حاکمیت داده
Data Governance

گروه آمار

Table of Contents

حاکمیت داده Data Governance

بسیاری از سازمان‌ها می دانند که داده‌ها، دارایی حیاتی سازمان است. داده و اطلاعات می‌تواند به آن‌ها در مورد مشتریان، محصولات و خدماتشان، بینش‌های جدیدی بدهد. همچنین می‌تواند به نوآوری و تحقق اهداف استراتژیک کمک کند. باوجود این، سازمان‌های معدودی به‌طور فعال این ارزش نهفته داده ها را استخراج می کنند. خلق ارزش از داده‌ها به‌صورت تصادفی رخ نمی‌دهد بلکه تحقق این مهم نیاز به تصمیم‌گیری، برنامه‌ریزی، هماهنگی و مهمتر از همه مدیریت و رهبری دارد. در واقع داده و اطلاعات دارایی های بالفعل سازمان نیستند و سازمان‌ها از طریق روش‌ها و برنامه‌هایی برای **مدیریت داده** و **حاکمیت داده** می توانند ارزش بالقوه آنها را بالفعل کنند.

از بطن یک برنامه حاکمیت داده مجموعه‌ای از اصول، قواعد و فرآیندها متولد می‌شود که قرار است همسو با اهداف کلان سازمان، موضوع خلق ارزش از داده‌ها را در قالب یک مدل عملیاتی هدایت کند. همچنین همسو با برنامه حاکمیت داده، برنامه مدیریت داده به روال‌های توسعه و اجرا در طول چرخه عمر داده می‌پردازد.

در این دوره به‌صورت کلی به تعریف و کارکرد برنامه حاکمیت داده و مدیریت داده و ارتباط بین این‌ها پرداخته می‌شود.

حاکمیت داده چیست؟

در جواب اینکه حاکمیت داده چیست باید گفت حاکمیت داده نوعی فرآیند مدیریتی در سیستم‌های سازمانی است. این مفهوم میزان دسترسی، یکپارچگی، امنیت و کاربردی بودن سیستم‌های سازمانی را مدیریت می‌کند. این کار بر اساس استانداردهای داده‌های داخلی و سیاست‌های سازمانی خاصی قابل انجام است .استانداردها و سیاست‌هایی که وظیفه نظارت بر استفاده از داده را نیز به عهده دارند. در حاکمیت داده افراد وظیفه دارند از قابل اعتماد بودن و ثبات داده‌ها اطمینان حاصل کنند. همچنین باید طوری از داده‌ها حفاظت کنند که از آن‌ها سوء استفاده نشود. این موضوع در حوزه داده‌ها بسیار اهمیت دارد، زیرا سازمان‌ها روز به روز باید مقررات جدیدی برای حفظ حریم خصوصی داده‌ها وضع کنند. مقرراتی که عملکرد تجارت را بهبود می‌دهند و منجر به تصمیم‌گیری‌های بهتر خواهند شد.

- Data Governance پایه و اساس مدیریت تمامی داده‌های سازمان را تشکیل می‌دهد.
- Data Governance استفاده کارآمد از داده‌های قابل‌اعتماد را فراهم می‌کند.
- Data Governance شامل افراد، فرآیندهای کسب‌وکاری و فناوری‌های موردنیازی است که برای مدیریت و محافظت از دارایی‌های داده سازمان به‌کار می‌روند.
- Data Governance قصد دارد تضمین کند که داده‌های سازمان قابل‌درک، درست، کامل، قابل‌اعتماد، امن و قابل‌دسترسی و استفاده باشند.
- مدیریت کارآمد داده‌ها کاری بسیار مهم است که نیازمند مکانیسم‌های کنترلی متمرکز می‌باشد.

به‌طورکلی مباحث موجود در Data Governance عبارت‌اند از:

- معماری داده Data Architecture
- کیفیت داده Data Quality
- متا داده (فرادادە) Meta-data
- انبار داده و هوش تجاری Data Warehousing & Business Intelligence
- داده‌های اصلی و مرجع Reference & Master Data
- مستندات و محتوای داده Documents & Content
- یکپارچه‌سازی و تعامل با داده Data Integration & Interoperability
- امنیت داده Data security
- محل ذخیره‌سازی و عملیاتی داده Data Storage & Operation
- مدل‌سازی و طراحی داده Data Modeling & Design

حاکمیت داده ها یا **data governance** پایه‌ای برای **مدیریت داده** در کل شرکت است و استفاده کارآمد از داده های قابل اعتماد را امکان پذیر می‌کند. مدیریت کارآمد داده ها وظیفه مهمی است که به مکانیزم‌های کنترل متمرکز نیاز دارد. **حاکمیت داده** شامل افراد، فرایندها و فنآوری‌های مورد نیاز برای مدیریت و محافظت از دارایی‌های داده شرکت برای تضمین داده های شرکتی قابل درک، صحیح، کامل، قابل اعتماد، امن و قابل کشف است. در هسته اصلی آن، **data governance** ایجاد روش‌ها و سازمانی با مسئولیت‌ها و فرایندهای مشخص برای استاندارد سازی، تلفیق، محافظت و ذخیره داده های شرکتی است.

