

در طول این سری آموزش‌های MDX (البته هنوز نمی‌دانم چند قسمت خواهد بود) تلاش خواهیم کرد تمامی موارد موجود در MDX ها را به طور کامل با شرح و توضیح مناسب پوشش دهیم.

امیدوارم شما دوستان عزیز پس از مطالعه‌ی این مجموعه مقالات به دانش کافی در خصوص MDX Query ها دست پیدا کنید.

در قسمت اول این آموزش‌ها در نظر دارم در ابتدا مفاهیم اولیه OLAP و همچنین مفاهیم مورد نیاز در Multi Dimentional Data Base ها برای شما عزیزان توضیح دهم و در قسمت‌های بعدی این مجموعه در خصوص MDX Query ها صحبت خواهیم کرد.

انبار داده (Data Warehouse)

عملاً یک یا چند پایگاه داده می‌باشد که اطلاعات جمع شده از دیگر پایگاه‌های داده را در خود نگه داری می‌کند. برای آرایه گزارشاتی که از پایگاه داده‌های OLTP نمی‌توانیم به راحتی بگیریم.

OLTP (Online transaction processing)

سیستم پردازش تراکنش برخط می‌باشند. که عملاً همان سیستم‌هایی می‌باشند که در طول روز دارای تغییرات بسیار زیادی می‌باشند (مانند سیستم‌های حسابداری، انبار داری و ... که در طول روز دایماً دارای تغییرات در سطح داده می‌باشند).

OLAP (Online Analysis Processing)

این سیستم‌ها خدماتی در نقش تحلیل‌گر داده و تصمیم گیرنده ارائه می‌کند. چنین سیستم‌هایی می‌توانند، داده را در قالب‌های مختلف برای هماهنگ کردن نیازهای مختلف کاربران مختلف، سازماندهی کنند.

تفاوت انبار داده (Data Warehouse) و پایگاه داده (Data Base)

وظیفه اصلی سیستم‌های پایگاه داده کاربردی Online، پشتیبانی از تراکنش‌های برخط و پردازش کوئری است. این سیستم‌ها، سیستم پردازش تراکنش برخط (OLTP) نامیده می‌شوند و بیشتر عملیات روزمره یک سازمان را پوشش می‌دهند. از سوی دیگر انبار داده، خدماتی در نقش تحلیل‌گر داده و تصمیم گیرنده ارائه می‌کند. چنین سیستم‌هایی می‌توانند داده را در قالب‌های مختلف برای هماهنگ کردن نیازهای مختلف کاربران مختلف، سازماندهی و ارائه می‌کند. این سیستم‌ها با نام سیستم‌های پردازش تحلیلی برخط (OLAP) شناخته می‌شوند.

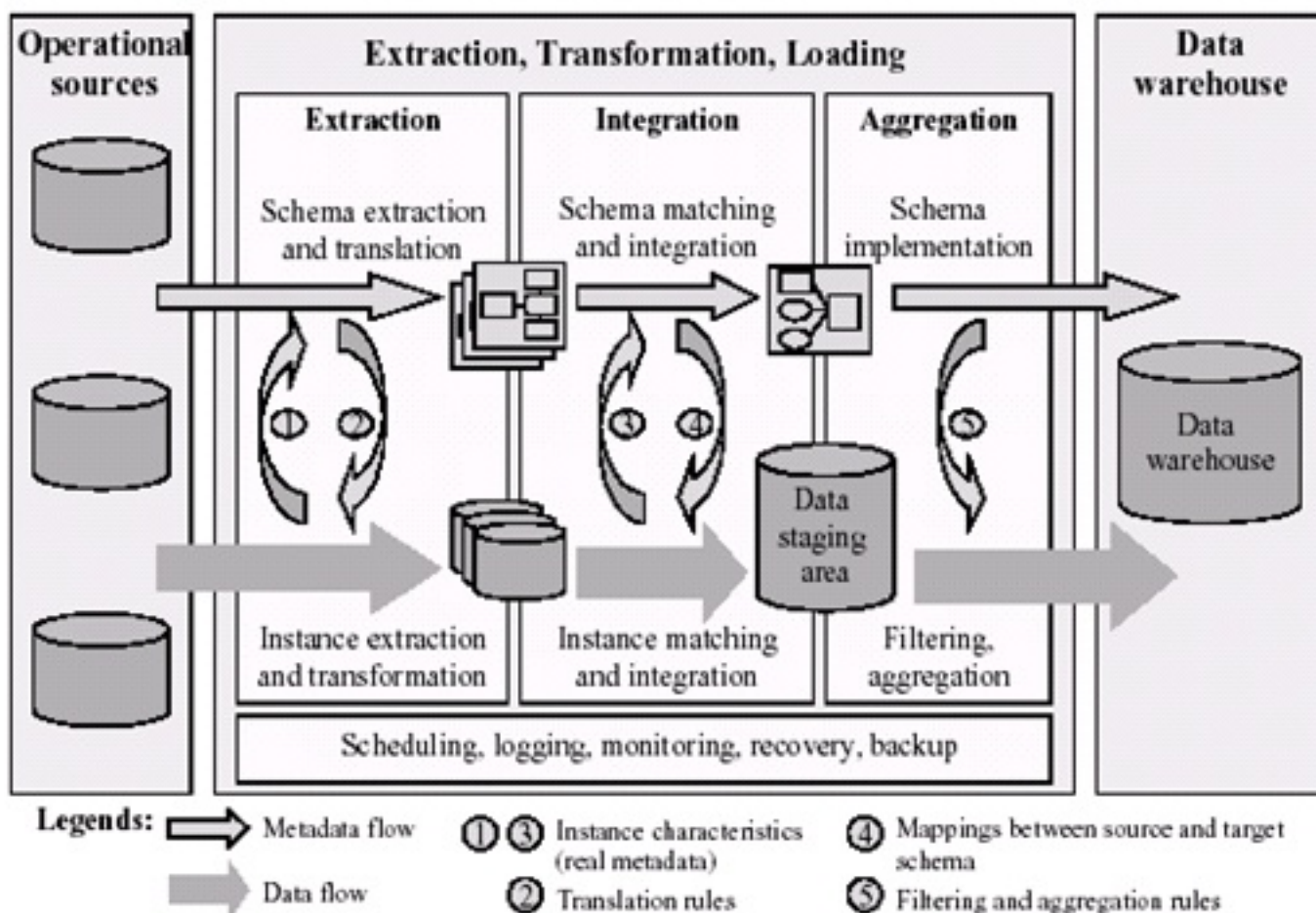
موارد تفاوت انبار داده (Data Warehouse) و پایگاه داده (Data Base)

