

# انبار داده (Data Warehouse) چیست؟ +

## مراحل، انواع و کاربردهای آن

در حال حاضر بیشتر فعالیتهای روزمره و حرفه‌ای به صورت الکترونیکی و دیجیتال انجام می‌شود و همراه با تولید داده‌های فراوانی است که برخی ذخیره‌سازی و نگهداری از این داده‌ها ضرورت دارد. داده‌ها در انبار داده Data Warehouse ذخیره می‌شوند که در واقع سیستمی برای جمع‌آوری، ذخیره و تحلیل داده‌هاست.

با پیشرفت تکنولوژی و به‌کارگیری ابزار دیجیتال در کسب‌وکارها در دهه 1980 میلادی مدیران شرکت‌های بزرگ و پیشرو متوجه شدند که پایگاه‌های داده سنتی برای ذخیره‌سازی اطلاعات کافی نیستند و باید یک سیستم مرکزی قوی برای این کار در نظر گرفته شود.

در هر سازمان نرم‌افزارهای متنوعی برای امور مختلف به تولید یا جمع‌آوری داده می‌پردازند. در فرایند استفاده از هوش تجاری و داشبوردهای مختلف مدیریتی استفاده از این داده‌ها به عنوان Query خواهد بود و باید وقت و زمان زیادی برای آن صرف شود. انبار داده کمک می‌کند تا داده‌ها از نظر موضوعی در طبقه‌بندی خود قرار گیرند و به راحتی قابل فراخوانی باشند.

کلیک کنید

برای دریافت دمو

## معرفی انبار داده

انبار داده را به اختصار با DW یا DWH نشان می‌دهند که در واقع پایگاه یا مخزنی برای ذخیره‌سازی داده‌های الکترونیکی یک سازمان است که از طریق سیستم‌های عملیاتی آن گردآوری شده‌اند. داده‌های ذخیره‌شده در این انبار به گونه‌ای سازمان‌دهی می‌شوند که برای ایجاد گزارش‌ها و آمار و عملیات روزمره به راحتی قابل استفاده باشند.

این مرکز ذخیره‌سازی داده تمام اطلاعات را از منابع پراکنده و مختلف جمع و یکپارچه‌سازی می‌کند تا برای امور مدیریتی و تصمیم‌گیری در مراحل بعد از آنها استفاده شود.



## ویژگی‌های انبار داده

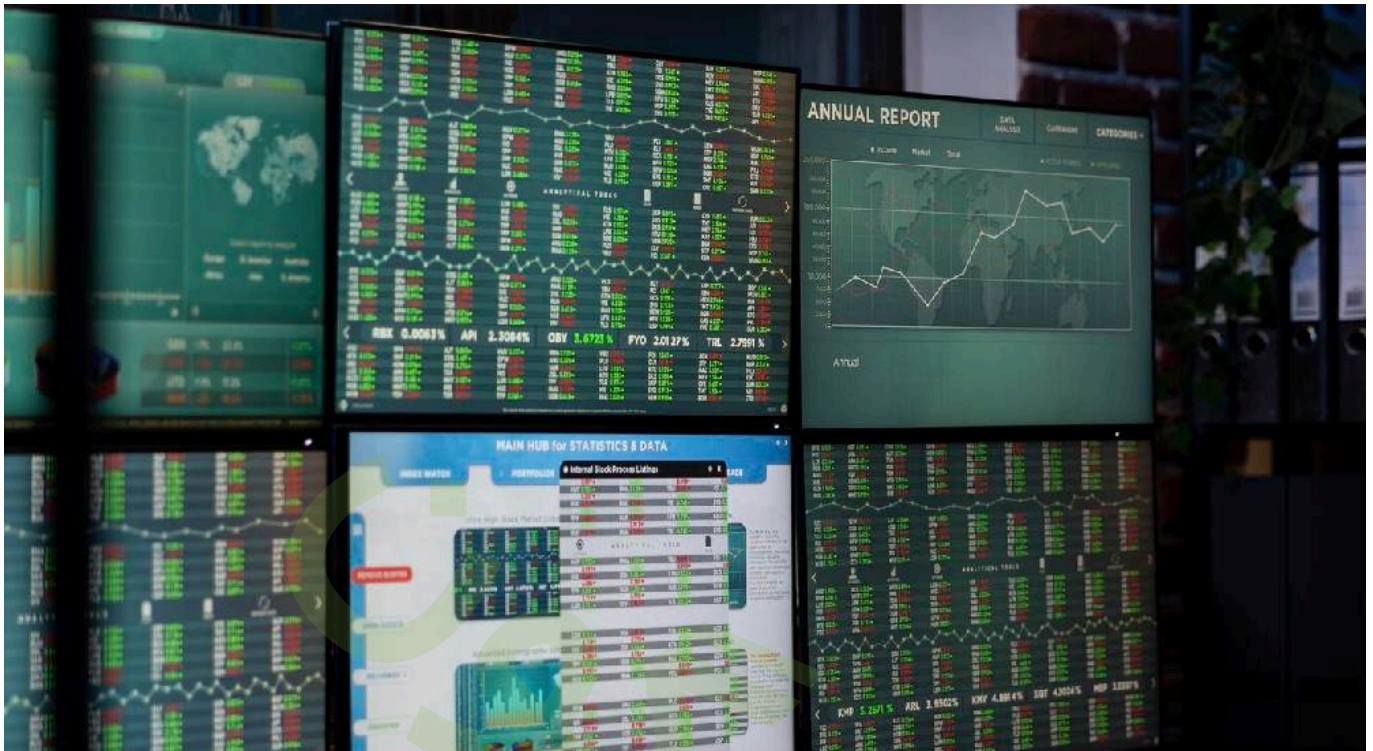
انبار داده یک مجموعه از داده‌های موضوع‌گرا (Subject Oriented)، یکپارچه (Integrated)، متغیر با زمان (Time-Varying) و غیرفرار (None-Volatile) است که در تصمیم‌سازی‌های مدیریت نقش اصلی را ایفا می‌کند. در ادامه هر یک از این ویژگی‌ها را بیشتر توضیح می‌دهیم.