اهداف Data Governance

هدف اصلی Data Governance ایجاد روش‌ها و چارچوبی با مسئولیت‌ها و فرآیندهای روشن برای استانداردسازی، یکپارچه‌سازی، محافظت و ذخیره‌سازی از داده‌های کسب‌وکار است. اهداف اصلی Data Governance عبارت‌اند از:

- کاهش ریسک و خطرات ناشی از داده
- تبیین و تعیین قوانین داخلی برای استفاده از داده
- پیاده‌سازی نیازمندی‌های انطباق داده
- بهبود ارتباطات داخلی و خارجی
- ایجاد ارزش از داده‌های سازمان
- کاهش هزینه‌های سازمان
- کمک به اطمینان از ادامه حضور سازمان در میدان رقابت بوسیله مدیریت ریسک و بهینه‌سازی
- تسهیل اجرای موارد فوق‌الذکر

برنامه های حاکمیت داده ها همیشه بر سطح استراتژیک، تاکتیکی و عملیاتی در شرکت‌ها تأثیر می‌گذارند. به منظور سازماندهی و استفاده بهینه از داده ها در متن سازمان و در هماهنگی با سایر پروژه‌های داده، برنامه‌های حاکمیت بر داده باید به عنوان یک روند تکرار شونده مداوم تلقی شوند. علاوه بر مسئولیت‌ها، جنبه‌های زیر در هر برنامه **data governance** باید روشن شود:

- سازمان («کجا» و «چه کسی»).
- جنبه‌های تجاری.
- جنبه‌های فنی.

بخش‌های مختلف حاکمیت داده و وظایف آنها

برنامه حاکمیت داده شامل گروه‌های مختلفی است. این گروه‌ها عبارتند از: کمیته راهبری (هیئت مدیران)، تیم حاکمیت و گروهی از مسئولان داده.کار این گروه‌ها ایجاد استانداردها و سیاست‌های لازم برای حاکمیت داده است. علاوه بر این مسئولان داده وظیفه دارند دستورالعمل‌های ضروری در حاکمیت داده را اجرا کنند. سایر افراد و گروه‌هایی که در حاکمیت داده نقش دارند شامل مدیران اجرایی، نمایندگان تجاری، گروه IT و مدیریت داده می‌شوند.

چرا حاکمیت داده مهم است؟

حاکمیت داده از مباحث مهم سازمانی و تجاری است. بدون وجود حاکمیت موثر، داده‌ها در سیستم‌های مختلف سازمان دچار ناسازگاری و تناقض می‌شوند. برای مثال، فهرست اسامی مشتریان در بخش‌های فروش، تدارکات و خدمات به هم می‌ریزد.

این مشکلات یکپارچگی داده‌ها را از بین می‌برند. درنتیجه دقت هوش تجاری (BI) پایین می‌آید و سیستم گزارش‌دهی و برنامه‌های تحلیلی نیز به مشکل برمی‌خورند. همچنین سیستم حاکمیت داده ضعیف منجر به نقض مقررات لازم برای حفظ حریم خصوصی داده‌ها خواهد شد.

حاکمیت داده سازمانی معمولاً تعاریف بهتری را برای توضیح داده‌ها ارائه می‌دهد. با وجود حاکمیت، داده‌ها به شکل استاندارد در می‌آیند. این داده‌های استاندارد در همه سیستم‌های تجاری به کار می‌روند.

هرچند Data Governance هنوز در سازمان‌ها نهادینه و فراگیر نشده است اما بیشتر سازمان‌ها برنامه Data Governance را برای برخی از اداره‌ها و یا بعضی از برنامه‌های کاربردی خود اجرا کرده‌اند. بنابراین ایجاد Data Governance به‌صورت روش مند در سازمان‌ها یعنی تحولی بزرگ در زمینه داده از قوانین غیررسمی به کنترل‌های رسمی در سازمان می‌باشد.

معمولاً Data Governance رسمی زمانی در شرکت یا سازمانی اجرا می‌شود که آن شرکت به‌اندازه‌ای رسیده باشد که دیگر نمی‌توان وظایف کارکردی را بطور مؤثری پیاده‌سازی کرد.

