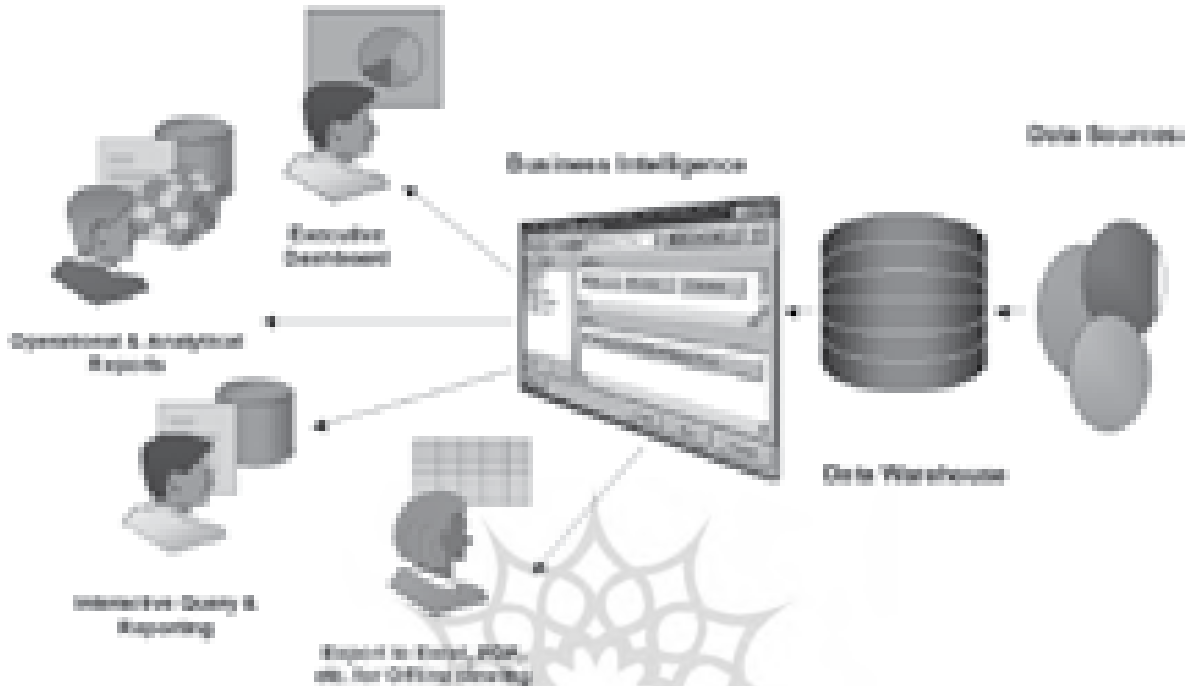


آشنایی با انبار داده (Data Warehouse)

تهیه کنندگان: مریم گلبند توکلی، معصومه کریمی مدیریت امور فناوری اطلاعات و ارتباطات



مقدمه

توسعه بکارگیری سیستمهای اطلاعاتی در سازمانها و گسترش انتقال فرآیندهای سازمانی به سامانههای الکترونیکی و نیز وابستگی روزافزون مدیران و تصمیم سازها به اطلاعات و سیستمهای اطلاعاتی، مباحثی نظیر یکپارچه سازی (Integration)، داده کاوی (Data mining) و مخزن داده (Warehouse Data) را مطرح ساخته است. مدیران و تصمیم گیران سازمانها در وهله اول نیازمند دسترسی به داده ها در هر محل بوده و در مرحله بعد به تحلیل آنها می پردازند تا بتوانند به مزیت رقابتی در بازار در مقابل رقیبان دست یابند.

بار بسیار سنگینی برای سرورهای پایگاه داده می باشند، وجود انبار داده سبب می گردد که اینگونه عملیات تأثیری بر فعالیت برنامه های کاربردی سازمان (OLTP) نداشته باشد.

همانگونه که پایگاه داده سیستمهای عملیاتی سازمان (برنامه های کاربردی) به گونه ای طراحی می شوند که انجام تغییر و حذف و اضافه داده به سرعت صورت پذیرد، در مقابل انبار داده ها دارای معماری ویژه ای می باشند که موجب تسریع انجام عملیات آماری و گزارش گیری می شود (OLAP). لازم به ذکر است:

۱- داده ها در مخزن داده ها بر اساس موضوعات خاص سازماندهی می شوند مانند مشتریان، فروشندگان،

تعریف انبار داده (Data Warehousing)

انبار داده به مجموعه ای از داده ها گفته می شود که از منابع مختلف اطلاعاتی سازمان جمع آوری، دسته بندی و ذخیره می شود. در واقع یک انبار داده مخزن اصلی کلیه داده های حال و گذشته یک سازمان است که برای همیشه جهت انجام عملیات گزارش گیری و آنالیز در دسترس مدیران می باشد. انبارهای داده حاوی داده هایی هستند که به مرور زمان از سیستم های عملیاتی آنلاین سازمان (OLTP) استخراج می شوند، بنابراین سوابق کلیه اطلاعات و یا بخش عظیمی از آنها را می توان در انبار داده ها مشاهده نمود. از آنجائیکه انجام عملیات آماری و گزارشات پیچیده

تهیه کنندگان، محصولات، حسابها

۲- داده ها در مخزن داده ها به لحاظ شکل و ساختار آنها و نحوه نامگذاری شان با اینکه از منابع متفاوت استفاده شده است، دارای ساختار یکپارچه هستند.

