

<p>معاونت برنامه ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور</p>	 <p>نهاد کتابخانه‌های عمومی کشور</p>	<p>نهاد آمار نهاد کتابخانه‌های عمومی کشور</p>
<p>عنوان سند:</p> <p><b>دستورالعمل ایجاد انبار داده</b></p>		

<p>کد سند: <b>IPLF-SS-W-14</b> <b>01:1397</b></p>	
<p>تاریخ بازنگری:</p>	<p>تاریخ تهیه: ۱۳۹۶/۱۰/۱۵</p>

<p>تصویب کننده:</p>	<p>تایید کننده: معاونت برنامه ریزی، پژوهش و فناوری اطلاعات نهاد</p>	<p>تهیه کننده: پژوهشکده‌ی آمار</p>
---------------------	---	--

### ۱- هدف

هدف از تهیه‌ی این سند علاوه بر ارایه‌ی مطالبی برای درک شفاف‌تر بر الزام پیاده‌سازی یک انبارداده در سازمان‌های عملیاتی بزرگ امروز، تعیین ساختار انبارداده و مراحل طراحی و پیاده‌سازی آن است.

### ۲- دامنه‌ی کاربرد

تمام کتابخانه‌های عمومی (نهادی، مشارکتی و مستقل) کشور.

### ۳- تعاریف

**مرکز داده (Data Center):** مرکز داده (DC) مجموعه‌ی بزرگی از سرورهای شبکه‌های رایانه‌ای است که معمولاً برای ذخیره‌سازی، پردازش یا توزیع حجم زیادی از داده‌ها از راه دور توسط سازمان‌ها مورد استفاده قرار می‌گیرد.

**اجرای انبارداده (Data Warehousing):** این اصطلاح به فرایند طراحی، ایجاد و نگهداری یک سیستم انبارداده اطلاق می‌شود.

**دادگان (Database):** دادگان (DB) مجموعه‌ای ساختاریافته از داده‌هاست که روی یک رایانه نگهداری می‌شود و از راه‌های مختلفی در دسترس قرار می‌گیرد.

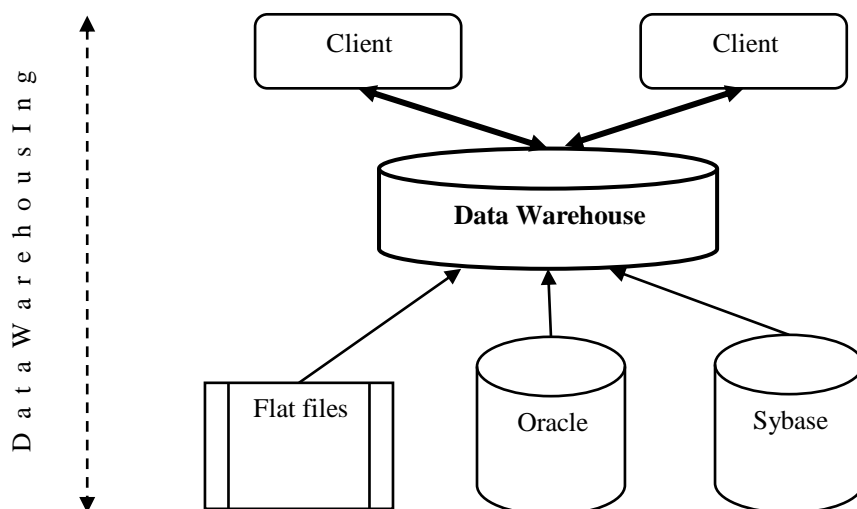
**تلخیص (Aggregation):** تلخیص، روشی برای افزایش سرعت اجرای پرس و جوهاست. در این روش اطلاعات مرجع با توجه به ابعاد انتخاب‌شده توسط کاربر خلاصه می‌شوند.

**استخراج؛ تبدیل؛ بارگذاری (ETL):** فرایندی است که طی آن داده‌ها از یک محیط به محیط دیگر جابجا می‌شوند.

**پردازش تحلیل برخط (OLAP):** پردازش تحلیلی برخط یا Online Analytical Processing روش‌های سریعی را برای به‌دست آوردن برشی از داده‌ها به کاربران نهایی ارایه می‌کند.

**مدیریت اطلاعات به صورت غیرفعال (Passive) و فعال (Active):** یک راه برای دستیابی به اطلاعات چندین دادگان توزیع‌شده‌ی ناهمگون، استفاده از روش «غیرفعال» است. در این روش وقتی که کاربر یک عملیات پرس و جو (Query) را اجرا می‌کند، سیستم منابع داده‌ای را تشخیص و برای هر منبع یک دستور پرس و جوی جداگانه را صادر می‌کند. برای رسیدن به پاسخ نهایی پس از استخراج نتایج از منابع مختلف، لازم است تا اطلاعات ترجمه، فیلتر و ترکیب شوند. در روش «غیرفعال» با وجود منابع داده‌ای توزیع‌شده‌ی مختلف و ناهمگون، اجرای عملیات پرس و جو بسیار پیچیده و زمان‌بر خواهد بود.