Data Governance پیش‌نیاز وظایف و پروژه‌های بی‌شماری است و فواید ارزشمندی نیز خواهد داشت. در زیر به برخی از فواید Data Governance اشاره می‌کنیم:

- داده‌ها و فرآیندهای سازگار و یکسان در سراسر سازمان شرط لازم برای پشتیبانی تصمیم‌گیری بهتر و جامع‌تر می‌باشند.
- رشد و افزایش چشم‌اندازهای فناوری اطلاعات در سطوح عملیاتی، تاکتیکی و استراتژیک به‌وسیله وضع قوانین جدید برای تغییر فرآیندها و داده‌ها.
- بهینه‌سازی هزینه‌های مدیریت داده به‌وسیله ساز وکارهای کنترل مرکزی (این سازوکارها به‌طور فزاینده‌ای در عصر انفجار داده‌ها بسیار کاربرد دارد)
- افزایش بهره‌وری از طریق سینرژی ایجادشده (به‌عنوان‌مثال با استفاده مجدد از فرآیندها و داده‌ها)
- ایجاد اعتمادبه‌نفس بالاتر نسبت به داده‌ها از طریق کیفیت داده تضمین‌شده، داده‌های دارای مهر تأیید و همچنین مستندات کامل فرآیندهای داده
- توانایی رسیدن به قوانین و استانداردهای تعریف‌شده جهانی یا در سطح کشور
- حفظ امنیت داده‌های داخلی و خارجی سازمان با نظارت و بررسی سیاست‌های حفظ حریم خصوصی
- افزایش کارآیی فرآیندها به‌وسیله کاهش فرآیندهای طولانی هماهنگی
- ایجاد ارتباطات شفاف و صحیح به‌وسیله استانداردسازی (این مورد پیش‌شرط اصلی برای فعالیت‌های کلان سازمانی مبتنی برداده می‌باشد)

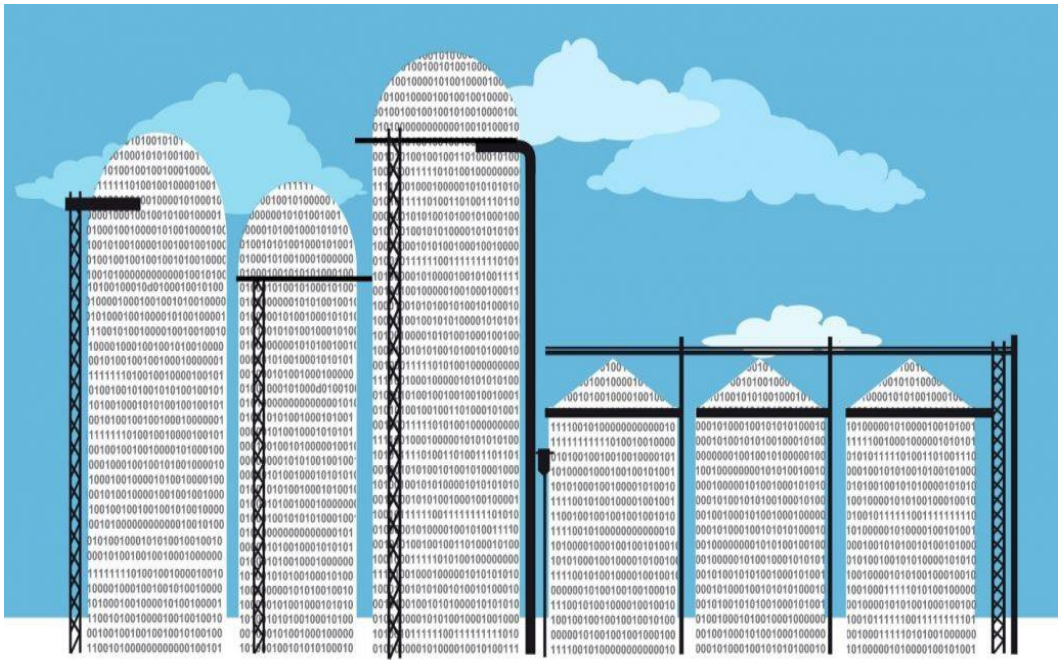
data governance بیش از هر زمان دیگری برای پاسخگویی شرکت‌ها حیاتی است. همچنین گشودن زمینه‌های جدید و نوآورانه تجارت مهم است، به عنوان مثال با تجزیه و تحلیل داده های بزرگ، که تداوم تفکر عقب مانده و ساختارهای اساسی را اجازه نمی‌دهد.

برنامه Data Governance همیشه تأثیر خود را در سطوح استراتژیک، تاکتیکی و عملیاتی سازمان‌ها بجای خواهد گذاشت. برنامه Data Governance به‌منظور سازماندهی و استفاده کارآمد از داده‌ها در متن شرکت و در راستای اجرای سایر پروژه‌های داده محور به‌عنوان فرآیندی مداوم و تکراری اجرا گردد.

