



Namatek
True Education

www.namatek.com

Data Warehouse

انبار داده چیست؟

فهرست مطالب

۱. انبار داده چیست؟
۲. ویژگی‌های انبار داده چیست؟
۳. ساختار انبار داده
۴. نحوه کارکرد انبار داده
۵. کاربرد انبار داده چیست؟
۶. طراحی انبار داده
۷. انواع انبار داده
۸. مزایا، معایب و چالش‌های استفاده از انبار داده چیست؟

وجود داده‌ها و تجزیه و تحلیل آن‌ها در هر کسب و کاری برای ماندن آن در بازار رقابتی، ضروری است. کاربران تجاری به منظور استخراج بینش از داده‌های خود، نظارت بر عملکرد کسب و کار و پشتیبانی از تصمیم‌گیری‌ها به گزارش‌ها، داشبوردها و ابزارهای تحلیلی متکی هستند. انبار داده، این نوع داده‌ها و ارائه سریع گزارش‌ها، داشبوردها و ابزارهای تحلیلی را با ذخیره سازی کارآمد داده‌ها برای به حداقل رساندن ورودی و خروجی نتایج کاوش به تعداد زیادی از کاربران به صورت همزمان، تقویت می‌کند.

در این مقاله به بررسی موارد زیر خواهیم پرداخت:

- انبار داده چیست؟
- اجزای انبار داده
- ویژگی‌های انبار داده چیست؟
- ساختار انبار داده چیست؟
- انواع انبار داده
- کاربرد انبار داده چیست؟

انبار داده چیست؟



انبار داده یک مخزن مرکزی اطلاعات است که می‌توان برای تصمیم‌گیری آگاهانه‌تر به تجزیه و تحلیل آن پرداخت. از طرف دیگر می‌توان گفت که انبار داده نوعی سیستم مدیریت داده است که به منظور فعال کردن و پشتیبانی از فعالیت‌های هوش تجاری، به ویژه در بخش تجزیه و تحلیل استفاده می‌شود. انبار داده صرفاً به منظور پرس و جو و تجزیه و تحلیل در نظر گرفته می‌شود و اغلب حاوی مقادیر زیادی از داده‌های تاریخی است. داده‌های موجود در انبار داده معمولاً از طیف گسترده‌ای از منابع، مانند فایل‌های گزارش برنامه و برنامه‌های تراکنشی مشتق می‌شوند. یک انبار داده، مقادیر زیادی از داده‌هایی را که از منابع متعدد به دست آمده‌اند، متمرکز و یکپارچه می‌کند.

قابلیت تحلیلی این نوع سیستم مدیریتی داده به سازمان‌ها کمک می‌کند تا بینش‌های تجاری ارزشمندی را از داده‌های خود به منظور بهبود در تصمیم‌گیری‌های جاری سازمان استخراج کنند. انبار داده با گذشت زمان یک رکورد تاریخی ایجاد می‌کند که می‌تواند برای دانشمندان داده و تحلیلگران تجاری بسیار با ارزش باشد. به دلیل وجود این نوع از قابلیت‌ها، یک انبار داده را می‌توان به عنوان منبع واحد حقیقت یک سازمان نیز در نظر گرفت.

عناصر انبار داده چیست؟



یک انبار داده معمولی شامل عناصر زیر است:

- یک پایگاه داده رابطه‌ای به منظور ذخیره سازی و مدیریت داده
- یک راه حل استخراج، بارگذاری و تبدیل (ETL) به منظور آماده سازی داده‌ها برای تجزیه و تحلیل
- قابلیت تجزیه و تحلیل آماری، گزارش دهی و داده کاوی
- ابزارهای تجزیه و تحلیل مشتری به منظور تجسم و ارائه داده به کاربران تجاری
- سایر برنامه‌های کاربردی و تحلیلی پیچیده‌تر (اطلاعات عملی با استفاده از الگوریتم‌های علم داده، هوش مصنوعی یا نمودار ویژگی‌های فضایی که امکان تجزیه و تحلیل بیشتر داده‌ها را در مقیاس مناسب فراهم می‌کند).