در روش «فعال»، وقتی عملیات پرس و جو اجرا می‌شود نیازی به ترجمه و ارسال آن عملیات به منابع داده‌ای مختلف وجود ندارد و اطلاعات در این روش دارای یک قالب واحد است. به‌طور ویژه این روش یا روش اجرای



نگاره ۲-۳- روش اجرای انبار داده

یک انبار داده (Data warehousing) نگاره ۲-۳، دارای مزایای زیادی برای آن دسته از کاربران است که نیازمند پیش‌بینی با استفاده از داده‌های موجود یا اجرای پرس و جوهای بالا هستند. انبار داده (DW) چیست؟

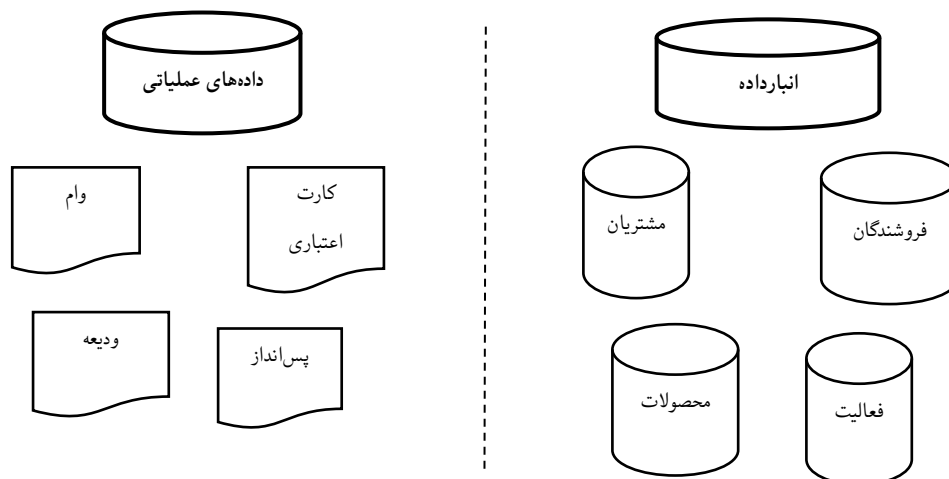
تعاریف مختلفی برای یک انبار داده وجود دارد. رالف کیمبال - کارشناس زبده‌ی انبار داده - تعریف خلاصه‌ای از یک انبار داده ارائه داده است. هرچند که در این تعریف وی بر روی وظیفه‌ی یک انبار داده متمرکز شده اما مانند بیل اینمون به چگونگی ایجاد یک انبار داده نپرداخته است: یک انبار داده، نسخه‌ای از داده‌های تراکنشی است که به‌طور ویژه برای انجام پرس و جو و تحلیل ساخت‌یافته شده است.

یکی از متداول‌ترین تعاریف انبار داده توسط بیل اینمون - دانشمند آمریکایی رایانه - ارائه شده است که بسیاری از صاحب‌نظران وی را به‌عنوان پدر این دانش به رسمیت می‌شناسند:

یک انبار داده، مجموعه‌ای از داده‌ها است که به‌صورتی «موضوع‌گرا»، «یکپارچه»، «متغیر با زمان» و «از دست نرفتنی» گرد آمده‌اند و در پشتیبانی فرایند تصمیم‌سازی‌های مدیریتی جای می‌گیرد.

○ موضوع‌گرا (subject-oriented): یک انبار داده باید بتواند برای تحلیل یک حوزه‌ی موضوعی خاص مورد استفاده قرار بگیرد. برای مثال، «خدمات»، «متقاضیان»، «پیمان‌کاران»... می‌توانند هر کدام یک موضوع خاص باشند. روش‌های زیادی برای طبقه‌بندی یا موضوعی کردن داده‌ها وجود دارد اما چیزی که در پیاده‌سازی و به‌کارگیری یک انبار داده اهمیت دارد این است که تصمیم‌سازان به‌عنوان کاربران نهایی به دنبال تحلیل چه چیز هستند و منابع اطلاعاتی توسط چه بخش(هایی) تولید می‌شود. برای مثال دادگان استاندارد یک سازمان، ممکن است وظایفی را برای ذخیره‌سازی اطلاعات وام‌ها،

