

# به نام خدا

امیر مهدی شاه حیدری

ازمایشگاه پایگاه داده

استاد میثاق یاریان

Data warehouse

## انبار داده چیست؟

انبار داده یا Data Warehouse مجموعه بزرگی از داده‌های تجاری است که به سازمان‌ها و کسب‌وکارها کمک می‌کند تا در تصمیم‌گیری‌های خود دقیق‌تر و هوشمندانه‌تر عمل کنند. انبار داده مفهوم جدیدی نیست و از دهه ۱۹۸۰ وجود داشته است. حجم زیادی از داده‌های موجود در انبارهای داده از منابع مختلفی جمع‌آوری می‌شوند؛ برنامه‌های کاربردی داخلی مانند بازاریابی، فروش و امور مالی نمونه‌هایی از این منابع هستند.

## چگونه می‌توانیم یک انبار داده بسازیم؟

برای ساخت انبار داده باید ۷ مرحله زیر را به ترتیب طی کنید:

### ۱- تعیین اهداف تجاری

- تعیین اهداف تجاری (تاکتیکی و استراتژیک)
  - شناسایی و اولویت‌بندی انتظارات و نیازهای شرکت، بخش‌ها و کاربران تجاری از پروژه
  - بررسی معماری تکنولوژیک فعلی شرکت، برنامه‌های کاربردی در حال استفاده و...
  - انجام یک تجزیه و تحلیل اولیه
  - تشریح محدوده انبار داده
- این مرحله به ۳ تا ۲۰ روز زمان نیاز دارد.

### ۲- مفهوم‌سازی و انتخاب پلتفرم

- تعریف مجموعه ویژگی‌های راحل انبار داده مورد نظر
  - انتخاب گزینه بهینه استقرار (در محل/فضای ابری/هیبرید)
  - انتخاب رویکرد طراحی معماری بهینه برای ساخت انبار داده
  - انتخاب فناوری‌های انبار داده با در نظر گرفتن تعداد منابع داده و حجم داده برای بارگیری در انبار داده
  - جریان‌های داده برای پیاده سازی
- این مرحله بین ۳ تا ۱۵ روز زمان می‌برد.

### ۳- ایجاد نقشه راه پروژه

- تعریف محدوده پروژه توسعه انبار داده، برنامه‌ریزی بودجه، جدول زمانی و...
- برنامه‌ریزی طراحی، توسعه و آزمایش

- تدوین اسناد محدوده پروژه انبار داده، سند چشم‌انداز معماری راهکار انبار داده، استراتژی استقرار انبار داده، استراتژی تست، نقشه راه اجرای پروژه
  - تدوین برنامه مدیریت ریسک
  - برآورد تلاش‌ها برای پروژه توسعه انبار داده، ROI و TCO
- زمان تقریبی این مرحله ۴ تا ۱۵ روز است.

#### ۴- تجزیه و تحلیل سیستم و طراحی معماری انبار داده

- تجزیه و تحلیل دقیق هر منبع داده
  - نوع و ساختار داده
  - حجم داده تولید شده روزانه
  - درجه حساسیت داده‌ها و رویکرد دسترسی به داده کاربردی
  - کیفیت داده، داده‌های از دست رفته/ کم‌ارزش، امکان انجام پاکسازی داده‌ها در سیستم منبع داده
  - ارتباط با سایر منابع داده
  - طراحی سیاست‌های پاکسازی داده‌ها
  - ایجاد خط‌مشی‌های امنیتی داده‌ها (سیاست‌های دسترسی به داده‌ها بر اساس محدودیت‌های قانونی و قوانین امنیت داده‌ها، سیاست‌های رمزگذاری داده‌ها، سیاست‌های نظارت بر دسترسی به داده‌ها و انطباق داده‌ها، راهبرد پشتیبان‌گیری از داده‌ها و...)
  - طراحی مدل‌های داده برای انبار داده و دیتا مارت
  - طراحی فرآیندهای ETL/ELT برای یکپارچه‌سازی و کنترل جریان داده‌ها
- انجام این مرحله دست‌کم ۱۵ روز زمان نیاز دارد.

