

چکیده- برای بهبود عملکرد هر سازمان مدیران سازمان با توجه به نحوه عملکرد آن سازمان اقدام به تصمیم گیری های مناسب می کنند. از طرفی شرکت های امروزی از نرم افزارهای مختلفی برای انجام کارهای روزانه خود استفاده می کنند، برخی از آن ها عملیاتی هستند و برخی نیز تولید محتوا و داده دارند و با استفاده از آن ها بانک های اطلاعاتی از داده های مربوط به سازمان یا شرکت و عملکرد آن به وجود می آید. لذا استفاده از داده های نرم افزارهای سازمان در تهیه گزارش سازمان و بهبود عملکرد آن می تواند مفید واقع شده و به تصمیم گیری های مدیران کمک می کند و در برخی موارد مبنای تصمیم گیری قرار گیرد.

با گسترش فعالیت های سازمان ها و بزرگ تر شدن آن ها، حجم داده های تولید شده نیز افزایش یافته و تصمیم گیری بر اساس آن ها دشوارتر خواهد شد. لذا تحلیل چنین داده های عظیمی به راحتی امکان پذیر نبوده و یا این که مستلزم نفرساعت زیادی کار است.

به همین منظور، برنامه های مختلفی برای کمک به تحلیل گران گسترش یافته اند و شرکت های مختلفی در این زمینه، با عنوان Business Intelligence یا هوش تجاری فعالیت می کنند. در این نوشتار بررسی اجمالی به مبحث هوش تجاری خواهد شد.

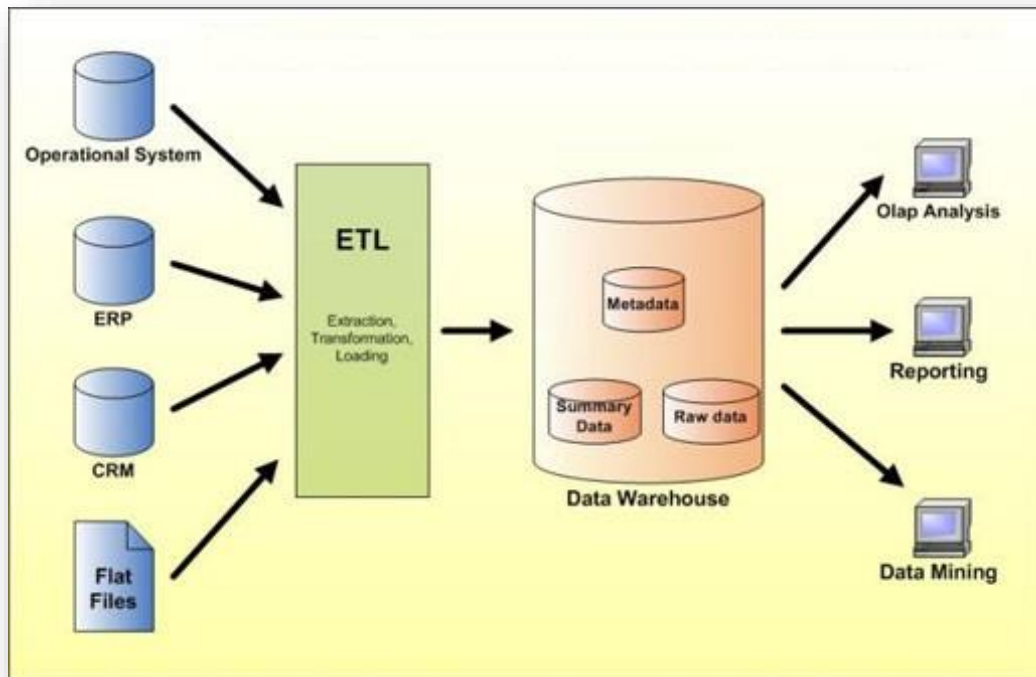
### مفهوم (BI) Business Intelligence

از نگاه Gartner<sup>1</sup>، BI به مجموعه ای از ابزارها، برنامه های کاربردی و شیوه های کسب و کار گفته می شود که هدف آن کمک به بهبود کسب و کار است. در BI با استفاده از تعدادی ابزار و Application هدف این است که به بهبود Business کمک شده و برای Business سود آوری به دست آید.

