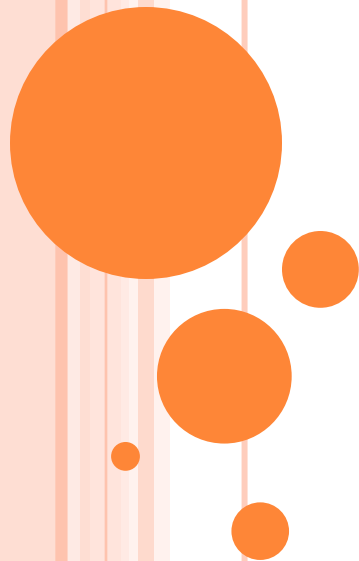


هوش تجاری

نقشه راه پروژه های هوش تجاری
(طراحی : مرحله چهارم)

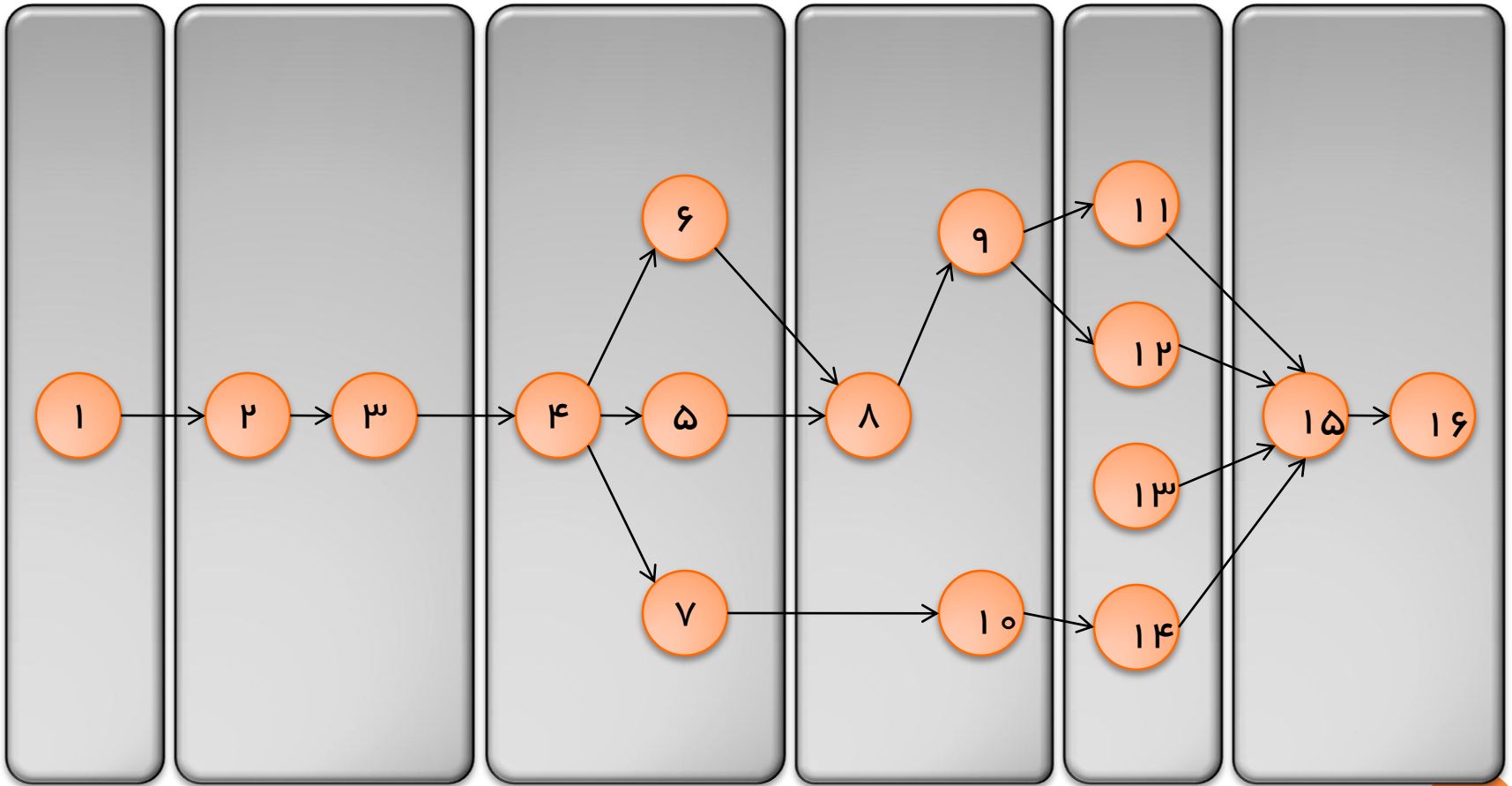


نقشه راه هوش تجاری

○ این نقشه ۱۶ گام برای سافت یک پروژه هوش تجاری را بر طبق همان ۶ مرحله انجام هر پروژه مهندسی که قبلاً ذکر شد، به شرح زیر بیان می کند:

- گام ۱: ارزیابی وضعیت تجاری
- گام ۲: ارزیابی زیرساخت سازمان
- گام ۳: طرح ریزی پروژه
- گام ۴: تعریف نیازمندی ها
- گام ۵: تحلیل داده ها
- گام ۶: نمونه سازی برنامه کاربردی
- گام ۷: تحلیل مؤزن فرا داده
- گام ۸: طراحی پایگاه داده عملیاتی
- گام ۹: طراحی ETL
- گام ۱۰: طراحی مؤزن فرا داده
- گام ۱۱: تولید ETL
- گام ۱۲: تولید برنامه کاربردی
- گام ۱۳: داده کاوی
- گام ۱۴: تولید مؤزن فرا داده
- گام ۱۵: پیاده سازی
- گام ۱۶: ارزیابی نسخه