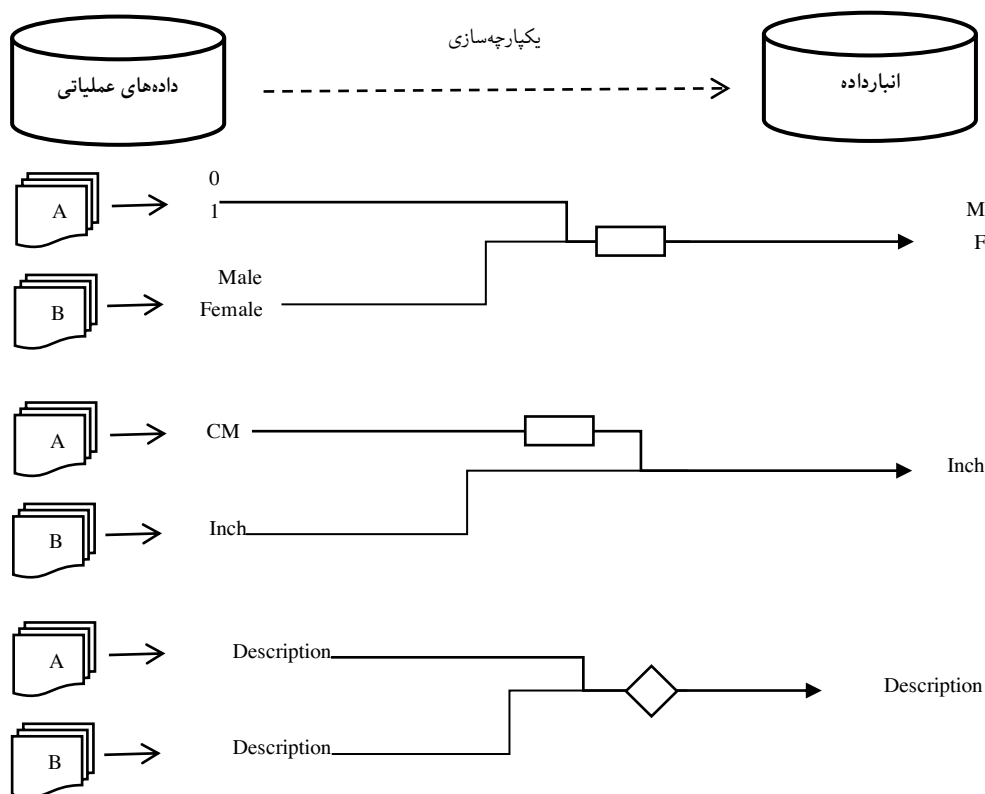
کارت‌های اعتباری، پس‌انداز و ودیعه بر عهده داشته باشد (عملیات‌گرا) اما یک تحلیل‌گر خواهد اطلاعاتی در مورد مشتریان، فروشندگان، محصولات و فعالیت آن‌ها داشته باشد. جهت‌بخشی به داده‌ها به‌سوی موضوعی خاص، منجر به انجام طبقه‌بندی‌هایی مفید برای یک تحلیل‌گر سازمانی خواهد شد



(موضوع‌گرا - نگاره ۲-۴).

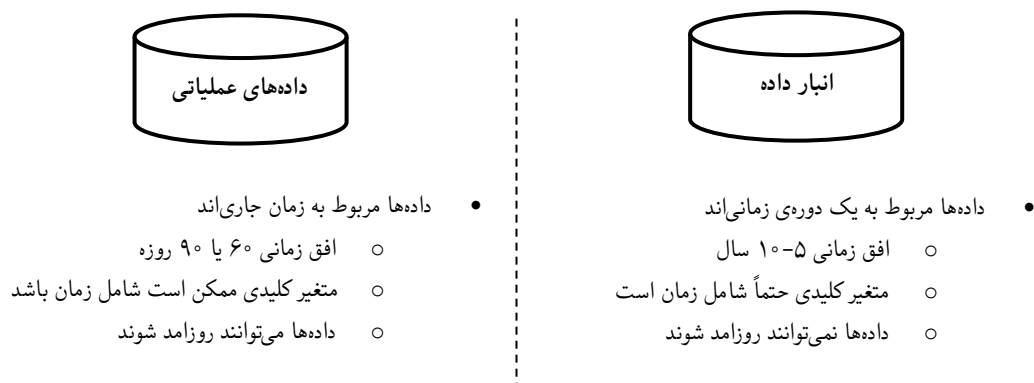
نگاره ۲-۴ - موضوع‌گرا بودن انبار داده

- یکپارچه (integrated): منابع داده‌ای متعدد درون یک انبار داده به‌صورت یک منبع اطلاعاتی یکپارچه در می‌آیند. برای مثال ممکن است منابع اطلاعاتی A و B حاوی روش‌های مختلفی برای شناسایی یک محصول یا ویژگی باشند. اما در یک انبار داده تنها یک روش برای شناسایی یک محصول یا ویژگی وجود خواهد داشت. به‌عبارتی دیگر خاصیت یکپارچگی یعنی با یک مدل واحد برای ذخیره‌سازی داده‌ها روبه‌رو هستیم. از آنجا که داده‌های عملیاتی یک سازمان در منابع گوناگونی ذخیره می‌شوند مدل داده‌ای هر کدام نیز متفاوت خواهد بود. هدف از پیاده‌سازی یک انبار داده این است که کاربر هنگام اجرای یک پرس و جو با مدل‌های داده‌ای مختلف سر و کار نداشته باشد. یکپارچه‌سازی داده‌ها در یک مکان و در یک مدل داده‌ای یکی از خواص اصلی یک انبار داده یا یکی از اهداف پیاده‌سازی آن است (نگاره ۲-۵).



