Install postgres on docker:

```
docker pull postgres
```

Run postgres container:

```
docker run –name my_postgres -e POSTGRES_PASSWORD=123456 -d postgres
```

Enable psql:

```
docker exec -it my_postgres psql -U  postgres
```

Transformation process:

Star Schema:

dim_furtniture (furniture_name)

dim_time(dim_time_id, date, month, year)

dim_room(room_number, name)

fact_city_info(dim_time_id, dim_furniture_name, dim_room_number, number_of_furniture, average_salary, number_of_employee, number_of_event)

Snowflake schema:

dim_year (year)

dim_month(month_id, month, *year*)

dim_date(date, *month_id*)

dim_furtniture (furniture_name)

dim_room(room_number, name)

fact_city_info(dim_date, dim_furniture_name, dim_room_number, number_of_furniture, average_salary, number_of_employee, number_of_event)

Explanation:
Generally data warehousing is the process of collecting and managing data from varied sources to provide meaningful business insights. Most often these are statistical data collected from OLTP (Online Transaction Processing) databases.

In data warehouse systems generally there are some dimension tables and one or more fact tables. For star schema the dimensions are de-normalized for the sake of fastness of the queries. However in snowflake schema there’s been normalization to some extend.

For the time dimension it has been assumed that we need statistics on daily, monthly and yearly basis only. So week, season has been skipped.

As in the above transformation process the time dimension are de-normalized in star schema but are normalized to year, month and date for the snowflake schema.

Creating database and tables:

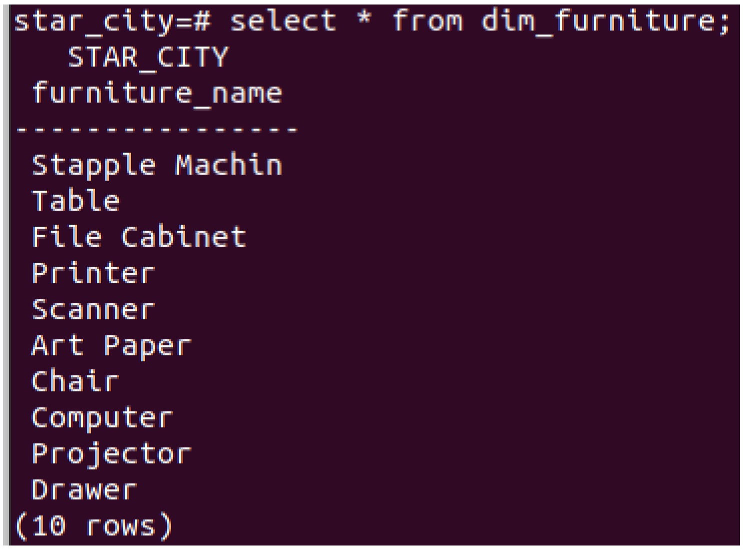
Star schema database:

create database star_city;

\c star_city;

create table dim_furniture as select distinct name as furniture_name from furniture;

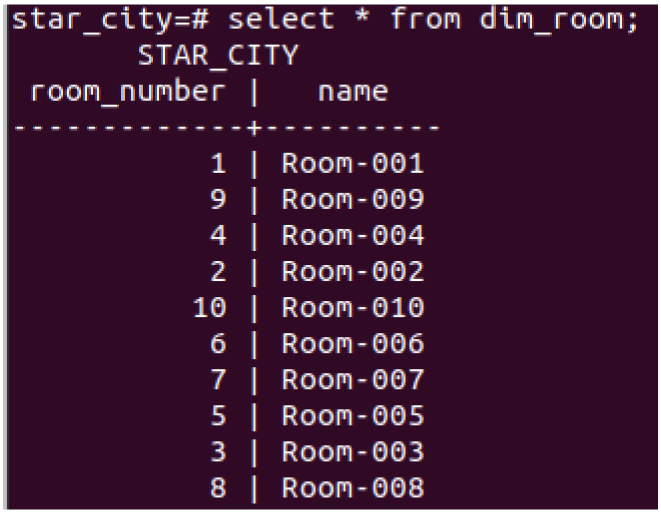
Tuples:



```
star_city=# select * from dim_furniture;
   STAR_CITY
 furniture_name
----------------
 Stapple Machin
 Table
 File Cabinet
 Printer
 Scanner
 Art Paper
 Chair
 Computer
 Projector
 Drawer
(10 rows)
```