مخزن داده چیست؟

هدف کارکردی حاکمیت داده در سازمان‌ها، تجزیه و تحلیل داده‌های موجود در **سیلوی اطلاعات (Data Silos)** است:
مخزن یا انباره داده

مخزن/انباره داده همان طور که از اسمش پیداست جایگاهی برای جمع‌آوری داده است که در سازمان‌ها قرار دارد .



مخزن داده زمانی تشکیل می‌شود که واحدهای تجاری، سیستم‌های جداگانه‌ای را برای پردازش معاملات در سازمان مستقر کرده، بدون اینکه از قبل هماهنگی‌های لازم را ایجاد و داده‌های سازمانی را طراحی کنند. اینجاست که حاکمیت داده نقش خود را ایفا می‌کند. حاکمیت داده، بین داده‌های موجود در این سیستم‌ها و سهامداران واحدهای تجاری مختلف هماهنگی به وجود می‌آورد. حاکمیت داده همچنین ضمانتی برای استفاده مناسب از داده‌هاست. این شیوه مدیریتی اجازه نمی‌دهد در سیستم خطا به وجود آید و از سوء استفاده احتمالی از اطلاعات مشتریان و اطلاعات مهم دیگر جلوگیری می‌کند.

این هدف با تنظیم سیاست‌های یکپارچه برای استفاده از داده و نظارت بر اجرای درست آن سیاست‌ها قابل دستیابی است. علاوه بر موارد گفته شده، حاکمیت داده از روش‌هایی برای جمع‌آوری داده استفاده می‌کند که طبق مقررات حفظ حریم خصوصی باشند.

دلایل استفاده از Data Governance

امروزه بیش از هر زمان دیگری Data Governance کمک فراوانی جهت پاسخگو بودن به سازمان‌ها خواهد کرد. همچنین Data Governance جهت ایجاد زمینه‌های جدید و نوآورانه در کسب‌وکار مانند تجزیه‌وتحلیل کلان داده‌ها که با تفکر سنتی رویکردهای جدید بسیار سخت خواهند بود.

در حال حاضر، مهم‌ترین عوامل پیشران که سازمان‌ها را مجبور خواهد کرد رویکردهای فعلی خود را مورد تجدیدنظر قرار دهند عبارت‌اند از:

- ایجاد نمایی داده محور برای پشتیبانی از مدل‌های کسب‌وکاری دیجیتالی
- بالا بردن کیفیت داده‌های سازمانی و مدیریت داده‌های اصلی
- ایجاد قابلیت مدیریت داده در محیط‌های کلان داده
- ایجاد استانداردهایی جهت افزایش توانایی واکنش به تأثیرات خارج از سازمان
- هوش تجاری سلف‌سرویس، کاربران می‌خواهند مستقل از فناوری اطلاعات آنالیز انجام دهند.
- انطباق داده: فرآیندهای داده شفاف و قابل‌فهم برای رعایت الزامات قانونی

مزایای حاکمیت داده در سازمان‌ها

در بالا درباره اینکه حاکمیت داده چیست، چرا مهم است و چه اهدافی دارد اطلاعات جامعی به شما دادیم. حال می‌خواهیم بررسی کنیم که وجود حاکمیت داده چه مزیت‌هایی دارد. ویژگی‌های فوق‌العاده حاکمیت داده آن را تبدیل به رکن مهمی در سازمان‌ها کرده است. مزایای این فرآیند شامل موارد زیر می‌شود:

- **رعایت بهتر مقررات**
- **تجزیه و تحلیل دقیق‌تر**
- **داده‌های با کیفیت‌تر**
- **دسترسی بهتر به داده‌ها**
- **کاهش هزینه مدیریت داده**
- **دادن اطلاعات بهتر به مدیران اجرایی**
- **تصمیم‌گیری تجاری بهتر**

در نهایت همه این ویژگی‌ها و مزایا باعث سودآوری بیشتر برای سازمان و افزایش درآمد می‌شود.