ابزارهای انبار داده

ابزار انبار داده، مجموعه ای یکپارچه از سخت افزارها و نرم افزارها، مانند موارد زیر است که با استفاده از آن یک کسب و کار می‌تواند به شبکه خود متصل شود و همان گونه که هست از آن استفاده کند:

• Amazon Redshift

• Microsoft Azure

• Google BigQuery

• Snowflake

یک ابزار انبار داده از نظر موارد زیر در جایی مابین پیاده سازی‌های ابری و درون پایگاه قرار می‌گیرد:

• هزینه اولیه

• سرعت استقرار

• سهولت مقیاس پذیری

• کنترل مدیریت داده

مجوز انبار داده

یک کسب و کار می‌تواند مجوز انبار داده را خریداری کند و سپس آن را در زیرساخت داخلی خود مستقر کند. اگرچه این کار معمولاً گران‌تر از سرویس انبار داده ابری است؛ اما ممکن است برای نهادهای دولتی، مؤسسات مالی یا سایر سازمان‌هایی که می‌خواهند کنترل بیشتری بر داده‌های خود داشته باشند یا نیاز به رعایت استانداردها یا مقررات امنیتی یا حریم خصوصی داده‌ها دارند، انتخاب بهتری باشند.

ویژگی‌های انبار داده چیست؟



انبار داده به منظور ذخیره سازی داده‌های مدرن ضروری است و پایه‌ای قوی برای سازمان‌ها به منظور ادغام و تجزیه و تحلیل استراتژیک داده‌ها فراهم می‌کند. ویژگی‌های متمایزی که این کسب و کارها با استفاده از ابزارهایی برای تصمیم‌گیری‌های آگاهانه و استخراج بینش‌های ارزشمند از داده‌ها دارند، آنها را توانمند می‌سازد. این ویژگی‌ها را در ادامه بررسی خواهیم کرد.

مخزن داده متمرکز

انبار داده یک مخزن متمرکز برای تمامی داده‌های سازمانی از منابع مختلف مانند پایگاه‌های داده‌ای تراکنشی، سیستم‌های عملیاتی و منابع خارجی فراهم می‌کند. این امر به سازمان‌ها کمک می‌کند تا دید جامعی از داده‌های خود داشته باشند که می‌تواند به تصمیم‌گیری‌های آگاهانه تجاری کمک کند.

یکپارچه سازی داده‌ها

انبار داده‌ها داده‌ها را از منابع مختلف در یک نمای واحد و یکپارچه ادغام می‌کند که می‌تواند به حذف سیلوی داده و کاهش تناقضات داده کمک کند.

پرس و جو و تجزیه و تحلیل

انبار داده قابلیت‌هایی قدرتمند به منظور پرس و جو و تجزیه و تحلیل داده فراهم می‌کند که کاربران را قادر می‌سازد تا داده‌ها را به روش‌های مختلف کاوش و تجزیه و تحلیل کنند. این موضوع به شناسایی الگوها و روندها کمک می‌کند و همچنین می‌تواند در تصمیم‌گیری‌های آگاهانه تجاری به صورتی مفید و سودمند عمل کند.

تبدیل داده

انبار داده شامل فرآیند تبدیل داده است که شامل پاکسازی، فیلتر کردن و قالب بندی داده‌ها از منابع مختلف برای سازگار و قابل استفاده کردن آنها است. این امر می‌تواند به بهبود کیفیت داده‌ها و کاهش تناقضات داده ای کمک کند.

داده کاوی

انبار داده قابلیت‌هایی را برای داده کاوی فراهم می‌کند که سازمان‌ها را قادر می‌سازد تا الگوها و روابط پنهان در داده‌ها را کشف کنند. این موضوع به شناسایی فرصت‌های جدید، پیش‌بینی روندهای آینده و کاهش خطرات کمک می‌کند.

امنیت داده‌ها

انبار داده ویژگی‌های امنیتی قوی داده‌ها مانند کنترل دسترسی‌ها، رمزگذاری داده‌ها و پشتیبان‌گیری از داده‌ها را فراهم می‌کند که تضمین‌کننده داده‌های ایمن است و از دسترسی‌های غیرمجاز، آنها را محافظت می‌کند.