۳- داده ها در مخزن داده متعلق به دوره های زمانی مختلف هستند و از این رو می توان تغییرات و الگوهای که این تغییرات از آن پیروی کرده اند را مورد مطالعه قرار دهد.

۴- داده ها در مخزن داده ها در حالت ایستا بوده و به روز نمی شوند و همواره رو به فزونی هستند.

۵- بازار داده ها یک زیرمجموعه کوچک و محدود از مخزن داده هاست که تهیه گزارش از داده ها و یا تحلیل بر روی یک بخش یا یک واحد و یا یک دپارتمان و یا عملیات در سازمان را امکان پذیر می کند. مانند فروش، حقوق و دستمزد و تولید. در بعضی موارد یک بازار داده به وسیله یک مخزن از داده های شخصی که معمولا کوچکتر از مخزن داده های سازمان است، تکمیل می شود.

تاریخچه و دلایل استفاده از انبار داده

از اواخر سال ۱۹۸۰ میلادی، انبار های داده به عنوان نوع متمایزی از پایگاه های داده مورد استفاده اغلب سازمانها و شرکت های متوسط و بزرگ واقع شدند. انبار های داده جهت رفع نیاز رو به رشد مدیریت داده ها و اطلاعات سازمانی که توسط پایگاه های داده سیستم های عملیاتی غیر ممکن بود، ساخته شدند. سیستم های عملیاتی سازمان دارای نقاط ضعفی می باشند که انبار های داده آنها را رفع می کنند. از جمله:

- بار پردازش گزارشات موجب کندی عملکرد برنامه های کاربردی می گردد.

- پایگاه های داده برنامه های کاربردی دارای طراحی مناسب جهت انجام عملیات آماری و گزارش نیستند.

- بسیاری از سازمانها دارای بیش از یک برنامه کاربردی (منابع اطلاعاتی) می باشند، بنابراین تهیه گزارشات در سطح سازمان غیر ممکن می شود.

- تهیه گزارشات در سیستم های عملیاتی غالبا نیازمند نوشتن برنامه های مخصوص می باشد که معمولا کند و پرهزینه هستند.

مراحل و نحوه ایجاد انبار داده در سازمان

بسیاری از شرکت ها و سازمانها به این باور رسیده اند که گردآوری، سازمان دهی و یکپارچه سازی داده ها در یک مخزن داده برای مدیریت بهینه و اتخاذ تصمیمات کلان یک ضرورت می باشد.

به طور کلی ساخت یک انبار داده، به شکل یک پروژه شامل مراحل اصلی زیر می باشد:

۱- استخراج داده های تراکنشی از پایگاه های داده به یک مخزن واحد

شناخت منابع داده های سازمان و استخراج داده های ارزشمند از آنها یکی از اصلی ترین مراحل ایجاد انبار داده می باشد.

۲- تبدیل داده ها

از آنجائیکه سیستم های اطلاعاتی و برنامه های کاربردی یک سازمان غالبا توسط افراد و پروژه های مختلف به مرور زمان در مواجهه با نیازهای جدید ساخته یا تغییر شکل داده می شوند، یکسان سازی آنها امری ضروری می باشد. در بسیاری از موارد نیز سیستم های اطلاعاتی در بستر های مختلف پایگاه داده مانند Microsoft SQL Server، Oracle، Sybase، Microsoft Access و غیره طراحی گردیده اند. بررسی جداول، برقراری ارتباط بین فیلدها و یک شکل سازی داده ها در این مرحله صورت می پذیرد.

۳- بارگذاری داده های تبدیل شده به یک پایگاه داده چند بعدی

بر خلاف پایگاه داده سیستم های عملیاتی که دارای معماری رابطه ای می باشند و از اصول نرمالیزه استفاده می کنند، طراحی انبار داده به شکلی ویژه بدون بهره گیری از اصول نرمالیزاسیون می باشد. در انبار داده فیلدها در جاهای مختلفی تکرار می شوند و روابط بین جداول کمتر به چشم می خورند. علت آن هم افزایش سرعت پردازش اطلاعات هنگام گزارشات و عملیات آماری می باشد.

۴- تولید مقادیر از پیش محاسبه شده جهت افزایش سرعت گزارش گیری

مقادیر از پیش محاسبه شده را تراکم نیز می نامند. این مرحله توسط سیستم های نظیر Microsoft SQL Server Analysis Services بسیار ساده تر شده است. این تراکم ها که در ابعاد مختلف انبار داده ساخته می شوند، موجب میشوند که سرعت انجام عملیات گزارش گیری به شکل محسوسی افزایش یابد. باید توجه داشت که عملیات ساخت این مقادیر بسیار زمان گیر بوده و نیازمند حافظه زیادی بر روی سرور است.

۵- ساخت (یا خرید) یک ابزار گزارش گیری

پس از انجام مراحل فوق، شما می توانید نسبت به ساخت یا خرید یک نرم افزار گزارش گیری تصمیم گیری نمایید. به طور معمول هزینه ساخت یک نرم افزار گزارش گیری، بالاتر از هزینه خرید آن می باشد. ■

منابع:

<http://www.sqliran.com>

شرکت دناپرداز

<http://www.mbaforum.ir>

دانشگاه مدیریت و اقتصاد دانشگاه صنعتی شریف