نقش حاکمیت داده و معماری داده در سازمان‌ها

روندهای نوین فناوری‌ها و توسعه سریع نرم‌افزارهای جدید منجر به تولید جزایر داده‌ای آشفته و پیچیده در سازمان‌ها شده است که خود باعث افت کیفیت داده و هزینه بالای به‌اشتراک‌گذاری داده شده است. ضرورت تبادل داده در مقیاس بین‌سیستمی یا بین‌سازمانی خود انبوهی از چالش‌ها و نیازمندی‌ها را در سطوح فنی-قانونی و کسب‌وکاری به همراه آورده است؛ همچنین مسائل مربوط به حریم خصوصی، سطوح دسترسی و ریسک‌های امنیتی نیز همگی اهمیت ساماندهی به **"معماری داده"** را بیش از پیش ضروری ساخته است. در حالی که تعاریف قدیمی از داده بر مفهوم بازنمایی واقعیت‌های جهان تکیه داشته، در ادبیات چند دهه اخیر داده به معنای اطلاعاتی که در فرمت دیجیتال ذخیره می‌شود در نظر گرفته شده است (اگرچه لزومی ندارد هر داده‌ای در فرمت دیجیتال ثبت و ذخیره شود) .طبیعتاً امروزه به دلیل قابلیت‌های فناوری برای ثبت و ذخیره انبوهی از چیزها به صورت الکترونیکی، به بسیاری چیزها «داده» می‌گوییم- مانند اسامی، آدرس‌ها، تاریخ‌ها و غیره- در حالی که چند دهه قبل به این چیزها عنوان داده اطلاق نمی‌شد. در تعاریف رسمی داده را **«ماده خام اطلاعات»** و اطلاعات را **«داده دارای زمینه»** می‌نامند. در این نوشته به بررسی مفاهیم حاکمیت داده و طراحی معماری داده می‌پردازیم و با مرور تجارب دولت‌ها در استقرار نظام حاکمیت داده، چارچوبی مفهومی برای تبیین جوانب مختلف حاکمیت داده تدوین و ارائه خواهد شد.

مروری بر مفاهیم و تعاریف

بر اساس تعریف DMBOK (Data Management Body of Knowledge)، **"مدیریت داده"** دربردارنده عملیات توسعه، اجرا و نظارت بر مجموعه طرح‌ها، خط‌مشی‌ها، برنامه‌ها و فعالیت‌هایی است که برای تحویل، کنترل، محافظت و ارتقای ارزش داده و دارایی‌های اطلاعاتی طی چرخه حیات داده انجام می‌شود.

"حاکمیت داده" به معنای اعمال اقتدار و کنترل بر مدیریت دارایی‌های داده‌ای است. از طرف دیگر، منشاء بسیاری از هزینه‌ها، دوباره‌کاری‌ها و ناکارآمدی‌ها به نوعی به کیفیت نامناسب داده برمی‌گردد. **«کیفیت داده»** شامل همه اقدامات برای مدیریت کیفیت داده است به طوری که اطمینان حاصل شود داده قابل استفاده است.

در تعریف ویکی‌پدیا **«حاکمیت داده»** شامل ترکیبی از فرایندها، افراد و فناوری‌ها شمرده شده که برای کار کردن صحیح و منسجم با داده‌های یک تشکیلات مورد نیاز است. این تعریف با برجسته کردن مفهوم قابلیت (ترکیب فرایندها، افراد و فناوری) این مفهوم را قابلیتی سازمانی و پایدار دانسته و نه صرفاً یک پروژه یا محصول قابل تامین. **«معماری داده»** درباره فهم جامع‌نگر و بنیادین از وضعیت ایستا و پویای موجودیت‌های داده‌ای سازمان و روابط و استانداردهای حاکم بر آن است که شامل نقشه‌های طراحی، مدل‌های مرجع، استانداردها، قالب‌های مستندسازی و برنامه تحقق وضعیت مطلوب می‌شود.

معماری داده و معماری سازمانی

دارایی (Asset)یک منبع اقتصادی است که قابلیت مالکیت یا کنترل دارد و ارزش تولید می‌کند و می‌تواند به «پول» تبدیل شود. داده در عصر جدید به عنوان یک دارایی سازمانی به رسمیت شناخته شده اگرچه هنوز نحوه مدیریت بر این دارایی ارزشمند نوین کاملاً شفاف و استاندارد نشده است. در ادبیات معماری سازمانی، داده یک دارایی مهم ولی ناملموس (Intangible) سازمانی است که در فرایندها و سیستم‌ها پردازش شده و منجر به خلق ارزش-سرویس می‌شود. چارچوب زکمن به عنوان اولین چارچوب شناخته‌شده معماری، ستون «چه چیز» یا "اشیای کسب‌وکار" را مختص داده در نظر گرفته بود. بعدها در چارچوب‌های دیگر مانند فدرال، توگف، وزارت دفاع و غیره نیز یک لایه مستقل و مهم برای داده-اطلاعات منظور شد. در **چارچوب ملی معماری سازمانی ایران (IEAF)** نیز یکی از دامنه‌های اصلی به داده و اطلاعات تخصیص یافته است.

گرچه معماری داده و مباحث آن چارچوبی مستقل از معماری سازمانی نیست و تقریباً در همه چارچوب‌های معماری سازمانی به داده و اطلاعات به صورت یک لایه مهم توجه شده، آنچه نیازمند توجه جدی است رشد نمایی کاربردها و ابزارهای لایه داده طی چند سال اخیر است به طوری که اهمیت این لایه معماری سازمانی را به عنوان یک لایه اولویت‌دار که باید برای آن سریعاً معماری و نظام مدیریت منسجم دیده شود تقویت کرده است. فناوری و راهکارهایی مانند **انباره داده(Data Warehouse)** ، **دریاچه داده(Data Lake)** ، **کلان‌داده(Big Data)** ، **داده باز(Open Data)** ، **مدیریت داده‌های کلیدی(MDM)** ، **سرویس‌های داده(Data Services)** ، **هوش تجاری (Business Intelligence)**، **جریان داده (Data Streaming)** و نظایر اینها تنها نمونه‌هایی از مصادیق رشد کاربردهای این لایه است.