#### ۵- توسعه و تثبیت

- سفارشی‌سازی پلتفرم انبار داده
- پیگیربندی نرم‌افزار امنیت داده‌ها و اجرای سیاست‌های امنیت داده
- توسعه و تست ETL/ELT
- تست عملکرد انبار داده

#### ۶- راه‌اندازی

- انتقال داده‌ها، ارزیابی کیفیت داده‌ها
- معرفی انبار داده به کاربران تجاری
- برگزاری جلسات و کارگاه‌های آموزشی برای کاربران

#### ۷- پشتیبانی پس از راه‌اندازی

- پیگیربندی عملکرد ETL / ELT

- پیکربندی عملکرد و در دسترس بودن انبار داده
- پشتیبانی از کاربران نهایی

بسیاری از کاربران و متخصصان فنی ممکن است بپرسند تفاوت انبار داده و مرکز داده (Data Center) در چیست. در ادامه به این موضوع می‌پردازیم.



### تفاوت انبار داده و مراکز داده چیست؟

مراکز داده، مکان‌های فیزیکی هستند که سرورها در آنها نگهداری می‌شوند؛ در حالی که انبار داده، یک مفهوم نرم‌افزاری و در واقع یک داده ساختار بر روی یک یا چند سرور است.

به بیان ساده‌تر، مرکز داده یک اتاق فیزیکی یا ساختمانی است که سرورهای داده و کامپیوترها در آن قرار می‌گیرند. در حالی که یک انبار داده فقط نوعی پایگاه داده نرم‌افزاری است که برای گزارش‌گیری و تجزیه و تحلیل داده‌ها استفاده می‌شود و یکی از اجزای اصلی هوش تجاری به شمار می‌آید.

از دیگر چیزهایی که ممکن است با انبار داده اشتباه گرفته شود پایگاه داده (Database) است. د ادامه تفاوت انبار داده و پایگاه داده را توضیح خواهیم داد.

## تفاوت انبار داده و پایگاه داده چیست؟

پایگاه داده سامانه‌ای اطلاعاتی است که وضعیت حال حاضر یک سامانه نرم‌افزاری را ثبت می‌کند و مقادیر داده‌ها به صورت مرتب و مکرر در حال تغییر و به روز رسانی است. در حالی که انبار داده یک سامانه اطلاعاتی است که داده‌های تاریخی را از منابع مختلف در خود گردآوری و جمع می‌کند.

پایگاه داده برای ذخیره و بازیابی مکرر داده‌های معین و انجام تراکنش‌های آنلاین (OLTP) طراحی شده است در حالی که انبار داده برای تجزیه و تحلیل تجمیعی داده‌ها (OLAP) کاربرد دارد.

## انبار داده شامل چه بخش‌هایی می‌شود؟

یک انبار داده معمولی اغلب شامل عناصر زیر است:

- یک پایگاه داده رابطه‌ای برای ذخیره و مدیریت داده‌ها
- یک راحل استخراج، بارگذاری و تبدیل (ELT) تا داده‌ها را برای تجزیه و تحلیل آماده کند
- قابلیت‌های تجزیه و تحلیل آماری، گزارش‌دهی و داده‌کاوی
- ابزارهای تجزیه و تحلیل مشتری برای مصورسازی و ارائه داده‌ها به کاربران تجاری

## دلایل استفاده از انبار داده

انبارهای داده این مزیت کلی را ارائه می‌کنند که به سازمان‌ها اجازه می‌دهند تا حجم زیادی از داده‌های مختلف را تجزیه و تحلیل کرده و اطلاعات ارزشمند و قابل توجهی از آن‌ها استخراج کنند.

چهار ویژگی منحصر به فرد زیر به انبارهای داده اجازه می‌دهند تا این مزیت کلی را در اختیار سازمان‌ها قرار دهند:

- **موضوع محور:** انبارهای داده می‌توانند داده‌های مربوط به یک موضوع خاص یا حوزه عملکردی (مانند فروش) را تجزیه و تحلیل کنند.
- **یکپارچه:** انبارهای داده بین انواع داده از منابع مختلف سازگاری ایجاد می‌کنند.
- **تغییرناپذیر:** هنگامی که داده‌ها در یک انبار داده ذخیره می‌شوند، پایدار می‌مانند و تغییر نمی‌کنند.
- **پنجره زمانی متغیر:** تجزیه و تحلیل در انبار داده در پنجره‌های زمانی مختلف قابل انجام است. انبار داده‌ای که به خوبی طراحی شده باشد، کوئری‌ها را سریع پاسخ می‌دهد، توان عملیاتی بالایی دارد و انعطاف کافی را برای کاربران نهایی فراهم می‌کند تا بتوانند حجم زیادی از داده‌ها را به آسانی و با سرعت بالا تحلیل کنند.