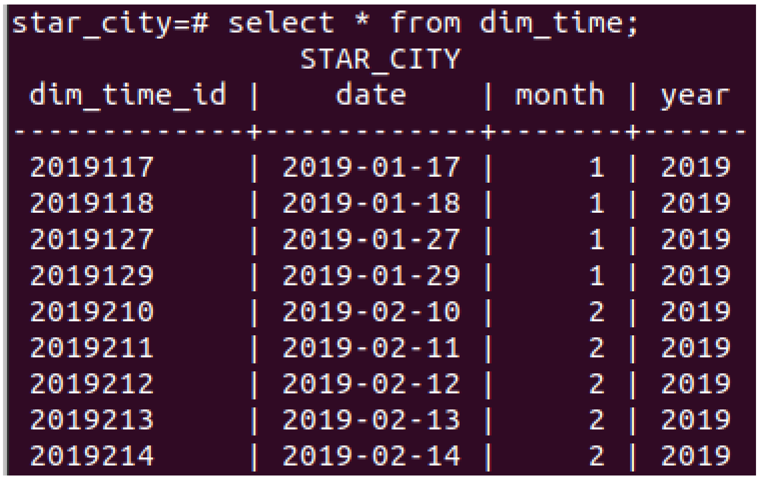
create table dim_room as select distinct number as room_number, name as name from room;

Tuples:



```
star_city=# select * from dim_room;
       STAR_CITY
 room_number |   name
-------------+----------
           1 | Room-001
           9 | Room-009
           4 | Room-004
           2 | Room-002
          10 | Room-010
           6 | Room-006
           7 | Room-007
           5 | Room-005
           3 | Room-003
           8 | Room-008
```

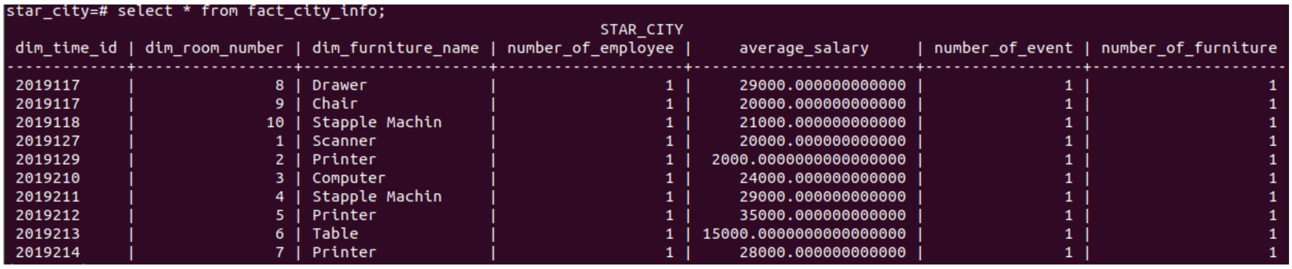
```
create table dim_time as select distinct  concat(date_part('year', date)::text, date_part('month',
date)::text, date_part('day', date)::text) as dim_time_id, date, date_part('month', date) as month,
date_part('year', date) as year from event;
```



```
star_city=# select * from dim_time;
                STAR_CITY
 dim_time_id |    date    | month | year
-------------+------------+-------+------
 2019117     | 2019-01-17 |     1 | 2019
 2019118     | 2019-01-18 |     1 | 2019
 2019127     | 2019-01-27 |     1 | 2019
 2019129     | 2019-01-29 |     1 | 2019
 2019210     | 2019-02-10 |     2 | 2019
 2019211     | 2019-02-11 |     2 | 2019
 2019212     | 2019-02-12 |     2 | 2019
 2019213     | 2019-02-13 |     2 | 2019
 2019214     | 2019-02-14 |     2 | 2019
```

```
create table fact_city_info as select concat(date_part('year', date)::text, date_part('month',
date)::text, date_part('day', date)::text) as dim_time_id, room.number dim_room_number,
furniture.name as dim_furniture_name, count(distinct employee.id) as number_of_employee,
avg(employee.salary) as average_salary, count(event.*) as number_of_event, count(furniture.name)
as number_of_furniture  from room inner join event on room.number = event.roomnumber inner
join furniture on room.furniturename = furniture.name inner join employee on room.employeeid =
employee.id group by concat(date_part('year', date)::text, date_part('month', date)::text,
date_part('day', date)::text), room.number, furniture.name;
```