در عصر دیجیتالی اهمیت لایه داده تا آنجاست که بر اساس برآوردها، حجم داده‌های الکترونیکی (دیجیتالی) در جهان تا سال ۲۰۲۵ به ۱۷۵ زتابایت (هر زتابایت معادل ۱۰ به توان ۲۱ بایت است) خواهد رسید. طبق تحقیق موسسه DATAVERSITY ، بیش از ۸۰ درصد از عملیات داده در سازمان‌ها «کاملاً خودکار» یا «نیمه خودکار» شده است و بیش از ۶۷ درصد سازمان‌ها به درجاتی نظام‌های مرتبط با حاکمیت داده را پیاده‌سازی کرده‌اند. برخی متخصصان برآورد کرده‌اند که سازمان‌ها حدود ۱۰ تا ۳۰ درصد از درآمد خود را صرف مدیریت داده می‌کنند .تخمین زده که در سال ۲۰۱۶ حدود ۳۱۰۰۰ میلیارد دلار هزینه صرف کیفیت پایین داده‌ها در جهان شده است.

حوزه‌های دانش مدیریت داده درDMBOK

حوزه‌های دانشی مرتبط با حاکمیت داده طیف وسیعی از موضوعات و تکنیک‌ها را در بر می‌گیرد. در آخرین نسخه پیکره دانش مدیریت داده(DMBOK) ، ۱۰ حوزه دانشی به اضافه حاکمیت داده سازماندهی شده است که دربردارنده مجموعه دانش و تکنیک‌های لازم برای مدیریت موثر داده است. هر کدام از این حوزه‌ها به صورت جداگانه در کتاب DMBOK تشریح و تبیین شده است .در ادامه توضیح مختصری درباره هر کدام از این حوزه‌ها ارائه می‌شود:

- **"معماری داده"** درباره فهم جامع‌نگر و بنیادین از وضعیت ایستا و پویای موجودیت‌های داده‌ای سازمان و روابط و استانداردهای حاکم بر آن است که شامل نقشه‌های طراحی، مدل‌های مرجع، استانداردها، قالب‌ها و برنامه تحقق وضعیت مطلوب می‌شود.
- **"طراحی و مدل‌سازی داده"** درباره ساختاردهی و بازنمایی گرافیکی موجودیت‌های داده‌ای در زمینه و محدوده مورد نظر است.

- **"ثبت و عملیات داده"** درباره مدیریت طراحی پایگاه داده، روش‌های پیاده‌سازی و پشتیبانی در راستای وظایف محوله و محافظت از ارزش داده است.
- **"امنیت داده"** شامل همه نظام‌ها و اقدامات لازم برای تامین امنیت تایید هویت، مجوزدهی و دسترسی به داده‌هاست اعم از اقدامات پیشگیرانه، ممیزی و تخفیف اثر ریسک.
- **"تعامل‌پذیری و یکپارچگی داده"** دربردارنده نظام‌ها و اقدامات لازم برای انتقال، تجمیع و تبدیل داده است که در جابه‌جایی داده از یک زمینه به زمینه دیگر لازم می‌شود.
- **"داده‌های مرجع و اصلی (Master)"** درباره مدیریت داده‌های اصلی (بحرانی) است برای اطمینان از دسترس‌پذیری، دقت، امنیت، قابلیت اعتماد.
- **"مدیریت مستندات و محتوا"** درباره فرایند مدیریت داده‌های غیرساخت‌یافته- از ایجاد و ذخیره‌سازی تا جست‌وجو و آرشیو- است.
- **"مدیریت فراداده"** شامل همه اقدامات لازم برای مدیریت اطلاعات درباره داده‌هاست. فراداده زمینه‌ساز فهم و استفاده مناسب داده، قابلیت یکپارچگی و امنیت داده است.
- **"کیفیت داده"** عبارت از همه اقدامات برای مدیریت کیفیت داده است به طوری که اطمینان حاصل شود داده قابل استفاده است.
- **"انباره‌داده"** همان مخزن متمرکز و یکپارچه داده های حوزه های مختلف سازمان است.
- **"هوش تجاری BI "** عبارت از همه اقدامات لازم برای فراهم‌سازی بینش کسب‌وکاری و پشتیبانی تصمیم‌گیری است.