نقشه راه هوش تجاری





گام ۱ (طراحی پایگاه داده

تفاوت پایگاه داده هوش تجاری با پایگاه داده های معمول

- گزارشات پایگاه داده‌ی عملیاتی ثابت و روتین است اما در هوش تجاری روزانه نیستند و از تنوع خیلی بالایی برخوردارند
- گزارشاتی که از هوش تجاری گرفته میشود به دلیل بزرگ بودن پروژه‌های هوش تجاری زمان پاسخ به این گزارشات خیلی طولانیتر خواهد بود
- در سیستم‌های پایگاه داده عملیاتی بیشتر بر روی کاهش افزونگی تمرکز میشود. در هوش تجاری با قرار گرفتن جداول مختلف کنار هم بدون شک نرمال سازی اتفاق نمیافتد
- در پایگاه داده‌های عملیاتی اطلاعات خیلی قدیمی نگهداری نمیشوند اما در DB هوش تجاری از داده‌های گذشته برای تحلیل بسیار استفاده میشود.
- معمولاً داده‌ها در پایگاه داده‌های عملیاتی خلاصه سازی شده نیستند اما در پایگاه هوش تجاری اعداد و ارقام خلاصه سازی شده اند

۱۸) طراحی پایگاه داده

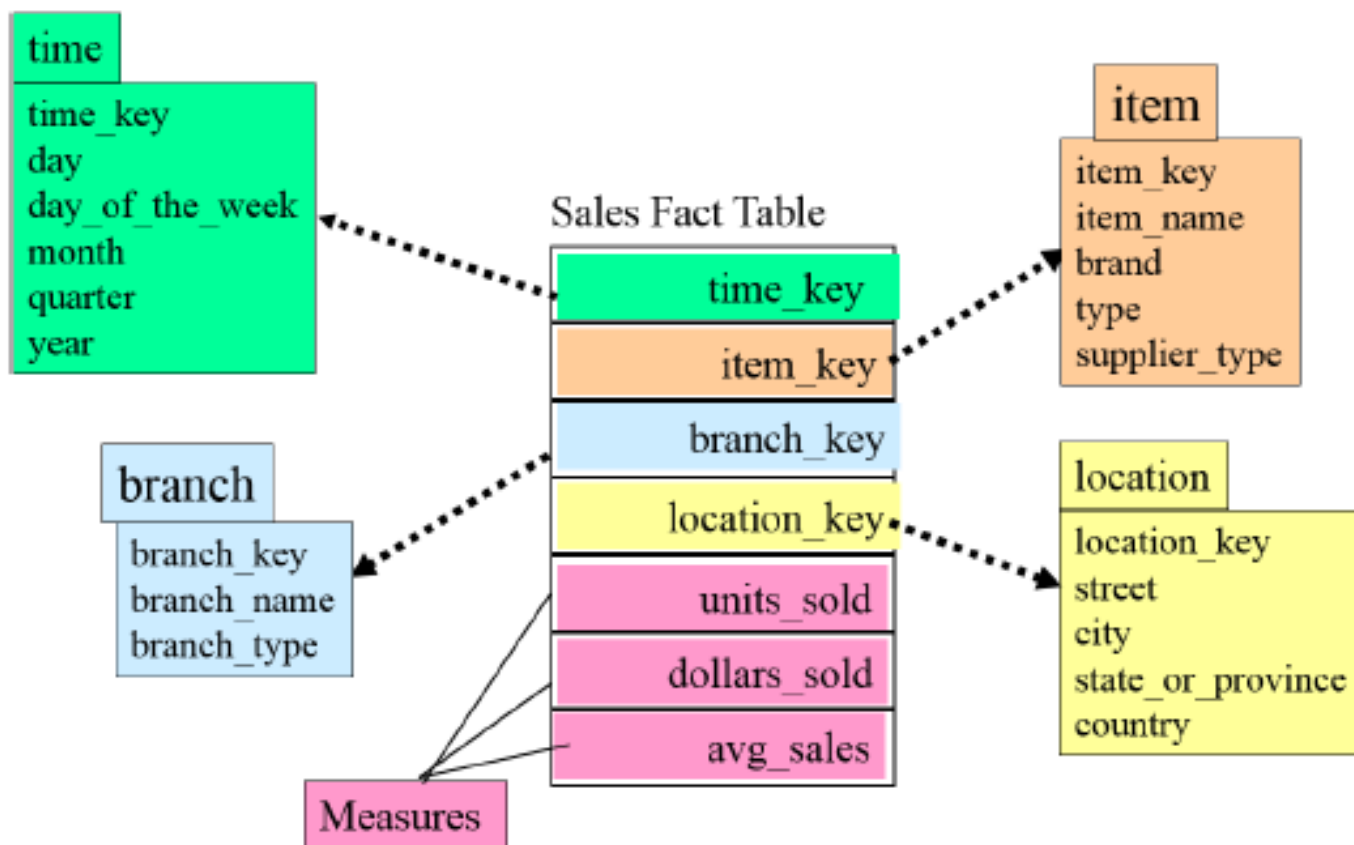
○ در این بخش پایگاه داده ی عملیاتی طراحی می شود.

