Faculty of Computers and Artificial Intelligence

Cairo University

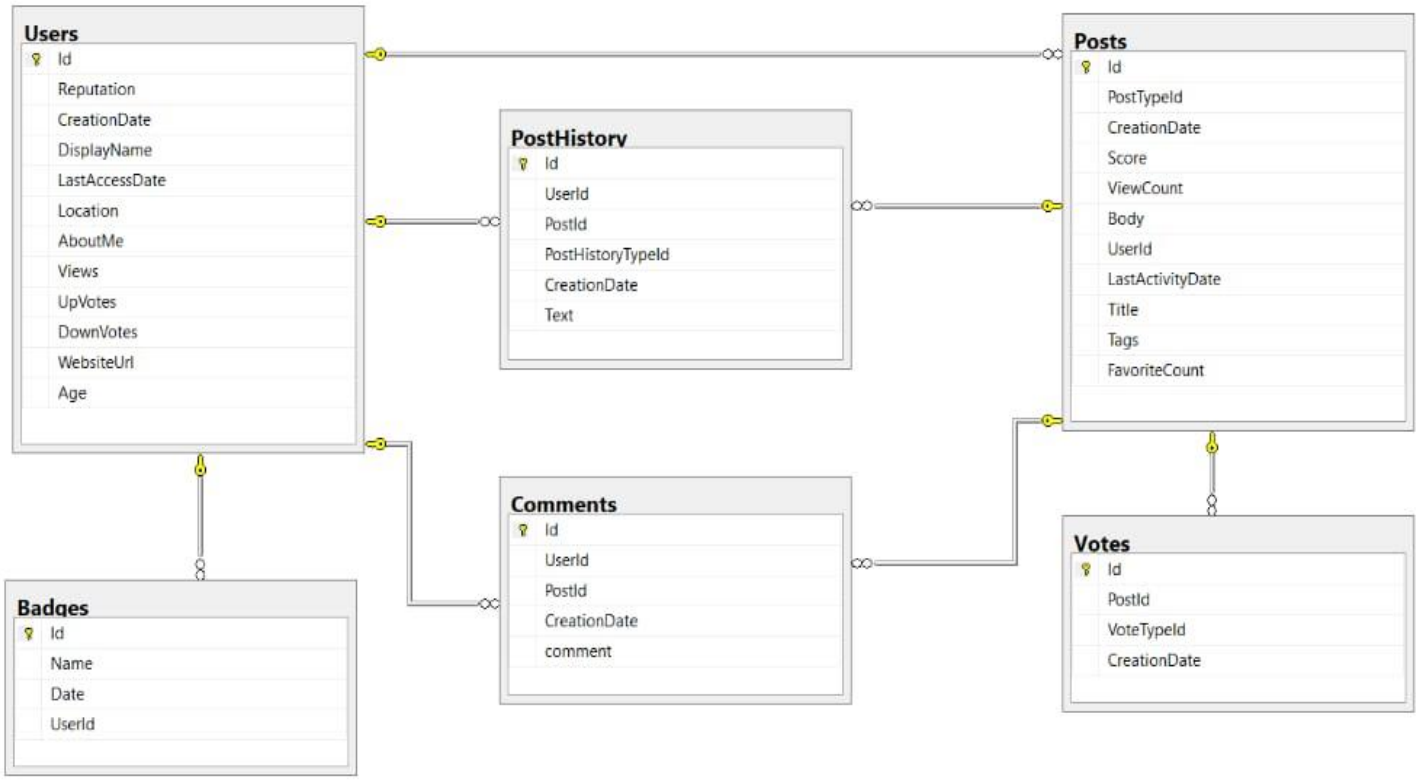
Department: Information Systems

Semester: Second Semester

Course Name: Data Warehousing **(IS313 and IS422)**

Name	ID	Email
Omneya Eid Mohamed	20180054	omneyaeid827@gmail.com
Sohila Abdalla Atea Abdalla	20180127	Sohilaabdalla113@gmail.com
Islam Ramadan Abdo	20180044	
Mahasen Mohamed Ahmed	20180211	

