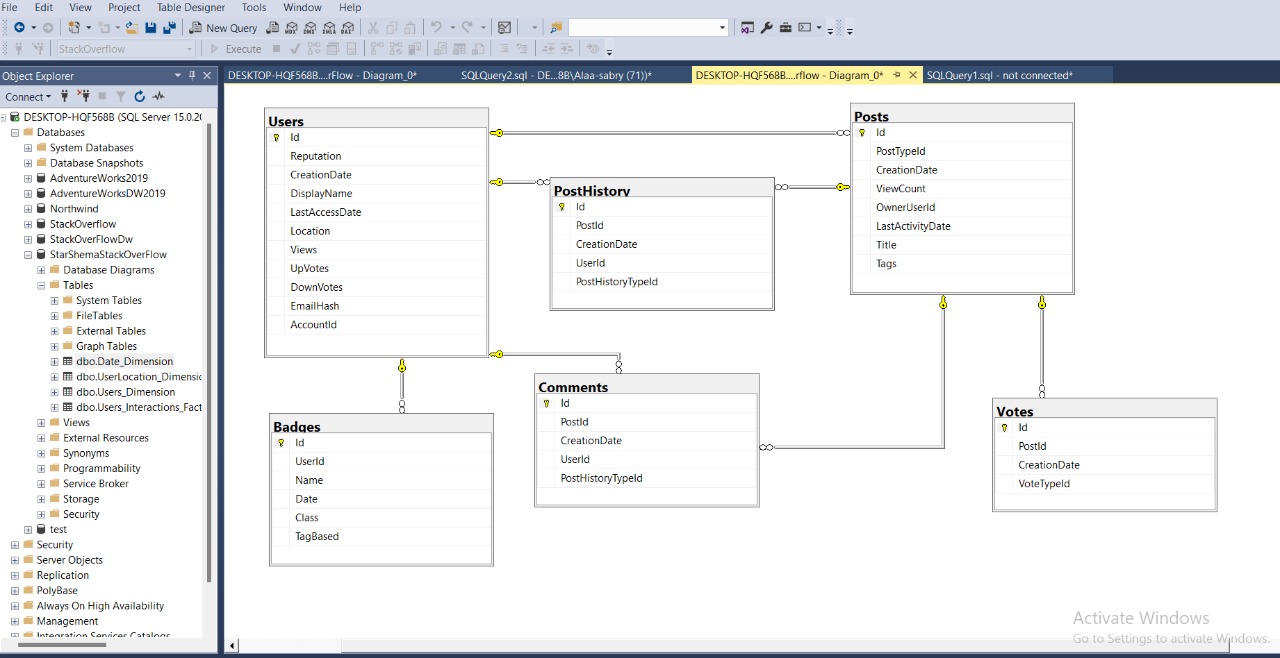


|  |  |  |  |
| --- | --- | --- | --- |
| names | id | emails | Group |
| Hassan Ragab | 20190177 | hassanragab776@gmail.com | 5 |
| Alaa Sabry | 20190335 | am5814476@gmail.com | 5 |
| Zainab medhat | 20170454 |  |  |
| Nouran ahmed | 20190596 |  |  |
| Shammakh salah | 20190793 |  |  |

DB Source Name: Stack Over Flow

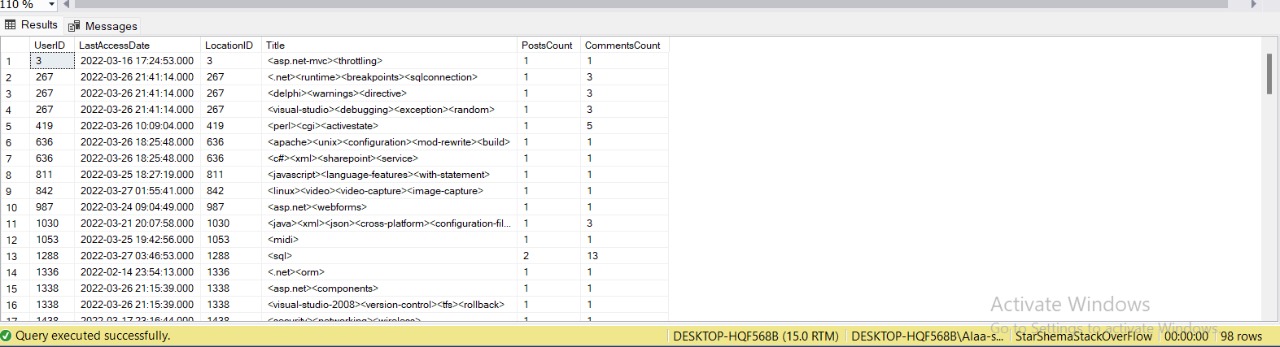
DB Source Link : https://data.stackexchange.com/stackoverflow/query/new