Tuples:

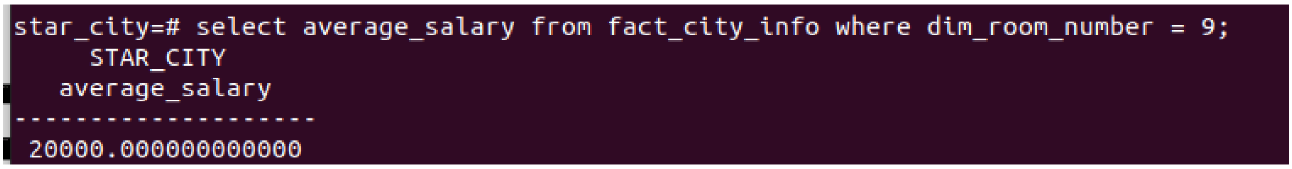


```
star_city=# select * from fact_city_info;
                                                  STAR_CITY
 dim_time_id | dim_room_number | dim_furniture_name | number_of_employee |      average_salary      | number_of_event | number_of_furniture
-------------+-----------------+--------------------+--------------------+--------------------------+-----------------+---------------------
 2019117     |               8 | Drawer             |                  1 |     29000.000000000000 |               1 |                   1
 2019117     |               9 | Chair              |                  1 |     20000.000000000000 |               1 |                   1
 2019118     |              10 | Stapple Machin     |                  1 |     21000.000000000000 |               1 |                   1
 2019127     |               1 | Scanner            |                  1 |     20000.000000000000 |               1 |                   1
 2019129     |               2 | Printer            |                  1 |  2000.0000000000000000 |               1 |                   1
 2019210     |               3 | Computer           |                  1 |     24000.000000000000 |               1 |                   1
 2019211     |               4 | Stapple Machin     |                  1 |     29000.000000000000 |               1 |                   1
 2019212     |               5 | Printer            |                  1 |     35000.000000000000 |               1 |                   1
 2019213     |               6 | Table              |                  1 | 15000.0000000000000000 |               1 |                   1
 2019214     |               7 | Printer            |                  1 |     28000.000000000000 |               1 |                   1
```

Queries from task 1:

Original query: SELECT AVG(Salary) FROM Employee WHERE ID IN
(SELECT EmployeeID FROM Room WHERE Number = 9);

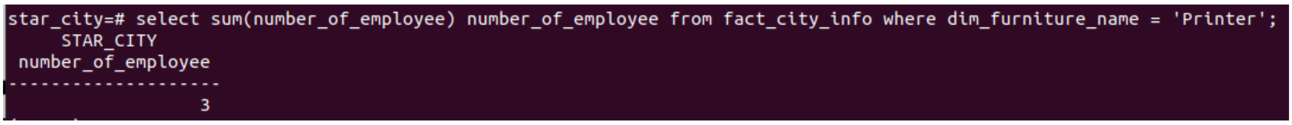
Adapted query: select average_salary from fact_city_info where dim_room_number = 9;



```
star_city=# select average_salary from fact_city_info where dim_room_number = 9;
     STAR_CITY
   average_salary
--------------------
 20000.000000000000
```

Original query: SELECT COUNT(EmployeeID) FROM Room WHERE FurnitureName = 'Printer';