ساختار انبار داده



ساختار یا معماری انبار داده بر اساس نیازهایی مخصوص هر سازمان معین می‌شود. ساختارهای رایج انبار داده به صورت زیر هستند:

- **ساده:** تمامی انبارهای داده یک طرح مشترک دارند که در آن ابر داده، داده‌های خام و داده‌های خلاصه در مخزن مرکزی انبار ذخیره می‌شوند. مخزن از یک طرف توسط منابع داده تغذیه می‌شود و از طرف دیگر توسط کاربران نهایی به منظور تجزیه و تحلیل، گزارش دهی و استخراج قابل دسترسی است.
- **ساده به همراه یک منطقه نمایشی:** داده‌های عملیاتی باید قبل از قرار گرفتن در انبار پردازش شوند. اگرچه این کار را می‌توان با استفاده از

برنامه نویسی انجام داد؛ اما در بسیاری از انبارهای داده و قبل از ورود به انبار، یک منطقه مرحله بندی برای داده‌ها اضافه می‌شود تا فرآیند آماده سازی داده تا حد امکان ساده شود.

• **هاب و اسپوک (Hub and Spoke):** اضافه کردن داده‌ها بین مخزن مرکزی و کاربر نهایی به سازمان کمک می‌کند تا انبار داده خود را برای خدمات رسانی به بخش‌های مختلف کسب و کار، شخصی سازی کند. هنگامی که داده‌ها آماده استفاده هستند، به دیتا مارت (Data Mart) مناسب منتقل خواهند شد.

• **Sandbox:** سندباکس‌ها مناطق خصوصی و ایمن هستند که به شرکت‌ها اجازه می‌دهند تا به سرعت و به صورت غیررسمی، مجموعه داده‌ها یا روش‌های تجزیه و تحلیل داده‌های جدید را بدون نیاز به انطباق با قوانین و پروتکل رسمی انبار داده بررسی و ارزیابی کنند.

اجزای ساختار انبار داده شامل چه مواردی است؟



اجزای ساختار یا معماری انبار داده به صورت زیر هستند.

ETL (1)

هنگامی که تحلیلگران پایگاه داده می‌خواهند داده‌ها را از یک منبع داده به انبار داده خود منتقل کنند، ETL داده‌ها را به یک فرمت قابل استفاده تبدیل می‌کند تا زمانی که در انبار داده قرار گرفتند، بتوان آنها را تجزیه و تحلیل و کاوش کرد.

(2) فراداده

فراداده، اطلاعاتی در مورد داده است. اساساً تمامی داده‌هایی که در یک سیستم ذخیره می‌شوند تا قابل کاوش باشند را توصیف می‌کند. برخی از نمونه‌های ابر داده شامل موارد زیر است:

- نویسندگان
- تاریخ ایجاد یک فایل
- تاریخ یا مکان یک مقاله
- اندازه یک فایل
- و مواردی از این قبیل

فراداده را همانند عناوین یک ستون در صفحه در نظر می‌گیرند و به کاربران این امکان را می‌دهد که داده‌های خود را سازماندهی کنند تا قابل استفاده باشد.

بنابراین می‌توان از آنها به منظور ایجاد داشبورد و گزارش تجزیه و تحلیل استفاده کرد.

(3) پردازش کاوش در SQL

SQL زبان واقعی کاوش داده‌ها است. SQL همان زبانی است که تحلیلگران به منظور استخراج اطلاعات از داده‌های ذخیره شده در انبار داده استفاده می‌کنند. معمولاً انبارهای داده دارای فناوری‌های پردازش کاوش اختصاصی

هستند که با محاسباتی همراه است. این موضوع کمک می‌کند تا SQL ، عملکرد بسیار بالایی در تجزیه و تحلیل کاربر داشته باشد. با این حال، نکته ای که باید به آن توجه کرد، این است که هر چه داده‌ها و منابع محاسباتی بیشتری وجود داشته باشد، هزینه انبار داده ممکن است گران‌تر شود.