با توجه به گزارشات Gartner رهبران BI کمپانی هایی نظیر Microsoft و Qlik و Tableau و Oracle و Information Business هستند که با قدرت در این زمینه فعالیت می کنند و مطابق با آمار سال ۲۰۱۶ رقابت اصلی بین شرکت Microsoft و Qlik می باشد.

---

<sup>1</sup> Gartner یک شرکت پژوهشی و مشاوره آمریکایی در زمینه تحقیق و پژوهش و مشاوره فناوری اطلاعات است.



شکل ۱- اجزای تشکیل دهنده پروژه BI

### اجزای تشکیل دهنده پروژه BI

به صورت کلی اجزای تشکیل دهنده یک پروژه BI از نظر Microsoft به این صورت است:

۱. Data source
۲. Extract Data Management
۳. Master Data Management
۴. Data Quality System
۵. Data Warehouse
۶. OLAP Cube
۷. Data Mining
۸. Reports & Dashboard

در ادامه اجزای معرفی شده تشریح می شود.

## Data Source

در یک پروژه BI معمولاً یکی از اهداف شناسایی منابعی که دارای اطلاعات هستند می باشد که می تواند جداول موجود در بانک های اطلاعاتی، فایل های اطلاعاتی برای مثال فایل هایی با فرمت xml و json و CSV و txt و...، وب سرویس ها یا ... باشد که باید داده های موجود در آن ها جمع آوری شده تا بتوان از آن ها استفاده نمود.

## Extract Transform Loading (ETL)

در مرحله بعدی داده ها از منابع اطلاعاتی فوق استخراج شده و تبدیلات لازم بر روی آن ها انجام می شود و در نهایت در یک DW<sup>2</sup> بارگزاری می شود. هنگام استخراج داده از منابع اطلاعاتی مختلف ابتدا اتصال با استفاده از ابزار ETL به منبع مورد نظر انجام می شود. ابزار ETL این قابلیت را دارد که به یک فایل اطلاعاتی یا جدولی که در بانک اطلاعاتی دیگری وجود دارد و... متصل شود و اطلاعات را واکنشی کرده و آن ها را تمیز و در صورت لزوم تبدیل نماید. برای مثال ممکن است در بانک اطلاعاتی سیستم فروش تاریخ به صورت شمسی ذخیره شده باشد و نیاز باشد که این تاریخ را به صورت میلادی ذخیره نمایید. در این صورت ابزار ETL به راحتی و با سرعت هر چه تمام این تبدیل را انجام خواهد داد و اطلاعات مورد نظر را در DW ذخیره خواهد نمود.

ابزار ETL شرکت Microsoft، "SSIS" نام دارد.

## Master Data Management

از ابزارهایی که در یک پروژه BI با آن کار می شود، Master Data Management است که یکی از سرویس های Microsoft است که با نام Master Data Service نیز شناخته می شود. هدف این است که داده های مشترک در سیستم های مختلف مدیریت شود. برای مثال ممکن است سیستم های CRM، حسابداری، انبارداری وجود داشته باشد و در صورت جدا بودن این سیستم ها از یکدیگر، اطلاعات مشتری در هر سیستم به صورت جدا تعریف شده است برای مثال فرد خاصی ممکن است در انواع مختلف این سیستم ها که توسط شرکت های جداگانه ای توسعه داده شده است رکورد اطلاعاتی داشته باشد. در این صورت می توان کلیه اطلاعات فرد مورد نظر را یکپارچه نموده و در Master Data Service قرار داد، داده های تکراری را حذف نمود، آن ها را تمیز نموده و در پروسه ETL استفاده نموده و در DW قرار داد. چنین پروسه ای توسط ماژول های این ابزار انجام خواهد شد و بنابراین نیاز به استخراج داده ها به از سیستم های مختلف و تمیز و یکپارچه سازی آن ها توسط کاربر وجود نخواهد داشت.

