**لایه های معماری انباره داده**

**لایه پایینی:**

سرور معماری انبار داده، شامل سرور پایگاه داده رابطه‌ای است که از ابزارهای  Back-End و دیگر ابزارهای کاربردی برای انتقال اطلاعات از منابع مختلف داده‌ای مانند پایگاه داده‌های تراکنشی و غیره، به لایه پایینی استفاده می‌شود. این ابزارهای کاربردی و ابزارهای  Back-End عملکردهای Extract، Clean ،Load و Refresh را انجام می‌دهند.

**لایه میانی:**

لایه‌ی میانی یک سرور OLAP را در اختیار می‌گیرد که به وسیله‌ی آن داده‌ها را به یک ساختار مناسب‌تر تبدیل می‌کند تا بتوان به کوئری‌های پیچیده بر روی داده‌ها و تحلیل آن‌ها دسترسی داشت. این سرور به دو روش می‌تواند کار کند:

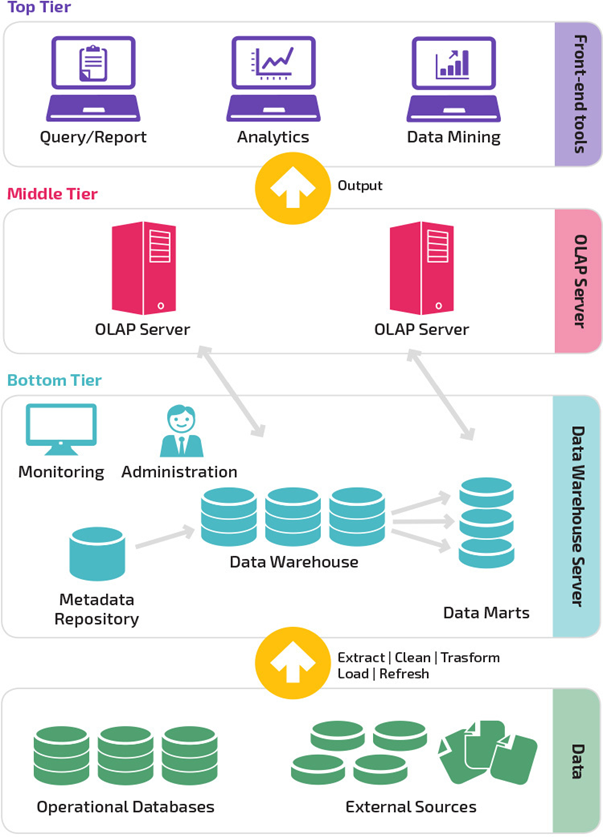
الف) Relational OLAP (ROLAP): یک سیستم مدیریت پایگاه داده رابطه‌ای گسترده است. ROLAP عملیات بر روی داده‌های چند بعدی را به عملیات‌های رابطه‌ای استاندارد تبدیل می‌کند.

ب) Multidimensional OLAP (MOLAP): که به طور مستقیم داده‌های چند بعدی و عملیات را اجرا می کند.

**لایه بالایی:**

لایه بالایی، لایه client یا front-end این لایه، ابزارهایی را برای استفاده در زمینه‌های تجزیه و تحلیل داده، پرس وجو (کوئری) گزارش‌گیری و داده کاوی فراهم می‌آورد.

**نمودار زیر نشان دهنده معماری سه لایه‌ای انباره داده می‌باشد.**



**مدل‌های انباره داده**

**از منظر معماری‌های انباره داده، ما سه مدل انباره داده داریم:**

1. انباره داده مجازی
2. دیتا مارت
3. انباره داده سازمانی

**انباره داده مجازی**

تعداد کمی از انباره داده‌های عملیاتی به عنوان انباره داده مجازی شناخته می شوند. ساخت آن نیازمند به یک ظرفیت مازاد بر روی سرور های پایگاه داده های عملیاتی است.

**دیتا مارت**

دیتا مارت، شامل یک زیر مجموعه از داده‌های سازمان می‌باشد. این زیر مجموعه از داده‌ها برای گروه‌های خاصی از سازمان (یک یا چند واحد سازمانی) ارزشمند است. به عبارت دیگر، ما می‌توان ادعا کرد که دیتا مارت حاوی داده‌های خاصی برای یک گروه یا واحد خاص از سازمان است. به عنوان مثال، دیتا مارت بازاریابی ممکن است حاوی اطلاعات مربوط به اقلام، مشتریان و فروش باشد. و نهایتا این که دیتا مارت‌ها به موضوعات محدود می‌شوند.

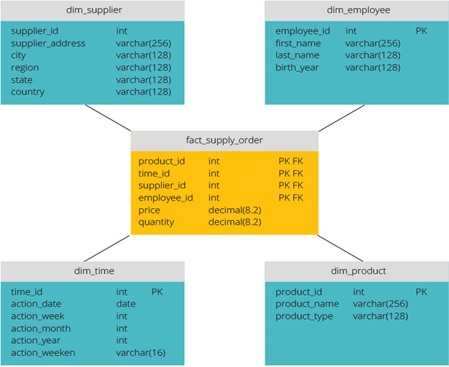
**انباره داده‌های سازمانی**

یک انباره سازمانی، همه‌ی اطلاعات و موضوعاتی که در کل سازمان وجود دارد را جمع‌آوری می‌کند. این نوع انباره داده‌، داده‌های سراسر سازمان را به صورت یکپارچه در اختیار ما قرار می‌دهد.داده‌ها از سیستم‌های عملیاتی داخلی و تامین کنندگان اطلاعات خارجی، فراهم می‌آید و نهایتا این که در این نوع انباره داده‌ها، حجم و گوناگونی اطلاعات می‌ تواند از چند گیگابایت تا چند صد گیگابایت یا چند ترابایت و حتی فراتر باشد.

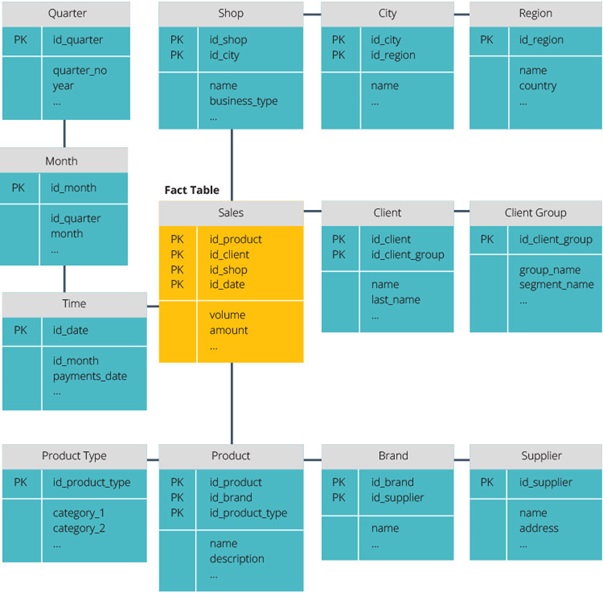
**شِماهای(Schema) ستاره‌ای و گلوله برفی**

شمای ستاره‌ای و شمای گلوله برفی دو روش برای ساختار دادن به انباره داده می‌باشد.

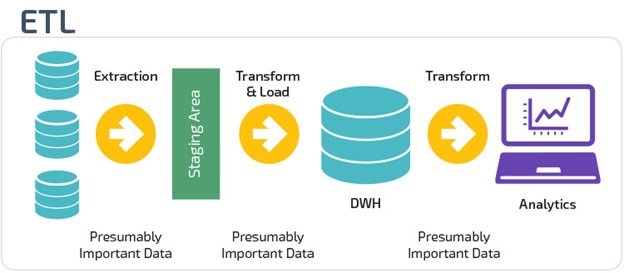
* شمای ستاره‌ای یک مخزن داده‌ای متمرکز دارد و ذخیره‌سازی آن در جدول‌های fact table انجام می‌شود. این شِما، جدول fact table را به یک سری از جداول dimension tables که نرمال‌سازی نشده‌اند، تقسیم می‌کند. fact table شامل داده های تجمعی است و برای اهداف گزارش‌دهی مورد استفاده قرار می‌گیرند، در حالی که dimension tables،داده‌های ذخیره شده را توصیف می‌کند.
* طراحی نرمال‌سازی نشده، دارای پیچیدگی‌های کمتری است، به این دلیل که داده‌ها به صورت گروه‌بندی شده هستند. جدول fact table تنها از یک لینک برای پیوستن (Join) به هر جدول dimension table استفاده می‌کند. طراحی انباره داده با شِمای ستاره‌ای، نوشتن کوئری‌های پیچیده را آسان‌تر می‌کند. نمایی از این شِما در شکل زیر قابل مشاهده است.