(4) لایه داده

لایه داده، لایه دسترسی است که به کاربران امکان دسترسی به داده‌ها را می‌دهد. از این نوع داده معمولاً در جاهایی که دیتا مارت وجود دارد، استفاده می‌شود. این لایه، بخش‌هایی از داده‌ها را بسته به افرادی که کاربر قصد دارد به آنها دسترسی داشته باشد، تقسیم می‌کند، بنابراین کل سازمان بسیار منظم و دقیق می‌تواند به کار خود ادامه دهد. به عنوان مثال، ممکن است شما در کسوت مدیر اجازه ندهید که تیم فروش شما اجازه دسترسی به اطلاعات مربوط به تیم منابع انسانی را داشته باشد یا بالعکس.

(5) امنیت

این بخش به لایه داده مربوط می‌شود؛ زیرا باید بتوان خط مشی‌های امنیتی و دسترسی‌های دقیق را در تمامی داده‌های سازمانی ارائه داد. معمولاً انبارهای داده دارای قابلیت مدیریت داده و امنیت بسیار خوبی هستند، بنابراین برای گنجاندن آن نیازی به انجام وظایف مربوط به مهندسی داده سفارشی نخواهد بود. با اضافه کردن داده‌های بیشتری به انبار و رشد سازمان، بسیار مهم است که برای امنیت سازمان برنامه ریزی‌های لازم را انجام داد.

نحوه کارکرد انبار داده

یک انبار داده ممکن است حاوی چندین پایگاه داده باشد. در هر پایگاه داده، داده‌ها در جدول‌ها و ستون‌هایی سازماندهی می‌شوند. در هر ستون، کاربر می‌تواند توضیحاتی را در مورد داده مانند عدد صحیح، فیلد داده یا رشته تعریف کند. جدول‌ها را می‌توان داخل طرحواره‌هایی سازماندهی کرد و آنها را به عنوان پوشه در نظر گرفت. هنگامی که داده به دست می‌آید در جدول‌های مختلفی که توسط طرح توصیف شده اند، ذخیره می‌شود. ابزارهای پرس و جو نیز از طرحواره‌ها استفاده می‌کنند تا نحوه دسترسی و نوع تجزیه و تحلیل هر یک از جدول‌ها را تعیین کنند.

کاربرد انبار داده چیست؟



از انبار داده می‌توان زمانی که با حجم عظیمی از داده‌ها رو به رو هستیم و می‌خواهیم نتایج آماری حاصل از آنها را ببینیم تا تصمیم‌گیری بهتر و آگاهانه‌تری داشته باشیم، استفاده کرد؛ همانند:

- **وبسایت رسانه‌های اجتماعی:** وبسایت‌های شبکه‌های اجتماعی مانند فیس بوک، توییتر و لینکدین بر اساس تجزیه و تحلیل داده‌های زیادی شکل گرفته‌اند. این سایت‌ها، داده‌هایی که مربوط به اعضا،

گروه‌ها و مکان‌ها هستند را جمع آوری می‌کنند و در یک مخزن مرکزی واحد ذخیره سازی می‌کنند. به دلیل حجم بالای داده به انبار داده برای پیاده سازی داده‌ها نیاز است.

- **بانکداری:** امروزه بسیاری از بانک‌ها از انبارها برای مشاهده الگوی مخارج حساب‌ها و دارندگان کارت استفاده می‌کنند. آنها از انبار داده به منظور ارائه پیشنهادات ویژه، معاملات و مواردی از این قبیل استفاده می‌کنند.

- **دولت:** دولت‌ها از انبار داده به منظور ذخیره و تجزیه و تحلیل پرداخت‌های مالی استفاده می‌کنند که برای کشف سرقت‌های مالی بسیار مناسب و کارآمد هستند.

طراحی انبار داده



هنگامی که سازمانی قصد طراحی یک انبار داده را دارد، باید پس از تعریف الزامات تجاری خاص خود، توافق بر سر محدوده و تهیه پیش نویس طرح مفهومی کار خود را آغاز کند. سپس، سازمان می‌تواند یک طراحی منطقی و فیزیکی را برای انبار داده ایجاد کند.

