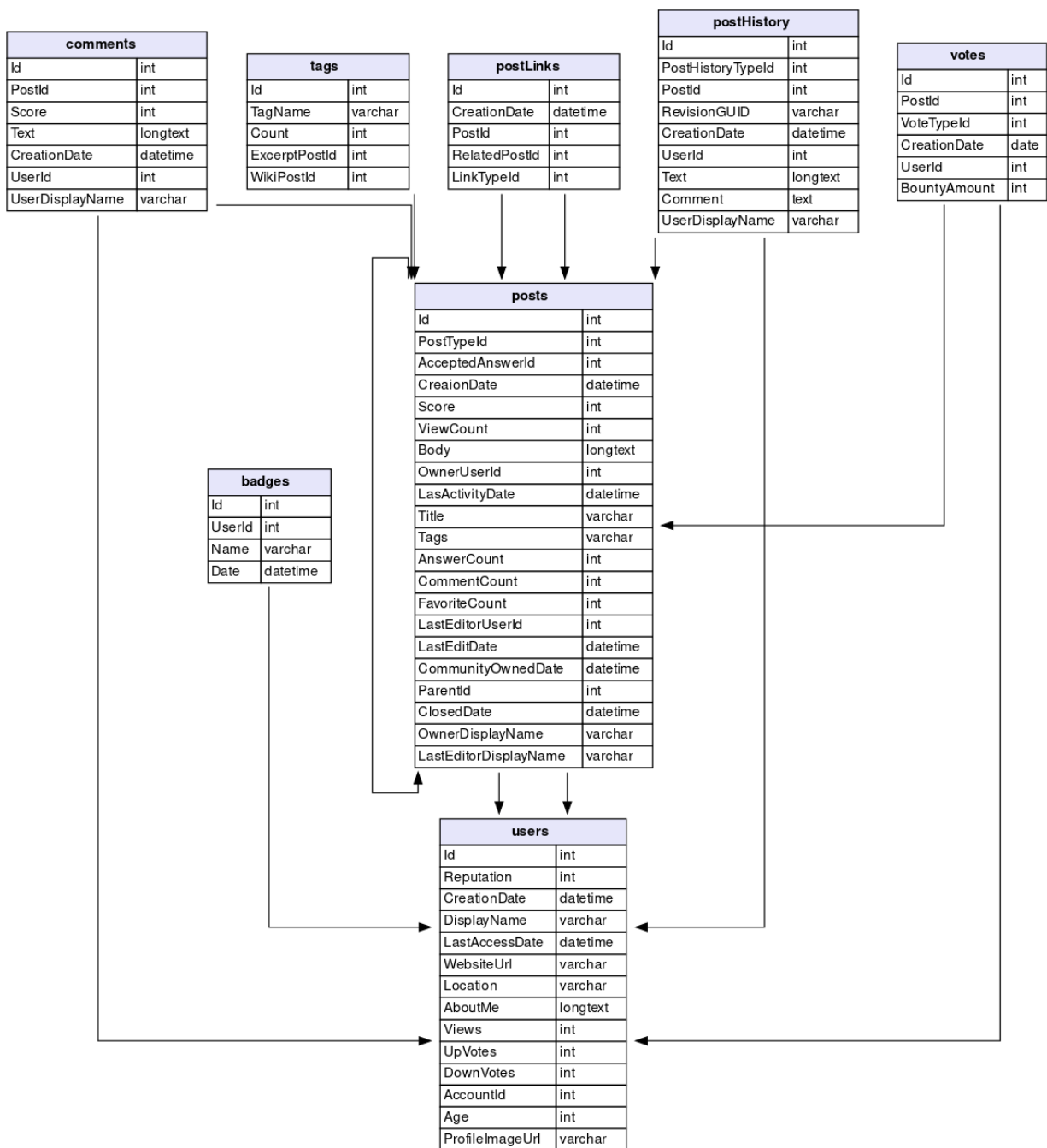


גישה לנתונים | - מסדי נתונים ושפת SQL

Databases for data analytics

מועד א - מילואימניקים

השאלות בבחינה תתייחסנה לסכמה של [StackOverflow](https://stackoverflow.com) המובאת בתרשים למטה.