• **از لحاظ مدل‌های داده:** پایگاه‌های داده برای مدل OLTP بهینه سازی شده‌است. که بر اساس مدل داده رابطه‌ای امکان پردازش تعداد زیادی تراکنش همروند، که اغلب حاوی رکوردهای اندکی هستند را دارد. اما در انبارهای داده که برای پردازش تحلیلی بر خط، طراحی شده‌اند امکان پردازش تعداد کمی کوئری پیچیده بر روی تعداد بسیار زیادی رکورد داده فراهم می‌شود. سرورهای OLAP می‌توانند از دو نوع رابطه‌ای (ROLAP) یا چندبعدی باشند (MOLAP).

• **از لحاظ کاربران:** کاربران پایگاه داده کارمندان دفتری و مسؤولان هستند در حالی که کاربران انبار داده مدیران و تصمیم‌گیرنده‌ها هستند.

• **از لحاظ عملیات قابل اجرا بر روی آن‌ها:** عملیات انجام شده بر روی پایگاه‌های داده عمدتاً عملیات (Select/Insert/Update/Delete) می‌باشد، در حالی که عملیات روی انبار داده عمدتاً Select ها می‌باشند.

- از لحاظ مقدار داده‌ها: مقدار داده‌های یک پایگاه داده در حدود چند مگابایت تا چند گیگابایت است در حالی که این مقدار در انبار داده در حدود چند گیگابایت تا چند ترابایت است.
 - از لحاظ زمان پرس و جو : به طور کلی سرعت پرس و جو ها روی انباری داده بسیار بالاتر از کوئری مشابه آن روی پایگاه داده می‌باشد.
- مراحل ساخت یک انباری داده (Data Warehouse) به شرح زیر می‌باشد



• پاکسازی داده (Data Cleansing)

پاکسازی داده‌ها عبارت است از شناسایی و حذف خطاها و ناسازگاریهای داده ای به منظور دستیابی به داده‌هایی با کیفیت بالاتر.

اگر داده‌ها از منابع یکسان مثل فایل‌ها یا پایگاه‌های داده ای گرفته شوند خطاهایی از قبیل اشتباهات تایپی، داده‌های نادرست و فیلدهای بدون مقدار را خواهیم داشت و چنانچه داده‌ها از منابع مختلف مثل پایگاه داده‌های مختلف یا سیستم اطلاعاتی مبتنی بر وب گرفته شوند. با توجه به نمایش‌های داده‌های مختلف خطاها بیشتر بوده و پاکسازی داده‌ها اهمیت بیشتری پیدا خواهد کرد. برای دستیابی به داده‌های دقیق و سازگار، بایستی داده‌ها را یکپارچه نموده و تکرارهای آنها را حذف نمود.

وجود خطاهای نویزی، ناسازگاری در داده‌های انبار داده و ناقص بودن داده‌ها امری طبیعی است. فیلدهای یک جدول ممکن است خالی باشند و یا دارای داده‌های خطا دار و ناسازگار باشند. برای هر کدام از این حالت‌ها روشهایی جهت پاکسازی و اصلاح داده‌ها ارائه می‌شود.

در این بخش عملیات مختلفی برای پاکسازی داده‌ها قابل انجام است:

- نادیده گرفتن تاپل‌های نادرست

- پرکردن فیلدهای نادرست به صورت دستی

- پرکردن فیلدهای نادرست با یک مقدار مشخص

- پرکردن فیلدها با توجه به نوع فیلد و داده‌های موجود

- پرکردن فیلدها با نزدیکترین مقدار ممکن (مثلاً میانگین فیلد تاپل‌های دیگر می‌تواند به عنوان یک مقدار مناسب در نظر گرفته شود) • **یکپارچه‌سازی (Integration)**

این فاز شامل ترکیب داده‌های دریافتی از منابع اطلاعاتی مختلف، استفاده از متاداده‌ها برای شناسایی و حذف افزونگی داده‌ها، تشخیص و رفع برخوردهای داده‌ای می‌باشد.

یکپارچه‌سازی داده‌ها از سه فاز کلی تشکیل شده است:

- **شناسایی فیلدهای یکسان:** فیلدهای یکسان که در جدول‌های مختلف دارای نامهای مختلف می‌باشند.

- **شناسایی افزونگی‌های موجود در داده‌های ورودی:** داده‌های ورودی گاهی دارای افزونگی است. مثلاً بخشی از رکورد در جدول مختلف وجود دارد.

- **مشخص کردن برخوردهای داده‌ای:** مثالی از برخوردهای داده‌ای یکسان نبودن واحدهای نمایش داده‌ای است. مثلاً فیلد وزن در یک جدول بر حسب کیلوگرم و در جدولی دیگر بر حسب گرم ذخیره شده است.