---

<sup>2</sup> Data Warehouse

<sup>3</sup> SQL Server Integration Service

## Data Quality System

از سرویس های دیگر موجود در یک پروژه BI، Data Quality System یا از نظر Microsoft، Data Quality Service است. یکی از ابزار های عالی که همراه با SQL Server برای BI نصب می شود. این سرویس برای تمیز کردن داده ها به کار برده می شود. داده های نامعتبر را حذف و ویرایش نموده و یک KB<sup>4</sup> ایجاد می کند. وظیفه این پایگاه دانش این است که خود به خود با توجه به شرایط و وضعیت داده ها، هوشمند شود. برای مثال ممکن است در پایگاه های داده مختلف نام استان وجود داشته باشد، نام استان آذربایجان شرقی در پایگاه های داده مختلف به صورت " آذربایجان شرقی" یا " آذر شرقی" یا "آذر شرقی" یا... ذخیره شده باشد. این سرویس می تواند از روی ترکیب داده ها الگوی صحیح را تشخیص داده و داده های نامعتبر را شناسایی نماید و پیشنهاد جایگزین نمودن داده های نامعتبر با مقدار معتبر را بدهد. بنابراین کاربر درگیر نوشتن کدهای پیچیده برای انجام این تبدیلات نخواهد شد.

## Data Warehouse

انباره داده یک پایگاه داده معمولی است که در SQL Server ایجاد می شود اما کاملاً De-Normal است. داده های موجود در انباره داده، دارای چهار ویژگی اساسی هستند:

- ✓ Subject Oriented: موضوع گرا هستند، به این معنی که تمام داده هایی که دارای یک موضوع مشترک هستند کنار هم قرار می گیرند. اطلاعات جداول Master و Detail به صورت De-Normal شده و در یک جدول ادغام می شود. برای مثال در در سیستم فروش اطلاعات جداول سربرج و اقلام و مشتری ها و شهر ها و ... همه در کنار یکدیگر قرار می گیرند و مفهوم Subject Oriented را به وجود می آورند.
- ✓ Integrated: یکپارچه هستند، داده هایی که مربوط به یک موضوع خاص هستند از سیستم ها و بانک های اطلاعاتی مختلف جمع شده و در یک مدل واحد که در DW بوجود آمده است ذخیره می شوند. بنابراین با ETL داده ها استخراج شده، تبدیلات لازم را انجام می شود و در DW کنار هم قرار می گیرد.
- ✓ Non-Volatile: از بین نرفتن داده، داده های موجود در DW تغییر ناپذیر هستند و نمی توانند از بین بروند. زمانی که داده ای در DW درج شد، دیگر حذف نخواهد شد. زیرا ارزش اطلاعاتی دارد و باید وجود داشته باشد، در صورتی هم که نیاز باشد که چنین داده ای وجود نداشته باشد، می توان توسط یک Flag آن را غیرفعال نمود.
- ✓ Time-Variant: داده های درج شده در DW شامل بازه های زمانی مختلف هستند. ممکن است داده های چندین سال در DW وجود داشته باشد. لذا حجم برخی از جداول Warehouse

---

<sup>4</sup> Knowledge Base

به شدت بالا رفته و باید کارهای خاصی در حوزه Tuning در SQL Server روی آن ها انجام شود. در یک DW داده ها به صورت جداول Dimension و جداول Fact وجود دارد.

نکته: طراحی اصولی DW، مطابق با آنچه به آن ها اشاره شد، تاثیر بسیار مناسبی در کارایی خواهد داشت.

## OLAP Cube

هرچند داده ها در DW به صورتی قرار گرفته است که دیگر نیازی به Join های پیچیده نیست، اما زمانی که حجم داده های موجود در DW بالا می رود، دیگر ad-hoc query جوابگو نخواهد بود. مثلا ممکن است نیاز به گزارش اطلاعات از ابعاد مختلف وجود داشته باشد، در این صورت ساختار Warehouse قادر به پاسخگویی با سرعت نخواهد بود، چرا که ممکن است محاسبات سنگینی روی حجم عظیمی از اطلاعات انجام شود.