1. שליפות

במהלך הבחינה מותר, ואפילו רצוי, להשתמש בשאלות שכתבתם קודם. כדאי לעשות זאת אנא השתמשו ב views וציינו בשאלתה המשתמשת את המקום בו הגדרתם אותה.

במידה ויש סעיף שהתקשתם בפתרונו, אתם יכולים לציין שם של view שפותר אותו, לכתוב מה השדות בו, ולהשתמש בו בסעיפים אחרים.

יש בבחינה 20 סעיפים, משקל כל סעיף 5 נקודות. בנוסף יש **בונסים** של חמש נקודות האחד שיכולים רק להעלות את הציון.

- a. שילפו את מספר ה tag ששמו 'Sql'
- b. שילפו את ה comments שה score שלהם הוא 100
- c. שילפו את ה comments שה text שלהם מכיל את המילה 'Sql'. יש לאתר את המילה בכל case אך לא בתוך מילה כמו MySQL. לצורך התרגיל ניתן להניח שמילה מוקפת ברווחים, מתחילה טקסט או מסיימת אותו.
- d. בטבלת posts יש שדה tags. השדה מכיל רשימה של אינדקסי tags (לדוגמה [4,7], [82]). שילפו את כל ה post שיש להם את ה tag מספר 7. שימו לב לא לשלוף את אלו עם 17 או 70.
- e. מיצאו את ה users ה"שנויים במחלוקת". אלו users שיש להם יותר מ 100 UpVotes ו 100 DownVotes ושאוחד ה UpVotes מכלל ה votes שלהם הוא בין 45 ל 55. **בונס:** שילפו, עבור ה users שעומדים במספר ה votes המינימלי את ממוצע ה UpVotes לשנויים במחלוקת ולשאינם כאלה.
- f. שילפו את כל ה post של OwnerUserId מספר 7, מסודרים לפי CreationDate
- g. שילפו את כל ה votes של ה Posts של OwnerUserId מספר 7, מסודרים לפי CreationDate של ההצבעה בסדר יורד
- h. שילפו את כל ה posts שה OwnerDisplayName שלהם אינו זהה לזה של ה user שהוא ה OwnerUserId שלהם. **בונס:** הכפילות בהחזקת OwnerDisplayName היא פתח לבעיות. תנו נימוק אפשרי לבחירה להשתמש בכפילות.
- i. שילפו את עשרת ה tags עם מספר ה count הגבוה ביותר. אין חשיבות לסדר הבחירה במקרה של שוויון ב count
- j. שילפו את מספר ה users
- k. שילפו את ה score הממוצע של post
- l. עבור כל score שילפו את מספר ה posts שקיבלו אותו, ממוינים לפי score
- m. שילפו את כל ה WebsiteUrl שהופיעו אצל יותר מ user יחיד. **בונס:** האם כדאי למנוע אפשרות זאת? אם כן, כיצד ניתן לממש זאת?
- n. שילפו את ה users שהם לא ה owners של אף post
- o. שילפו את כל ה users שאף post שלהם לא קיבל אף vote
- p. שילפו את מספר ה comments לכל post, בעזרת טבלת comments
- q. עבור posts שקיבלו הערות, שילפו את ה comments per post distributions.
- r. כלומר, מספר הפעמים ש post קיבל הערה 1, 2, 3 וכדומה. הרחיבו את הסעיף הקודם ל posts שלא קיבלו הערות.

- s. הציגו את זוגות ה post ids שיותר מ 10 אנשים זהים הצביעו לשניהם (כלומר אותו אדם הצביע לשני ה posts). שילפו באופן סימטרי, כלומר (a,b) וגם (b,a) בנוסף: האם כדאי לשלוף זוגות רפלקסיבים (כלומר post עם עצמו)?
- t. בטבלה posts יש שדה בשם CommentsCount. בנוסף ניתן לחשב את השדה בעזרת טבלת comments. כיתבו שאילתה המוצאת אי הסכמות בין השדה לטבלת comments. השתמשו בשאילתה מסעיף קודם המוצאת את מספר ה comments בעזרת טבלת comments.