**What Is warehouse ?**

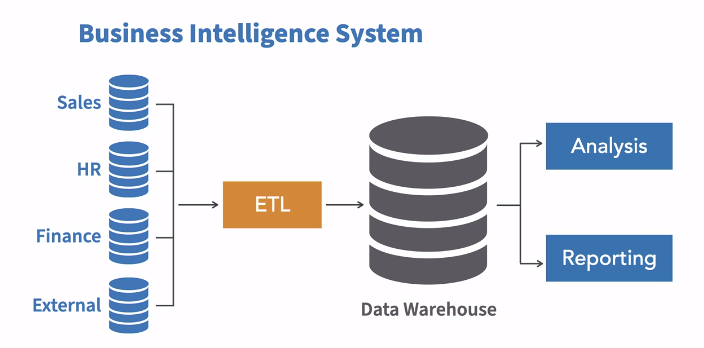
A data warehouse is a large and centralized repository of data that is specifically designed to support **business intelligence (BI)** activities such as reporting, data analysis, and data mining. It is a system that stores and manages data from multiple sources in a way that is optimized for querying and analysis.

The data within a data warehouse is typically organized into a dimensional model, which is a database design optimized for reporting and analysis. This model consists of a set of fact tables that contain the quantitative data about the business, such as sales, revenue, or customer behavior, and a set of dimension tables that provide context for the facts, such as customer, product, or time.

The process of building a data warehouse typically involves extracting data from various sources, transforming it into a consistent format, and loading it into the data warehouse. This process is known as **ETL (extract, transform, and load)** and is critical to ensuring that the data is accurate, consistent, and complete.

Data warehouses are an important tool for organizations that need to analyze large amounts of data to make informed business decisions. They allow organizations to consolidate data from multiple sources, provide a single source of truth for reporting and analysis, and support complex queries and analysis that would be difficult or impossible with other systems.

یک انبار داده، یک مخزن داده بزرگ و مرکزی است که برای هوش تجاری طراحی شده است. این سیستم، داده ها را از منابع مختلف جمع‌آوری، مدیریت و ذخیره می‌کند و برای پرس‌وجو و تجزیه و تحلیل بهینه شده است. داده ها **در انبار داده، به صورت مدل بعدی سازماندهی می‌شوند که یک طراحی پایگاه داده بهینه شده برای گزارش دهی و تجزیه و تحلیل است**. برای ساخت انبار داده، داده ها از منابع مختلف استخراج، به یک فرمت یکسان تبدیل و در انبار داده بارگذاری می‌شوند. انبار داده، به سازمان‌ها این امکان را می‌دهد که داده ها را از منابع مختلف تجمیع کرده، یک منبع یکتا برای گزارش دهی و تجزیه و تحلیل فراهم کنند و پرس‌وجوهای پیچیده و تحلیل‌هایی را که با سیستم‌های دیگر سخت یا غیرممکن است، پشتیبانی کنند



**Benefits of data warehouse**

There are several benefits of having a data warehouse, including:

**Improved data quality:** Data warehouses typically involve a process of data cleaning and transformation, which can improve data quality and consistency. This means that decision-makers can rely on accurate and trustworthy data for making informed business decisions.

**Enhanced reporting and analysis**: Data warehouses are optimized for reporting and analysis, which means that decision-makers can easily access and analyze the data they need to make informed decisions. This can lead to better insights and more effective decision-making.

**Centralized data management:** Data warehouses consolidate data from multiple sources into a single location, which makes it easier to manage and maintain data. This can save time and resources compared to managing data in multiple locations.

**Historical data analysis:** Data warehouses can store historical data over a long period of time, which means that decision-makers can analyze trends and patterns over time. This can help identify opportunities for improvement or areas where the business is performing well.

**Scalability:** Data warehouses are designed to handle large volumes of data and can scale to accommodate growing data volumes. This means that organizations can continue to use the data warehouse as their data needs grow.

Overall, data warehouses provide a powerful tool for organizations to manage and analyze their data, which can lead to better business decisions and improved performance.

