	Lehrveranstaltung	**Databases and Information Systems 2020**		
	Aufgabenzettel	**6**		
	STiNE-Gruppe 14	**Simon Weidmann, Aram Yesildeniz**		
	Ausgabe	**23. Juni 2020**	Abgabe	**7. Juli 2020**

6.1 ETL Process

Scheme Decisions

We use a Star Schema with the dimensions Product, Geography and Time.

The Product Dimension includes: id, article-id, article-name, productgroup-id, productgroup-name, productfamily-id, productfamily-name, productcategory-id, productcategory-name and price.

The Geography Dimension includes: id, shop-id, shop-name, city-id, city-name, region-id, region-name, country-id, and country-name.

The Time Dimension includes: id, day, month, year and quarter.

The Fact Table consists of the keys of the dimension tables as well as the measures Sold and Revenue.

CSV Transformation

- The Date string "01.01.2020"needs to be split up into its parts Day, Month, Year and Quarter.
- The comma in revenue needs to be replaces by a dot
- Corrupt data needs to be ignored

Log Outputs

Amount of Imported Tuples

Time Dimension

Data Output | Explain | Messages | Notifications

	id [PK] integer	day integer	month integer	year integer	quarter integer
1	1	1	1	2019	1
2	2	2	1	2019	1
3	3	3	1	2019	1
4	4	4	1	2019	1
5	5	5	1	2019	1
6	6	6	1	2019	1
7	7	7	1	2019	1
8	8	8	1	2019	1
9	9	9	1	2019	1
10	10	10	1	2019	1

	Lehrveranstaltung	**Databases and Information Systems 2020**		
	Aufgabenzettel	**6**		
	STiNE-Gruppe 14	**Simon Weidmann, Aram Yesildeniz**		
	Ausgabe	**23. Juni 2020**	Abgabe	**7. Juli 2020**

Product Dimension

Data Output | Explain | Messages | Notifications

	id [PK] integer	article_id integer	article_name character varying (255)	productgroup_id integer	productgroup_name character varying (255)	productfamily_id integer	productfamily_name character varying (255)
1	1	1	Pioneer DVR-550HX	1	HD-Rekorder	1	Video
2	2	2	LG RH-T 298	1	HD-Rekorder	1	Video
3	3	3	Samsung DVD-SR275	1	HD-Rekorder	1	Video
4	4	4	BenQ DE350P	1	HD-Rekorder	1	Video
5	5	5	Panasonic HDC-SD707	2	Camcorder	1	Video
6	6	6	Sony HDR-CX115	2	Camcorder	1	Video
7	7	7	Kodak Zx3 Playsport	2	Camcorder	1	Video
8	8	8	Toshiba Camileo S20	2	Camcorder	1	Video
9	9	9	Onkyo DX-7355	3	CD-Player	2	Audio
10	10	10	Yamaha CDX-497	3	CD-Player	2	Audio

Geo Dimension

Data Output | Explain | Messages | Notifications

	id [PK] integer	shop_id integer	shop_name character varying (255)	city_id integer	city_name character varying (255)	region_id integer	region_name character varying (255)	country_id integer
1	1	1	Superstore Stuttgart	1	Stuttgart	1	Baden-Württemberg	
2	2	2	Superstore München	2	München	2	Bayern	
3	3	3	Superstore Berlin	3	Berlin	3	Berlin	
4	4	4	Superstore Potsdam	4	Potsdam	4	Brandenburg	
5	5	5	Superstore Bremen	5	Bremen	5	Bremen	
6	6	6	Superstore Hamburg	6	Hamburg	6	Hamburg	
7	7	7	Superstore Wiesbaden	7	Wiesbaden	7	Hessen	
8	8	8	Superstore Schwerin	8	Schwerin	8	Mecklenburg-Vorpommern	
9	9	9	Superstore Hannover	9	Hannover	9	Niedersachsen	
10	10	10	Superstore Düsseldorf	10	Düsseldorf	10	Nordrhein-Westfalen	

Schema Creation

```
String createSQL = "CREATE TABLE PRODUCT_DIMENSION"
 + "("
 + "id serial NOT NULL,"
 + "article_id integer NOT NULL,"
 + "article_name character varying(255) NOT NULL,"
 + "productgroup_id integer NOT NULL,"
 + "productgroup_name character varying(255) NOT NULL,"
 + "productfamily_id integer NOT NULL,"
 + "productfamily_name character varying(255) NOT NULL,"
 + "productcategory_id integer NOT NULL,"
 + "productcategory_name character varying(255) NOT NULL,"
 + "price double precision NOT NULL,"
 + "PRIMARY KEY (id)"
 + ")";

String createSQL = "CREATE TABLE TIME_DIMENSION"
 + "("
 + "id serial NOT NULL,"
 + "day integer NOT NULL,"
 + "month integer NOT NULL,"
 + "year integer NOT NULL,"
 + "quarter integer NOT NULL,"
 + "PRIMARY KEY (id)"
```

```
 + ")";

String createSQL = "CREATE TABLE GEOGRAPHY_DIMENSION"
+ "("
+ "id serial NOT NULL,"
+ "shop_id integer NOT NULL,"
+ "shop_name character varying(255) NOT NULL,"
+ "city_id integer NOT NULL,"
+ "city_name character varying(255) NOT NULL,"
+ "region_id integer NOT NULL,"
+ "region_name character varying(255) NOT NULL,"
+ "country_id integer NOT NULL,"
+ "country_name character varying(255) NOT NULL,"
+ "PRIMARY KEY (id)"
+ ")";


String createSQL = "CREATE TABLE FACT_TABLE"
+ "("
+ "product_id integer NOT NULL,"
+ "time_id integer NOT NULL,"
+ "geography_id integer NOT NULL,"
+ "sold integer NOT NULL,"
+ "revenue numeric(10, 2) NOT NULL,"
+ "PRIMARY KEY (product_id, time_id, geography_id),"
+ "FOREIGN KEY (product_id) REFERENCES PRODUCT_DIMENSION(id),"
+ "FOREIGN KEY (time_id) REFERENCES TIME_DIMENSION(id),"
+ "FOREIGN KEY (geography_id) REFERENCES GEOGRAPHY_DIMENSION(id)"
+ ")";
```

6.2 Data Analysis

TBD

Database Queries

Log Output

Lowest Granularity Level

Highest Granularity Level