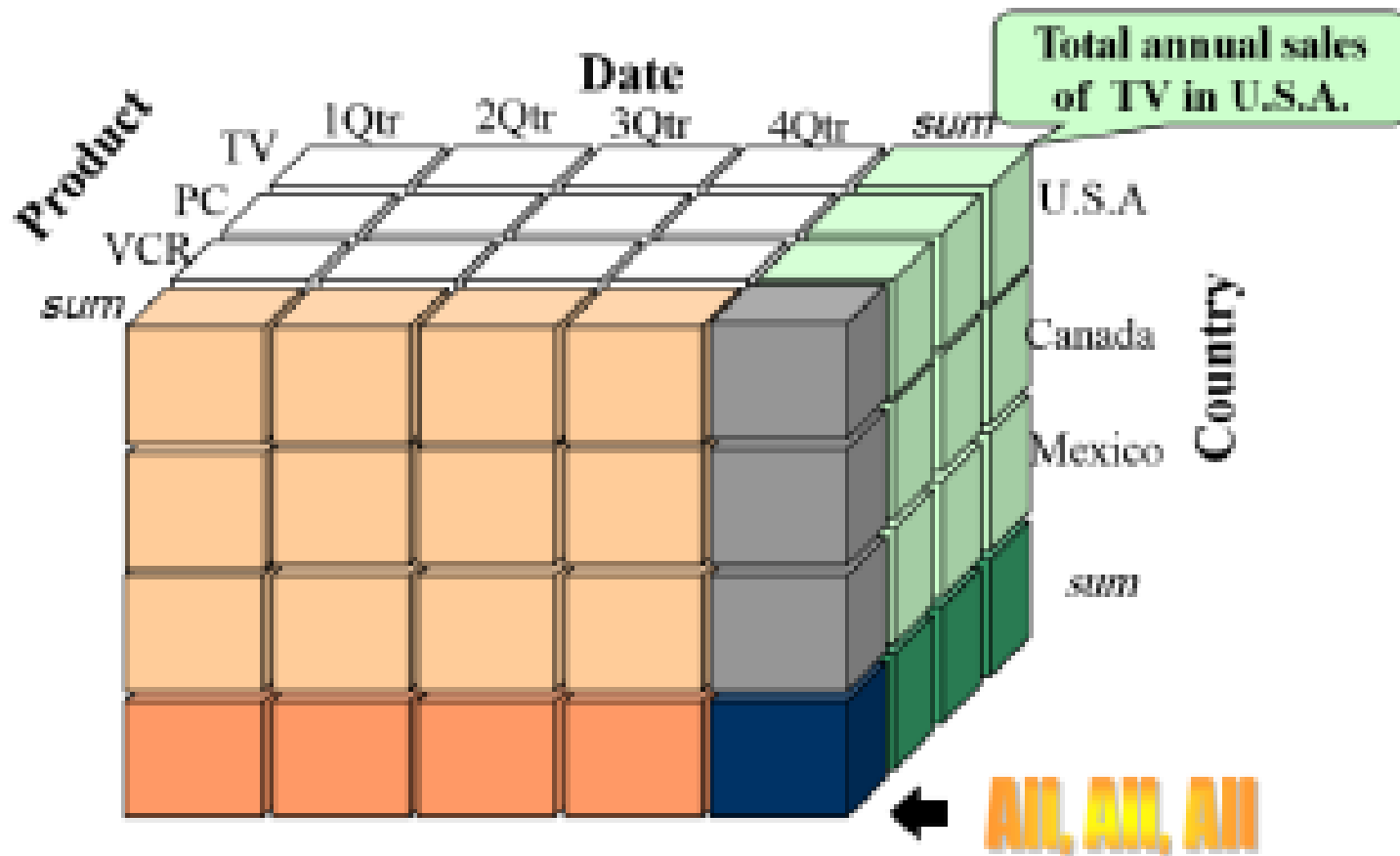
○ طراحی منطقی پایگاه داده:

- مدل ستاره ای
- مدل دانه برفی
- مدل مدل منظومه ای

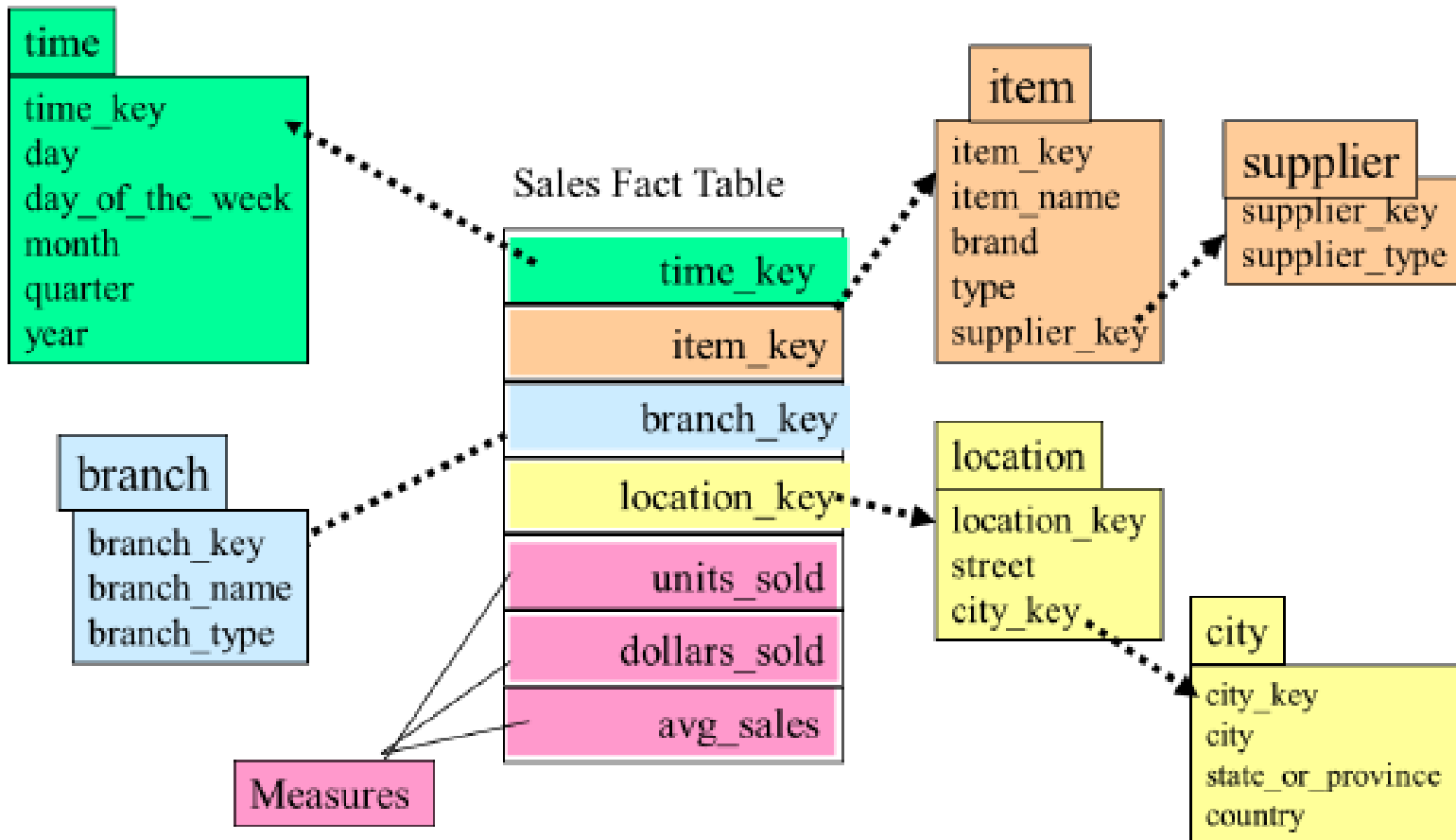
مدل ستاره ای

درشمای Star یک جدول در وسط قرار میگیرد که به آن (Fact Table) جدول اصلی (میگویند و بقیه جدولها دور آن قرار میگیرند)