* شِمای گلوله برفی متفاوت است به این دلیل که داده‌ها به صورت [نرمال‌سازی](https://keysun-co.com/%d8%af%d8%a7%d8%af%d9%87%e2%80%8c%d9%87%d8%a7%db%8c-%d8%aa%d9%85%db%8c%d8%b2-%da%a9%d8%ab%db%8c%d9%81/) شده هستند. نرمال‌سازی به معنای سازماندهی موثر داده‌ها است تا همه وابستگی‌های داده به خوبی تعریف شوند و هر جدول شامل حداقل انحرافات باشد. در این نوع ساختار جداول dimension tables به صورت واحد هستند بنابراین شاخه‌ها درون dimension tables جداگانه قرار می‌گیرد. شِمای گلوله برفی از فضای دیسک کمتری استفاده می‌کند و موجب می شود حفظ یکپارچگی داده بهتر صورت بگیرد.عیب اصلی این روش پیچیدگی کوئری‌ها برای دستیابی به داده‌هاست.(برای به دست آوردن داده‌ها، معمولا در کوئری‌ها باید از multiple joins استفاده کرد) نمایی از این شِما در شکل زیر قابل مشاهده است.



**ETL چیست؟**



استخراج، تبدیل، بارگزاری (ETL) ، یک فرایند خودکار است که داده‌های خام را دریافت می‌کند، سپس از آن‌ها اطلاعات مورد نیاز برای تجزیه و تحلیل را استخراج می‌نماید، سپس آن‌ها را به یک فرمت تبدیل می‌کند که می‌تواند به نیاز‌های کسب و کار را تامین کرده و نهایتا آن را به در یک انباره‌داده بارگیری کند. ETL عموما داده‌ها را خلاصه می‌کند تا اندازه آن‌ها را کاهش دهد و همچنین عملکرد را برای نوع خاصی از تجزیه و تحلیل‌ها بهبود می‌بخشد.

**چرا ETL مهم است؟**

1. هنگامی که با انبار داده‌های سازمانی (اطلاعات در حالت استراحت) استفاده می‌شود ETL زمینه‌های تاریخی عمیقی برای کسب و کار فراهم می‌آورد.
2. با ارائه یک نمایه تلفیقی، ETL باعث می‌شود که تجزیه و تحلیل و گزارش داده‌ها و ابتکارات در ارتباط با آن‌ها، برای کاربران کسب و کار آسان‌تر شود.
3. ETL می‌تواند بهره‌وری حرفه‌ای داده‌ها را بهبود ببخشد؛ زیرا این پروتکل‌ها رمزهای پردازش شده را مجددا مورد استفاده قرار می‌دهد که این امر باعث می‌شود داده‌ها را بدون نیاز به مهارت‌های فنی برای نوشتن کد یا اسکریپت انتقال داد.
4. ETL در طول زمان تکامل یافته است تا از نیازهایی که در ارتباط با یکپارچگی برای مواردی مانند جریان داده‌ای بروز می‌کنند، پشتیبانی کند.
5. سازمان‌ها به ETL نیاز دارند تا داده‌ها را با هم تجمیع کنند، از دقت داده‌ها اطمینان حاصل کنند و حسابرسی مورد نیاز برای انباره داده‌ها، گزارش‌دهی و تجزیه و تحلیل را فراهم آورند.

**ETL چگونه کار می کند؟**

ETL درارتباط نزدیکی با برخی مفاهیم دیگری از قبیل توابع، فرآیندها و تکنیک‌های یکپارچه‌سازی اطلاعات است. درک این مفاهیم، دیدگاه واضحتری از نحوه عملکرد ETL فراهم می‌کند.

**ساخت ETL با پردازش دسته‌ای**

از طریق پردازش زیر شما می‌توانید یک ETL ایجاد نماید.در این مدل از ETL پردازش و انتقال به صورت دسته‌ای از منابع پایگاه داده به منابع انباره داده صورت می‌گیرد.برای ساخت ETL با روش پردازش دسته‌ای به صورت زیر عمل می‌کنیم:

1. **داده‌های مرجع**

مجموعه‌ای از داده‌ها را ایجاد کنید که مجموعه‌ای از مقادیر مجاز را تعریف می‌کند و ممکن است داده‌های شما را نیز در بر بگیرد.

1. **استخراج از منابع داده‌ای**

پایه‌ی موفقیت در مراحل بعدی ETL، استخراج داده‌ها به صورت درست است. بیشتر سیستم‌های ETL ترکیبی از داده‌ها از منابع داده‌ای مختلف هستند، که هر کدام از منابع، سازماندهی داده‌ای و فرمت خود را دارند.(که شامل پایگاه دادهای رابطه‌ای، پایگاه داده‌های غیر رابطه‌ای،XML ، JSON ، CSV، فایل‌ها و .. می‌باشند) استخراج موفق ، داده‌ها را به یک فرمت واحد تبدیل می‌کند تا پردازش استاندارد باشد.

1. **اعتبارسنجی داده‌ها**

اعتبار سنجی داده‌ها یک فرایند خودکار است که تایید می‌کند که آیا داده‌های بدست آمده و استخراج شده  شده از منابع داده‌ای مختلف دارای مقادیر مورد انتظار هستت یا نه. – برای مثال، در پایگاه داده معاملات مالی از سال گذشته، فیلد تاریخ باید حاوی تاریخ معتبر در طی ۱۲ ماه گذشته باشد. اگر داده‌ها قوانین اعتبارسنجی را رعایت نکنند، موتور اعتبار سنجی داده ها را نمی‌پذیرد.

1. **تبدیل داده‌ها**

حذف اطلاعات اضافی و غیر اصلی یا حذف اطلاعات نادرست (تمیز کردن)، اعمال قوانین کسب و کار، بررسی یکپارچگی داده‌ها (اطمینان حاصل شود که داده‌ها در منبع داده‌ای خراب نبوده یا توسط ETL خراب نشده‌اند و اطلاعات در مراحل قبلی حذف نشده‌اند)

1. **مرحله‌ی نمایش داده‌ها در پایگاه داده**

شما به طور معمول داده‌های تبدیل شده را مستقیما در انبار داده‌های هدف قرار نمی‌دهید. ابتدا داده‌ها باید به یک پایگاه اطلاعاتی متصل شوند و در معرض نمایش قرار بگیرد تا اگر چیزی اشتباه باشد، roll back راحتر باشد.

1. **انتشار داده‌ها در انبار داده**

داده‌ها را در جداول هدف بارگذاری کنید. برخی از انباره داده‌ها هر زمان که ETL یک دسته جدید را بارگذاری می‌کند، اطلاعات موجود را بازنویسی می‌کند.این ممکن است روزانه، هفتگی یا ماهانه رخ دهد. در موارد دیگر، ETL می‌تواند داده‌های جدید را بدون تغییر مجدد اضافه کند، که این کار با نشانه‌گذاری بر روی آن‌ها صورت می‌گیرد. شما باید این کار را با دقت انجام دهید تا از انفجار انباره داده‌ها به علت کمبود فضای دیسک و محدودیت‌های عملکرد جلوگیری شود.

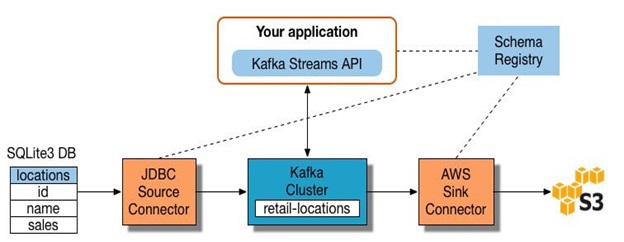
* **ساخت ETL با پردازش جریان داده‌ای**

اغلب پردازش داده‌ای مدرن، شامل داده‌های بلادرنگ نیز می‌شوند. به عنوان مثال تحلیل داده‌ی وب از یک وبسایت تجارت الکترونیک بزرگ.در این شرایط نمی‌توان تبدیل و استخراج داده‌ها را در یک دسته بزرگ، انجام داد.پس این نیاز بوجود می‌آید که ETL در جریان داده‌ها صورت بگیرد. این امر به این معنی است که Client applicationها که  داده‌ها را به منابع داده‌ای ارسال می‌کنند، باید داده‌ها را بلافاصله به پایگاه داده‌های هدف منتقل و ذخیره کنند.

امروزه ابزارهای پردازش جریان داده‌ای زیادی در دسترس هستند که برخی از آن‌ها در زیر آمده است:

Apache Samza, Apache Storm, and Apache Kafka .

