

## بسمه تعالی

### نکات مهم در طراحی انبار داده ها

در سطح کلان و کاملاً مدیریتی، ایجاد انبار داده ها یکی از مهم ترین و زیر ساختی ترین مراحل استقرار هوشمندی کسب و کار می باشد. سئوالات مهمی که باید در طراحی انبار داده ها پاسخگویی و تحلیل شوند، عبارتند از:

**سؤال اول:** چه سیستم های عملیاتی و یا مکانیزه در سازمان باید بعنوان منابع اطلاعاتی در انبار داده ها، ETL شوند؟

**سؤال دوم:** انتقال داده های سیستم های مکانیزه و عملیاتی سازمان به انبار داده ها چگونه باید انجام شود؟

**سؤال سوم:** چه اطلاعات و آمارهایی و با چه سطح دسترسی و با چه جزئیاتی (صفات و مقادیر صفات) باید استخراج شوند؟

**سؤال چهارم:** کیفیت داده های ETL شده (Data Quality) در چه سطحی خواهد بود؟

در چرخه عمر هوشمندی کسب و کار، سیستم های عملیاتی نقطه آغازین برای ارائه داده ها هستند که در ادامه، مورد تحلیل قرار خواهند گرفت. اگر داده هایی که در سیستم های عملیاتی ذخیره شده اند، به درستی در انبار داده ها تجمیع نشوند، امکان تحلیلی صحیح و جامع از آنها میسر نخواهد بود. همچنین اگر سیستم های عملیاتی دارای خطای داده ای باشند، در زمان نگاشت (Mapping) و تجمیع (Aggregate) داده ها، مشکلات زیادی بروز خواهند کرد. لذا کیفیت داده ها موضوع مهمی در ایجاد انبار داده ها و کیفیت تصمیمات حاصله خواهد بود.

چنانچه داده ای کاملاً اهداف خود را پوشش دهد، آنگاه آن داده ها دارای کیفیت شناخته می شوند. عدم کیفیت داده در سیستم های عملیاتی و مکانیزه، یکی از مهمترین عوامل شکست پروژه های BI می باشد. چنانچه کاربران احساس کنند که داده ها و اطلاعات دریافتی آنها از صحت و دقت کافی برخوردار نیست، اعتماد خود را نسبت به سیستم های هوشمندی کسب و کار از دست می دهند.

**اهمیت کیفیت داده ها را می توان در موارد زیر خلاصه نمود:**

- اطمینان مضاعف در تصمیم گیری ها
- کاهش ریسک تصمیم گیری ها و در نتیجه کاهش هزینه ها
- بهبود در تصمیمات استراتژیک
- پرهیز از تاثیرات مرکب داده های آلوده
- دلایل ناپاکی و یا بی کیفیت بودن داده ها را می توان بصورت زیر عنوان نمود:
- اشکالات سیستمی برنامه های عملیاتی و مکانیزه سازمان
- در برنامه نرم افزاری
- در طراحی بانک اطلاعاتی
- استفاده مقادیر پایه های متفاوت

- وجود شناسه های غیر یکتا
- وجود معانی نامفهوم در مقادیر پایه ای
- وجود فیلدهای چند منظوره
- تبدیل داده ها از سیستم های قدیمی و انتقال نادرست اطلاعات از بانک اطلاعاتی قدیمی به جدید
- وجود تعاریف متفاوت یا تعریف نشده یا شفاف نشده از داده ها
- گردآوری نادرست داده های چند سیستم
- ترکیب نادرست انتظارات کاربران
- داده های ورودی ناقص
- فقدان سیاست های مدیریت و کنترل صحت داده ها (عدم وجود تاییدکنندگان در سیستم های مکانیزه و عملیاتی)
- وجود چند کپی از داده ها در دست چندین نفر و بروز رسانی شخصی آن داده ها توسط هر فرد
- ورود سهوی داده های اشتباه
- ورود عمدی داده های اشتباه (مثلاً کلاهبرداری)

برای کنترل کیفیت داده ها باید به حاکمیت داده ها (Data Governance) در سه سطح زیر توجه ویژه شود :

- مالکیت داده یا Data Ownership که به عهده واحدهای متولی داده در سازمان است. واحدهای مختلف سازمان در رابطه به داده های خود، سطوح دسترسی، نحوه توزیع و قوانین مرتبط سازمانی را بهتر از دیگران می دانند.
- نظارت بر داده یا Data Stewardship که به عهده مدیران واحدهای متولی داده یا واحدهای نظارتی یا ستادی سازمان است. در این سطح از اختیارات، امکان ارائه پیشنهادهای در خصوص دسترسی، امنیت، توزیع و نگهداشت ارائه می شود.
- نگهداشت داده ها یا Data Custodianship که بعهد مدیریت زیر ساخت فناوری اطلاعات سازمان است. از آنجایی که این مدیریت صلاحیت لازم در خصوص مهارتهای فنی لازم، شناخت مفاهیم و رویه های امنیت داده، رویه های بازیابی داده ها و ... را دارا می باشد، مسئولیت ذخیره سازی، در دسترس قرار دادن، پشتیبان گیری، آرشیو داده و مواردی اینگونه را در اختیار دارد.