- **موضوع‌گرا:** انبارهای داده به شکلی بهینه‌سازی می‌شوند که برای پاسخ‌گویی به پرسش‌های مختلف درباره یک موضوع آماده شوند. به عبارت دیگر این راهکار به جای آنکه تمام فعالیت‌های سازمان را در خود داشته باشد به یک موضوع خاص اختصاص داده می‌شود. مثلاً اطلاعاتی نظیر محصول، فروش، مشخصات مشتری و تأمین‌کننده را خواهد داشت.

• **یکپارچه:** داده‌های مختلفی که در یک انبار داده گردآوری می‌شوند ممکن است از منابع مختلفی جمع‌آوری شده باشند. به همین منظور داده‌ها پیش از آنکه در انبار ذخیره‌سازی شوند از تکنیک‌های پاکسازی و مجتمع‌سازی در آنها استفاده می‌شود. مثلاً زمانی داده‌ها از منابع ناهمگنی مثل فایل‌ها و پایگاه‌های داده به دست آمده باشند، یکپارچه‌سازی رابطه‌ای بین داده‌ها ایجاد می‌کند تا تجزیه و تحلیل آنها را بهبود دهد.

• **متغیر با زمان:** داده‌ها در انبارهای داده با شناسه زمانی ذخیره‌سازی می‌شوند. داده‌های قدیمی‌تر همچنان با همان هویت قبلی در این انبار نگهداری و در صورت تغییر با شناسه زمانی جدید ذخیره می‌شوند. به این ترتیب هیچ داده‌ای پس از ثبت شدن از بین نمی‌رود و نسخه پیشین هر به‌روزرسانی و تغییری در این انبار باقی خواهد ماند.

• **غیرفراز:** داده‌ها همیشه در این انبار ماندگار خواهند بود و اگر داده‌ای ثبت شود امکان تغییر در آن وجود ندارد. ممکن است با جایگزینی به‌روزرسانی، داده‌ای جدید از ورژن قدیمی به وجود آید، اما ورژن قدیمی همچنان به قوت خود باقی و تمام و کمال در پایگاه داده حفظ می‌شود.



## معماری انبارهای داده

معماری انبارهای داده یک ساختار دقیق و متناسب با نیازهای کسب و کار و سازمان است. در معماری انبارهای داده از مدل‌های مختلفی استفاده می‌شود تا بهترین روش استخراج اطلاعات از داده‌ها قابل انجام باشد. سه نوع معماری اصلی در طراحی این انبارها وجود دارد: معماری تک لایه، دو لایه و سه لایه.

- **معماری تک لایه:** هدف از این معماری به حداقل رساندن مقدار داده‌هایی است که ذخیره می‌شوند. از این معماری استفاده چندانی نمی‌شود.
- **معماری دو لایه:** در این معماری منابع فیزیکی و انبار داده از هم جدا می‌شوند. این

معماری قابلیت گسترش ندارد و کاربران چندانی را پشتیبانی نمی‌کند.

- معماری سه لایه: پرکاربردترین نوع معماری است و از سه لایه زیرین، میانی و بالایی تشکیل شده است.