**شکل زیر نشان دهنده ی ETL بر مبنای Kafka است.**



در علم محاسبات یا رایانش، یک انبار داده (DW یا DWH) که به نام انبار داده سازمانی (EDW) نیز نامیده می‌شود، سیستمی است که برای گزارش‌دهی و تحلیل داده استفاده می‌شود و به عنوان جزء اصلی هوش تجاری محسوب می‌شود. انبارهای داده، مخازن مرکزی داده‌های یکپارچه از یک یا چند منبع پراکنده هستند. انبارهای داده، داده‌های کنونی و سابق را در یک محل گرد هم می‌آورند که متعاقباً برای ایجاد گزارش‌های تحلیلی برای کارمندان بخش‌های مختلف سازمان مورد استفاده قرار می‌گیرد.

داده‌های ذخیره شده در انبار داده از سیستم‌های عملیاتی مختلف مانند بازاریابی یا فروش بارگذاری می‌شوند. این داده‌ها ممکن است از یک data store عملیاتی بگذرند و ممکن است نیازمند پاک‌سازی داده‌ای برای عملیات‌های اضافی باشند. بدین ترتیب مطمئن شویم که پیش از استفاده در انبار داده‌ای یا گزارش‌گیری، کیفیت مورد نظر را دارند.

data warehouse

انبارهای داده معمولی مبتنی بر استخراج (Extract)، تبدیل (Transform) و بارگذاری (Load) هستند که به اختصار (ETL) نامیده می‌شوند و از سه لایه staging، یکپارچه‌سازی داده و دسترسی برای ایفای کارکردهای اصلی خود تشکیل می‌یابند. لایه staging یا در واقع پایگاه‌های داده staging، داده‌های جدید استخراج شده از هر سیستم داده منبع جداگانه را ذخیره می‌کند. لایه یکپارچه‌سازی مجموعه داده‌های پراکنده را با تبدیل داده‌ها از لایه staging مسنجم می‌سازد و غالباً داده‌های تبدیل یافته را در یک پایگاه داده data store عملیاتی (ODS) ذخیره می‌کند. سپس داده‌های یکپارچه‌سازی شده به یک پایگاه داده دیگر انتقال می‌یابند و در آن جا داده‌ها به گروه‌های سلسله مراتبی که غالباً بُعد، واقعیت‌ و واقعیت‌های تجمیعی (aggregate facts) نام دارند، تقسیم می‌شوند. ترکیب واقعیت‌ها و ابعاد گاهی اوقات شِمای ستاره‌ای نامیده می‌شود. لایه دسترسی به کاربران به بازیابی داده‌ها کمک می‌کند.

منبع اصلی داده، پاک‌سازی، تبدیل و کاتالوگ بندی شده و از سوی مدیران و دیگر متخصصین حرفه‌ای جهت داده‌کاوی، پردازش تحلیلی آنلاین، تحقیقات بازار و پشتیبانی تصمیم مورد استفاده قرار می‌گیرد. با این حال ابزار بازیابی و تحلیل، استخراج، تبدیل و بارگذاری داده‌ها و همچنین مدیریت دیکشنری داده‌ها نیز جزو اجزای اصلی سیستم انبار داده تلقی می‌شود. بسیاری از منابع برای سیستم‌های انبار داده از این معنی گسترده‌تر استفاده می‌کنند. از این رو یک تعریف جامع‌تر از انبار داده به این صورت است: ابزارهای هوش تجاری، استخراج، تبدیل و بارگذاری داده‌ها در مخازن و ابزارهایی برای مدیریت و بازیابی فراداده.

## مزیت‌ها

انبار داده شامل یک کپی از اطلاعات است که از سیستم‌های تبادل داده‌ای به دست می‌آید. این معماری فرصتی برای موارد زیر به دست می‌دهد:

* یکپارچه‌سازی داده‌ها از چند منبع در یک پایگاه داده و مدل داده‌ای واحد/ تجمیع بیشتر داده‌ها در یک پایگاه داده منفرد به طوری که با یک موتور کوئری بتوان داده‌ها را در یک ODS عرضه کرد.
* رفع مشکل رقابت بر سر قفل در سطح ایزولاسیون پایگاه داده در سیستم‌های پردازش تراکنش که ناشی از کوئری‌های تحلیلی بزرگی است که به مدتی طولانی در پایگاه‌های داده پردازش تراکنش اجرا می‌شوند.
* نگهداری سابقه داده حتی در مواردی که سیستم‌های تراکتش منبع این کار را انجام نمی‌دهند.
* یکپارچه‌سازی داده‌ها از سیستم‌های منبع چندگانه و ایجاد امکان مشاهده مرکزی در کل سطح سازمان. این مزیت همواره ارزشمند است؛ اما زمانی که سازمان از طریق ادغام با شرکت یا سازمانی دیگر بزرگ می‌شود، اهمیتی دوچندان می‌یابد.
* بهبود کیفیت داده از طریق ارائه کدها و توصیف‌های مداوم و رفع مشکل داده‌های نامناسب
* ارائه مداوم اطلاعات سازمان
* ارائه یک مدل داده ستون مفرد برای همه داده‌های مورد نظر صرف‌نظر از منبع داده
* بازسازی داده‌ها به طرزی که برای کاربران تجاری معنی‌دار باشد.
* بازسازی داده‌ها به طوری که عملکرد کوئری حتی برای کوئری‌های تحلیلی بدون تأثیر بر سیستم‌های عملیاتی بهبود یابد.
* افزودن ارزش به برنامه‌های تجاری عملیاتی و به طور قابل توجه به سیستم‌های مدیریت رابطه مشتری (CRM)
* آسان‌تر ساختن نوشتن کوئری‌های پشتیبان تصمیم
* سازمان‌دهی و ابهام‌زدایی از بازاریابی داده‌ها.



## محیط معمول

محیط مناسب برای انبار داده یا داده‌گاه (Data mart) شرایط زیر را دارد:

* سیستم‌های منبع که داده‌ها را در اختیار انبار داده یا داده‌گاه قرار می‌دهند،
* فناوری یکپارچه‌سازی داده‌ها و پردازش‌هایی که برای قابل استفاده ساختن داده‌ها لازم هستند،
* معماری‌های مختلف برای ذخیره‌سازی داده‌های یک سازمان در یک انبار داده یا داده‌گاه،
* ابزارها و برنامه‌های مختلف برای کاربران متفاوت،
* پردازش‌های فراداده، کیفیت داده و حاکمیتی باید اعمال شوند تا مطمئن شویم که انبار داده‌ها یا داده‌گاه‌ها می‌توانند اهداف خود را برآورده سازند.

«آر. کلی رینر» (R. Kelly Rainer) در خصوص سیستم‌های منبع که در بخش فوق مورد اشاره قرار گرفت، بیان می‌کند: «یک منبع معمول برای داده‌ها در انبار داده، پایگاه‌های داده عملیاتی شرکت هستند که می‌توانند پایگاه داده رابطه‌ای باشند».

رینر در خصوص یکپارچه‌سازی داده بیان می‌کند: «لازم است که داده‌ها از سیستم‌های منبع استخراج شوند و آن‌ها را در یک داده‌گاه یا انبار داده بارگذاری کرد».

رینر ذخیره‌سازی داده‌ها در انبار داده سازمان یا داده‌گاه‌ها را نیز بررسی کرده است. وی فراداده و داده‌هایی در مورد داده‌ها را این گونه توضیح می‌دهد: «کارکنان فناوری اطلاعات باید اطلاعاتی در مورد منابع داده، پایگاه داده، جدول و نام‌های ستون، زمان‌بندی‌های رفرش و معیارهای کاربرد داده داشته باشند».

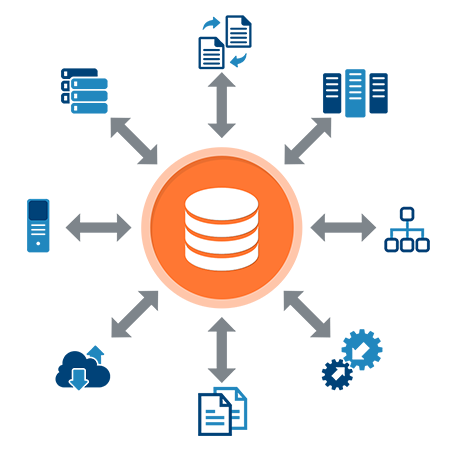
امروزه موفق‌ترین شرکت‌ها آن‌هایی هستند که بتوانند به سرعت و با انعطاف بالا به تغییرات و فرصت‌ها پاسخ دهند. کلید پاسخ به این مسئله کاربرد مؤثر و کارآمد داده‌ها و اطلاعات از سوی تحلیل‌گران و مدیران است. یک انبار داده مخزنی از داده‌های تاریخی است که بر اساس موضوع سازمان‌ یافته‌اند تا به تصمیم‌گیران سازمان کمک کند. زمانی که داده‌ها در یک داده‌گاه یا انبار داده ذخیره می‌شوند، می‌توان آن‌ها را مورد ارزیابی قرار داد.