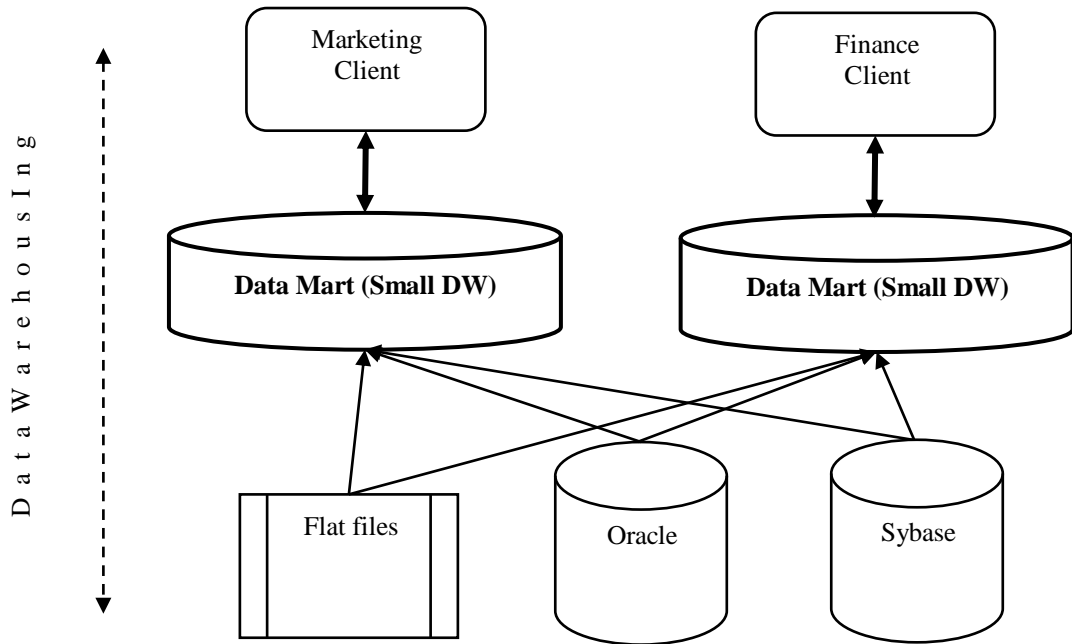
نگاره ۲-۵- یکپارچگی انبار داده

- متغیر با زمان (time-variant): داده‌های موجود در یک انبار داده مبتنی بر زمان نگهداری می‌شوند و می‌توان داده‌های مربوط به سه، شش، دوازده ماه پیش یا حتی قدیمی‌تر را از آن به دست آورد. این خاصیت در تضاد با سیستمی است که به دلایلی جدیدترین داده‌ها در آن نگهداری می‌شود. برای مثال یک سیستم ثبت معاملات ممکن است شامل آدرس‌های اخیر یک مشتری باشد در صورتی که یک انبار داده می‌تواند حاوی تمامی آدرس‌های مرتبط با آن مشتری باشد. داده‌های موجود در یک انبار داده می‌توانند در برگیرنده‌ی اطلاعات مربوط به یک دهه باشد در حالی که دادگان استاندارد معمولاً فقط شامل داده‌های یک ماه گذشته هستند. در یک انبار داده، زمان یک عامل کلیدی است چون به کاربر اجازه می‌دهد تا برای تحلیل خود یک روند در نظر داشته باشد. برای مثال با چنین قابلیت می‌توان یک ویژگی را برای یک فصل از سال استخراج و مورد بررسی قرار داد و آن را دوره‌های زمانی دیگر مقایسه کرد (نگاره ۲-۶).



نگاره ۲-۶- متغیر بودن انبارداده با زمان

- از دست نرفتگی (non-volatile): هنگامی که داده‌ای در انبارداده قرار می‌گیرد تغییر نخواهد کرد. بنا بر این داده‌های موجود در یک انبارداده که مبتنی بر تاریخ هستند هرگز نباید تغییر یابند. در یک دادگان استاندارد، داده‌ها می‌توانند اضافه، حذف یا روزآمد شوند و معمولاً حالت فعلی سازمان را نشان می‌دهند. در یک انبارداده، داده‌های جدید فقط می‌توانند اضافه شوند و کاربران نمی‌توانند داده‌های موجود در آن را تغییر دهند. در انبارداده برخی سازمان‌ها حتی اطلاعات ۱۰ تا ۲۰ سال ذخیره می‌شود.
- انبارداده موضوعی (Data Mart): یک انبارداده موضوعی (DM) انبارداده‌ی است که برای یک واحد ویژه در یک سازمان ایجاد شده است. برای مثال واحدهای مالی، فروش و بازاریابی یک سازمان می‌توانند انبارداده موضوعی خود را داشته باشند (نگاره ۲-۷). هم‌چنین تعریف دیگری برای انبارداده موضوعی به این صورت بیان می‌شود که یک انبارداده موضوعی یک لایه‌ی دسترسی به محیط انبارداده است و از آن برای استخراج داده برای کاربران استفاده می‌شود. به عبارت دیگر انبارداده موضوعی زیرمجموعه‌ای از یک انبارداده است که برای گروه مخصوصی در کسب و کار مورد نظر جهت دار شده است. عملاً انبارداده موضوعی قسمت‌های کوچکی از یک انبارداده هستند.



نگاره ۲-۷- انبارداده موضوعی

## ۴-روش کار

## • اجرای انبارداده (Data Warehousing)

در اجرای یک انبارداده دو مرحله وجود دارند که عبارتند از: مرحله تعیین ساختار و مرحله طراحی (پیاده‌سازی)

## ○ ساختار انبارداده (Data Warehouse Architecture)

سیستم‌های مختلف انبارداده ساختارهای متفاوتی دارند. بعضی از آنها می‌توانند دارای یک انبارداده عملیاتی (ODS: Operational Data Store) باشند. بعضی دیگر می‌توانند دارای انبارداده موضوعی (Data Marts) باشند. برخی حاوی چند منبع اطلاعاتی و برخی حاوی تعداد زیادی از آنها هستند.