- **تبدیل داده‌ها (Data Transformation)**

در این فاز، داده‌های ورودی طی مراحل زیر به شکلی که مناسب عمل داده‌کاوی باشند، در می‌آیند:

- از بین بردن نویز داده‌ها (Smoothing)

- تجمیع داده‌ها (Aggregation)

- کلی‌سازی (Generalization)

- نرمال‌سازی (Normalization)

- افزودن فیلدهای جدید

در ادامه به شرح هر یک می‌پردازیم: 1. **از بین بردن نویزهای داده‌ای (Smoothing):** منظور از داده‌های نویزی، داده‌هایی هستند که در خارج از بازه مورد نظر قرار می‌گیرند. مثلاً اگر بازه حقوقی کارمندان بین یک صد هزار تومان و یک میلیون تومان باشد، داده‌های خارج از این بازه به عنوان داده‌های نویزی شناخته شده و در این مرحله اصلاح می‌گردند. برای اصلاح داده‌های نویزی از روشهای زیر استفاده می‌شود:

- استفاده از مقادیر مجاور برای تعیین یک مقدار مناسب برای فیلدهای دارای نویز

- دسته‌بندی داده‌های موجود و مقداردهی فیلد دارای داده نویزی با استفاده از دسته نزدیکتر

• ترکیب روشهای فوق با ملاحظات انسانی، در این روش، اصلاح مقادیر نویزی با استفاده از یکی از روشهای فوق انجام می‌گیرد اما افرادی برای بررسی و اصلاح نیز وجود دارند

2. **تجمیع داده‌ها (Aggregation):** تجمیع داده‌ها به معنی بدست آوردن اطلاعات جدید از ترکیب داده‌های موجود می‌باشد. به عنوان مثال بدست فروش ماهانه از حساب فروش‌های روزانه. 3. **کلی سازی (Generalization):** کلی سازی به معنی دسته بندی داده‌های موجود براساس ماهیت و نوع آنها است. به عنوان مثال می‌توان اطلاع رده‌های سنی خاص (جوان، بزرگسال، سالخورده) را از فیلد سن استخراج کرد.

4. **نرمال سازی (Normalization):** منظور از نرمال سازی، تغییر مقیاس داده‌ها است. به عنوان مثالی از نرمال سازی، می‌توان به تغییر بازه یک فیلد از مقادیر موجود به بازه 0 تا 1 اشاره کرد.

5. **افزودن فیلدهای جدید:** گاهی اوقات برای سهولت عمل داده کاوی می‌توان فیلدهایی به مجموعه فیلدهای موجود اضافه کرد. مثلا می‌توان فیلد میانگین حقوق کارمندان یک شعبه را به مجموعه فیلدهای موجود اضافه نمود.

• کاهش داده‌ها (Reduction)

در این مرحله، عملیات کاهش داده‌ها انجام می‌گیرد که شامل تکنیکهایی برای نمایش کمینه اطلاعات موجود است

. این فاز از سه بخش تشکیل می‌شود:

- **کاهش دامنه و بعد:** فیلدهای نامربوط، نامناسب و تکراری حذف می‌شوند. برای تشخیص فیلدهای اضافی، روشهای آماری و تجربی وجود دارند؛ یعنی با اعمال الگوریتمهای آماری و یا تجربی بر روی داده‌های موجود در یک بازه زمانی مشخص، به این نتیجه می‌رسیم که فیلد یا فیلدهای خاصی کاربردی در انبار داده ای و داده کاوی نداشته و آنها را حذف می‌کنیم.
- **فشرده سازی داده‌ها:** از تکنیکهای فشرده سازی برای کاهش اندازه داده‌ها استفاده می‌شود.
- **کد کردن داده‌ها:** داده‌ها در صورت امکان با پارامترها و اطلاعات کوچکتر جایگزین می‌شوند.

مدل داده‌ای رابطه‌ای (Relational) و چند بعدی (Multidimensional):

1. **مدل داده رابطه‌ای (Relational data modeling)** بر اساس دو مفهوم اساسی موجودیت (entity) و رابطه (relation) بنا نهاده شده است. از این رو آن را با نام مدل ER نیز می‌شناسند.

• **موجودیت (entity):** نمایانگر همه چیزهایی که در پایگاه داده وجود خارجی دارند یا به تصور در می‌آیند. پدیده‌ها دارای مشخصاتی هستند که به آن‌ها صفت (attribute) گفته می‌شود.

• **رابطه (relation):** پدیده‌ها را به هم می‌پیوندد و چگونگی در ارتباط قرار گرفتن آن‌ها با یکدیگر را مشخص می‌کند.

2. **مدل داده چندبعدی (Multidimensional modeling)** یا MD بر پایه دو ساختار جدولی اصلی بنا نهاده شده است:

• جدول حقایق (Fact Table)

• جداول ابعاد (Dimension Table)

این ساختار امکان داشتن یک نگرش مدیریتی و تصمیم‌گیری به داده‌های موجود در پایگاه داده را تسهیل می‌کند.

جدول حقایق: قلب حجم داده‌ای ما را تشکیل می‌دهد و شامل دو سری فیلد است: کلیدهای خارجی به ابعاد و شاخص‌ها (Measure).

