

جلسه 1:

مرور دیبی 1:

برای پروژه: هر فرد یک data mart باید داشته باشد

پروژه: طراحی و پیاده سازی یک انبار داده است

Stored procedure: جایی که می خواهیم یک عملیات داده ای تکراری انجام بدیم استفاده می کنیم

Stored procedure می تونیم داشته باشیم که فقط سلکت داخلش باشه؟ توی گزارش اگر بخوایم استفاده بکنیم اونجا کاربرد داره --> میشه به عبارتی این کارو کرد ولی ما به این فضا خیلی کاری نداریم توی این درس و ما Stored procedure به این معنا می خوایم استفاده بکنیم که داده ها رو بتونیم تغییر بدیم

اگر یک تغییر داده ای بخوایم انجام بدیم و بخوایم اینو تکرارش بکنیم --> اینجا با تکرارش کار داریم و Stored procedure وقتی که بخواد تکرار بشه چی کارش بکنیم؟ می تونیم توی job فراخوانیش بکنیم

job: مثلا همیشه ساعت 2 شب برو اون Stored procedure که می خواستم دستی انجامش بدم رو به صورت اتوماتیک انجامش بده

توی فضای انبار داده ایا به دونه Stored procedure داریم؟ میشه این کارو کرد ولی این کار خیلی منطقی نیست --> بسته به طراحی که انجام میدیم توی انبار داده که چه جدول هایی داریم معمولا اون ها تعیین کننده هستن که چه Stored procedure داشته باشیم

Lag lead:

مثل همون چیزی که توی آژ داشتیم که می اومد مثلا چندتا سطر قبل رو با هم می گرفت یا چندتا سطر بعدش رو که به این lag lead می گن

توی این می تونیم بحث پارتیشن بای هم داشته باشیم

مثلا 10 lag یی برای اون سطر باید 10 تا رکورد بره قبلش بگیره و جلوی این سطر قرار بده

اینجا حتما باید ORDER BY داشته باشیم

پارتیشن بای:

توی این تعداد رکوردها هیچ فرقی نمی کنه و همون می مونه و چیزی کم نمیشه
توی این فرض میکنیم که داخل جدول دانشجو 10000 رکورد دانشجو وجود داره بعد میایم
مثلا روی دپارتمنش پارتیشن می زنیم به ما 10000 تا سطر برمیگردونه
همین مثال برای گروپ بای باشه و روی دپارتمنتش گروپ بزیم 10 تا سطر به ما برمیگردونه

گروپ بای:

توی این تعداد رکورد های ما متاثر از گروپ بای است

توی این درس پرفرمنس خیلی مهم است

union all : پرفرمنس این بهتر است چون داخلش چیزی رو چک نمیکنه در مقایسه با union
در چه شرایطی از union all استفاده میکنیم؟ اگر مطمئن باشیم که دوتا جدول رکورد تکراری
ندارن یا تکرار برامون مهم نباشه

نکاتی که راجع به پرفرمنس مطرح میشه چه الان چه بعدا توی تمرین ها و پروژه ها می بینه
و اگر رعایت نشده باشه کم میکنه پس مهمه که از چی استفاده میکنیم

With

:join

فرق این که ما دوتا جدول رو جویین بکنیم و یک شرطی رو روی on بذاریم یا روی where بذاریم
فرقشون چیه؟ توی where شرطه چک میشه و اگر اون شرط برقرار نبود سطر توی خروجی
نمایش داده نمی شود ولی برای on این بستگی به جویین دارد

اگر یک دیتایی تکرار بشه توی خروجی این می تونه از چی باشه؟ یا به عبارت دیگر در چه
صورت هایی چه تیپ کوئری هایی داشته باشیم احتمال اینکه دیتای تکراری داشته باشیم
توی خروجی اون از چی هست یه مثلا انتظار ما 10 میلیون رکورد است ولی شده 15 میلیون
رکورد این بخاطر چیه؟ یکی از جاها جویین است --> جویین اگر می زنیم باید خیلی دقت بکنیم
مخصوصا توی فضای انبار داده که اگر داریم جویین می زنیم باید درست بشناسیم دیتاهامون
رو ممکنه که داده های ما اشکالی داخلشون وجود داشته باشه یا ممکنه داخل جویین اشکالی
وجود داشته باشه