- لایه زیرین یا مرحله استخراج، تبدیل و بارگذاری داده ETL: در این مرحله داده‌ها

از منابع مختلف جمع‌آوری و بارگذاری می‌شوند. این منابع می‌تواند شامل پایگاه‌های داده و فایل‌ها باشند. بانک اطلاعاتی سرورها در این لایه قرار می‌گیرد و داده‌ها با استفاده از ابزار back-end پاکسازی، تبدیل و بارگذاری می‌شوند.

- لایه میانی یا تحلیل داده: در این مرحله داده‌ها به صورت پیوسته آنالیز می‌شوند تا

اطلاعات و آمار مورد نیاز مدیران و کاربران را استخراج کنند. در این مرحله از تحلیل داده از تکنولوژی OLAP استفاده می‌شود. OLAP مخفف Online

analytical processing است که به آن پردازش تحلیلی آنلاین نیز گفته می‌شود.

- لایه بالایی: شامل لایه کاربری یا front-end است که می‌تواند از ابزار جانبی نیز استفاده کند.

## مزایای ابزار داده

- دسترسی به داده‌های یکپارچه
- افزایش سرعت تصمیم‌گیری
- بهبود کیفیت تصمیم‌گیری
- سازماندهی بهتر داده‌ها

- افزایش قابلیت تحلیل
- کاهش هزینه‌ها
- افزایش توانایی پیش‌بینی
- ارتباط بهتر با مشتریان
- کاهش خطاها و اشتباهات

## مدل‌های انبار داده

دو مدل انبار برای داده وجود دارد. مجازی و دیتا مارت.

- **مجازی:** انبار داده مجازی نمای کاملی از موجودی داده‌ها را ارائه می‌کند و انبار یا مخازن داده‌های عملیاتی کمی به عنوان انبار داده مجازی شناخته می‌شوند. داده‌های متا در این انبارها قرار می‌گیرند. در انبارهای داده مجازی کیفیت و سازگاری داده تضمین نمی‌شود. زمان دسترسی کاربر نهایی نیز با توجه به وجود منابع عملیاتی، بار شبکه و پیچیدگی درخواست غیرقابل پیش‌بینی است.
- **دیتا مارت:** Data Mart یک مدل ویژه است که برای بازیابی داده‌های مشتری مداری استفاده می‌شود. دیتا مارت زیرمجموعه انبارهای داده است که داده‌های یک فعالیت مشخص را در خود دارد. انبارهای داده از داده‌هایی با گستردگی بیشتر پشتیبانی می‌کنند اما دیتا مارت تنها از داده‌های یک واحد سازمانی با موضوعاتی محدود پشتیبانی می‌کند. دیتا مارت یک نمونه کوچک و زیرمجموعه انبار داده است. دیتا مارت سه نوع مختلف دارند: وابسته (Dependent)، مستقل (Independent) و ترکیبی (Hybrid).



## تکنیک‌های مدل‌سازی انبار داده

- مدل ستاره‌ای
- مدل دانه برفی
- طراحی بالا به پایین
- طراحی پایین به بالا

## مراحل ساخت انبار داده

- **تعیین اهداف تجاری:** شامل شناسایی و اولویت‌بندی نیازهای کسب‌وکار و کاربران، تعیین اهداف به طور شفاف، بررسی زیرساخت‌های فناوری در حال استفاده، بررسی‌های اولیه و مشخص کردن گنجایش انبار می‌شود.

- **مفهوم‌سازی و انتخاب پلتفرم:** شامل تعیین ویژگی‌های انبار، انتخاب بهترین گزینه استقرار انبار، رویکرد طراحی معماری، انتخاب فناوری و جریان داده می‌شود.
- **ایجاد نقشه راه پروژه ساخت:** شامل تعریف حدود پروژه، بودجه‌بندی و زمان‌بندی، طراحی و توسعه و آزمایش، تهیه مستندات لازم برای ساخت انبار و مدیریت ریسک است.
- **تجزیه و تحلیل سیستم و طراحی معماری انبار داده:** شامل بررسی کامل هر منبع داده، شناسایی نوع و ساختار داده‌ها، اندازه‌گیری و تخمین حجم داده‌های تولیدشده، تعیین میزان اهمیت و حساسیت داده‌ها و سطح دسترسی، کیفیت داده، ارتباط با منابع داده دیگر، طراحی شیوه پاکسازی داده‌ها و تأمین امنیت داده‌ها می‌شود.
- **توسعه و تثبیت:** شامل قابلیت ایجاد تنظیمات پلتفرم داده، امنیت داده و تست عملکرد است.
- **راه‌اندازی:** در این مرحله داده‌ها به انبار ایجادشده، منتقل می‌شوند و برنامه‌های معرفی و تبلیغات و آموزش و وبینارهای معرفی انجام می‌شود.
- **پشتیبانی:** در این مرحله از عملکردهای مختلف و کاربران پشتیبانی می‌شود.