شاخص‌ها (Measure): معیارهایی هستند که بر روی آن‌ها تحلیل انجام می‌گیرد و درون جدول حقایق قرار دارند. شاخص‌ها قبل از شکل‌گیری انبار داده توسط مدیران و تحلیل‌گران به دقت مشخص می‌شوند. چون در مرحله کار با انبار اطلاعات اساسی هر تحلیل بر اساس همین شاخص‌ها شکل می‌گیرد. شاخص‌ها تقریباً همیشه مقادیر عددی را شامل می‌شوند. مثلاً برای یک فروشگاه

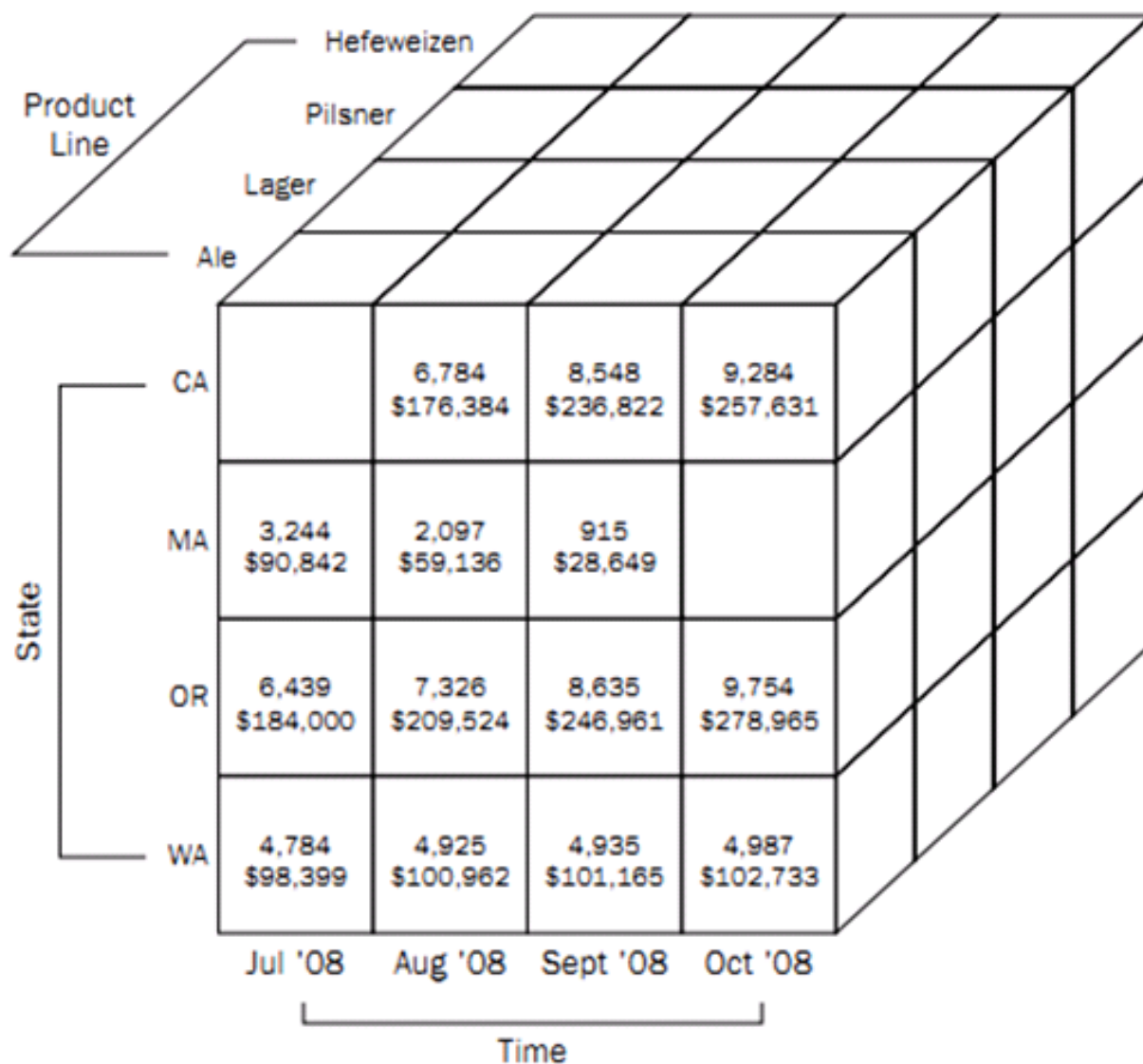
زنجیره‌ای این شاخص‌ها می‌توانند واحدهای فروخته‌شده کالاها و مبلغ فروش به تومان باشند.

بعد (Dimension): هر موجودیت در این مدل می‌تواند با یک بعد تعریف شود. ولی بعدها با موجودیت‌های مدل ER متفاوتند زیرا آن‌ها سازمان شاخص‌ها را تعیین می‌کنند. علاوه بر این دارای یک ساختار سلسله مراتبی هستند و به طور کلی برای حمایت از سیستم‌های تصمیم‌گیری سازمان‌دهی شده‌اند.

اجزای بعدها member نام دارند و تقریباً همه بعدها، memberهای خود را در یک یا چند سطح سلسله مراتبی (hierarchies) سازمان‌دهی می‌نمایند، که این سلسله مراتب نمایانگر مسیر تجمیع (integration) و ارتباط بین سطوح پایین‌تر (مثل روز) و سطوح بالاتر (مثل ماه و سال) است. وقتی یک دسته از memberهای خاص با هم مفهوم جدیدی را ایجاد می‌کنند، به آنها یک سطح (Level) می‌گوییم. (مثلاً هر سی روز را ماه می‌گوییم. در این حالت ماه یک سطح است.)

حجم‌های داده‌ای (Data Cube)

حجم‌های داده‌ای یا Cube از ارتباط تعدادی بعد با تعدادی شاخص تعریف می‌شود. ترکیب memberهای هر بعد از حجم داده‌ای فضای منطقی را تعریف می‌کند که در آن مقادیر شاخص‌ها ظاهر می‌شوند. هر بخش مجزا که شامل یکی از memberهای بعد در حجم داده‌ای است، سلول (cell) نامیده می‌شود. سلول‌ها شاخص‌های مربوط به تجمیع‌های مختلف را در خود نگهداری می‌نمایند. در واقع مقادیر مربوط به حقایق (Fact) که در جدول حقایق (Fact) تعریف می‌شوند در حجم داده‌ای (Data Cube) در سلول‌ها (Cell) نمایان می‌گردند.