## سیستم‌های مرتبط (داده‌گاه، OLAP، OLTP، تحلیل پیش‌بینی)

داده‌گاه شکل ساده‌ای از یک انبار داده است که روی یک موضوع منفرد متمرکز شده است. از این رو داده‌گاه، داده‌ها را از تعداد محدودی از منابع مانند بخش فروش، مالی و بازاریابی دریافت می‌کند. داده‌گاه‌ها غالباً از سوی بخش‌های منفرد سازمان ساخته شده و مورد کنترل قرار می‌گیرند. این منابع می‌توانند سیستم‌های عملیاتی داخلی، انبار داده مرکزی و یا داده‌های بیرونی باشند. نرمال‌زدایی (Denormalization) در تکنیک‌های مدل‌سازی داده در این سیستم کاملاً طبیعی است. با توجه به این که داده‌گاه‌ها به طور کلی تنها زیرمجموعه‌ای از داده‌های موجود در یک انبار داده را شامل می‌شوند، در اغلب موارد پیاده‌سازی آن‌ها ساده‌تر و سریع‌تر است.

تفاوت‌های بین انبار داده و داده‌گاه

|  |  |  |
| --- | --- | --- |
| خصوصیت | انبار داده | داده‌گاه |
| حوزه داده‌ها | در سطح سازمان | در سطح یک بخش سازمان |
| تعداد حوزه‌های موضوعی | چندگانه | منفرد |
| دشواری ساختن | زیاد | کم |
| زمان مورد نیاز برای ساخت | زیاد | کم |
| مقدار حافظه | زیاد | محدود |



## انواع داده‌گاه‌ها: مستقل، وابسته و هیبرید

پردازش تحلیلی آنلاین (OLAP) از روی میزان تراکنش‌های نسبتاً پایین مشخص می‌شود. کوئری‌ها غالباً بسیار پیچیده هستند و شامل تجمیع هستند. در سیستم‌های OLAP زمان پاسخ‌دهی یک معیار مؤثر است. برنامه‌های OLAP به طور گسترده‌ای از سوی تکنیک‌های داده‌کاوی مورد استفاده قرار می‌گیرند. پایگاه‌های داده OLAP داده‌های تجمیع یافته و تاریخی را در طرح‌های چندبعدی (غالباً شِمای ستاره‌ای) ذخیره می‌کنند. سیستم‌های OLAP به طور معمول تأخیر داده‌ای به مدت چند ساعت دارند که از این جهت در مقابل داده‌گاه‌ها قرار می‌گیرند که تأخیر مورد انتظارشان در حد یک روز است. از رویکرد OLAP برای تحلیل داده‌های چندبعدی از منابع چندگانه استفاده می‌شود. سه عملیات مقدماتی در OLAP به صورت Roll-up (Consolidation)، Drill-down و Slicing & Dicing است.

پردازش تراکنش آنلاین (OLTP) از روی تعداد بالای تراکنش‌های آنلاین (INSERT, UPDATE, DELETE) مشخص می‌شود. سیستم‌های OLTP بر روی پردازش بسیار سریع کوئری تأکید دارند و از حفظ یکپارچگی داده‌ها در محیط‌های چند-دسترسی مشخص می‌شوند. در مورد سیستم‌های OLTP کارآمدی بر اساس تعداد تراکنش‌ها در ثانیه مشخص می‌شود. پایگاه‌های داده OLTP شامل داده‌های دقیق و جاری هستند. طرح کلی مورد استفاده برای پایگاه‌های داده، ذخیره‌سازی تراکنش مدل موجودیت (معمولاً 3NF) است. نرمال‌سازی در تکنیک‌های مدل‌سازی داده در چنین سیستم‌هایی یک هنجار محسوب می‌شود.

روش‌های تحلیلی پیش‌بینی به یافتن و کمّی‌سازی الگوهای پنهان در داده‌ها با استفاده از مدل‌های ریاضیاتی پیچیده مربوط است و از آن می‌توان برای پیش‌بینی خروجی‌های آینده استفاده کرد. روش‌های تحلیلی پیش‌بینی از OLAP متفاوت هستند چون OLAP بر روی تحلیل داده‌های تاریخی متمرکز می‌شود و ماهیتی واکنشی دارد؛ در حالی که تحلیل پیش‌بینی بر روی آینده تمرکز دارد. این سیستم‌ها برای مدیریت رابطه با مشتری (CRM) استفاده می‌شوند.



## تاریخچه

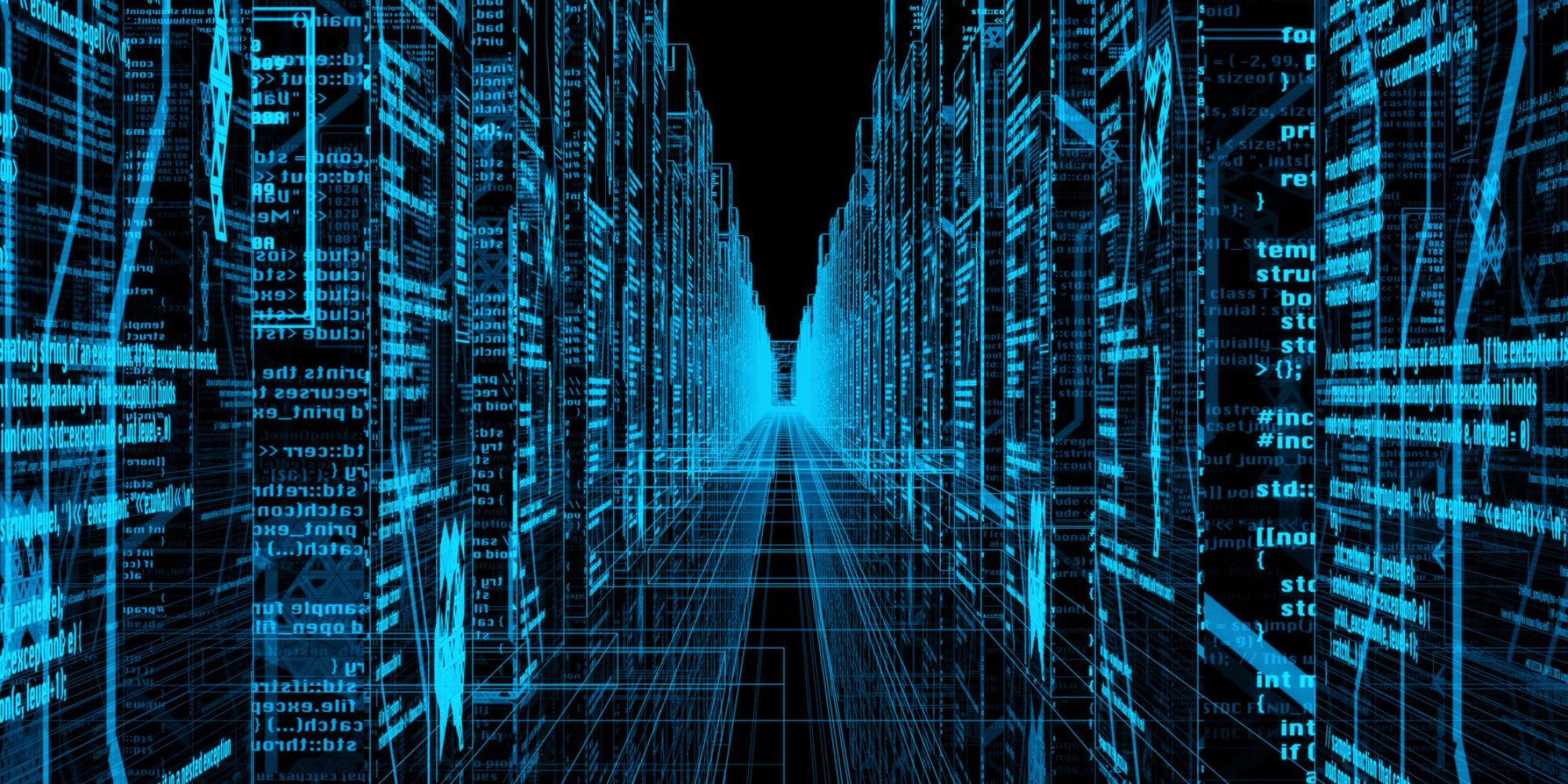
مفهوم انبارسازی داده‌ها به دهه 1980 میلادی باز می‌گردد که محققان IBM به نام بِری دلوین و پائول مورفی «انبار داده تجاری» را توسعه دادند. انبارسازی داده در ابتدا برای ارائه یک مدل معماری برای گردش داده‌ها از سیستم‌های عملیاتی به محیط‌های پشتیبانی تصمیم طرحی شد. این مفهوم تلاش می‌کرد که مسائل مختلف مرتبط با این گردش را رفع کند که غالباً به صورت هزینه بالای محاسباتی ظاهر می‌شد. در غیاب معماری انبار داده حجم بالایی از افزونگی برای استفاده از محیط‌های پشتیبانی تصمیم چندگانه مورد نیاز بود. در شرکت‌های بزرگ‌تر معمولاً محیط‌های پشتیبان تصمیم به طور مستقل از هم عمل می‌کردند.

