

به نام خدا

شهاب الدین لاهوتی

آزمایشگاه پایگاه داده ها

استاد یاریان

Data Warehouse

Data Warehouse چیست؟

انبار داده (Data Warehouse) یک پایگاه داده رابطه ای (Relational Database) است که برای هدف پرس و جو و تحلیل طراحی شده است و با پایگاه داده (Database) که مدیریت ثبت و پردازش تراکنش های کسب و کار است، کاملاً متفاوت است.

انبار داده شامل داده های تاریخی (Historical) است که از داده های تراکنشی (Transactional) مشتق می شود اما می تواند منابع داده ای دیگر را نیز شامل شود. انبار داده تحلیل را از تراکنش جدا می کند و به سازمان اجازه می دهد که منابع مختلف داده ای را با یکدیگر تلفیق و یکپارچه کند. به علاوه انبار داده شامل عملیات استخراج (Extract)، تبدیل (Transform) و بارگذاری داده (Load) است که به اختصار به آن ETL گفته می شود که طی این فرآیند جمع آوری، پاکسازی و انتقال داده از پایگاه داده به انبار داده صورت می گیرد.

4 Data Warehouse ویژگی اصلی دارد:

1. موضوع گرا (Subject-Oriented)

2. یکپارچه (Integrated)

3. غیر فرار (Non-Volatile)

4. متغیر با زمان (Time-Variant)

تفاوت طراحی انبار داده با رویکرد Star و Snow Flake

طراحی ستاره ای و دانه برفی دو نوع از مهمترین طراحی های انبار داده به شمار می روند. مهمترین تفاوت این دو نوع مدل سازی در میزان نرمال سازی جداول است به طوری که جداول در طراحی ستاره ای کاملاً غیر نرمال طراحی می شوند اما در طراحی دانه برفی درجه نرمال سازی کمتر است. در واقع نرمال سازی با هدف کاهش تکرار صورت می پذیرد. در طراحی مدل داده با ۲ نوع جدول روبرو هستیم که شامل جداول Fact و Dimension می باشند. طراحی پایگاه داده رابطه ای که معمولاً با شناسایی موجودیت ها و ساخت جدول از این موجودیت ها و ایجاد روابط (Relationship) بین آن ها و کاملاً نرمال صورت می پذیرد، برای پردازش تراکنش های روزانه مناسب است و برای مقاصد تحلیلی توصیه نمی شود. در واقع انبار داده که برای مقاصد تحلیلی مورد استفاده قرار می گیرد، باید با درجه ی نرمال سازی کمتر و موضوع گرا طراحی شود. در واقع جداول Fact و Dimension حول یک موضوع خاص طراحی شوند (به طور مثال فروش)

طراحی Star

طراحی ستاره ای به طراحی گفته می شود که جدول **Fact** در مرکز قرار گرفته و توسط جداول **Dimension** محاصره می شود. در این مدل جداول **Dimension** با یکدیگر ارتباط ندارند و فقط از طریق کلید با جدول **Fact** ارتباط دارند. این نوع طراحی برای گزارش گیری بسیار بهینه است.

جدول **Fact** شامل رویداد هاست و اعداد و ارقامی که آن رویداد را توصیف می کنند. به طور مثال فروش یک کالا یک رویداد است که اتفاق می افتد و اعداد و ارقامی مانند تعداد فروش و مبلغ فروش یا تخفیف اعداد و ارقامی هستند که این رویداد را توصیف می کنند. همچنین برداشت وجه نقد از دستگاه عابر بانک یک رویداد است که اتفاق می افتد و اعداد و ارقامی چون مبلغ برداشتی از حساب اعداد و ارقامی هستند که این رویداد را توصیف می کنند.

جداول **Dimension** ، موجودیت هایی هستند که یک رویداد را شرح می دهند. در واقع مشتری، زمان، کارمند، فروشگاه و ... موجودیت هایی هستند که هر کدام به صورت جداگانه در قالب یک جدول ظاهر می شوند و رویداد فروش را شرح می دهند. جداول **Fact** معمولاً شامل اعداد و ارقام هستند در حالی که جداول **Dimension** معمولاً رشته ها را در خود جای می دهند.

از آنجا که در طراحی ستاره ای جدول **Fact** با **Dimension** ها در ارتباط است می توان گفت جداول **Fact** شامل کلید جداول **Dimension** و اعداد و ارقام مربوط به رویداد است. جداول **Fact** معمولاً شامل اعداد و ارقام هستند در حالی که جداول **Dimension** معمولاً رشته ها را در خود جای می دهند.