معمولاً در معماری تمامی سیستم‌های انبارداده لایه‌های زیر وجود دارند (نگاره ۲-۸):

- ✓ لایه منبع داده‌ها (Data Source Layer)
- ✓ لایه استخراج داده‌ها (Data Extraction Layer)
- ✓ مرکز استقرار داده‌ها (Staging Area)
- ✓ لایه ETL یا لایه استخراج؛ تبدیل؛ بارگزاری (ETL Layer)
- ✓ لایه ذخیره‌سازی داده‌ها (Data Storage Layer)
- ✓ لایه منطق داده‌ها (Data Logic Layer)
- ✓ لایه آرایه‌ی داده‌ها (Data Presentation Layer)
- ✓ لایه فراداده (Metadata Layer)
- ✓ لایه عملیات سیستمی (System Operations Layer)

Metadata Layer						
Data Source Layer	Data Extraction Layer	Staging Area	ETL Layer	Data Storage Layer	Data Logic Layer	Data Presentation Layer
System Operations Layer						

نگاره ۲-۸- معماری انبارداده

### ✓ لایه‌ی منبع داده‌ها (Data Source Layer)

این لایه نمایان‌گر منابع داده‌ای مختلفی است که داده‌های انبارداده را تأمین می‌کنند. یک منبع داده‌ای می‌تواند دارای هر نوع قالبی باشد. برای مثال یک فایل متنی ساده (Plain text file)، دادگان رابطه‌ای، انواع دیگری از دادگان، فایل اکسل و سایر می‌توانند نقش یک منبع داده‌ای را ایفا نمایند. انواع بسیار زیادی از داده‌ها می‌توانند یک منبع داده‌ای باشند:

- داده‌های عملیاتی؛ از قبیل داده‌های فروش، داده‌های تولید، داده‌های فهرست اموال، داده‌های بازاریابی
- گزارش‌های ثبت شده‌ی سرور وب حاوی اطلاعات مرور شده توسط کاربران
- داده‌های پژوهشی بازار داخلی
- داده‌های شخص ثالث مانند داده‌های سرشماری، اطلاعات جمعیت‌شناسی یا داده‌های آمارگیری

تمامی این منابع داده‌ای با یکدیگر لایه‌ی منبع داده‌ها را تشکیل می‌دهند.

### ✓ لایه‌ی استخراج داده‌ها (Data Extraction Layer)

در این لایه، داده‌ها برای ورود به سیستم انبارداده از منابع داده‌ای مختلف استخراج می‌شوند. گاهی در این لایه تا حدی داده‌ها مورد پاک‌سازی قرار می‌گیرند.

### ✓ مرکز استقرار داده‌ها (Staging Area)

این لایه همان جایی است که داده‌ها پیش از پاک‌سازی و تبدیل به داده‌های انبارداده یا داده‌های داده انبار موضوعی در آن قرار می‌گیرند. داشتن یک فضای مشترک، پردازش و یک‌پارچه‌سازی‌های بعدی داده‌ها را آسان‌تر می‌نماید.

### ✓ لایه‌ی ETL یا لایه‌ی استخراج؛ تبدیل؛ بارگزاری (ETL Layer)

در این لایه، به‌منظور تبدیل یک ساختار تراکنشی به یک ساختار تحلیلی، منطقی روی داده‌ها اعمال می‌شود. این لایه همچنین جایی است که پاک‌سازی داده‌ها انجام می‌گیرد. معمولاً در پیاده‌سازی پروژه‌های انبارداده، مرحله‌ی ETL مرحله‌ای زمان‌بر است و برای این منظور در این مرحله از یک ابزار ETL استفاده می‌شود.

### ✓ لایه‌ی ذخیره‌سازی داده‌ها (Data Storage Layer)

داده‌ها پس از تبدیل و پاک‌سازی در این لایه قرار می‌گیرند. این لایه بر اساس حوزه و عملکرد داده‌ها می‌تواند دربرگیرنده‌ی سه نوع موجودیت باشد: انبارداده (DW)، انبارداده‌ی موضوعی (DM)، و انبارداده عملیاتی (ODS). هر سیستمی می‌تواند یک یا تمامی این موجودیت‌ها را با هم داشته باشد.



### ✓ لایه‌ی منطق داده‌ها (Data Logic Layer)

قواعد کسب و کار در این لایه ذخیره می‌شوند. این قواعد تأثیری بر قواعد اساسی مربوط به تبدیل داده‌ها ندارند اما شکل گزارش را تحت تأثیر خود قرار می‌دهند.

### ✓ لایه‌ی ارائه‌ی داده‌ها (Data Presentation Layer)

عملکرد این لایه به اطلاعات دریافت شده از کاربر بستگی دارد. این اطلاعات می‌توانند جدول یا گزارش گرافیکی تحت وب باشد یا گزارش روزانه‌ای که به‌طور خودکار تولید و ایمیل می‌شود. در این لایه معمولاً یک ابزار OLAP و یک ابزار گزارش‌گیری مورد استفاده قرار می‌گیرد.

### ✓ لایه‌ی فراداده (Metadata Layer)

اطلاعات مربوط به داده‌های ذخیره شده در انبارداده در این لایه نگهداری می‌شود. یک مدل برای داده‌های منطقی می‌تواند مثالی برای موجودیت‌های این لایه باشد. اغلب برای مدیریت فراداده از یک ابزار فراداده استفاده می‌شود.