Stack OverFlow ERD

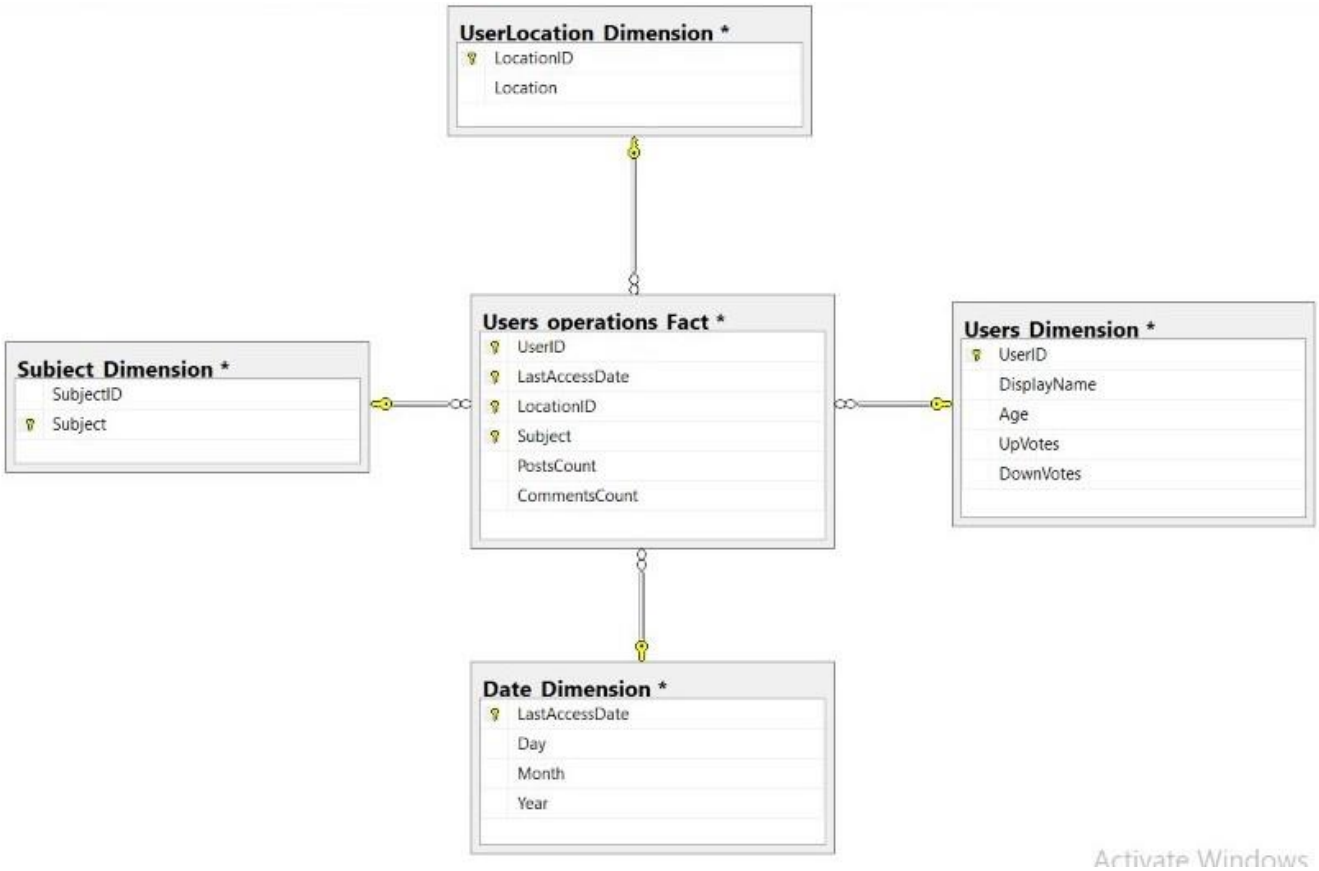


The motivation:

We are creating the Stack Over Flow star schema so that we creating this schema for analyzing and count the number of post and comment that user done in specific time at specific place in specific topic. we would be able to analyze the way users interact with each other within a published problem in an interval of time. The Below table is the fact table, we can analyze each user`s actions with
A specific time.

	UserID	LastAccessDate	Locatio...	Subject	PostsCount	CommentsCo...
1	351	2012-12-05T14:38:31.410	351	<mysql><performance>< storage>	1	3
2	171	2013-09-06T02:25:45.260	171	<sql-serv er-2008><sql-server><shrink>	1	215
3	174	2013-05-14T13:45:40.800	174	< explain-p lan><optimization>	1	15
4	120	2013-08-15T05:34:08.687	120	<aggregate><oracle10g-r2>	1	15
5	44	2012-10-07T02:55:59.473	44	<architecture><database-engine><database-structu...	1	4
6	2660	2013-09-06T04:00:54.977	2660	<backup><duplication><sql-azure>	1	20
7	81	2013-09-06T10:23:50.143	81	<backup><sql-server-2008-r2><database-administr...	1	42
8	521	2013-03-04T14:54:38.543	521	<backup><sql-server-2008-r2><disk-space>	1	5
9	1342	2013-09-05T17:46:55.633	1342	<best-practices><db2><etl>	2	4
10	180	2013-09-05T20:49:39.607	180	<best-practices><development>	1	35
11	521	2013-03-04T14:54:38.543	521	<best-practices><join>	1	5
12	1342	2013-09-05T17:46:55.633	1342	<best-practices><physical-design><multi-tenant>	1	4
13	81	2013-09-06T10:23:50.143	81	<best-practices><scripting>	1	42
14	6219	2013-09-06T10:23:05.897	6219	<best-practices><unit-test>	1	1
15	12	2012-01-11T21:57:29.910	12	<books><lear ning>	1	4

Star Schema Model



Schema description

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.

- **Subject_Dimension**

  select Id as TitleID , Title from posts
  Transform posts data (Id, Tags) to Subject dimension (SubjectID, Subject)

- **UserLocation_Dimension**

  select Users.Id as LocationID , Location from Users

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

- **Users_Dimension**

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

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

- **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.

2) Fact Table

- Users_Operations_Fact

```
select UserID1 AS UserID ,Date1 as LastAccessDate ,
LocationID ,Subject ,PostsCount,CommentsCount  from
(select Posts.UserId 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 UserId , 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.