برای رفع این مشکل ساختاری به نام OLAP Cube ایجاد خواهد شد. یک سری Cube یا مکعب ایجاد می شود که شامل ساختار لازم برای استخراج داده است. برای مثال در یک مکعب یک بخش از اطلاعات، گزارش را بر اساس مشتری، یک بخش بر اساس زمان، یک بخش بر اساس شهر یا هر چیز دیگری که نیاز باشد در مکعب قرار می گیرد. می توان ویژگی های دیگری نیز به مکعب اضافه نمود. بنابراین مکعب ایجاد شده گزارشی را باید تحویل دهد که شامل چندین بعد است و هر لحظه ممکن است روابط بین ابعاد نیز تغییر کند، برای مثال ممکن است یک بار گزارش را بر اساس "شهر و زمان و مشتری" و یا بار دیگر بر اساس "مشتری و زمان و شهر" تهیه نمایید. بنابراین زمانی که Cube ایجاد می شود هنگام Process آن خودش همه حالت ها و جایگشت های مختلف فیلدها برای تهیه گزارش را در نظر گرفته و داده های آن را از DW استخراج می نماید و محاسبات لازم که اصطلاحا به آن Pre-calculation گفته می شود، روی آن ها را انجام داده و ساختار فایلی Cube ذخیره می شود. حال می توان به جای استفاده مستقیم ad-hoc query روی DW، به راحتی با ابزارهای گزارش گیری به Cube مورد نظر متصل شده و گزارشات مورد نیاز را در کسری از ثانیه استخراج نمود. زیرا Cube همه حالت های ممکن گزارش را از قبل محاسبه نموده است و نتیجه را دارد و می تواند خیلی سریع آن را در اختیار قرار دهد. هنگام استفاده از داشبوردهای مدیریتی همیشه نیاز نیست که داده های Real time در اختیار قرار گیرد و اگر یک سری گزارشات درست و ارزشمند در اختیار گذاشته شود، هر چند تاخیر نیز وجود داشته باشد، بسیار ارزشمند خواهد بود.

مدل OLAP در سیستم Microsoft به دو صورت Tabular و Multi Dimensional پیاده سازی می شود.

## Data Mining

هدف از این مرحله استخراج اطلاعات پنهان مابین داده های موجود است. روابط مشخصی که بین داده ها وجود داشته است و به آن ها توجه نشده است. اصطلاحا به آن کشف و استخراج دانش از داده می گویند. SQL Server از ۹ الگوریتم قوی برای Data Mining پشتیبانی می کند و از آن جا که SQL با زبان R، Integrate شده است، در نسخه 2016 آن پکیج های زبان R در اختیار قرار گرفته است تا با قدرت بتوان عمل استخراج را انجام داد. برای مثال در سیستم فروشگاه اینترنتی می توان با استخراج داده های پنهانی کاربران به کاربر کالای خاصی را پیشنهاد داد، برای مثال رفتارها و سبدهای خرید کاربران را بررسی نموده و کالاهای پر مصرف و مشابه آن ها را به کاربر نمایش داد و یا با تحلیل سبد خرید کاربران مشابه کاربر مورد نظر از نظر سن، جنسیت، منطقه جغرافیایی و... کالاهای مناسب کاربر را پیشنهاد کرد. استفاده از ۹ الگوریتم فوق زبان R باعث انجام سریع و دقیق پردازش ها خواهد شد.

## Reports & Dashboard

در آخرین مرحله پروژه BI، پس از جمع آوری و استخراج داده ها و ایجاد Cube و... باید داده ها به صورت Visualize روی داشبورد نمایش داده شود. برای این منظور ابزارهایی مانند Reporting Service ، Power BI ، Mobile Report Publisher ، Excel ، SharePoint و... مورد استفاده قرار می گیرد که توسط آن ها به راحتی می توان گزارش های مختلف را در بستر دلخواه تهیه نمود.