### ✓ لایه‌ی عملیات سیستمی (System Operations Layer)

این لایه حاوی اطلاعاتی در خصوص چگونگی عملکرد سیستم انبارداده است. برای مثال وضعیت کاری ETL، عملکرد سیستم و تاریخ دسترسی کاربر.

### ○ طراحی انبارداده (Data Warehouse Design)

طراحی انبارداده می‌تواند پس از انتخاب ابزار و اعضای گروه آغاز شود. فازهای مرسوم که در چرخه‌ی پیاده‌سازی یک انبارداده درگیر هستند عبارتند از: جمع‌آوری اطلاعات مورد نیاز؛ راه‌اندازی محیط فیزیکی؛ مدل‌سازی داده‌ها؛ ETL؛ طراحی مکعب OLAP؛ پیش‌توسعه؛ توسعه‌ی گزارش؛ تنظیم عملکرد؛ بهینه‌سازی پرس و جو؛ تضمین کیفیت؛ معرفی محصول؛ نگهداری محصول؛ ارتقاء تدریجی.

### ✓ جمع‌آوری اطلاعات مورد نیاز (Requirement Gathering)

هدف عمده‌ی این فاز تشخیص هر آن چیزی است که به‌منزله‌ی موفقیت در پیاده‌سازی پروژه‌ی انبارداده محسوب می‌شود. در این فاز به‌خصوص گزارش‌ها و تحلیل‌های مورد نیاز کاربران نهایی تشخیص داده می‌شوند و همراه این نیازها تعریفی واقعی‌تر برای دیگر جزئیات از قبیل اطلاعات مربوط به اندازه‌های سخت‌افزاری، ملزومات آموزشی، تشخیص منبع داده‌ای و از همه مهم‌تر برنامه‌ای که نمایان‌گر زمان پایان پروژه‌ی پیاده‌سازی انبارداده باشد به دست می‌آید. معمولاً زمان مورد نیاز برای انجام این فاز ۲ تا ۸ هفته است و خروجی آن از مستندات زیر تشکیل می‌شود:

- فهرستی از گزارش‌ها که در پایان این فاز به کاربران نهایی تحویل داده می‌شود.
- یک برنامه‌ی کاری روزآمد که به‌طور روشن مشخص‌کننده‌ی مسئولیت‌های مرجع و زمان تحویل مراحل اصلی است.

### ✓ راه‌اندازی محیط فیزیکی (Physical Environment Setup)

پروژه‌های انبارداری این سه فضا هستند: توسعه، آزمون، و تولید. باید توجه کرد که استقرار نهایی فضاهای فیزیکی مختلف کافی نیست و علاوه بر آن استقرار مناسب فرایندهای مختلفی از قبیل ETL، مکعب OLAP، گزارش‌گیری، برای فضاهای مختلف نیز مورد نیاز است. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۱ هفته است و خروجی آن عبارت است از مستندات استقرار

و نصب سخت‌افزار و نرم‌افزار برای تمامی فضاها شامل مشخصات سخت‌افزاری، کدها و تنظیمات نرم‌افزاری.

#### ✓ مدل‌سازی داده‌ها (Data Modeling)

این قسمت مرحله‌ی بسیار مهمی در پیاده‌سازی یک پروژه‌ی انبارداده است و یک سیستم انبارداده بر مدل‌سازی داده‌ها استوار است. یک مدل داده‌ای مناسب علاوه بر عملکرد خوب یک سیستم انبارداده به آسانی به آن اجازه‌ی رشد می‌دهد. در پروژه‌ی پیاده‌سازی یک انبارداده منطقی داده‌ها (Logical Data Model) بر اساس نیازهای کاربر ایجاد شده و سپس به‌صورت یک مدل فیزیکی (Physical Data Model) تبدیل می‌شود. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۲ تا ۶ هفته است و خروجی آن از مستندات زیر تشکیل می‌شود:

- تشخیص منابع داده‌ای
- مدل منطقی داده‌ها
- مدل فیزیکی داده‌ها

#### ✓ استخراج؛ تبدیل؛ بارگزاری (ETL: Extraction, Transformation, Loading)

فرایند استخراج؛ تبدیل و بازنگری یا ETL بیشترین زمان را به خود اختصاص می‌دهد و این زمان چیزی در حدود ۵۰ درصد چرخه‌ی پیاده‌سازی انبارداده یا حتی طولانی‌تر از آن است. دلیل این امر زمان بر بودن دریافت داده‌های مرجع، درک متغیرهای ضروری، قواعد کسب و کار و مدل‌های منطقی و فیزیکی است. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۱ تا ۶ هفته است و خروجی آن از مستندات زیر تشکیل می‌شود:

- جزئیات اتصال مدل‌های داده‌ای مختلف به یکدیگر

#### (Data Mapping Document)

- کدهای برنامه‌ی ETL / بسته‌ی ETL

#### ✓ طراحی مکعب OLAP (OLAP Cube Design)

طراحی مکعب OLAP یا پردازش تحلیلی برخط (Online Analytical Processing) معمولاً بر اساس فاز اول یا «جمع‌آوری اطلاعات مورد نیاز» انجام می‌شود. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۱ تا ۲ هفته است و خروجی آن از مستندات زیر تشکیل می‌شود:

- توضیحات مربوط به ابعاد و مقادیر مکعب OLAP
- مکعب OLAP تهیه شده

#### ✓ پیش‌توسعه (Front End Development)

پیش‌توسعه بخش مهمی از راه‌اندازی اولیه‌ی انبارداده محسوب می‌شود. در این خصوص باید امکان قرار گرفتن گزارش‌ها بر روی وب در نظر گرفته شود. برای این منظور گزینه‌های متعددی پیش روست. استفاده از زبان‌های برنامه‌نویسی مانند ASP، PHP، یا Perl تا محصولات نرم‌افزاری آماده مانند Seagate Crystal Reports و محصولات حرفه‌ای‌تر مانند Actuate. باید پیش از انتخاب محصول از وجود امکان سفارشی شدن ساده‌ی آن برای سازمان اطمینان

حاصل کرد. به خصوص امکاناتی که برای ایجاد تغییرات در گزارش‌گیری در نظر گرفته شده است. این تغییرات نباید فقط دربرگیرنده‌ی شکل ظاهری گزارش باشد، بلکه باید امکان ایجاد تغییرات در محتوای گزارش نیز وجود داشته باشد. برای مثال امکان تغییر سرورها از Solaris/Oracle به SQL وجود داشته باشد. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۱ تا ۴ هفته است و خروجی آن مستندات مربوط به چگونگی عملیاتی نمودن پیش توسعه است.

#### ✓ توسعه‌ی گزارش (Report Development)

خصوصیات گزارش عموماً از «فاز جمع‌آوری اطلاعات مورد نیاز» به دست می‌آید. تنها نقطه‌ی اشتراک مستقیم کاربران نهایی با سیستم انبارداده گزارش‌هایی است که مشاهده می‌کنند. هر چند مرحله‌ی توسعه‌ی گزارش مانند مراحل دیگر مثل ETL و مدل‌سازی داده‌ها زمان‌بر نیست اما با این وجود نقش بسیار مهمی را در تشخیص موفقیت‌آمیز بودن پروژه‌ی انبارداده بازی می‌کند. پیش از انتشار گزارش لازم است که گروه پیاده‌سازی انبارداده به مواردی توجه داشته باشند. گزارش‌های منتشر شده باید قابلیت سفارشی‌سازی شدن داشته باشند. هم‌چنین روش‌های مختلفی برای تحویل گزارش مانند تحت وب بودن یا ارسال آن از طریق رایانامه وجود داشته باشد و از همه مهم‌تر تعیین سطح دسترسی به اطلاعات است. پس از پیاده‌سازی کامل سیستم انبارداده نیز فاز توسعه‌ی گزارش همچنان قابلیت اجرایی دارد. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۱ تا ۲ هفته است و خروجی آن از مستندات زیر تشکیل می‌شود:

#### ▪ خصوصیات گزارش

- استقرار جلوتر گزارش‌ها / گزارش‌های تحویل داده شده به کاربران تعیین شده

#### ✓ تنظیم عملکرد (Performance Tuning)

سه قسمت اصلی وجود دارند که سیستم انبارداده تا اندازه‌ای می‌تواند از تنظیم عملکرد آن‌ها بهره‌مند شود. این سه قسمت عبارتند از: ETL، پردازش پرس و جو، و تحویل گزارش. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۳ تا ۵ روز است و خروجی آن شامل مستندات تنظیم عملکرد با ذکر اهداف و نتایج است.

#### ✓ بهینه‌سازی پرس و جو (Query Optimization)

در هر دادگان تولیدی، عملکرد پرس و جوی SQL دیر یا زود بحث‌برانگیز خواهد شد. داشتن پرس و جوهای طولانی نه تنها منابع سیستم را مورد استفاده قرار داده و عملکرد سرور و اجرای برنامه را کند می‌نماید، بلکه ممکن است موجب گردد تا داده‌های جدول نتایج نیز آسیب ببینند. بنا بر این بهینه‌سازی پرس و جو یک کار مهم به حساب می‌آید.

برای این منظور رعایت اصول زیر در بهینه‌سازی پرس و جوها توصیه می‌شود:

#### ▪ درک چگونگی اجرای پرس و جوی مورد نظر کاربر توسط دادگان

دادگان‌های مختلف روش‌های مختلفی برای این کار دارند. برای مثال در MySQL می‌توان با استفاده از کد "EXPLAIN [SQL Query]" چگونگی عملکرد پرس و جو را مشاهده کرد.

#### ▪ فراخوانی اندک داده‌ها در حد امکان

فراخوانی داده‌های بیشتر یعنی احتیاج به منابع بیشتر برای پردازش و ذخیره‌ی آن‌ها. بنا بر این برای مثال وقتی به مقادیر یک متغیر از جدولی احتیاج است نباید از "SELECT\*" استفاده کرد.