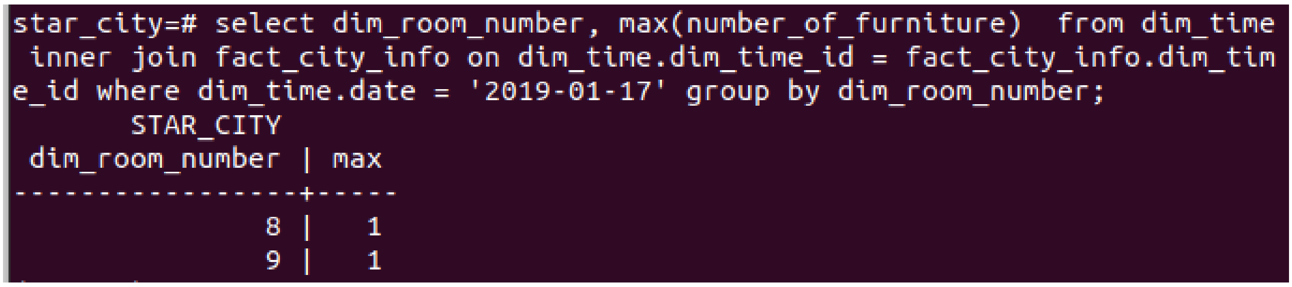
Adapted query: select sum(number_of_employee) number_of_employee from fact_city_info where dim_furniture_name = 'Printer';



```
star_city=# select sum(number_of_employee) number_of_employee from fact_city_info where dim_furniture_name = 'Printer';
     STAR_CITY
 number_of_employee
--------------------
                  3
```

Original query: SELECT Number RoomNumber, COUNT(FurnitureName)
NumberOfMaxFurniture FROM Room
WHERE Number IN (SELECT RoomNumber FROM Event WHERE Date = '2019-01-17')
GROUP BY Number ORDER BY 2 DESC LIMIT 1;

Adapted query: select dim_room_number, max(number_of_furniture) from dim_time inner join fact_city_info on dim_time.dim_time_id = fact_city_info.dim_time_id where dim_time.date = '2019-01-17' group by dim_room_number;



```
star_city=# select dim_room_number, max(number_of_furniture)  from dim_time
 inner join fact_city_info on dim_time.dim_time_id = fact_city_info.dim_tim
e_id where dim_time.date = '2019-01-17' group by dim_room_number;
       STAR_CITY
 dim_room_number | max
-----------------+-----
               8 |   1
               9 |   1
```

The other two queries are not statistical so has been omitted to adapt on data warehousing query.

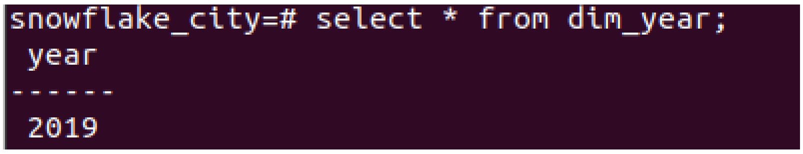
Snowflake schema database:

create database snowflake_city;

\c snowflake_city;

create table dim_year as select distinct date_part('year', date) as year from event;

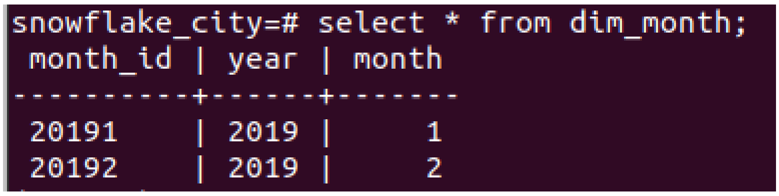
Tuples:



```
snowflake_city=# select * from dim_year;
 year
------
 2019
```

create table dim_month as select distinct concat(date_part('year', date)::text, date_part('month', date)) as month_id, date_part('year', date) as year, date_part('month', date) as month from event;