## استفاده از انبار داده در صنایع مختلف

[کلان‌داده‌ها](#) در برخی از صنایع نقش بسیار حیاتی و مهمی را بازی می‌کنند. این صنایع به انبار داده نیاز ویژه‌ای دارند. در ادامه برخی از صنایعی که بیشتر به سراغ استفاده از انبارهای داده می‌روند را معرفی می‌کنیم:

## صنعت سرمایه‌گذاری و بیمه

انبار داده بیش از هر صنعت دیگری برای تجزیه و تحلیل روندهای مشتری، بازار و سایر الگوهای داده در بخش سرمایه‌گذاری و بیمه استفاده می‌شود. بازارهای فارکس و سهام دو زیرمجموعه اصلی این صنعت هستند که انبارهای داده نقش مهمی در آنها ایفا می‌کنند.

## صنعت خرده‌فروشی

در صنعت خرده‌فروشی از انبار داده برای ردیابی کالاها، بررسی سیاست‌های قیمت‌گذاری و تجزیه و تحلیل روند خرید مشتری استفاده می‌شود.

## صنعت سلامتی و مراقبت‌های بهداشتی

در بخش مراقبت‌های بهداشتی، انبارهای داده برای پیش‌بینی نتایج، تولید گزارش‌های درمانی و به اشتراک‌گذاری داده‌ها با ارائه‌دهندگان خدمات بیمه، آزمایشگاه‌های تحقیقاتی و سایر واحدهای پزشکی استفاده می‌شود. انبارهای داده ستون فقرات سیستم‌های مراقبت‌های بهداشتی هستند زیرا دسترسی به جدیدترین و به‌روزترین اطلاعات درمانی برای نجات جان انسان‌ها بسیار مهم است.

## ابزارهای انبار داده

ابزارهای زیادی برای ذخیره‌سازی داده‌ها در بازار موجود است. در اینجا چند نمونه از معروف‌ترین ابزارهای انبار داده را به شما معرفی می‌کنیم:

### MarkLogic

این ابزار، راه‌حلی مفید برای ذخیره‌سازی داده است که با استفاده از برخی ویژگی‌های سازمانی، یکپارچه‌سازی داده‌ها را آسان‌تر و سریع‌تر می‌کند. این ابزار باعث می‌شود تا فرآیندهای جستجوی بسیار پیچیده، به شکل ساده‌تری انجام شوند MarkLogic. می‌تواند انواع مختلفی از داده‌ها مانند داکيومنت‌ها، روابط و ابرداده‌ها را به کوئری تبدیل کند.

### Oracle

اوراکل یک پایگاه داده پیشرو در صنعت داده است. این ابزار، طیف گسترده‌ای از راه‌حل‌های انبار داده را ارائه می‌دهد و با افزایش کارایی عملیاتی، به بهینه‌سازی تجربیات مشتری کمک می‌کند.

### Amazon RedShift

آمازون ردشیفت یکی از بهترین ابزارهای انبار داده است؛ ابزاری ساده و مقرون به صرفه برای تجزیه و تحلیل انواع داده‌ها با استفاده از SQL استاندارد و ابزارهای BI موجود. این ابزار، امکان اجرای کوئری‌های پیچیده را در برابر پتابایت داده‌های ساختاریافته با استفاده از تکنیک بهینه‌سازی کوئری در اختیار ما قرار می‌دهد.

## جمع‌بندی

داده‌ها برای تصمیم‌گیری آگاهانه سازمان‌ها بسیار ضروری‌اند، بنابراین منطقی است که انبارهای داده برای هر سازمانی مهم باشند، زیرا همه داده‌ها را در خود ذخیره می‌کنند. بدون انبار داده نمی‌توانید به جریان اطلاعات دسترسی داشته باشید و از هوش تجاری در کسب‌وکار خود بهره‌مند شوید.

انبارهای داده کمک می‌کنند تا حجم زیادی از داده‌ها را در یک پایگاه داده مرکزی ذخیره کنید، آن‌ها را در یک مکان امن نگهدارید و در زمان نیاز، داده‌ها را برای نیازهای تجاری خود تجزیه و تحلیل کنید.