#### ▪ ذخیره‌ی نتایج میانی یا واسطه

نتایج مورد نظر را می‌توان با استفاده از فنونی مثل **Inline View** و استفاده از عبارت‌های ترکیبی مانند **UNION** نگهداری کرد. در این موارد این نتایج در دادگان ذخیره نمی‌شوند اما می‌توانند به سرعت درون پرس و جو مورد استفاده قرار بگیرند. استراتژی‌های مختلف برای بهینه‌سازی پرس و جوها عبارتند از:

- افزایش سرعت اجرای پرس و جوها با استفاده از **Index**
- ایجاد جداول تلخیص شده از روی جداول اصلی
- تقسیم‌بندی سطری و ستونی داده‌ها برای کاهش حجم داده‌های در نظر گرفته شده برای پردازش
- ترکیب داده‌های جداول مختلف با هم درون یک جدول
- بهبود منابع سخت‌افزاری سرور

#### ✓ تضمین کیفیت (Quality Assurance)

وقتی گروه توسعه آمادگی پروژه را برای آزمون اعلام گردد گروه تضمین کیفیت کار خود را شروع می‌کند. این گروه از مشتری تشکیل می‌شود. مدیران پروژه باید توجه کنند که پیش از ورود به فضای تولید، اجرای فرایند تضمین کیفیت ضروری است. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۱ تا ۴ هفته است و خروجی آن از مستندات زیر تشکیل می‌شود:

- طرح و جزئیات آزمون تضمین کیفیت (QA)
- تأییدیه‌ی تضمین کیفیت مبنی بر آمادگی ورود سیستم انبارداده به فضای تولید

#### ✓ معرفی محصول (Rolling out to Production)

پس از تأیید گروه تضمین کیفیت، سیستم انبارداده کار خود را شروع می‌کند. بسته به تعداد برخظ کاربران نهایی شروع به کار انبارداده، تا یک هفته به طول می‌انجامد. بیشتر کاربران نهایی از طریق وب به انبارداده دسترسی دارند. معمولاً زمان مورد نیاز برای انجام این فاز حدود ۱ تا ۳ روز است و خروجی آن شامل مستندات تحویل سیستم انبارداده نهایی است.

#### ✓ نگهداری محصول (Production Maintenance)

وقتی انبارداده وارد فضای تولید می‌شود، لازم است تا از آن نگهداری شود. اقداماتی از قبیل تهیه‌ی نظام فایل‌های پشتیبان و مدیریت بحران بسیار حائز اهمیت بوده و باید دارای طرح و برنامه باشند. همچنین نظارت مداوم بر استفاده‌های کاربران نهایی دارای اهمیت به سزایی است. این کار دو هدف را دنبال می‌کند. اول این که به درخواست‌های از دست‌رفته رسیدگی شود و آن‌ها را پیش از این که موجب کندی سیستم شوند رفع کرد و دوم این که متوجه شد چه میزان از کاربران در حال استفاده از آن هستند.

خروجی این فاز در دسترس بودن همیشگی سیستم انبارداده برای کاربران نهایی است.

#### ✓ ارتقاء تدریجی (Incremental Enhancements)

وقتی سیستم انبارداده شروع به فعالیت می‌کند ارتقاء تدریجی مورد نیاز ضروری خواهد شد. منظور از این کار ایجاد فازهای جدید انبارداده نیست، بلکه تغییرات کوچکی است که خود سازمان آن را دنبال می‌کند. خروجی این فاز از مستندات زیر تشکیل می‌شود:

- مدیریت تغییرات
- تغییرات فعلی اعمال شده در سیستم انبارداده

#### ۵- مسئولیت‌ها

مدیریت ایجاد زیرساخت‌ها بعهدہ معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور و مدیریت بهره برداری و تهیه گزارش بعهدہ مدیریت آمار نهاد کتابخانه‌های عمومی کشور و مدیریت آمار سطح یک، دو و سه می‌باشد که به شرح زیر می‌باشد.

ردیف	شرح اقدام	مرجع سازمان مسئول	مرجع قانونی
۱	جمع‌آوری اطلاعات مورد نیاز	مدیریت آمار نهاد کتابخانه‌های عمومی کشور و مدیریت آمار سطح یک، دو و سه	ماده ۱۳، ۱۷، ۱۸، ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۲	راه اندازی محیط فیزیکی	معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۳	مدل‌سازی داده‌ها	مدیریت آمار نهاد کتابخانه‌های عمومی کشور	ماده ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۴	استخراج؛ تبدیل، بارگزاری	معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۵	طراحی مکعب OLAP	مدیریت آمار نهاد کتابخانه‌های عمومی کشور با همکاری معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۶	پیش‌توسعه	مدیریت آمار نهاد کتابخانه‌های عمومی کشور با همکاری معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۷	توسعه گزارش	مدیریت آمار نهاد کتابخانه‌های عمومی کشور با همکاری معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۸	تنظیم عملکرد	معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور با همکاری مدیریت آمار نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۹	بهینه‌سازی پرس و جو	معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات با همکاری مدیریت آمار نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۱۰	تضمین کیفیت	مدیریت آمار نهاد کتابخانه‌های عمومی کشور	ماده ۴۴ و ۵۵ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۱۱	معرفی محصول	مدیریت آمار نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۱۲	نگهداری محصول	مدیریت آمار نهاد کتابخانه‌های عمومی کشور با همکاری معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور
۱۳	ارتقا تدریجی	مدیریت آمار نهاد کتابخانه‌های عمومی کشور با همکاری معاونت برنامه‌ریزی، پژوهش و فناوری اطلاعات نهاد کتابخانه‌های عمومی کشور	ماده ۴۳ و ۴۴ نظام‌نامه نظام آمار نهاد کتابخانه‌های عمومی کشور