در هر محیط کاربران مختلفی فعالیت می‌کردند که غالباً نیازمند داده‌های تقریباً مشابهی بودند. فرایند جمع‌آوری، پاک‌سازی و یکپارچه‌سازی داده‌ها از منابع مختلف معمولاً از سیستم‌های بلندمدت عملیاتی موجود، در غالب موارد تا حدودی در هر یک از این محیط‌ها تکرار می‌شد. به علاوه سیستم‌های عملیاتی به طور مکرر مورد بازآزمایی قرار می‌گرفتند، چون الزامات جدیدی برای پشتیبانی از تصمیم ظاهر می‌شدند. این الزامات جدید در قالب موارد نیازمند گردآوری، پاک‌سازی و یکپارچه‌سازی داده‌ها از داده‌گاه‌هایی بودند که برای دسترسی بی‌واسطه کاربران طراحی شده بودند.



توسعه‌های کلیدی انبار داده در سال‌های اولیه به صورت زیر بوده است:

* دهه 1960: شرکت مواد غذایی جنرال میلز و دانشگاه دارتموث در یک پروژه تحقیقاتی مشترک اصطلاح‌های ابعاد و واقعیت‌ها را ارائه کردند.
* دهه 1970: شرکت‌های تحقیقات و بازاریابی نیسلون و IRI، داده‌گاه‌های ابعادی را برای خرده‌فروشی‌ها عرضه کردند.
* دهه 1970: بیل اینمون شروع به تعریف و بررسی اصطلاح انبار داده کرد.
* 1975: شرکت Sperry Univac مفهوم MAPPER (حفظ، آماده‌سازی و تولید گزارش‌های اجرایی) را معرفی کرد که سیستم مدیریت پایگاه داده و گزارشی بود که نخستین 4GL (زبان برنامه‌نویسی نسل چهارم) دنیا را شامل می‌شد. این اولین پلتفرم طراحی شده برای ساخت مراکز اطلاعات بود.
* 1983: شرکت ترادیتا رایانه پایگاه داده DBC/1012 را معرفی کرد که به طور اختصاصی برای پشتیبانی از تصمیم، طراحی شده بود.
* 1984: شرکت سیستم‌های رایانه‌ای متافور از سوی دیوید لیدل و دون ماسارو پایه‌گذاری شده و نخستین بسته سخت‌افرازی/نرم‌افزاری و GUI را برای کاربران تجاری عرضه کرد تا به کمک آن سیستم مدیریت و تحلیل پایگاه داده بسازند.
* 1985: شرکت Sperry مقاله‌ای در خصوص مراکز اطلاعاتی منتشر ساخت که در آن اصطلاح انبار داده MAPPER در چارچوب معنایی مراکز اطلاعاتی معرفی شده بود.
* 1988: بری دالوین (Barry Devlin) و پائول مورفی (Paul Murphy) مقاله «مقدمه‌ای بر سیستم تجاری و اطلاعاتی» را منتشر کرده و در آن اصطلاح «انبار داده تجاری» را معرفی کردند.
* 1990: شرکت سیستم‌های «رد بریک» که از سوی رالف کیمبال بنیان‌گذاری شده بود، انبار داده رد بریک را معرفی کرد که یک سیستم مدیریت پایگاه داده اختصاصی برای انبارسازی داده‌ها محسوب می‌شد.
* 1991: شرکت «پریسم سولوشنز» که از سوی بیل اینمون پایه‌گذاری شده بود، نرم‌افزار مدیریت انبار داده پریسم را معرفی کرد که نرم‌افزاری برای توسعه یک انبار داده بود.
* 1992: بیل اینمون کتاب «ساخت انبار داده» را منتشر ساخت.
* 1995: موسسه انبارسازی داده‌ها که یک سازمان انتفاعی بود با هدف ارتقای انبارسازی داده‌ها تأسیس شد.
* 1996: رالف کیمبال کتاب «کیت ابزار انبار داده» را منتشر کرد.
* 2012: بیل اینمون یک فناوری که به نام «ابهام‌زدایی متنی» (textual disambiguation) نامیده می‌شد را توسعه داده و به صورت عمومی عرضه کرد. این تکنیک بر روی متون خام استفاده شده، آنها را مورد قالب‌بندی مجدد قرار داده و به یک قالب پایه داده‌ای استاندارد تبدیل می‌کرد. زمانی که متن خام از مرحله ابهام‌زدایی متنی رد می‌شود، می‌تواند به سادگی و به طور کارآمدی از سوی فناوری‌های هوش تجاری استاندارد مورد دسترسی و تحلیل واقع شود. ابهام‌زدایی متنی از طریق اجرای ETL متنی اجرا می‌شد. این فناوری در مواردی که متون خامی مانند اسناد، هادوپ، ایمیل و مواردی از این دست وجود داشته باشند، روشی متداولی به حساب می‌آید.



## ذخیره اطلاعات

ذخیره اطلاعات در انبارهای داده شامل موضوعات مختلفی است که در ادامه به تفصیل توضیح داده شده است.

### واقعیت‌ها (Facts)

واقعیت به معنی مقدار یا اندازه‌ای نشان‌دهنده حقیقتی در مورد یک موجودیت یا سیستم مدیریت شده است.

واقعیت‌ها زمانی که از سوی موجودیت گزارش دهنده، گزارش می‌شوند در سطح خام هستند. مثلاً در یک سیستم تلفن همراه، BTS (آنتن) 1000 درخواست را برای تخصیص کانال ترافیک دریافت می‌کند، 820 مورد را قبول کرده و بقیه را رد می‌کند و سه واقعیت یا اندازه را به سیستم مدیریت گزارش می‌دهد:

* tch\_req\_total = 1000
* tch\_req\_success = 820
* tch\_req\_fail = 180

واقعیت‌های در سطح خام در ابعاد مختلف به سطوح بالاتر تجمیع می‌شوند تا اطلاعات مرتبط‌تر با خدمات یا تجارت از آن‌ها استخراج شود. این وضعیت تجمیع یا خلاصه یا واقعیت‌های تجمعی نامیده می‌شوند.

برای نمونه اگر سه آنتن BTS در یک شهر وجود داشته باشند، در این صوت واقعیت‌های فوق را می‌توان از BTS های در سطح شهر در بعد شبکه تجمیع کرد. برای مثال:

* tch\_req\_success\_city = tch\_req\_success\_bts1 + tch\_req\_success\_bts2 + tch\_req\_success\_bts3
* avg\_tch\_req\_success\_city = (tch\_req\_success\_bts1 + tch\_req\_success\_bts2 + tch\_req\_success\_bts3) / 3