شماهای داده‌ای (Data Schema) : سه نوع Schema در طراحی Data Warehouse وجود دارد

1. Stare

2. Snowflake

3. Galaxy

1. شماهای ستاره‌ای (Star Schema) : متداولترین شما، همین شماهای ستاره‌ای است. که در آن انبارداده با استفاده از اجزای زیر تعریف می‌شود:

- یک جدول مرکزی بزرگ به نام جدول حقایق که شامل حجم زیادی از داده‌های بدون تکرار است.

- مجموعه‌ای از جدول‌های کمکی کوچک‌تر به نام جدول بعد ، که به ازای هر بعد یکی از این جداول موجود خواهد بود.

- شکل این شما به صورت یک ستاره است که جدول حقایق در مرکز آن قرار گرفته و هر یک از جداول بعد به وسیله شعاع‌هایی به آن مربوط هستند.

مشکل این مدل احتمال پیشامد افزونگی در آن است.

2. **شمای دانه‌برفی (Snowflake Schema)** : در واقع شمای دانه‌برفی، نوعی از شمای ستاره‌ای است که در آن بعضی از جداول بعد نرمال شده‌اند. و به همین خاطر دارای تقسیمات بیشتری به شکل جداول اضافی می‌باشد که از جداول بعد جدا شده‌اند.

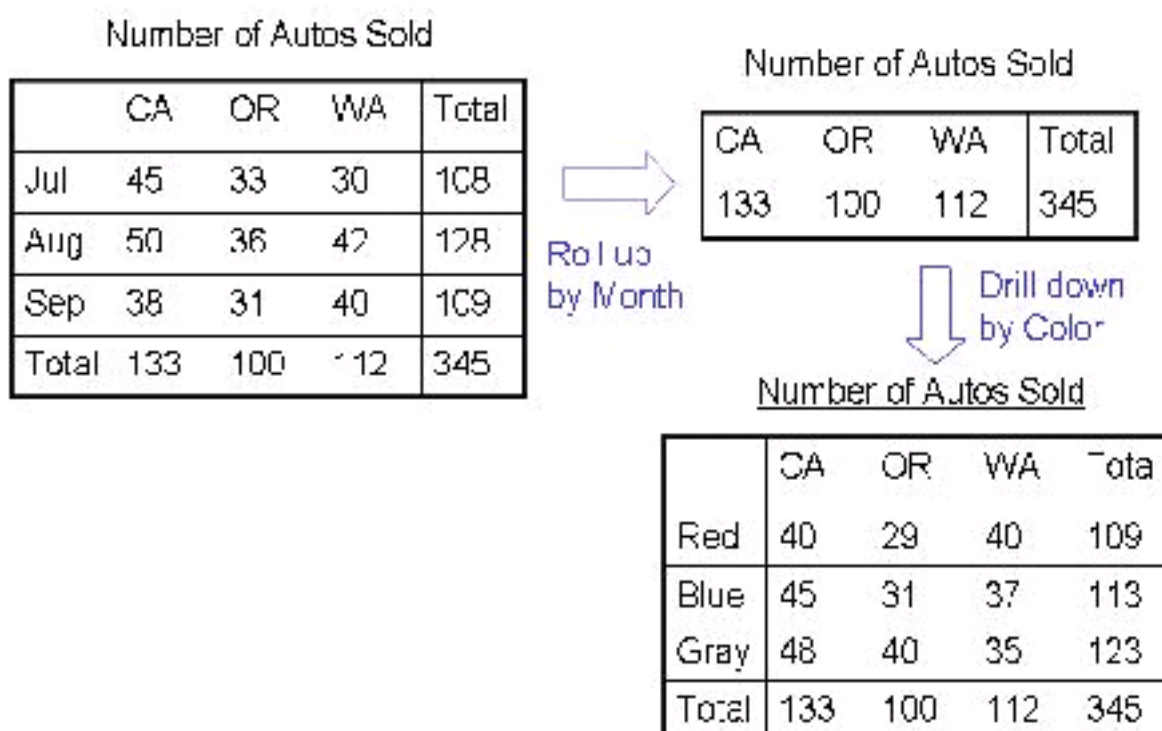
تفاوت این دو شما در این است که جداول شمای دانه برف نرمال هستند و افزونگی در آن‌ها کاهش یافته است. که این برای کار کردن با داده‌ها و از لحاظ فضای ذخیره‌سازی مفید است. ولی در عوض کارایی را پایین می‌آورد، زیرا در محاسبه کوئری‌ها به joinهای بیشتری نیاز داریم.

3. **شمای کهکشانی (galaxy schema)** : در کاربردهای پیچیده برای به اشتراک گذاشتن ابعاد نیاز به جداول حقایق چندگانه احساس می‌شود که یک یا چند جدول بعد را در بین خود به اشتراک می‌گذارند. این نوع شما به صورت مجموعه‌ای از شماهای ستاره‌ای است و به همین دلیل شمای کهکشان یا شمای منظومه‌ای نامیده می‌شود. این شما به ما این امکان را می‌دهد که جداول بعد بین جداول حقایق مختلف به اشتراک گذاشته شوند.

عملیات بر روی حجم‌های داده‌ای :

• **Roll Up (یا Drill-up)** : با بالا رفتن در ساختار سلسله مراتبی مفهومی یک حجم داده‌ای، یا با کاهش دادن بعد، یک مجموعه با جزئیات کمتر (خلاصه شده) ایجاد می‌نماید. بالا رفتن در ساختار سلسله مراتبی به معنای حذف قسمتی از جزئیات است. برای مثال اگر قبلاً بعد مکان بر حسب شهر بوده آن را با بالا رفتن در ساختار سلسله مراتبی بر حسب کشور درمی‌آوریم. ولی وقتی با کاهش دادن بعد سروکار داریم منظور حذف یکی از ابعاد و جایگزین کردن مقادیر کل است. در واقع همان عمل تجمیع (aggregation) است.