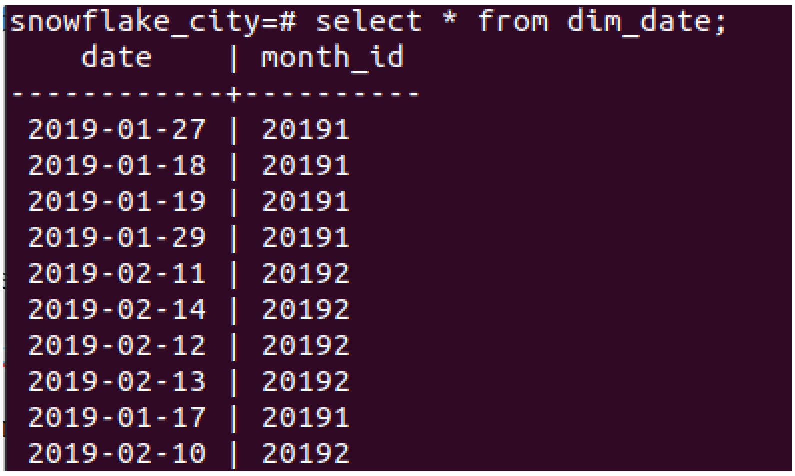
Tuples:



```
snowflake_city=# select * from dim_month;
 month_id | year | month
----------+------+-------
 20191    | 2019 |     1
 20192    | 2019 |     2
```

create table dim_date as select distinct date, concat(date_part('year', date)::text, date_part('month', date)) as month_id from event;

Tuples:

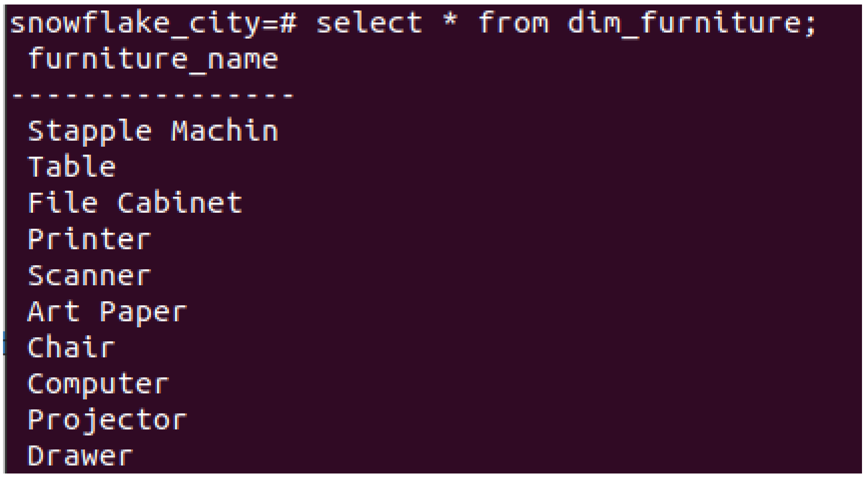


```
snowflake_city=# select * from dim_date;
    date     | month_id
------------+----------
 2019-01-27 | 20191
 2019-01-18 | 20191
 2019-01-19 | 20191
 2019-01-29 | 20191
 2019-02-11 | 20192
 2019-02-14 | 20192
 2019-02-12 | 20192
 2019-02-13 | 20192
 2019-01-17 | 20191
 2019-02-10 | 20192
```

creating dim_furniture and dim_room table is same as the star schema.

create table dim_furniture as select distinct name as furniture_name from furniture;

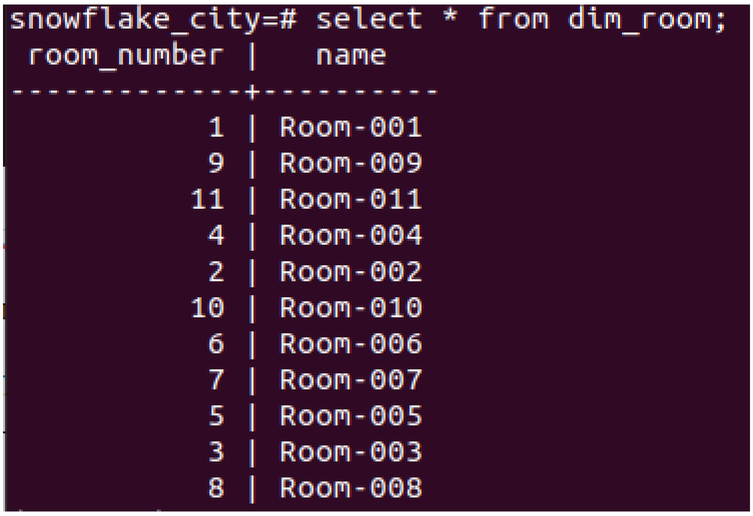
Tuples:



```
snowflake_city=# select * from dim_furniture;
 furniture_name
----------------
 Stapple Machin
 Table
 File Cabinet
 Printer
 Scanner
 Art Paper
 Chair
 Computer
 Projector
 Drawer
```

create table dim_room as select distinct number as room_number, name as name from room;