## رویکرد ابعادی در برابر رویکرد نرمال‌سازی برای ذخیره‌سازی داده‌ها

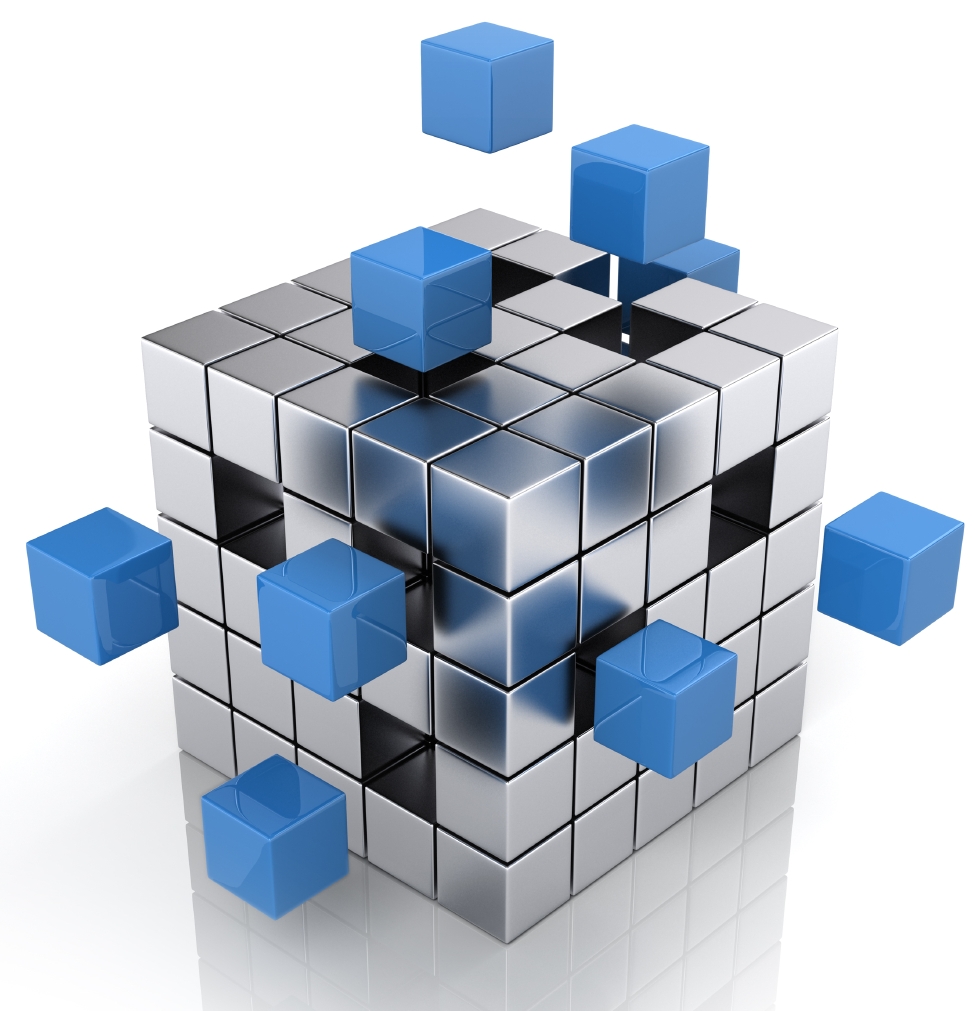
سه یا چند رویکرد عمده برای ذخیره‌سازی داده‌ها در یک انبار داده وجود دارد که مهم‌ترین رویکردها به صورت رویکرد ابعادی و نرمالیزه هستند.

### رویکرد ابعادی

رویکرد ابعادی اشاره به رویکرد رالف کیمبال دارد که در آن بیان شده است که انبار داده باید با استفاده از مدل ابعادی (شِمای ستاره‌ای) مدل‌سازی شود. رویکرد نرمالیزه که مدل 3NF (Third Normal Form) نیز نامیده می‌شود، اشاره به رویکرد بیل اینمون دارد که در آن انبار داده‌ باید با استفاده از مدل E-R/ مدل نرمال‌سازی شده، مدل‌سازی شود.

در یک رویکرد ابعادی، داده‌های تراکنش‌ها به صوت «واقعیت»‌هایی که به طور کلی داده‌های تراکنش عددی هستند و «ابعاد»‌ی که اطلاعات مرجع مشخص‌کننده مفهوم واقعیت‌ها هستند، افراز می‌شوند. برای نمونه یک تراکنش فروش را می‌توان به واقعیت‌هایی مانند تعداد محصولات سفارش داده شده و قیمت کلی پرداخت شده برای محصولات و ابعادی به صورت تاریخ سفارش، نام مشتری، تعداد محصول، آدرس‌های ارسال صورتحساب و کالا و فروشنده مسئول دریافت سفارش، تجزیه کرد.

مزیت کلیدی رویکرد ابعادی آن است که در آن درک و استفاده از انبار داده برای کاربر آسان‌تر است. ضمناً بازیابی داده‌ها از انبار داده بسیار سریع‌تر صورت می‌گیرد. درک ساختارهای ابعادی برای کاربران تجاری آسان‌تر است، چون ساختار به اندازه‌ها/واقعیت‌ها و زمینه/ابعاد تقسیم می‌شود. واقعیت‌ها با فرایندهای تجاری سازمان و سیستم‌های عملیاتی مرتبط هستند؛ در حالی که ابعاد شامل معنای این اندازه‌گیری‌ها هستند. مزیت دیگر مدل ابعادی آن است که نیازی به پایگاه‌های داده رابطه‌ای ندارد. از این رو این نوع از تکنیک مدل‌سازی برای کوئری‌های کاربر نهایی در انبار داده بسیار مناسب است.



مدل واقعیت‌ها و ابعاد می‌تواند به صورت مکعب داده نیز نگریسته شود. در این مدل ابعاد مختصات دسته‌ای در یک مکعب چندبعدی هستند و واقعیت‌ها مقادیر متناظر با آن مختصات به حساب می‌آیند.

معایب اصلی رویکرد ابعادی به صورت زیر هستند:

1. حفظ یکپارچگی واقعیت‌ها و ابعاد، در صورت بارگذاری داده‌ها از سیستم‌های عملیاتی مختلف امری پیچیده خواهد بود.
2. اصلاح ساختار انبار داده در صورتی که سازمان در رویکرد ابعادی خود تغییراتی اعمال کند دشوار خواهد بود. چنین تغییراتی در صورت تغییر روش تجارت سازمان ناگزیر خواهند بود.

### رویکرد نرمالیزه

در رویکرد نرمالیزه داده‌ها در انبار داده به صورتی ذخیره می‌شوند که تا حدودی از قاعده نرمال‌سازی پایگاه داده تبعیت می‌کند. جداول بر اساس موضوع با هم گروه‌بندی می‌شوند که منعکس‌کننده دسته‌های عمومی داده (برای مثال، داده‌های مشتریان، محصولات، مالی و …) هستند. ساختار نرمال شده، داده‌ها را به صورت موجودیت‌هایی تقسیم می‌کند که چند جدول را در یک پایگاه داده رابطه‌ای تشکیل می‌دهند. وقتی از این رویکرد در یک سازمان بزرگ استفاده شود، نتیجه آن پدید آمدن تعداد بالایی از جداول خواهد بود که از طریق شبکه درهم‌تنیده‌ای از لینک‌ها به هم پیوند یافته‌اند.

به علاوه هر یک از این موجودیت‌های ایجاد شده در زمان تشکیل پایگاه داده به جداول فیزیکی جداگانه‌ای تبدیل می‌شود. مزیت اصلی این رویکرد آن است که فرایند اضافه کردن اطلاعات به پایگاه داده سرراست است. از جمله معایب این رویکرد عبارت از این است که به دلیل تعداد جداول ایجاد شده، الحاق داده‌ها از منابع مختلف برای ساخت اطلاعات معنی‌دار دشوار است و دسترسی به اطلاعات بدون داشتن درک دقیقی از منابع داده‌ای و ساختارهای داده‌ای در انبار داده نیز سخت خواهد بود.



هر دو مدل نرمالیزه و ابعادی را می‌توان در نمودارهای موجودیت-رابطه نمایش داد که هر دو شامل جداول رابطه‌ای الحاق یافته باشند. تفاوت بین دو مدل در میزان نرمالیزاسیون است. این رویکردها متقابلاً منحصربه‌فرد نیستند و رویکردهای دیگری نیز وجود دارند. رویکردهای ابعادی می‌توانند شامل حدودی از نرمال‌سازی داده‌ها باشند.

رابرت هیلارد در کتاب «کسب‌وکارهای مبتنی بر اطلاعات» رویکردی را ارائه می‌کند که این دو رویکرد را بر اساس نیازهای اطلاعاتی کسب‌وکار مقایسه می‌کند. این تکنیک نشان می‌دهد که مدل‌های نرمال شده حتی در مواردی که از حوزه‌های یکسانی استفاده می‌شود، اطلاعات بیشتری نسبت به معادل‌های ابعادی خود نگهداری می‌کنند؛ اما این اطلاعات بیشتر به قیمت کاهش کاربردپذیری به دست می‌آید. این تکنیک کمّیت اطلاعات را برحسب آنتروپی اطلاعات و کاربردپذیری بر اساس معیارهای تبدیل داده Small Worlds اندازه‌گیری می‌کند.