Stack OverFlow ERD

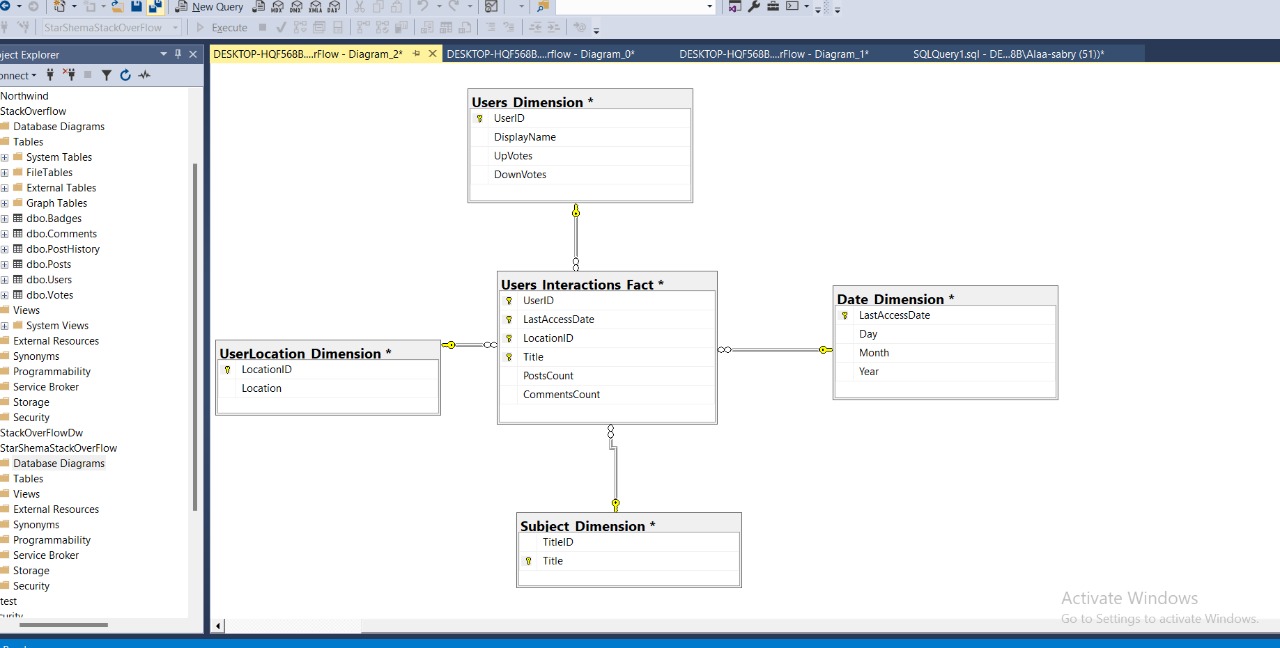


The motivation:

We're making the Stack Over Flow star schema in order to analyse and count the amount of posts and comments that a user made at a certain time and in a specific location on a specific topic. We'd be able to examine how users interact with one another in a published challenge over a period of time. The fact table shown below allows us to examine each user's actions over time.



Star Schema Model



Schema description

1. Dimensions

We decided to make for dimensions\_ Date dimension, Users dimensions, User Location dimension and Subject dimension, each one of them has attributes related to the schema`s measures as shown above.

a) Subject\_Dimension

select Id as TitleID , Tags as Title from posts

Transform posts data (Id, Tags) to Subject dimension (TitleId, Title)

b) UserLocation\_Dimension

select Users.Id as LocationID , Location from Users

Transform user location data (Users.Id, Location) to UserLocation dimension (LocationID, Location)

c) Users\_Dimension

select Users.Id as UserID , DisplayName ,UpVotes , DownVotes from Users

Transform user data (Users.Id, DisplayName, UpVotes, DownVotes) to Users dimension (UserId, Name , UpVotes, DownVotes)

d) Date\_Dimension

Select distinct (LastAccessDate), DAY(LastAccessDate) AS Day, MONTH(LastAccessDate) AS Month, YEAR(LastAccessDate) AS Year FROM Users Order by LastAccessDate

Transform Users date(LastAccessDate DAY(LastAccessDate), MONTH(LastAccessDate) YEAR(LastAccessDate)) to the Date dimension(LastAsccessDate, Day, Month, Year). The dimension levels are day, month year.

1. Fact Table

• Users\_Interaction\_Fact

select UserID1 AS UserID ,Date1 as LastAccessDate , LocationID ,Subject ,PostsCount,CommentsCount from

(select Posts.OwnerUserId as UserID1 , Posts.Tags as Subject ,Users.Location as location1 , Users.LastAccessDate as Date1 ,count(Posts.Tags) as PostsCount

from Posts , Users

where Posts.Tags != 'NULL'

Group by OwnerUserId , Posts.Tags,Users.Location ,Users.LastAccessDate) x1

INNER JOIN

(select Users.Id as UserID2 , Users.Id as LocationID , Users.Location as location2, Users.LastAccessDate as Date2 , COUNT(Comments.CreationDate) AS CommentsCount

from Users INNER JOIN Comments ON Comments.UserId = Users.Id

Group by Users.Id ,Users.LastAccessDate, Users.Location) x2

ON x1.UserID1 = x2.UserID2 and x1.location1 = x2.location2 and x1.Date1 = x2.Date2

In the fact table, we have a foreign key for each dimension table, and we have two measures (postCount and commentCount) to be able to analyze the user actions.