

پروژه پایانی پایگاه داده (بخش Data Warehouse)

امیرحسین رجبی (۹۸۱۳۰۱۳)

توضیح طراحی فعلی

همانگونه که در فایل schema.sql آمده است برای هر رابطه در پایگاه داده مبدا چون $R(A_1, A_2, \dots, A_n)$ که primary key آن A_1, A_2, \dots, A_p باشند، رابطه های زیر را در انبار داده در نظر میگیریم:

$R_{history}(id, operation_timestamp, operation_type, reason)$

$R_{primary\ key}(history_id, A_1, A_2, \dots, A_p)$

$R_{all}(history_id, A_1, A_2, \dots, A_n)$

هنگام اضافه شدن یک رکورد به جدول R یک رکورد که مشخص کننده اطلاعات عملیات (مانند نوع و زمان و دلیل عملیات) به $R_{history}$ و یک رکورد که همه مقادیر رکورد جدید را دارد به R_{all} اضافه میشود. اگر رکوردی آپدیت شود، علاوه بر اضافه شدن رکورد به $R_{history}$ رکوردی شامل اطلاعات سطر دچار تغییر (primary key آن) به $R_{primary\ key}$ و رکوردی شامل مقادیر جدید ستون ها به R_{all} اضافه میشود. در صورت انجام عملیات حذف، علاوه بر اضافه شدن رکوردی به $R_{history}$ رکوردی شامل اطلاعات سطر حذف شده به $R_{primary\ key}$ اضافه میشود. همه اینها به کمک TRIGGER انجام شده است.

چگونه به گذشته سفر می کنیم؟

برای مشاهده پایگاه داده در زمان خاصی از تاریخ، همه رکورد های موجود در جداول $R_{history}$ را که زمان انجام عملیات کوچکتر از زمان مد نظر است را در نظر میگیریم و براساس ستون $operation_timestamp$ مرتب میکنیم و به ترتیب از اولین عملیات انجام شده تا آخرین عملیات آنها را در یک دیتابیس دیگر شبیه سازی میکنیم. در پایان وضعیت پایگاه داده جدید ساخته شده، همان وضعیت پایگاه داده مبدا در زمان مد نظر است. همچنین میتوان این عملیات را به صورت in place نیز انجام داد یعنی ابتدا همه رکورد های موجود در جداول $R_{history}$ را براساس ستون زمان عملیات $operation_timestamp$ مرتب میکنیم. سپس با شروع از آخرین عملیات و ادامه تا اولین عملیات با زمان بزرگتر از زمان مدنظر معکوس عملیات ها را در پایگاه داده اجرا کنیم. یعنی اگه عملیات insert بوده است، آن را حذف کنیم و اگر عملیات delete بود آن رکورد را اضافه کنیم. برای بازیابی مقادیر ستون ها میتوان با تغییری اندک در رابطه $R_{primary\ key}$ ، مقادیر قریم ستون ها را نگه داریم. در نتیجه رکورد با primary key را به جدول اضافه کرده و مقادیر باقی ستون ها رو از مقادیر ذخیره شده در این رکورد بازیابی می کنیم. اگه عملیات update باشد. رکورد با مشخصات در رابطه نظیر R_{all} به رکورد با مشخصات نظیر $R_{primary\ key}$ تبدیل می شود. در صورت تغییر نکردن primary key این یک آپدیت خواهد بود. در غیر این صورت رکورد با اطلاعات در R_{all} حذف و رکوردی با اطلاعات

در $R_{primary\ key}$ اضافه میشود.

مدل داده ای

ارتباط بین سه رابطه $R_{history}$ ، $R_{primary\ key}$ و R_{all} را میتوان از نوع ارث بری تعبیر کرد: رابطه های $R_{primary\ key}$ و R_{all} هر دو از $R_{history}$ ارث می برند و این ارث بری به صورت $total$ و $overlapping$ است. زیرا هر رکورد باید یا $insert$ ، $delete$ و $update$ باشد و می توانند در هر دو رابطه قرار بگیرد مانند $update$. تعبیر غیر EER را در فایل `data_model.pdf` ببینید.