## عملکرد انبار داده در سازمان

1. **استخراج داده‌ها:** جمع‌آوری و استخراج داده‌ها از پایگاه‌های داده و انتقال آنها به یک مخزن یکپارچه اولین مرحله از ایجاد این انبار در سازمان است.
2. **تبدیل داده‌ها:** داده‌های مختلف در سیستم‌های متفاوت و با توجه به نیاز هر سازمان

ایجاد می‌شوند. یکسان‌سازی داده‌ها در این شرایط ضروری است. این داده‌ها معمولاً در بستریهای متفاوتی مانند، Oracle ، Sybase ، Microsoft SQL Server ، Microsoft Access طراحی شده‌اند. در این مرحله تمامی داده‌ها یکسان‌سازی می‌شوند و بین آنها ارتباط برقرار می‌شود.

3. **بارگذاری داده در یک پایگاه داده چندبُعدی:** برخلاف پایگاه داده‌های عملیاتی،

معماری این انبارها رابطه‌ای و با استفاده از اصول نرمالیزاسیون نیست. در انبارهای داده فیلدها در جاهای مختلفی تکرار می‌شوند و روابط بین جدولها کمتر به چشم می‌خورد.

4. **تولید مقادیر از پیش محاسبه‌شده:** برای افزایش سرعت گزارش‌گیری مقادیر از

پیش محاسبه‌شده یا تراکم تولید می‌شوند. این تراکم‌ها که در ابعاد مختلف انبار داده ساخته می‌شوند باعث سرعت گزارش‌گیری خواهند شد. البته ساخت این تراکم‌ها عملیات وقت‌گیری دارد و همچنین به حافظه زیادی بر روی سرور نیازمند است.

5. **ساخت یا خرید ابزار گزارش‌گیری:** در این مرحله با استفاده از ابزار گزارش‌گیری

می‌توان از داده‌های انبارهای داده استفاده کرد. انتخاب و ساخت ابزار به نوع فعالیت سازمان و کسب‌وکار و پیچیدگی فرایندهای آنها بستگی دارد.



## کاربرد انبارهای داده

انبار با ذخیره‌سازی داده‌ها تهیه گزارش، داشبورد و ابزار تحلیل را فراهم می‌کند. تحلیل‌گران کسب‌وکار، مهندسان داده و متخصصان ابزار هوش تجاری، کاربران برنامه‌های تحلیلی از داده‌ها استفاده می‌کنند تا در بازار به رقابت پردازند و موفقیت بیشتری نصیب کسب‌وکار متعلق به خود کنند.

## کاربردهای انبار داده در صنایع مختلف

صنایعی با بیگ دیتا سروکار دارند به انبارهای داده نیاز دارند. برخی از این صنایع در ادامه معرفی می‌شوند:

- سرمایه‌گذاری و بیمه: بررسی و تجزیه و تحلیل روندهای بازار، مشتریان، الگوهای داده در بخش سرمایه‌گذاری و بیمه به کاربردهای این ابزارها نیازمند هستند.
- خرده‌فروشی: در ردیابی کالا، بررسی سیاست‌های قیمت‌گذاری و آنالیز روندهای خرید مشتریان صنعت خرده‌فروشی از ابزارهای داده استفاده می‌شود.
- بهداشت و درمان: ابزارهای داده به پیش‌بینی نتایج، تولید گزارش‌ها، به اشتراک‌گذاری اطلاعات با سازمان‌های مختلف به کار می‌آید. این روش دسترسی به داده‌ها به پیشبرد اهداف پزشکی و سلامتی کمک بسیار زیادی می‌کند.

## انواع ابزار داده از نظر کاربرد

سه نوع اصلی برای این گونه ابزار وجود دارد: سازمانی، عملیاتی و دیتا مارت.

**انبار داده سازمانی Enterprise Data Warehouse:** یک انبار متمرکز برای داده‌هاست که خدمات کل سازمان را تأمین می‌کند. رویکرد این انبار واحد است و طبق آن داده‌ها را طبقه‌بندی می‌کند.

**انبار داده عملیاتی Operational Data Store:** در مواردی به کار می‌رود که داده‌هایی فراخوانی می‌شوند تا از آنها استفاده شود.