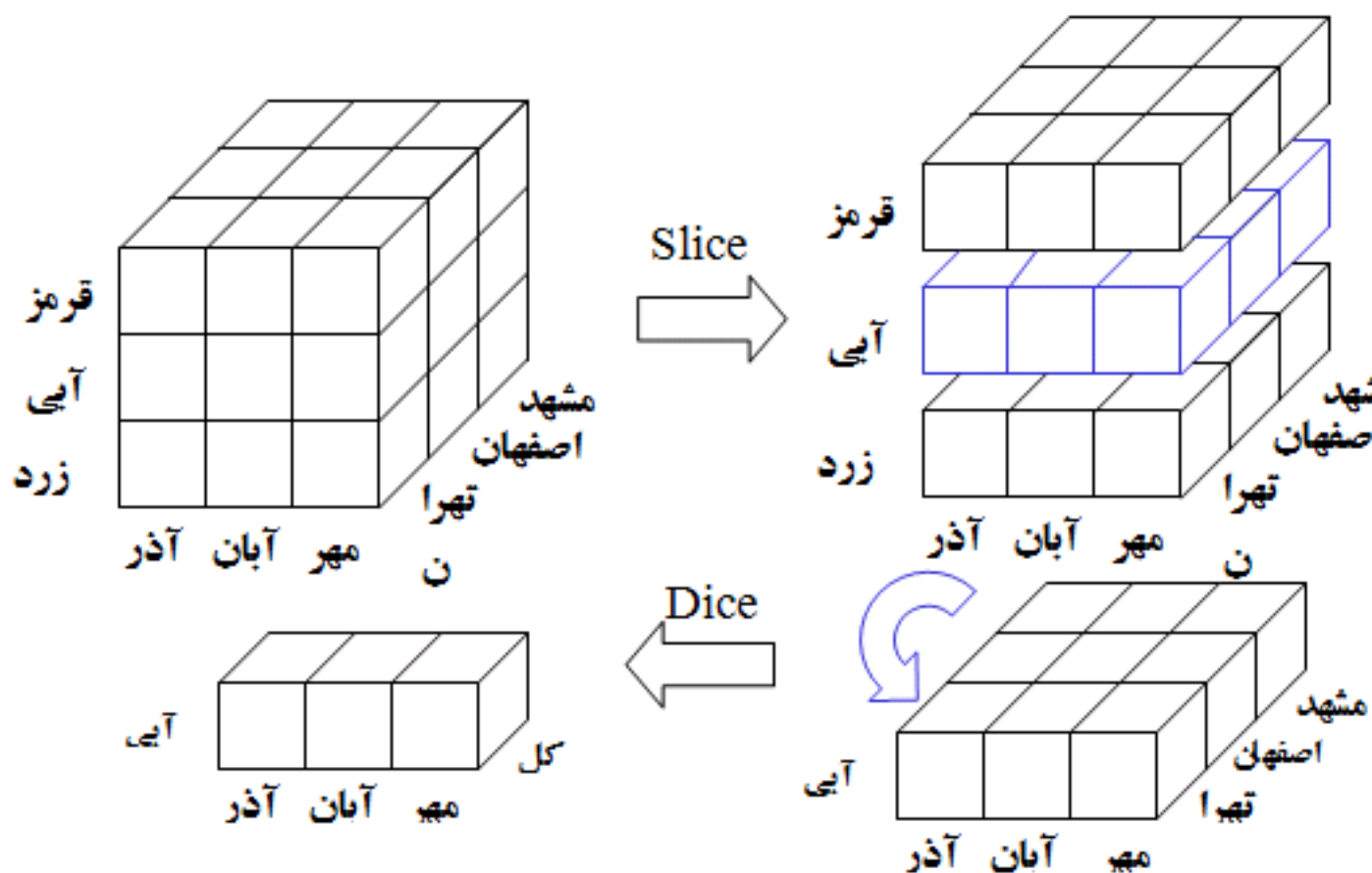
• **Drill Down** : بر عکس عمل Roll-up است و از موقعیتی با جزئیات داده‌ای کم به جزئیات زیاد می‌رود. این کار با پایین آمدن در ساختار سلسله مراتبی (به سمت جزئیات بیشتر) یا با ایجاد ابعاد اضافی انجام می‌گیرد.



نمونه‌ای از عملیات Roll Up و Drill Down

• **Slice** : با انتخاب و اعمال شرط بر روی یکی از ابعاد یک subcube به شکل یک برش دو بعدی ایجاد می‌کند. در واقع همان عمل انتخاب (select) است.

• **Dice** : با انتخاب قسمتی از ساختار سلسله مراتبی بر روی دو یا چند بعد یک subcube ایجاد می‌نماید.



نمونه‌ای از عملیات Slice و Dice

• **Pivot (یا Rotate)** : این عملیات بردارهای بعد را در ظاهر می‌چرخاند.

Region	Sales variance
Africa	105%
Asia	57%
Europe	122%
North America	97%
Pacific	85%
South America	163%

Nation	Sales variance
China	123%
Japan	52%
India	87%
Singapore	95%

نمونه‌ای از عملیات pivot

• **Drill-across** : نتیجه اجرای کوئری‌هایی که نتیجه اجرای آنها حجم‌های داده‌ایهای مرکب با بیش از یک fact-table است.