## روش‌های طراحی

سه روش عمده برای طراحی انبارهای داده وجود دارند که در ادامه به بررسی آنها پرداخته‌ایم:

### طراحی پایین به بالا

در رویکرد طراحی پایین به بالا، داده‌گاه‌ها به طور مقدماتی برای قابلیت‌های گزارش‌دهی و تحلیلی برای پردازش‌های خاص تجاری ایجاد می‌شوند. سپس این داده‌گاه‌ها را می‌توان یکپارچه ساخت تا یک انبار داده جامع ایجاد کرد. معماری گذرگاه انبار داده به طور ابتدایی یک پیاده‌سازی از «گذرگاه» به معنی مجموعه‌ای از ابعاد معین و واقعیت‌های معین است که در آن ابعاد بین واقعیت‌ها در یک یا چند داده‌گاه مشترک هستند.

### طراحی بالا به پایین

رویکرد طراحی بالا به پایین با استفاده از یک مدل داده سازمانی طراحی می‌شود. داده اتمیک (Atomic data) به معنی داده در بالاترین سطح از جزییاتش است که در انبار داده ذخیره می‌شود. داده‌گاه‌های ابعادی باید شامل داده‌های مورد نیاز برای پردازش‌های خاص تجاری یا بخش‌های مشخص از سازمان باشند و از انبارهای داده‌ای ساخته می‌شوند.

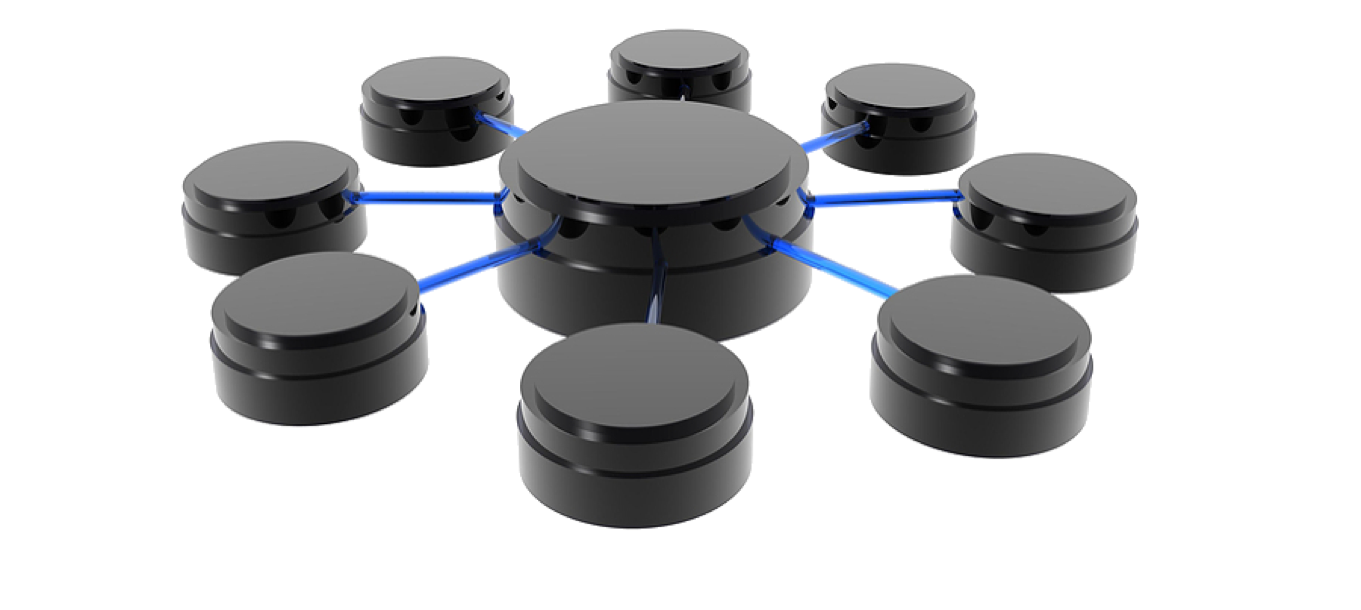


### طراحی ترکیبی

انبار داده غالباً به صورت معماری hub and spokes تشبیه می‌شود. سیستم‌های موروثی (Legacy systems) که انبارهای داده را تغذیه می‌کنند، اغلب شامل مدیریت رابطه با مشتری و برنامه‌ریزی منابع سازمانی هستند که حجم‌های بالاتری از داده را تولید می‌کنند. برای یکپارچه‌سازی این مدل‌های داده مختلف و تسهیل پردازش بار تبدیل استخراج، انبارهای داده در اغلب موارد از انبارهای داده‌ی عملیاتی استفاده می‌کنند، یعنی اطلاعاتی که متعاقباً به صورت انبارهای داده واقعی تجزیه می‌شوند. برای کاستن از افزونگی داده‌ها، سیستم‌های بزرگ‌تر غالباً داده‌ها را به روشی نرمال‌سازی شده ذخیره می‌کنند. سپس می‌توان داده‌گاه‌هایی برای گزارش‌های خاص بر روی این انبارهای داده‌ای ساخت.

یک پایگاه داده ترکیبی به صورت 3NF یعنی شکل نرمال سوم نگهداری می‌شود تا افزونگی داده‌ها حذف شود. با این حال، یک پایگاه داده رابطه‌ای نرمال برای گزارش‌های هوش تجاری که مدل‌سازی ابعادی در آن رایج است، چندان بهینه محسوب نمی‌شود. داده‌گاه‌های کوچک را می‌توان از روی انبارهای داده منسجم ساخت و از داده‌های فیلتر شده مشخص برای جداول واقعیت و ابعاد مورد نیاز استفاده کرد. انبار داده یک منبع منفرد اطلاعاتی ارائه می‌کند که از روی آن می‌توان داده‌گاه‌هایی را ایجاد کرد و طیف متنوعی از اطلاعات تجاری را ارائه می‌کند. معماری ترکیبی امکان جایگزینی مخزن مدیریت داده اصلی را با انبار داده مهیا می‌سازد که در آن می‌توان اطلاعات عملیاتی و نه استاتیک را ذخیره کرد.

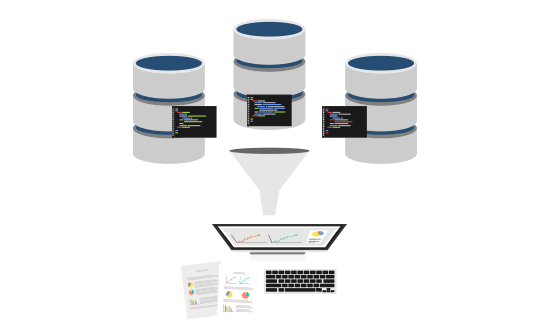
اجزای مدل‌سازی data vault از معماری hub and spokes تبعیت می‌کنند. این سبک مدل‌سازی یک طراحی هیبرید است که شامل بهترین رویه‌های «شکل نرمال سوم» و «شِمای ستاره‌ای» (star schema) است. مدل data vault یک شکل واقعی نرمال سوم نیست و برخی از قواعد آن را نقض می‌کند؛ اما یک معماری بالا به پایین با طراحی پایین به بالا محسوب می‌شود. مدل data vault طوری تنظیم شده است که به طور صریح یک انبار داده باشد. این تنظیم در جهت افزایش دسترس آن برای کاربر نهایی نبوده است و هنگامی که ساخته شود برای مقاصد تجاری همچنان نیازمند نواحی انتشاری بر مبنای شِمای ستاره‌ای یا داده‌گاه خواهد بود.



## خصوصیات انبار داده

برخی ویژگی‌های مقدماتی وجود دارند که داده‌ها را در انبار داده تعریف می‌کنند و شامل موضوع-محوری، یکپارچه‌سازی داده، زمان-متغیر، داده‌های غیر فرّار و دانه‌بندی (Granularity) داده است.