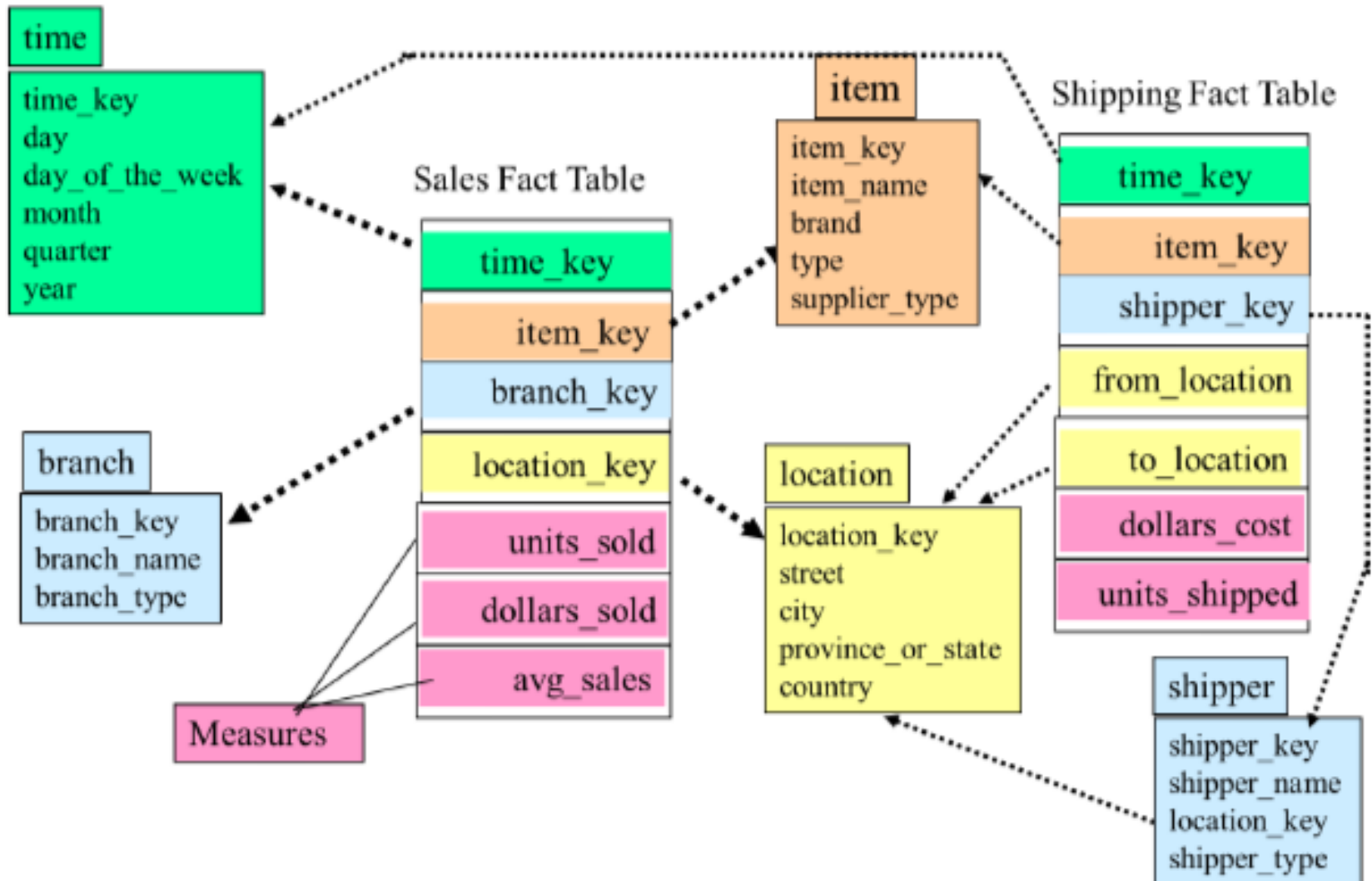




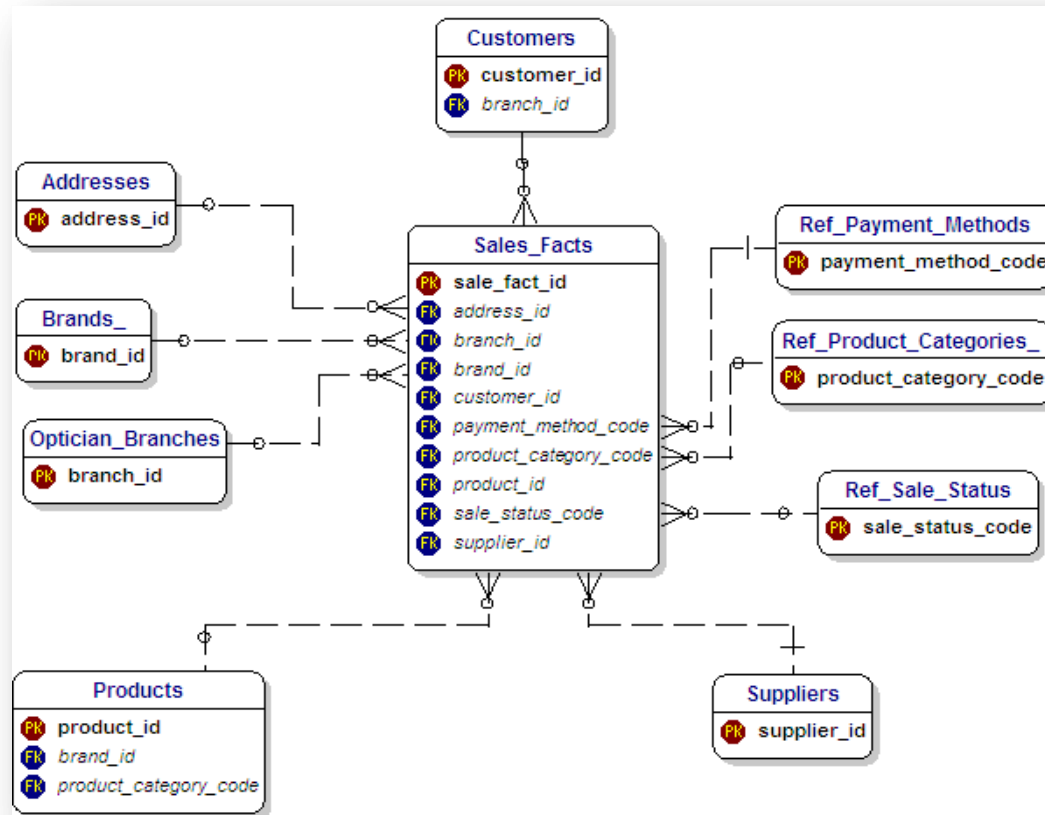
مدل برفگونه



مدل صورت فلکی



۱) طراحی پایگاه داده



○ برای طراحی پایگاه داده در این پروژه از مدل دانه برقی استفاده خواهیم نمود.

○ برای این منظور از *MySQL Server 2008* بهره خواهیم گرفت.

○ نمونه ای از مدل داده ای ایجاد شده در این پروژه.

۱۸) طراحی پایگاه داده

○ فعالیت های گام طراحی پایگاه داده عملیاتی:

- مرور نیازمندیهای دسترسی به داده
- مشخص کردن نیازمندیهای یکپارچه سازی و خلاصه سازی
- طراحی پایگاه داده عملیاتی
- طراحی سافتار فیزیکی پایگاه داده عملیاتی
- سافت پایگاه داده عملیاتی
- توسعه رویه های نگهداری پایگاه داده عملیاتی
- آماده سازی برای نظارت و تنظیم نمودن طراحی پایگاه داده عملیاتی
- آماده سازی برای نظارت و تنظیم نمودن طراحی پرس و جو

کام ۹ طراحی ETL

۹) طراحی ETL

- اطلاعات موجود در سیستم های عملیاتی سازمان، قابلیت استفاده جهت تصمیمات استراتژیک را ندارند.
- این اطلاعات می بایست تغییر شکل یافته و وارد سیستم DW شوند تا قابلیت استفاده جهت تصمیم گیری های استراتژیک را پیدا نمایند.
- عملیات انجام شده در ETL اطلاعات موجود در سیستم های عملیاتی سازمان را استخراج می کنند و آنها را به شکل و قالبی که قابل استفاده در DW شود، تبدیل کرده و تغییرات لازم را بر روی آنها اعمال می کنند.
- در صورتی که اطلاعات به نحوه صمیمی استخراج نشده و یا عملیات *cleansing* به درستی بر روی آنها صورت نگیرد، زیر ساخت سیستم DW دچار افتلال خواهد شد.

۹) طراحی ETL

○ تحت عملیات ETL اطلاعات از سیستم‌های عملیاتی سازمان استخراج شده، به شکل قابل استفاده در DW تبدیل شده و در آن بارگذاری می‌شوند.