• **Ranking** : سلول‌هایی را باز می‌گرداند که در بالا یا پایین شرط خاصی واقع هستند. مثلاً ده محصولی که بهترین فروش را داشته‌اند.

سرورهای OLAP :

در تکنولوژی OALP داده‌ها به دو صورت چندبعدی (Multidimensional OLAP) (MOLAP) و رابطه‌ای (Relational OLAP) (ROLAP) ذخیره می‌شوند. OLAP پیوندی (HOLAP) تکنولوژی است که دو نوع قبل را با هم ترکیب می‌کند.

MOLAP : روشی است که معمولاً برای تحلیل‌های OLAP در تجارت مورد استفاده قرار می‌گیرد. در MOLAP، داده‌ها با ساختار یک حجم داده‌ای (Data Cube) چند بعدی ذخیره می‌شوند. ذخیره‌سازی در پایگاه داده‌های رابطه‌ای انجام نمی‌گیرد، بلکه با یک فرمت خاص انجام می‌شود. اغلب محصولات موفق MOLAP از یک روش چندبعدی استفاده می‌نمایند که در آن یک سری حجم‌های داده‌ای کوچک، انبوه و از پیش محاسبه‌شده، یک حجم داده‌ای بزرگ (hypercube) را می‌سازند.

علاوه بر این MOLAP به شما امکان می‌دهد داده‌های دیدهای (View) تحلیل‌گران را دسته بندی کنید، که این در حذف اشتباهات و برخورد با ترجمه‌های پرغلط کمک بزرگی است.

گذشته از همه این‌ها از آن‌جا که داده‌ها به طور فیزیکی در حجم‌های داده‌ای بزرگ چندبعدی ذخیره می‌شوند، سرعت انجام فعالیت‌ها بسیار زیاد خواهد بود.

از آنجا که یک کپی از داده‌های منبع در کامپیوتر Analysis server ذخیره می‌شود، کوئری‌ها می‌توانند بدون مراجعه به منابع مجدداً

محاسبه شوند. کامپیوتر Analysis server ممکن است کامپیوتر سرور که تقسیم بندی‌ها در آن انجام شده یا کامپیوتر دیگری باشد. این امر بستگی به این دارد که تقسیم بندی‌ها در کجا تعریف شده‌اند. حتی اگر پاسخ کوئری‌ها از روی تقسیمات تجمیع (integration) شده قابل دستیابی نباشند، MOLAP سریع‌ترین پاسخ را فراهم می‌کند. سرعت انجام این کار به طراحی و درصد تجمیع تقسیم بندی‌ها بستگی دارد.

مزایا : کارایی عالی- حجم‌های داده‌ای MOLAP برای بازیابی سریع داده‌ها ساخته شده‌اند و در فعالیت‌های slice و dice به صورت بهینه پاسخ می‌دهند. ترکیب سادگی و سرعت مزیت اصلی MOLAP است.

در ضمن MOLAP قابلیت محاسبه محاسبات پیچیده را فراهم می‌کند. همه محاسبات از پیش وقتی که حجم‌های داده‌ای ساخته می‌شود، ایجاد می‌شوند. بنابراین نه تنها محاسبات پیچیده انجام شدنی هستند بلکه بسیار سریع هم پاسخ می‌دهند.

معایب : عیب این روش این است که تنها برای داده‌هایی با مقدار محدود کارکرد خوبی دارد. از آنجا که همه محاسبات زمانی که حجم‌های داده‌ای ساخته می‌شود، محاسبه می‌گردند، امکان این که حجم‌های داده‌ای مقدار زیادی از داده‌ها را در خود جای دهد، وجود ندارد. ولی این به این معنا نیست که داده‌های حجم‌های داده‌ای نمی‌توانند از مقدار زیادی داده مشتق شده باشند. داده‌ها می‌توانند از مقدار زیادی داده مشتق شده باشند. اما در این صورت، فقط اطلاعات level خلاصه (level 1) ای که دارای کمترین جزئیات است یعنی سطوح بالاتر) می‌توانند در حجم‌های داده‌ای موجود باشند.

ROLAP : محدودیت MOLAP در حجم داده‌های قابل پرس و جو و نیاز به روشی که از داده‌های ذخیره شده به روش رابطه‌ای حمایت کند، موجب پیشرفت ROLAP شد.

مبنای این روش کارکردن با داده‌هایی که در پایگاه داده‌های رابطه‌ای ذخیره شده‌اند، برای انجام اعمال slicing و dicing معمولی است. با استفاده از این مدل ذخیره سازی می‌توان داده‌ها را بدون ایجاد واقعی تجمیع در پایگاه داده‌های رابطه‌ای به هم مربوط کرد.

مزایا : با این روش می‌توان به حجم زیادی از داده‌ها را رسیدگی کرد. محدودیت حجم داده در تکنولوژی ROLAP مربوط به محدودیت حجم داده‌های قابل ذخیره سازی در پایگاه داده‌های رابطه‌ای است. به بیان دیگر، خود ROLAP هیچ محدودیتی بر روی حجم داده‌ها اعمال نمی‌کند.

معایب : ممکن است کارایی پایین بیاید. زیرا هر گزارش ROLAP در واقع یک کواری SQL (یا چند کواری SQL) در پایگاه داده‌های رابطه‌ای است و اگر حجم داده‌ها زیاد باشد ممکن است زمان پاسخ کواری طولانی شود. در مجموع ROLAP سنگین است، نگهداری آن سخت است و کند هم هست. بخصوص زمانی که نیاز به آدرس دهی جدول‌های ذخیره شده در سیستم چند بعدی داریم.