* **موضوع-محور:** داده‌ها در انبار داده بر خلاف سیستم‌های عملیاتی پیرامون موضوعات سازمان (نرمالیزاسیون پایگاه داده) گرد هم می‌آیند. جهت‌گیری موضوعی برای تصمیم‌گیری واقعاً مفید است. گردآوری اشیای مورد نیاز، موضوع-محوری نامیده می‌شود.
* **یکپارچه:** داده‌های موجود درون یک انبار داده یکپارچه هستند. از آنجا که این داده‌ها از چند سیستم عملیاتی می‌آیند همه ناسازگاری‌ها می‌بایست حذف شوند. خصوصیات آن شامل قراردادهای نام‌گذاری، اندازه‌گیری متغیرها، ساختارهای انکودینگ، خصوصیات فیزیکی داده و مواردی از این دست است.
* **زمان-متغیر:** با این که سیستم‌های عملیاتی به دلیل پشتیبانی از عملیات‌های هر روزه مقادیر فعلی را نشان می‌دهند؛ اما انبارهای داده، نشان‌دهنده داده‌هایی در افق زمانی بلندمدت‌تر (تا 10 سال) هستند. این بدان معنی است که انبار داده حاوی داده‌های تاریخی است. انبار داده به طور عمده برای داده‌کاوی و پیش‌بینی استفاده می‌شود، چون اگر کاربری در جستجوی یک الگوی خرید برای یک مشتری خاص باشد، می‌بایست به داده‌هایی در مورد خریدهای فعلی و گذشته نگاه کند.
* **غیر فرّار:** داده‌های موجود در انبار داده تنها خواندنی هستند، یعنی نمی‌توان آن‌ها را به‌روزرسانی، ایجاد یا حذف کرد.
* **جمع‌بندی:** داده‌ها در انبار داده در سطوح مختلفی جمع‌بندی می‌شوند. کاربر انبار داده به واحدهای فروش کلی یک محصول در کل یک منطقه نگاه می‌کند. سپس کاربر به آمار آن منطقه نگاه می‌کند. در نهایت ممکن است فروشگاه‌های منفردی را در منطقه‌ای خاص بررسی کند. بدین ترتیب معمولاً تحلیل در سطوح بالاتر آغازمی شود و برای یافتن جزییات به سطوح پایین‌تر منتقل می‌شود.



## معماری انبار داده

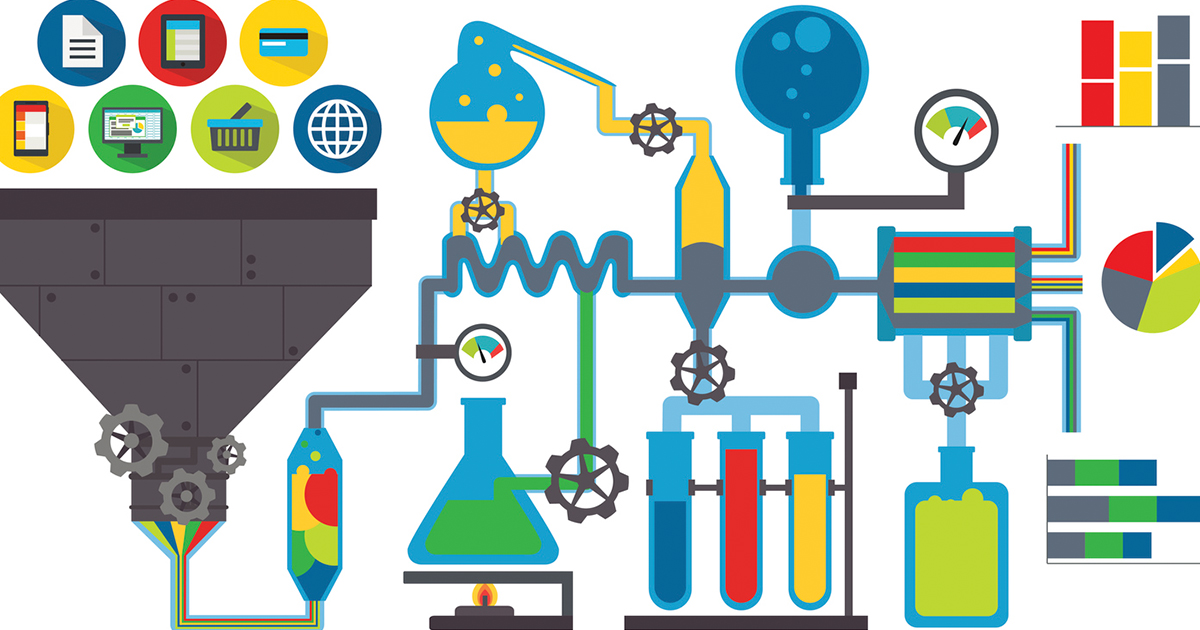
روش‌های مختلف مورد استفاده برای ساخت/سازمان‌دهی یک انبار داده بسیار متنوع هستند. سخت‌افزار مورد استفاده، نرم‌افزار ایجاد شده و منابع داده‌ای که به طور خاص برای کارکرد صحیح انبار داده الزامی هستند، اجزای اصلی معماری انباره داده محسوب می‌شوند. همه انبارهای داده چند مرحله دارند که در طی آن نیازمندی‌های سازمان تعدیل و تنظیم دقیق می‌شوند.

### انبار داده در برابر سیستم عملیاتی

سیستم‌های عملیاتی برای حفظ یکپارچگی داده و سرعت ثبت تراکنش‌های تجاری با استفاده از نرمالیزاسیون پایگاه داده و مدل موجودیت-رابطه بهینه‌سازی شده‌اند. طراحان سیستم‌های عملیاتی به طور کلی از قوانین دوازده‌گانه کاد (Codd’s 12 rules) برای نرمال‌سازی پایگاه داده در جهت تضمین یکپارچگی بهره می‌گیرند. طراحی‌های پایگاه‌های داده کاملاً نرمال‌سازی شده (یعنی آن‌هایی که همه قواعد کاد در آن‌ها رعایت شده‌اند) غالباً موجب می‌شوند که اطلاعات حاصل از تراکنش‌های مالی در ده‌ها تا صدها جدول ذخیره شوند. پایگاه‌های داده رابطه‌ای در مدیریت روابط بین این جدول‌ها مؤثر هستند. این پایگاه‌های داده عملکرد به‌روزرسانی بسیار سریعی دارند، زیرا در هر بار پردازش یک تراکنش تنها مقدار کمی از داده در این جداول تحت تأثیر قرار می‌گیرند. برای بهبود عملکرد، داده‌های قدیمی معمولاً به طور دوره‌ای از سیستم‌های عملیاتی تخلیه می‌شوند.

انبارهای داده برای الگوهای دسترسی تحلیلی بهینه‌سازی شده‌اند. الگوهای دسترسی تحلیلی به طور کلی شامل گزینش فیلدهای خاصی هستند و دستورهایی به صورت «\* select» در پایگاه‌های داده عملیاتی اگر کلاً بی‌استفاده نباشند، کاملاً نادر خواهند بود. به دلیل همین اختلاف‌ها در الگوهای دسترسی، پایگاه‌های داده عملیاتی (OLTP) از کاربرد DBMS ردیف-محور بهره می‌گیرند؛ اما پایگاه‌های داده تحلیلی (OLAP) از یک DBMS ستون-محور استفاده می‌کنند. انبارهای داده‌ای برخلاف سیستم‌های عملیاتی که تصویری لحظه‌ای از کسب‌وکار را در خود نگهداری می‌کنند، به طور کلی برای نگهداری یک تاریخچه نامتناهی استفاده می‌شوند که از طریق پردازش‌های ETL پیاده‌سازی شده و به طور دوره‌ای داده‌ها از سیستم‌های عملیاتی به انبارهای داده انتقال می‌یابند.

## تکامل در کاربرد سازمانی



این اصطلاح‌ها به سطح پیچیدگی یک انبار داده مربوط هستند:

* **انبارهای داده عملیاتی آفلاین:** انبارهای داده در این مرحله از تکامل برحسب یک چرخه زمانی منظم (معمولاً روزانه، هفتگی یا ماهانه) از سیستم‌های عملیاتی به‌روز می‌شوند و داده‌ها در قالب گزارش-محور یکپارچه‌ای ذخیره می‌شوند.
* **انبارهای داده آفلاین:** انبارهای داده در این مرحله از داده‌های سیستم‌های عملیاتی در یک مبنای منظم به‌روزرسانی می‌شوند و انبارهای داده آن‌ها را در ساختارهای داده‌ای ذخیره می‌کنند که جهت تسهیل گزارش‌گیری طراحی شده‌اند.
* **انبارهای داده لحظه‌ای:** انبارهای داده یکپارچه آنلاین نماینده مرحله طراحی انبارهای داده همزمان هستند که در این مرحله داده‌ها برای هر تراکنشی که روی داده‌های منبع انجام می‌یابد، به‌روزرسانی می‌شوند.
* **انبارهای داده یکپارچه:** این انبارهای داده‌ای، داده‌ها را از بخش‌های مختلف کسب‌وکار گردآوری می‌کنند و از این رو کاربران می‌توانند اطلاعاتی که از سیستم‌های دیگر نیاز دارند را نیز ملاحظه نمایند.