استانداردها و چارچوب‌های حاکمیت-مدیریت داده به سرعت جای خود را در دولت‌ها و نظام‌های دستگاه‌های اجرایی نیز باز کرده است. بر اساس قانون Evidence Act که در کنگره آمریکا و با هدایت پال رایان -(Paul Ryan) رئیس وقت مجلس نمایندگان آمریکا- در سال ۲۰۱۷ تهیه و نهایتاً سال ۲۰۱۹ تصویب شد، تکالیف مشخصی برای راه‌اندازی نظام حاکمیت داده در سازمان‌های فدرال تعیین شده است. دستگاه‌ها موظف شده‌اند مخزنی از دارایی‌های داده‌ای ایجاد کنند، استانداردهای داده را جاری کرده، دقت تصمیم‌گیری‌های مدیریتی را ارتقا بخشیده و شرایط به‌اشتراک‌گذاری داده **(داده باز)** را در سطح دولت فدرال تسهیل کنند.

همچنین نقش **مدیر ارشد داده (CDO)** نیز در مقام هدایت‌کننده موضوع و مسئول ظرفیت‌سازی و توانمندسازی مهارت‌های مورد نیاز داده در سازمان تعریف شده است. بنا بر تحقیقی که از مدیران ارشد داده (CDO) در دستگاه‌های اجرایی دولت آمریکا انجام شده است، بیشترین تعامل این نقش در درجه اول با مدیر ارشد اطلاعاتی (CIO)و در درجه دوم با مدیران عملیاتی و ریاست سازمان است. از جمله وظایف نقش CDO می‌توان به موارد ذیل اشاره کرد:

- احصا و مستندسازی نیازمندی‌های اولویت‌دار مرتبط با داده در سازمان
- راه‌اندازی ساختار و تشکیلات برای حاکمیت داده
- ارزیابی وضعیت جاری بلوغ داده و زیرساخت‌های داده‌ای
- شناسایی فرصت‌های توسعه مهارت‌ها و سواد داده‌ای پرسنل
- شناسایی موجودیت‌های داده‌ای کلیدی (اولویت‌دار) برای اشتراک‌گذاری
- انتشار و به‌روزرسانی مخزن موجودیت‌های داده‌ای سازمان
- تاسیس و شروع‌به‌کار شورای حاکمیت داده در سازمان
- تعیین انواع نقش‌ها و مسئولیت‌های مرتبط با داده

- تعیین داده‌های نیازمند همگام‌سازی (سینک یا آسینک)
- تعیین روش و فرمت نام‌گذاری موجودیت‌های داده‌ای
- انتخاب و تهیه کاتالوگ استانداردهای داده در سازمان
- تعیین سیاست‌های ناظر بر حفاظت از داده‌ها
- ایجاد چارچوب اصول حاکم بر مدیریت داده
- تامین و راه‌اندازی ابزارهای جمع‌آوری، تصفیه، نگهداشت و انتشار داده
- تهیه شاخص‌های ارزیابی کیفیت داده و اندازه‌گیری دوره‌ای شاخص‌ها.

بر اساس ادعای مایکروسافت، ۷۵ درصد از مدیران ارشد داده در آمریکا از پیشرفت قابل توجه در راه‌اندازی نظام حاکمیت داده و ایجاد مخزن دارایی‌های داده‌ای در راستای پشتیبانی از نیازمندی‌های سازمان خبر داده‌اند، اگرچه چالش‌ها و موانع نیز کم نبوده است؛ برای مثال عدم مشارکت کافی کارکنان در ایفای نقش‌های مرتبط با مدیریت داده و نیز محدودیت‌های بودجه از جمله موانع مهم در این راستا برشمرده شده است.

طبق گزارش سازمان همکاری و توسعه اقتصادی(OECD)، برنامه‌ریزی راهبردی در سطح ملی (دولتی) برای استقرار موثر نظام‌های حاکمیت-مدیریت داده به کشور آمریکا محدود نشده و در سایر کشورهای پیشرو از جمله کانادا، ژاپن، آلمان، ایرلند، فرانسه، کره‌جنوبی، ایتالیا، انگلیس، هلند و دانمارک نیز فعالیت‌های مشابهی دیده شده که البته مبتنی بر الگوبرداری از تجارب دولت آمریکا بوده است.

حاکمیت داده ، نظامی برای اطمینان بخشی کیفیت داده ها

در دنیای امروز داده‌ها و اطلاعات به عنوان ثروت سازمانی محسوب گشته و همواره شرکت‌ها و سازمان‌های بزرگ و موفق دنیا به دنبال استفاده مناسب‌تر و تجاری‌تر از این منابع مجازی می‌باشند. از سوی دیگر با پیچیده شدن محیط‌های کسب و کار، ماهیت و حجم داده‌های سازمانی بسیار متفاوت شده و نگاه یکپارچه و مدیریتی به آنها ضروری می‌گردد. یکی از راه‌حل‌هایی که اخیراً در این زمینه اتخاذ شده است، ایجاد یک نظام حاکمیت داده (Data Governance) می‌باشد. حاکمیت داده، سیستمی از اختیارات و مسئولیت‌های تصمیم‌گیری برای فرآیندهای مرتبط با اطلاعات است که براساس مدل‌های از پیش تعیین شده اجرا می‌شود و تعیین می‌کند که چه فردی می‌تواند چه اقدامی را بر روی چه اطلاعاتی، در چه زمانی و تحت چه شرایط و به چه روشی انجام دهد.