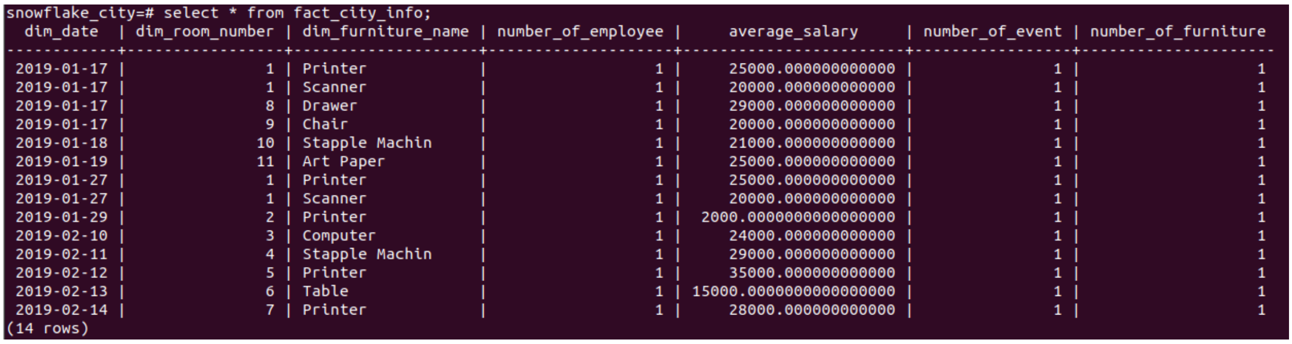
Tuples:



```
snowflake_city=# select * from dim_room;
 room_number |   name
-------------+----------
           1 | Room-001
           9 | Room-009
          11 | Room-011
           4 | Room-004
           2 | Room-002
          10 | Room-010
           6 | Room-006
           7 | Room-007
           5 | Room-005
           3 | Room-003
           8 | Room-008
```

create table fact_city_info as select event.date as dim_date, room.number dim_room_number, furniture.name as dim_furniture_name, count(distinct employee.id) as number_of_employee, avg(employee.salary) as average_salary, count(event.*) as number_of_event, count(furniture.name) as number_of_furniture from room inner join event on room.number = event.roomnumber inner join furniture on room.furniturename = furniture.name inner join employee on room.employeeid = employee.id group by event.date, room.number, furniture.name;

