látható

Automatikusan generált leírás

19. ábra 8.use\_case

# Kódok

## STAGE

|  |
| --- |
| CREATE TABLE [ath\_Customers] (  [customer\_id] varchar(50),  [first\_name] varchar(50),  [last\_name] varchar(50),  [email] varchar(50),  [phone] varchar(50),  [city] varchar(50),  [country] varchar(50),  [birth\_date] varchar(50)  )  CREATE TABLE [ath\_Shops] (  [city] varchar(50),  [country] varchar(50),  [shop\_id] varchar(50),  [name] varchar(50),  [service\_id] varchar(50)  )  CREATE TABLE [ath\_Models] (  [customer\_id] varchar(50),  [city] varchar(50),  [country] varchar(50),  [Model\_id] varchar(50),  [model] varchar(50),  [made\_date] varchar(50),  [brand\_id] varchar(50),  [shop\_id] varchar(50)  )  CREATE TABLE [ath\_Brands] (  [brand\_id] varchar(50),  [brand] varchar(50)  )  CREATE TABLE [ath\_Company] (  [company\_id] varchar(50),  [name] varchar(50),  [owner] varchar(50),  [model\_id] varchar(50)  )  CREATE TABLE [ath\_Employees] (  [employee\_id] varchar(50),  [name] varchar(50),  [shop\_id] varchar(50)  )  CREATE TABLE [ath\_Junkyard] (  [name] varchar(50),  [junkyard\_id] varchar(50),  [junked\_date] varchar(50),  [model\_id] varchar(50)  )  CREATE TABLE [ath\_Mechanics] (  [mechanic\_id] varchar(50),  [name] varchar(50),  [service\_id] varchar(50)  )  CREATE TABLE [ath\_serviceprice] (  [Price\_id] varchar(50),  [Usd] varchar(50),  [model\_id] varchar(50),  [service\_id] varchar(50)  )  CREATE TABLE [ath\_Services] (  [name] varchar(50),  [service\_id] varchar(50)  ) |

## HST

|  |
| --- |
| CREATE TABLE [hts\_Customers] (  [customer\_id] int,  [first\_name] varchar(50),  [last\_name] varchar(50),  [email] varchar(50),  [phone] varchar(50),  [city] varchar(50),  [country] varchar(50),  [birth\_date] date,  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Shops] (  [city] varchar(50),  [country] varchar(50),  [shop\_id] int,  [name] varchar(50),  [service\_id] int,  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Models] (  [customer\_id] int,  [city] varchar(50),  [country] varchar(50),  [Model\_id] int,  [model] varchar(50),  [made\_date] date,  [brand\_id] int,  [shop\_id] int,  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Brands] (  [brand\_id] int,  [brand] varchar(50),  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Company] (  [company\_id] int,  [name] varchar(50),  [owner] varchar(50),  [model\_id] int,  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Employees] (  [employee\_id] int,  [name] varchar(50),  [shop\_id] int,  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Junkyard] (  [name] varchar(50),  [junkyard\_id] int,  [junked\_date] date,  [model\_id] int,  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Mechanics] (  [mechanic\_id] int,  [name] varchar(50),  [service\_id] int,  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Serviceprice] (  [Price\_id] int,  [Usd] varchar(50),  [model\_id] int,  [service\_id] int,  [start\_date] date,  [end\_date] date  )  CREATE TABLE [hts\_Services] (  [name] varchar(50),  [service\_id] int,  [start\_date] date,  [end\_date] date  ) |

## DM

|  |
| --- |
| CREATE TABLE [fact\_Sales] (  [sale\_id] int,  [model\_id] int,  [customer\_id] int,  [shop\_id] int,  [sale\_date] date,  [day] int,  [month] int,  [year] int,  [price] int,  [date\_id] int  )  CREATE TABLE [dim\_Customers] (  [customer\_id] int,  [first\_name] varchar(50),  [last\_name] varchar(50),  [email] varchar(50),  [phone] varchar(50),  [city] varchar(50),  [country] varchar(50),  [birth\_date] date,  )  CREATE TABLE [dim\_Shops] (  [city] varchar(50),  [country] varchar(50),  [shop\_id] int,  [name] varchar(50),  [service\_name] varchar(50)  )  CREATE TABLE [dim\_Models] (  [city] varchar(50),  [country] varchar(50),  [Model\_id] int,  [model] varchar(50),  [made\_date] date,  [brand] varchar(50)  )  CREATE TABLE [dim\_Date] (  [date\_id] INT PRIMARY KEY,  [Date] DATE,  [Day] INT,  [Month] INT,  [Quarter] INT,  [Year] INT  )  USE [CarDealership]  GO  /\*\*\*\*\*\* Object: StoredProcedure [dbo].[LoadDimDate] Script Date: 2024. 05. 16. 14:07:47 \*\*\*\*\*\*/  SET ANSI\_NULLS ON  GO  SET QUOTED\_IDENTIFIER ON  GO  ALTER PROCEDURE [dbo].[LoadDimDate]  @StartDate DATE = '2000-01-01',  @EndDate DATE = '2024-05-14'  AS  BEGIN  SET XACT\_ABORT ON;  BEGIN TRANSACTION;  TRUNCATE TABLE dbo.dim\_Date;  DECLARE @date\_id INT = 1; -- Initialize date\_id to 1  WHILE @StartDate <= @EndDate  BEGIN  INSERT INTO dbo.dim\_Date (  [date\_id],  [Date],  [Day],  [Month],  [Quarter],  [Year]  )  SELECT  [date\_id] = @date\_id,  [Date] = @StartDate,  [Day] = DAY(@StartDate),  [Month] = MONTH(@StartDate),  [Quarter] = DATEPART(q, @StartDate),  [Year] = YEAR(@StartDate);  SET @StartDate = DATEADD(DD, 1, @StartDate);  SET @date\_id = @date\_id + 1; -- Increment date\_id  END;  SELECT \* FROM dbo.dim\_Date;  COMMIT;  END; |

# Ábrajegyzék

[1. ábra adatbázis 3](#_Toc166766448)

[2. ábra stage táblák 5](#_Toc166766449)

[3. ábra hst táblák 6](#_Toc166766450)

[4. ábra DM táblák 7](#_Toc166766451)

[5. ábra extract folyamat 8](#_Toc166766452)

[6. ábra transform folyamat 9](#_Toc166766453)

[7. ábra transform folyamat 9](#_Toc166766454)

[8. ábra load folyamat 10](#_Toc166766455)

[9. ábra dim tábla 10](#_Toc166766456)

[10. ábra dim tábla 11](#_Toc166766457)

[11. ábra fact tábla 11](#_Toc166766458)

[12. ábra 1.use\_case 12](#_Toc166766459)

[13. ábra 2.use\_case 12](#_Toc166766460)

[14. ábra 3.use\_case 13](#_Toc166766461)

[15. ábra 4.use\_case 13](#_Toc166766462)

[16. ábra 5.use\_case 14](#_Toc166766463)

[17. ábra 6.use\_case 14](#_Toc166766464)

[18. ábra 7.use\_case 15](#_Toc166766465)

[19. ábra 8.use\_case 15](#_Toc166766466)