○ عملیات ETL در DW شامل سه فعالیت می‌باشد:

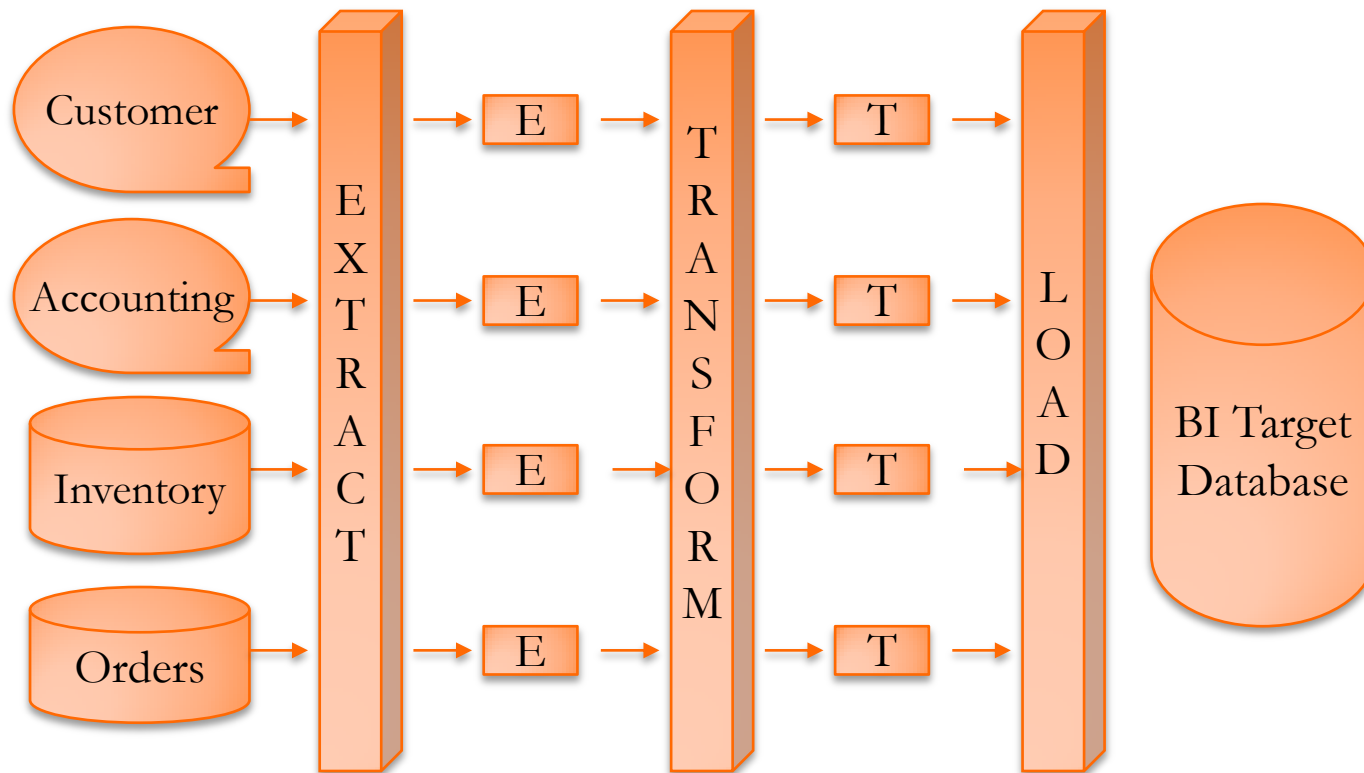
- استخراج اطلاعات از سیستم‌های عملیاتی موجود در سازمان،
- تغییر قالب آنها به طوری که قابل استفاده در DW باشند
- و در نهایت بارگذاری آنها در DW می‌باشد.

○ هر یک از فعالیت‌های انجام شده در ETL اهمیت فاصی دارد و نباید نادیده گرفته شوند.

در این گام یک قانون به نام ۸۰/۲۰ (یعنی ۸۰٪ و ۲۰٪) وجود دارد در این گام ۸۰٪ کارها Transform

است یعنی از سه کار ETL، ۸۰٪ کار روی تبدیل داده‌هاست. (فصل ۲ کتاب داده کلوی دارای تکنیک هایی برای تبدیل داده‌هاست مطالعه شود)

ETL طراحی (۹)



۹) طراحی ETL

○ مشکلاتی که در انجام عملیات ETL وجود دارد، عبارتند از:

- سیستم‌های منبع اطلاعات عموماً گوناگون و گسترده می‌باشند.
- سیستم‌های عملیاتی عموماً بر روی پلتفرم‌ها و سیستم‌های عامل گوناگون قرار دارند.
- پایگاه‌های داده سیستم‌های اطلاعاتی عموماً قدیمی بوده و با تکنولوژی روز تطابق ندارند.
- تغییرات بر روی داده‌ها که از اطلاعات اساسی در DW ها می‌باشد، در سیستم‌های اطلاعاتی ذخیره نمی‌شوند.
- کیفیت داده‌ها در بسیاری از پایگاه‌های داده قدیمی بسیار پایین می‌باشد.
- نحوه ذخیره داده‌ها در پایگاه‌های داده با یکدیگر متفاوت است.
- اطلاعات موجود در پایگاه‌های داده برای کاربران دارای مفهوم نمی‌باشد و به صورت رمزگونه و مبهم ذخیره می‌شوند.