**دیتا مارت Data Mart:** ابزارهای جانبی داده‌ای هستند که هر کدام به یک بخش اختصاص دارند و تغییراتی که در آنها اعمال می‌شود بر داده‌های دیگر تأثیر نخواهد داشت.

## فرق انبار داده و پایگاه داده

انبار داده و پایگاه داده Database شباهت‌های زیادی با یکدیگر دارند. تفاوت این دو زمانی آشکار می‌شود که با داده‌های بزرگ سروکار داشته باشیم و پایگاه داده قابلیت اداره آن را نداشته باشد.

- پایگاه داده برای مدل‌هایی بهینه‌سازی شده که بر اساس مدل رابطه‌ای تراکنش‌های همروند را پردازش می‌کنند. انبارهای داده برای پردازش تحلیلی آنلاین طراحی شده و امکان پردازش تعداد کمی کوئری پیچیده بر روی تعداد زیادی رکورد داده‌ای را به وجود می‌آورند.
- کاربران پایگاه داده کارمندان اداری هستند. کاربران انبارهای داده مدیران و تصمیم‌گیرندگان هستند.
- داده‌های یک پایگاه داده در بیشترین حالت به چند گیگا بایت می‌رسد. داده‌های انبارهای داده از چند گیگا بایت تا چند ترابایت هستند.

## فرق انبارهای داده و مرکز داده

مراکز داده مکان‌هایی فیزیکی و ساختمان‌هایی هستند برای نگهداری سرورها و کامپیوترها به کار می‌روند، انبار داده یک مفهوم انتزاعی و غیرملموس است که یک ساختار را بر روی یک یا چند سرور دربرمی‌گیرد.



## انبیاء داده و هوش تجاری

انبیاء داده در واقع تأمین‌کننده اصلی خوراک هوش تجاری است. انبیاء داده جریان اصلی اطلاعات در سیستم‌ها و داشبوردهای عملیاتی را فراهم می‌کنند. داده‌ها در این سیستم به بلوغ می‌رسند و قابل استفاده می‌شوند. در هوش تجاری وجود انبیاء داده با جمع‌آوری داده‌ها از منابع مختلف ضروری هستند.

انبیاء داده فرایند یکپارچه‌سازی داده را انجام می‌دهند و امکان ذخیره‌سازی داده‌ها بر اساس شناسه زمانی را نیز فراهم می‌کنند. به این ترتیب گزارش‌گیری و تصمیم‌گیری بر اساس تجزیه و تحلیل‌های انجام‌شده به صورت شفاف و کاملاً دقیق امکان‌پذیر می‌شود و تصمیمات آگاهانه و با ریسک کمتری برای پیشرفت هر کسب‌وکاری گرفته خواهد شد.

محیط کاربری هوش تجاری به دلیل سازگاری با انبارهای داده قابل انعطاف و پویا هستند. هوش تجاری به کمک نرم افزارهای مختلف قابلیت ارائه داشبوردهای مقایسه‌ای و مدیریت را دارد. مثلاً در حوزه منابع انسانی که یکی از مهم‌ترین بخش‌های هر سازمانی به شمار می‌آید می‌توان از مزیت‌های هوش تجاری در مدیریت و شفاف‌سازی داده‌ها بهره‌مند شد. داشبورد هوش تجاری چارگون در حوزه منابع انسانی قابلیت‌های متعددی دارد. سازمان‌ها می‌توانند در این داشبوردها آمار مربوط به مشخصات پرسنلی، سابقه، تاریخچه همکاری، پراکندگی مدرک تحصیلی و تمامی فاکتورهای مرتبط با منابع انسانی را مشاهده کنند.

داشبوردهای منابع انسانی شرایطی را فراهم می‌آورند که مدیران با دسترسی سریع و افزایش قدرت پیش‌بینی، مدیریت ریسک بهتری داشته باشند و با پایش فرایندهای منابع انسانی موجب بهبود ساختار و وضعیت سازمانی شوند و در نهایت به بهره‌وری بیشتر که از اهداف اصلی و اولیه هر سازمان است دست یابند.

## سخن پایانی

در اختیار داشتن امکاناتی برای ذخیره‌سازی داده‌ها و فراخوانی داده برای هر سازمانی ضروری است، زیرا از این طریق می‌توان به جریان اطلاعات دسترسی داشت و از مزایای هوش تجاری در موفقیت کسب‌وکار بهره‌مند شد. انبار داده یک بانک اطلاعاتی گسترده است که به مدیران و تصمیم‌گیرندگان امکان می‌دهد تا از تمام داده‌های مورد نیاز خود گزارش‌هایی جامع دریافت کنند.