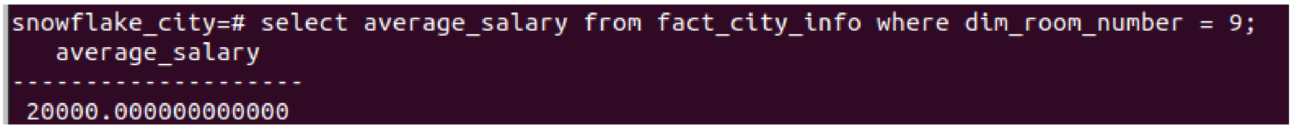
Tuples:



```
snowflake_city=# select * from fact_city_info;
  dim_date  | dim_room_number | dim_furniture_name | number_of_employee |     average_salary     | number_of_event | number_of_furniture
------------+-----------------+--------------------+--------------------+------------------------+-----------------+---------------------
 2019-01-17 |               1 | Printer            |                  1 |     25000.000000000000 |               1 |                   1
 2019-01-17 |               1 | Scanner            |                  1 |     20000.000000000000 |               1 |                   1
 2019-01-17 |               8 | Drawer             |                  1 |     29000.000000000000 |               1 |                   1
 2019-01-17 |               9 | Chair              |                  1 |     20000.000000000000 |               1 |                   1
 2019-01-18 |              10 | Stapple Machin     |                  1 |     21000.000000000000 |               1 |                   1
 2019-01-19 |              11 | Art Paper          |                  1 |     25000.000000000000 |               1 |                   1
 2019-01-27 |               1 | Printer            |                  1 |     25000.000000000000 |               1 |                   1
 2019-01-27 |               1 | Scanner            |                  1 |     20000.000000000000 |               1 |                   1
 2019-01-29 |               2 | Printer            |                  1 |  2000.0000000000000000 |               1 |                   1
 2019-02-10 |               3 | Computer           |                  1 |     24000.000000000000 |               1 |                   1
 2019-02-11 |               4 | Stapple Machin     |                  1 |     29000.000000000000 |               1 |                   1
 2019-02-12 |               5 | Printer            |                  1 |     35000.000000000000 |               1 |                   1
 2019-02-13 |               6 | Table              |                  1 | 15000.0000000000000000 |               1 |                   1
 2019-02-14 |               7 | Printer            |                  1 |     28000.000000000000 |               1 |                   1
(14 rows)
```

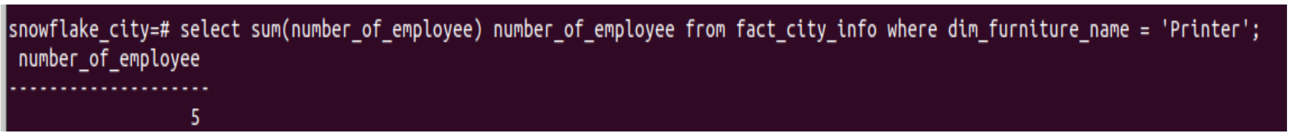
Adapted queries from Task 1:

select average_salary from fact_city_info where dim_room_number = 9;



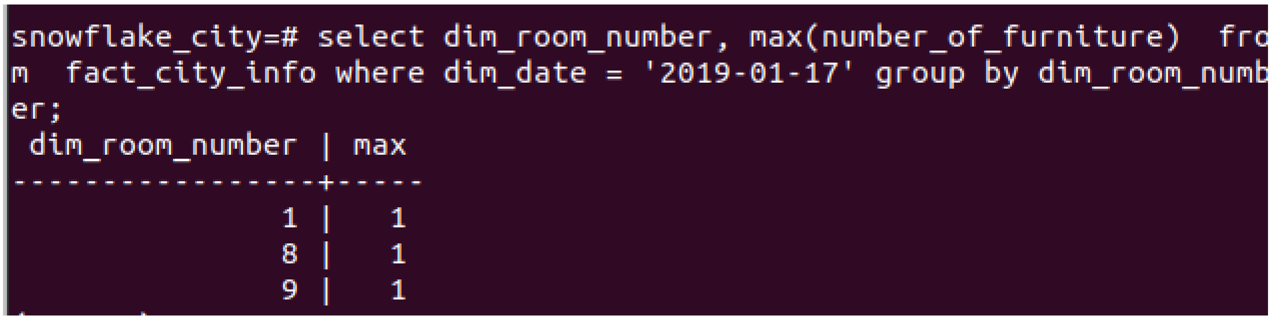
```
snowflake_city=# select average_salary from fact_city_info where dim_room_number = 9;
   average_salary
--------------------
 20000.000000000000
```

select sum(number_of_employee) number_of_employee from fact_city_info where dim_furniture_name = 'Printer';



```
snowflake_city=# select sum(number_of_employee) number_of_employee from fact_city_info where dim_furniture_name = 'Printer';
 number_of_employee
--------------------
                  5
```

select dim_room_number, max(number_of_furniture) from fact_city_info where dim_date = '2019-01-17' group by dim_room_number;



```
snowflake_city=# select dim_room_number, max(number_of_furniture)  fro
m  fact_city_info where dim_date = '2019-01-17' group by dim_room_numb
er;
 dim_room_number | max
-----------------+-----
               1 |   1
               8 |   1
               9 |   1
```