داشتن انبار داده چندین مزیت دارد، از جمله:

**بهبود کیفیت داده:**

انبار داده ها معمولاً شامل فرآیند پاکسازی و تبدیل داده ها هستند که می‌تواند کیفیت و پایداری داده ها را بهبود بخشد. به این معناست که تصمیم گیرندگان می‌توانند بر روی داده های دقیق و قابل اعتماد برای تصمیم گیری تجاری خود حساب کنند.

**بهبود گزارش دهی و تجزیه و تحلیل:**

انبار داده ها برای گزارشدهی و تجزیه و تحلیل بهینه شده اند، به این معناست که تصمیم گیرندگان می‌توانند به راحتی داده های مورد نیاز خود را برای تصمیم گیری بهتر و بازدهی بیشتر دسترسی داشته باشند.

**مدیریت داده مرکزی:**

انبار داده ها داده ها را از منابع مختلف به یک مکان متمرکز تجمیع می‌کنند، که این باعث سهولت مدیریت و حفظ داده ها می‌شود. این می‌تواند در مقایسه با مدیریت داده ها در چندین مکان، زمان و منابع را صرفه جویی کند.

**تجزیه و تحلیل داده های تاریخی:**

انبار داده ها می‌توانند داده های تاریخی را در طول یک دوره زمانی طولانی ذخیره کنند، که این باعث می‌شود تصمیم گیرندگان بتوانند روندها و الگوها را در طول زمان تجزیه و تحلیل کنند. این می‌تواند به شناسایی فرصت‌های بهبود یا حوزه هایی که کسب و کار در آن عملکرد خوبی دارد، کمک کند.

**قابلیت مقیاس پذیری:**

انبار داده ها برای مدیریت حجم بزرگی از داده ها طراحی شده اند و می‌توانند برای سازمان ها با رشد نیازهای داده‌ای، قابلیت مقیاس پذیری داشته باشند.

در کل، انبار داده ها ابزار قدرتمندی برای سازمان ها برای مدیریت و تجزیه و تحلیل داده های خود است، که می‌تواند منجر به تصمیم گیری های بهتر و بهبود عملکرد شود.

**انبار داده چیست ؟**

1 – یک بانک اطلاعاتی عادی است

2 – دارای ساختار D – Normal

3 – هسته اصلی یک سیستم BI

4 – مورد استفاده جهت تحلیل + اخذ گزارش

**انبار داده چه ویژگی هایی دارد ؟**

**1 – Subject Oriented ( موضوع گرا )**

مثلا من یک انبار داده دارم یک سری دیتا در مورد فروش یک کالا دارد یکی از موضوعات ( Subject ) هایی که در حوزه فروش وجود دارد بحث مشتری هست . یا موضوعاتی مثل زمان ، کالا و فروش ، قیمت و .... اگر من مشتری رو حذف کنم سیستم دچار بی ارزش شدن تحلیل ها می شود . یعنی ما نمی توانیم موضوعات اساسی را در یک انبار داده حذف کنیم .

**2 – Integrated ( یکپارچه )**

دیتا ها در انبار داده از جاهای مختلف می آید سیستم فروش مشهد ، سیستم فروش طرقبه ، سیستم فروش شاندیز قراره بیان یک جا . اگر این سیستم های فروش من چلو کباب بفروشم باید برای تحلیل چلو کباب به صورت کلی اگر کدینگ چلوکباب در سه سیستم فروش متفاوت هست برای تحلیل درست باید کدینگ ها را یکسان کنم .

**3 – Non – Volatile ( از بین نرفتنی ( غیر قابل تغییر + غیر قابل حذف ))**

فرض کنید سیستم فروش در یک شرکت مدت ها به صورت سنتی انجام می شود دیتا به یک شکل سنتی ذخیره می شده است . حال از تاریخ امروز سیستم فروش شرکت به کل به صورت تجارت الکترونیک انجام می شود به جای اینکه دیتا های قبل تغییر کند دیتا های جدید به صورت یک موجودیت جدید در نظر گرفته می شود   
موجودیت سیستم فروش سنتی و موجودیت سیستم فروش الکترونیک . تحلیل ها به دو صورت بیان می شود فروش های سنتی ، فروش های دیجیتال .