برای آماده سازی دادهها در T چه کارهایی میتوان انجام داد؟

○ نبودن مقدار برای داده:

○ برخی سیستمها بعضی از داده هایشان مقدار ندارد

○ مثلاً یک بیمار در دادههای بیمارستان وزن نداشته باشد یا

مثلاً آدرس یک دانشجو وجود نداشته باشد،

○ حال برای تحلیل این دادهها یا باید از آنها صرف نظر کرد یا

باید برای آنها تعداد در نظر گرفت.

برای آماده سازی داده‌ها در T چه کارهایی میتوان انجام داد؟

داده‌های پرت: داده‌هایی هستند که از نظر مقدار درست می‌باشند اما با محدوده‌ی اکثر داده‌های موجود تطابق ندارند. مثلاً می‌خواهیم بدانیم مردان ۳۰-۳۵ چه سایزی برای کفششان استفاده می‌کنند (این داده‌ها رنج ۴۱-۴۲ دارند) اما وجود یک داده مانند ۴۹ با این که واقعی است اما پرت است و برای میانگین گرفتن کار ما را سخت خواهد کرد و معدل را بالا خواهد برد.

برای آماده سازی داده‌ها در T چه کارهایی میتوان انجام داد؟

داده‌های نویز (اشتباه): داده‌هایی هستند که ناشی از اشتباه کاربران می‌باشند و باید تشخیص و حذف شوند. مثلاً بعد از جمع‌آوری داده‌ها یک جا شماره دانشجویی کلید است اما با ID اما جای دیگر با ST کلید شده است وقتی این اطلاعات کنار هم قرار گرفتند هوش تجاری باید متوجه باشد که این دو به یک چیز اشاره دارد.

برای آماده سازی داده‌ها در T چه کارهایی میتوان انجام داد؟

تغییر شکل داده‌ها: مثلاً برای مقایسه دو دانشجو بر اساس قد و معدل برای پی بردن به شباهت آدم‌ها مثلاً معدل یکی ۱۷ و دیگری ۱۸ است اما قد یکی ۱۷۲ و دیگری ۱۷۵ است (معدل ۱ واحد و قد ۳ واحد تفاوت دارد) حال ۱ واحد تفاوت در معدل ارزش بیشتری دارد یا ۳ واحد تفاوت در قد، که با تحلیل مشخص می‌شود که ۱ واحد معدل تفاوت زیادی است تا ۳ واحد در قد.

در Transform داده‌ها را نرمال سازی می‌کنند یعنی همه‌ی آن‌ها را در یک رنج خاص قرار می‌دهند. (معدل ۰-۲۰ را به ۰-۱ و قد ۰-۲۵۰ را به ۰-۱ تبدیل کرده و بعداً با هم مقایسه می‌کنند) (فصل ۲)

برای آماده سازی داده‌ها در T چه کارهایی میتوان انجام داد؟

کاهش ابعاد داده‌ها: گاهی داده‌ها آنقدر حجیم هستند که نمی‌توان از همه داده‌ها استفاده نمود و برخی ستون‌ها اضافه هستند. لذا داده‌های زیادی وجود دارد که همگی آن‌ها برای تحلیل مناسب نیستند (مثلاً شماره موبایل یا تلفن دانشجو) ، بنابراین باید داده‌ها را کاهش داد.

مثلاً داده‌های موجود در حد تراپیت است و زمان زیادی برای تحلیل وجود ندارد اینجا از Sampling یا نمونه‌گیری استفاده می‌شود. (تکنیک‌های آماده سازی داده‌ها برای تحلیل - فصل ۲ کتاب)

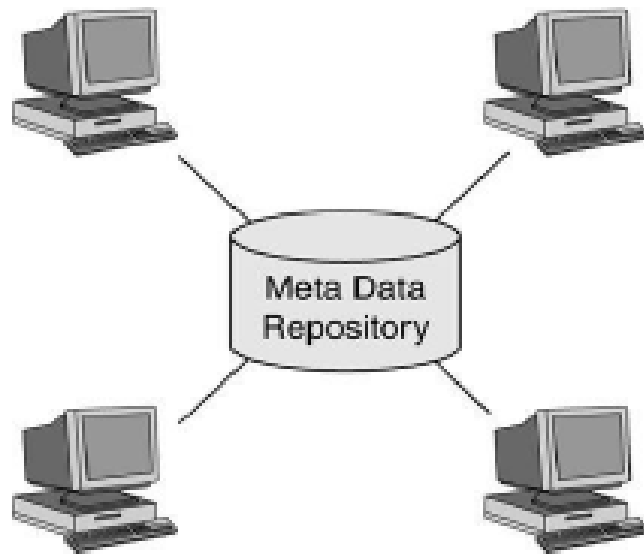
کام ۱۰) طراحی مفرن متا دیتا

۱۰) طراحی مخزن متادیتا

در این گام تصمیم گرفته می‌شود که متادیتا توسط خود سازمان ایجاد شود یا از منبع دیگری خریداری شود.