این محدودیت ناشی از عملکرد SQL است. زیرا تکنولوژی ROLAP بر پایه عبارات مولد SQL برای پرسش و پاسخ بر روی پایگاه داده رابطه‌ای است و عبارات SQL به همه نیازها پاسخ نمی‌دهند (مثلاً محاسبه حساب‌های پیچیده در SQL مشکل است)، بنابراین فعالیت‌های ROLAP به آن چه SQL قادر به انجام آن است محدود می‌گردد.

تفاوت MOLAP و ROALP : تفاوت اصلی این دو در معماری آن‌ها است. محصولات MOLAP داده‌های مورد نیاز را در یک حافظه نهان (cache) مخصوص می‌گذارد. ولی ROLAP تحلیل‌های خود را بدون استفاده از یک حافظه میانی انجام می‌دهد، بدون آن که از یک مرحله میانی برای گذاشتن داده‌ها در یک سرور خاص استفاده کند.

با توجه به کند بودن ROLAP در مقایسه با MOLAP، باید توجه داشت که کاربرد این روش بیشتر در پایگاه داده‌های بسیار بزرگی است که گاه‌گاهی پرس و جویی بر روی آن‌ها شکل می‌گیرد، مثل داده‌های تاریخی و کمتر جدید سال‌های گذشته.

نکته: اگر از Analysis Services که به وسیله Microsoft OLE DB Provider مهیا شده استفاده می‌کنید، تجمیع‌ها نمی‌توانند برای تقسیم بندی از روش ROLAP استفاده نمایند.

HOLAP : با توجه به نیاز رو به رشدی که برای کارکردن با داده‌های بلادرنگ (real time) در بخش‌های مختلف در صنعت و تجارت

احساس می‌شود، مدیران تجاری انتظار دارند بتوانند با دامنه وسیعی از اطلاعات که فوراً و بدون حتی لحظه‌ای تأخیر در دسترس باشند، کار کنند. در حال حاضر شبکه اینترنت و سایر کاربردها یی که به داده‌هایی از منابع مختلف مراجعه دارند و نیاز به فعالیت با یک سیستم بلادرنگ هم دارند، همگی از سیستم HOLAP بهره می‌گیرند.

: named set

Named Set مجموعه‌ای از memberهای بعد یا مجموعه‌ای از عبارات است که برای استفاده مجدد ایجاد می‌شود.

Calculated member

Calculated Memberها memberهایی هستند که بر اساس داده‌ها نیستند بلکه بر اساس عبارات ارزیابی MDX هستند. آنها دقیقاً به سبک سایر memberهای معمولی هستند. MDX یک مجموعه قوی از عملیاتی را تامین میکند که میتواند برای ساخت Calculated Memberها مورد استفاده قرار گیرند به طوری که به شما امکان داشتن انعطاف زیاد در کار کردن با داده‌های چند بعدی را بدهد.

امیدوارم در این قسمت با مفاهیم نخستین OLAP آشنا شده باشید.

تلاش خواهیم کرد در قسمت بعدی در خصوص نصب SQL Server Analysis Services و نصب پایگاه داده‌ی Adventure Work DW 2008 شرح کاملی را ارائه کنم.

نظرات خوانندگان

نویسنده: afshin

تاریخ: ۱۷:۲۶ ۱۳۹۲/۰۹/۲۰

خیلی ممنون ... لطفا ادامه بدید ... به بحث عملی هم میپردازید ؟ کلا بحث عملی بر پایه SQL Server هست یا از نرم افزارهای دیگه هم استفاده میشه ؟ در مورد ETS ها هم اگه میشه توضیح بدید باز هم تشکر

نویسنده: vahid

تاریخ: ۲۱:۱۶ ۱۳۹۲/۰۹/۲۰

ممنون از مطلب مفیدتون لطفا در مورد اینکه چگونه datawarehouse را پیاده سازی کنیم توضیح دهید ممنون

نویسنده: اردلان شاه قلی

تاریخ: ۲۳:۸ ۱۳۹۲/۰۹/۲۰

به طور کلی زبان MDX بین DBMS یک سان می باشد (تقریبا) و البته من در این مجموعه تلاش دارم در خصوص MDX Query ها آنچه را بلد هستم با شما در میان بگذارم. اگر عمری بود مطالبی در خصوص UI هم اضافه خواهم کرد.

نویسنده: اردلان شاه قلی

تاریخ: ۲۳:۹ ۱۳۹۲/۰۹/۲۰

تلاشم را می کنم بعد از آموزش MDX ها یک مطلب هم در این خصوص برای شما بگذارم.

نویسنده: محمد باقر سیف الهی

تاریخ: ۱۴:۳ ۱۳۹۲/۰۹/۲۱

سلام، لطفا تعریفی از Datamart و ارتباطش با Datawarehouse ارائه بدید. تشکر

نویسنده: اردلان شاه قلی

تاریخ: ۱۵:۴۹ ۱۳۹۲/۰۹/۲۱

اگر اجازه دهید تعاریف DataMart و ارتباطش با DW را در یک مقاله کوتاه انتشار می دهم . اینجا در کامنت بگذارم تعداد کمی بهش دسترسی پیدا خواهد کرد.

نویسنده: علاقه مند

تاریخ: ۱۱:۱۰ ۱۳۹۲/۱۰/۱۰

سلام،

آیا کتاب فارسی تالیفی (یا ترجمه بصورت مناسب و کامل) با موضوع انبار داده ها و OLAP در بازار سراغ دارید؟

نویسنده: اردلان شاه قلی

تاریخ: ۱۶:۱۷ ۱۳۹۲/۱۰/۱۰

نه من از منبع فارسی مطالعه نکردم و اطلاعی هم در این خصوص ندارم . البته انتشارات Wrox دو تا کتاب خود در خصوص SSAS دارد منبع اطلاعاتی هم که بنده اینجا ارایه می کنم همین مرجع ها می باشد.