**4 – Time – Varient ( متغیر با زمان )**

باید مشخص شود هر رکورد در انبار داده کی آمده است تا با استفاده از مورد سوم و مورد چهارم بتوانیم تحلیل ها رو ایجاد کنیم .

**انبار داده از دید فنی**

با استفاده از D Normal کردن Subject ها

**نقش انبار داده در پروژه های هوش تجاری :**

1 – جمع آوری مجموعه ای از داده ها ( انبار داده قلب پروژه های BI است )

2 – مورد استفاده در فرایند های تحلیل

3 – مورد استفاده فرایند های گزارش گیری

**ETL چیست ؟**

**ETL مخفف Extract, Transform, and Load است. این فرآیند در ادغام داده ها استفاده می شود تا داده ها را از منابع مختلف جمع آوری کرده، آن‌ها را به یک فرمت یکسان تبدیل کرده و در یک سیستم مقصد مانند دیتا ورهاوس یا دیتا مارت بارگذاری کند.**

**فاز استخراج**، شامل بازیابی داده از منابع مختلف مانند پایگاه داده ها، فایل ها یا API ها است. داده ها ممکن است در قالب ها و ساختارهای مختلف باشند. هدف این فاز، استخراج داده های لازم از این منابع و به صورت یک فرمت مشترک به آنها انتقال داده شود.

**فاز تبدیل**، شامل تبدیل و استانداردسازی داده های استخراج شده به یک فرمت یکسان است که می‌تواند توسط سیستم هدف استفاده شود. این ممکن است شامل پاکسازی داده ها، اعتبارسنجی داده ها، غنی سازی داده ها و تجمیع داده ها باشد. داده های تبدیل شده به طور معمول در یک مخزن میانی به نام ناحیه مرحله اولیه ذخیره می‌شوند، پیش از بارگذاری در سیستم مقصد.

**فاز بارگذاری**، شامل بارگذاری داده های تبدیل شده به سیستم مقصد، مانند دیتا ورهاوس ( انبار داده ) یا دیتا مارت است. داده ها می‌توانند به شکل‌های مختلفی بارگذاری شوند، بسته به نیاز سیستم مقصد. هدف این فاز، اطمینان از بارگذاری داده ها به صورت دقیق و کارآمد در سیستم مقصد است تا برای گزارش دهی، تجزیه و تحلیل و تصمیم گیری استفاده شوند.

در کل، ETL یک فرآیند اساسی در ادغام داده ها است که به سازمان ها امکان می دهد تا داده های خود را از منابع مختلف جمع آوری و تجمیع کنند، آنها را به یک فرمت یکسان تبدیل و در سیستم مقصد برای تجزیه و تحلیل و گزارشدهی بارگذاری کنند.

**نقش ETL در هوش تجاری**

استخراج داده ها از سیستم های عملیاتی ، مرتب سازی و استاندارد سازی دیتا ها و ذخیره سازی اطلاعات شسته و رفته شده در انبار های اطلاعاتی کل تعریف مکانیزم ETL هست .

**ابزاری که برای ETL مورد استفاده قرار می گیرد :**

1 – استفاده از SSIS

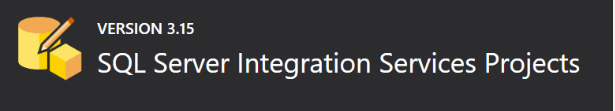
2 – با قابلیت نصب در یک سرور جداگانه

3 – توسعه با استفاده از Visual Studio Data Tools

4 – سرویس مورد استفاده در SQL Server : Sql Server Integration Service

برای داشتن این امکان در ویژال استادیو باید Extension مربوطه نصب شود

Sql Server Integration Services Projects

****