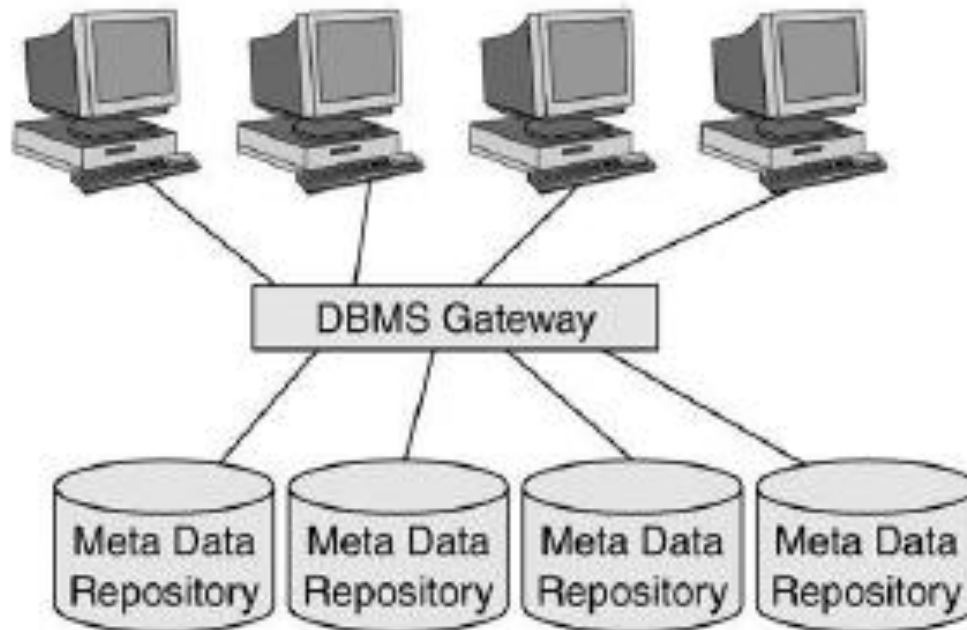
۱۰) طراحی مخزن متادیتا

راه حل اول استفاده از مخزن متمرکز:



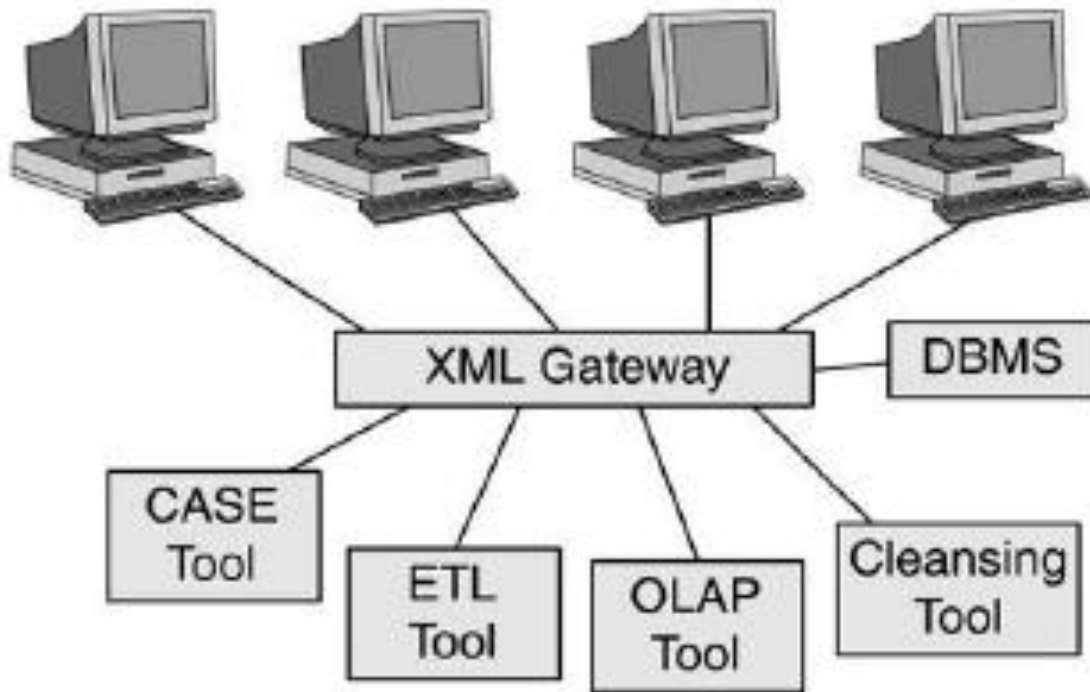
۱۰) طراحی مخزن متادیتا

راه حل دوم استفاده از مخزن غیر متمرکز:



۱۰) طراحی مخزن متادیتا

راه حل سوم استفاده از مخزن توزیع شده:



۱۰) طراحی مخزن متادیتا

معیارهای خرید یک مخزن متادیتا:

هفت معیار برای انتخاب یک مخزن متادیتا وجود دارد.

معیار ۱: تطبیق محصولاتی که می‌خواهیم خریداری کنیم با مدل منطقی داده‌های ماست یعنی اگر با ER یا هر چیزی جهت انتخاب محصول اقدام کردیم این کار چقدر به مدل منطقی داده‌های ما نزدیک است؟

معیار ۲: محصولی را که می‌خواهیم خریداری کنیم آیا نیازمندی‌های اصلی و مهم ما را پوشش می‌دهد؟

معیار ۳: آیا انتخاب‌های اختیاری در اختیار ما می‌گذارد؟ یعنی امکان دارد ما بخشی از آن محصول را فقط خریداری کنیم؟ و یا بعد از خرید فقط بخشی از آن را جهت استفاده فعال کنیم و بخش‌های دیگر را فعال نکنیم.

۱۰) طراحی مخزن متادیتا

معیارهای خرید یک مخزن متادیتا:

معیار ۴: قابلیت توسعه پذیری دارد. یعنی این محصول در آینده قابل تعمیم و گسترش خواهد بود و می توان چیزی را به آن اضافه کرد.

معیار ۵: این محصول واسط کاربری خاصی دارد و آیا نیاز به آموزش دارد و یا نه کاربران به راحتی قادرند با آن کار کنند. (خیلی مهم)

معیار ۶: آیا این محصولات قابلیت کار در شبکه های کامپیوتری را دارند.

معیار ۷: آیا این محصول، API است یا نه.