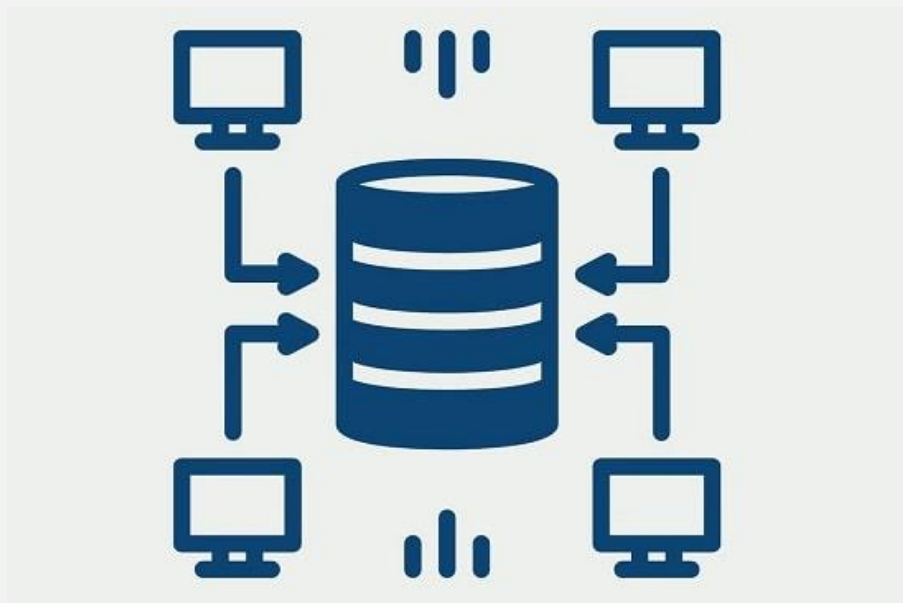
طراحی منطقی شامل روابط بین اشیا را نشان می‌دهد و طراحی فیزیکی شامل بهترین راه به منظور ذخیره سازی و بازیابی اشیا است. طراحی فیزیکی همچنین فرآیندهای حمل و نقل، پشتیبان‌گیری و بازیابی را در بر خواهد گرفت.

چه مواردی در طراحی انبار داده باید رعایت شوند؟

برای طراحی انبار داده باید موارد زیر را رعایت کرد:

- محتوای داده‌های خاص
 - روابط درون و بین گروه‌های داده
 - محیط سیستم پشتیبانی کننده از انبار داده
 - انواع تبدیل‌های داده‌های مورد نیاز
 - فرکانس تازه سازی داده‌ها
- یکی از عوامل اصلی در طراحی انبار داده، نیاز کاربر نهایی است. بیشتر کاربران نهایی به انجام تجزیه و تحلیل و ارزیابی داده‌ها به صورت انبوه علاقه مند هستند تا بررسی تراکنش‌های فردی.

انواع انبار داده



انواع انبار داده به صورت زیر هستند:

انبار داده ابری

انبار داده ابری، نوعی از انبارهای داده است که به صورت مخصوص برای اجرا در فضای ابری ساخته شده است و به عنوان یک سرویس مدیریت شده به مشتریان ارائه می‌شود. انبار داده مبتنی بر ابر، در سال‌های اخیر محبوبیت زیادی پیدا کرده است؛ زیرا سازمان‌های بیشتری از خدمات رایانش ابری استفاده می‌کنند و به دنبال کاهش مرکز داده داخلی خود هستند. با وجود یک انبار داده ابری، زیرساخت انبار داده فیزیکی توسط یک سازمان ابری مدیریت می‌شود، به این معنا که مشتری نیاز به سرمایه‌گذاری اولیه در بخش سخت‌افزاری یا نرم‌افزاری دارد و نیازی به مدیریت یا نگهداری راه حل انبار داده وجود ندارد. از جمله مزایای انبار داده ابری می‌توان به موارد زیر اشاره کرد:

• پشتیبانی الاستیک

- مقیاس پذیری برای محاسبات بزرگ، متغیرها یا نیازهای ذخیره سازی
- راحتی در استفاده
- سهولت در مدیریت
- صرفه جویی در هزینه

انبار داده مدرن

کاربران مختلف انبار داده در سراسر سازمان، نیازهای متفاوتی در ارتباط با آن دارند؛ آنها ممکن است بخشی از موارد زیر باشند:

- فناوری اطلاعات
- مهندسی داده
- تجزیه و تحلیل تجاری
- تیم علم داده

یک معماری داده مدرن با ارائه راهی به منظور مدیریت انواع داده‌ها، حجم کاری و تجزیه و تحلیل آن، نیازهای مختلفی را رفع می‌کند. این موضوع، شامل الگوهای معماری با اجزای ضروری است که به منظور کار با یکدیگر در راستای بهترین شیوه‌های مورد استفاده در صنعت، یکپارچه شده‌اند.

انبار داده مدرن شامل موارد زیر است:

- یک پایگاه داده همگرا (که مدیریت انواع داده را آسان تر می‌کند و راه‌های مختلفی به منظور استفاده از داده ارائه می‌دهد).
- خدمات جذب و تبدیل داده‌های سلف سرویس
- یادگیری ماشین، نمودار و پردازش فضایی
- پشتیبانی از SQL

- آسان کردن چندین گزینه تحلیلی به منظور استفاده از داده‌ها بدون نیاز به جابه‌جایی آنها
- مدیریت خودکار به منظور تهیه، مقیاس بندی و مدیریت ساده

انبار داده سازمانی

این نوع انبار داده، به منظور پرس و جوی سریع و تحقق اهداف تحلیلی بهینه سازی شده است. در این مدل، داده‌ها را از منابع مجزا مانند تنظیمات تراکنش، سیستم‌های عملیاتی و سایر پایگاه‌های داده بازیابی می‌کنند. انبار داده سازمانی بر اساس یک مدل داده چند بعدی است و به صورت ایده آل برای نیازهای یک سازمان، سفارشی شده است. مزایای ذخیره سازی انبار داده سازمانی عبارت است از پشتیبانی تصمیم گیری و ارائه یک رویکرد واحد برای تجسم داده‌ها است.

انبار داده عملیاتی

همان گونه که از نام آن مشخص است، این نوع انبار داده برای نیازهای عملیاتی یک سازمان استفاده می‌شود. این مدل، داده‌ها را معمولاً در زمان واقعی جمع آوری و تجزیه و تحلیل می‌کند. هنگامی که سیستم‌های انبار داده موجود فاقد نتایج گزارش دهی باشند، مدل انبار داده عملیاتی وارد خواهد شد. انبار داده عملیاتی به فعالیت‌های روزانه یک کسب و کار کمک می‌کند؛ همانند سوابق کارکنان، مدیریت ارتباط با مشتری و مواردی از این قبیل.

مزایا، معایب و چالش‌های استفاده از انبار داده چیست؟



مزایا، معایب و چالش‌های انبار داده به صورت زیر هستند.

مزایای انبار داده

مزایای انبار داده عبارت اند از:

- تصمیم‌گیری هوشمند
- هوش تجاری
- تجزیه و تحلیل تاریخی
- کیفیت داده‌ها
- مقیاس‌پذیری
- کاوش‌های مؤثر
- امنیت داده

معایب انبار داده

معایب استفاده از انبار داده به صورت زیر هستند:

- هزینه بالا
- پیچیدگی‌های زیاد
- نیاز به نیروی متخصص
- زمان‌بر بودن
- چالش‌های یکپارچه سازی داده‌ها
- خطرانی برای امنیت داده‌ها

چالش‌های انبار داده

چالش‌های انبار داده عبارت اند از:

- زمان و نحوه جمع آوری داده‌ها
- نوع طراحی مورد استفاده
- تبدیل و پاکسازی داده‌ها
- نحوه انتشار به روزرسانی‌ها
- انتخاب داده‌ها برای خلاصه کردن