سؤال پنجم: فراداده (Meta Data) های مورد نیاز برای مدیریت انبار داده ها چگونه شناسایی و مستند سازی شوند؟

انواع فراداده های مورد نیاز عبارتند از:

- فراداده کسب و کار (Business Metadata): به توصیف معانی هوشمندی کسب و کار و داده های انبار شده از دید سازمان می پردازد، داده هایی چون تعریف داده و سنج (metric) ها، قوانین کسب و کار، مدل های داده ای، تعریف مالک / ناظر داده و غیره.
- فراداده فرایندی (Process Metadata): به توصیف مکان، زمان و چگونگی مراحل دریافت، تبدیل و بارگذاری داده می پردازد، داده هایی چون نقشه های منبع / مقصد، قوانین تبدیل داده، قوانین پاکسازی داده و غیره.



- **فراداده کاربرد (Application Metadata):** به توصیف چگونگی دسترسی و استفاده از داده می پردازد، داده هایی چون تاریخچه دسترسی به داده، فرکانس، زمان و چگونگی دسترسی به داده و غیره.
- **فراداده فنی (Technical Metadata):** به توصیف مکان های فیزیکی داده، قالب های داده، حجم داده، نام های فنی، نوع داده، ساختارهای داده و غیره می پردازد.

**سؤال ششم:** مدیریت داده های اصلی یا **MDM (Master Data Management)** چگونه انجام خواهد شد؟

در هر سازمان، داده های اصلی زیادی وجود دارد که باید برای انتقال در انبار داده ها، یکپارچه و Unique شوند. این داده ها و اطلاعات اصلی می توانند انواع محصولات، انواع خدمات، تقسیمات جغرافیایی، سرفصلهای بودجه ای، اسامی پروژه ها و ... در پایگاههای مختلف سازمان باشند که باید یکپارچه سازی شوند. به عملیات یکپارچه سازی این داده های اصلی **MDM** یا **Master Data Management** گفته می شود. نمونه رایج و همیشگی مشکلات **MDM** را میتوان موارد بصورت زیر دسته بندی نمود:

- مدیریت صفات و ساختار فرزند ساختاری بین صفات مرتبط و همچنین یکپارچه سازی مقادیر و کدهای مرتبط با صفات در انبار داده ها
- مدیریت اطلاعات مرتبط به داده های اصلی در دو پایگاه اطلاعاتی مختلف
- مدیریت فرزند ساختارهای غیر یکسان در دو پایگاه اطلاعاتی مختلف
- و مواردی دیگر

البته مشکلات وجود داده های اصلی متناقض و یا تکراری به ورود اطلاعات های اشتباهی و همچنین خطاهای سیستمی نیز مرتبط می باشد.

**سؤال هفتم:** آیا داده ها در زمان درست (**Right Time Data**) به انبار داده ها منتقل می شوند؟

مدیریت زمانبندی انتقال داده ها به انبار داده ها یکی از تنظیمات مهم و تاثیر گذار در صحت تصمیم گیری های مرتبط به **BI** می باشد. به روز آوری ها، هر چه بیشتر در زمان درست خود انجام شوند، اجازه بینش به موقع و اقدام اصلاحی بهتری را میسر می سازند. از منظری متفاوت، هرچه تاخیرها در به روز آوری داده ها بیشتر باشد، تبعات بیشتری به سازمان منعکس می شود. در اینجاست که مولفه های تاخیر در داده ها باید بررسی و مدیریت گردد.

**تاخیر ها در داده ها که بر تصمیم گیری ها تاثیر می گذارند عبارتند از:**

- تاخیر دریافت داده از منبع
- تاخیر در تجزیه و تحلیل
- تاخیر در تصمیم گیری

شایان ذکر است که بلادرنگ بودن یا به روزآوری های سریع برای همه موضوعات **BI** ضرورت ندارد و فرکانس های به روزرسانی انبار داده جز مواردی هستند که باید برای هر منبع و هر موضوع به دقت بررسی و ارزیابی شوند. چه بسا داشتن داده

های روزانه و یا ساعتی، موجب بروز مشکلات Performance در BI و همچنین بروز مشکلات مالی در حوزه سخت افزاری، شبکه ای و نرم افزاری (هم در انبار داده ها و هم در منبع اطلاعاتی) شود.

**سئوال هشتم:** از کدام مدل طراحی انبار داده ها استفاده شود و جداول انبار داده ها (Data Warehouse Tables) چگونه طراحی شوند؟

فراتر از مباحث کیفیت داده ها، انتخاب بهترین معماری برای مدل کردن داده ها در انبار داده ها خیلی مهم و اساسی می باشد. امروزه از دو دیدگاه Top-Down و Bottom-Up برای مدل کردن انبار داده ها استفاده می شود که تفاوت آنها بیشتر در انبارش داده ها می باشد. در روش Top-Down که توسط Bill Inmon معرفی شده است، انبارش داده ها به روش ریزدانه (Granular) و نرمال شده و بصورت دانه برفی (Snowflake) طراحی می شوند. اما در روش Bottom-Up که توسط Ralph Kimball معرفی شده است از روش غیر نرمال بودن داده ها و بصورت Star استفاده می کند.

البته باید توجه کرد که در راهکارهای هوش کسب و کار، می توان راهکارهایی را انتخاب کرد که بدون نیاز به انبار داده ها راهکار نهایی ارائه شود!