همانطورکه وجود داده‌ها و اطلاعات به هنگام، صحیح، دقیق و قابل اطمینان نقش حیاتی در بقا و توسعه یک سازمان ایفا می‌نماید، عدم وجود ویژگی‌های فوق می‌تواند تاثیرات و عوارض نامطلوبی به همراه داشته باشد زیرا هرگاه مبنای تصمیم‌گیری، آمار و اطلاعات نارسا و غلط باشد در این صورت جز تصمیم‌گیری نادرست نتیجه‌ای دیگر عاید نخواهد شد و به میزان اهمیت این تصمیم‌گیری، میزان پیامد سوء این آمار و اطلاعات غلط شدت پیدا می‌کند. مدیران بخش‌های مختلف سازمان برای تصمیمات به موقع و صحیح نیازمند در اختیار داشتن اطلاعات تولید شده می‌باشند که این مستلزم اطلاعات دقیق و صحیح و یکپارچه می‌باشد .برای مدیریت ثمربخش فرآیند تولید و توزیع اطلاعات جهت تهیه آمارهای لازم، طراحی و استقرار نظام حاکمیت داده‌ها با هدف ایجاد بستر اطلاعاتی جهت فراهم نمودن آمار و اطلاعات صحیح و به موقع که پیش‌نیاز تصمیم‌گیری و برنامه‌ریزی است باید صورت گیرد. اشکالات متعدد در داده‌ها و اقلام اطلاعاتی موجود در سیستم‌های موجود موجب می‌شود که مدیران، اطلاعات و گزارش‌های مورد نیاز را نتوانند به راحتی بدست آورند.

اصول اساسی حاکمیت داده‌ای، به زعم موسسه حاکمیت داده، به شرح زیر می‌باشند. این اصول، در تمامی پروژه‌های حاکمیت داده برقرار بوده‌اند:

- **یکپارچگی :**افراد مختلف درگیر در جاری‌سازی حاکمیت داده در یک سازمان، باید در تعاملات با یکدیگر به اصل یکپارچگی پایبند بوده و در رابطه با محرک‌ها، محدودیت‌ها، گزینه‌ها و تاثیرات تصمیم‌های مرتبط با داده، صادق و متعهد باشند.
- **شفافیت :**برای افراد مختلف درگیر در جاری‌سازی حاکمیت داده در یک سازمان و ممیزین آن، شفاف‌سازی شده باشد که تصمیم‌های مرتبط با داده‌ها و کنترل‌های موردنیاز، چه زمانی و چگونه در فرایندها وارد می‌شوند.
- **قابلیت کنترل و ممیزی :**تصمیم‌ها، فرآیندها و کنترل‌های مرتبط با حاکمیت داده‌ها، باید قابل ممیزی و نظارت باشند .بدین‌منظور، باید مستندسازی لازم انجام پذیرد.
- **مسئولیت‌پذیری :**حاکمیت داده‌ای، مسئولیت کنترل‌ها، فرایندها و تصمیم‌های میان کارکردی مرتبط با داده را تعریف و تدوین می‌نماید.
- **حاکمان داده‌ای :**مسئولیت نظارت بر داده‌ها در حاکمیت داده‌ای باید به یک فرد یا مجموعه واگذار شود که حاکمان داده‌ای نامیده می‌شوند.
- **کنترل و ایجاد تعادل :**حاکمیت داده‌ای، مسئولیت‌ها را در میان افراد مختلف به‌گونه‌ای تعریف می‌نماید که تعادل مناسبی بین تیم‌های فنی و کسب و کار و همچنین بین افراد مختلفی که درگیر ایجاد و جمع‌آوری داده، مدیریت داده، استفاده‌کنندگان از داده و کسانی که استانداردها و نیازمندی‌های جدید را تبیین می‌کنند، برقرار شود.

- **استانداردسازی :**حاکمیت داده‌ای، به معرفی و پشتیبانی از استانداردسازی داده‌های سازمانی می‌پردازد.
- **مدیریت تغییر :**حاکمیت داده‌ای، از فعالیت‌های مدیریت تغییر کنشی و واکنشی برای مقادیر داده‌ای مرجع و ساختار داده‌های مبنا و فراداده پشتیبانی می‌نماید.

حاکمیت داده، روی ۶ حوزه اصلی تمرکز دارد که عبارت از موارد زیر می‌باشند:

۱) **راهبرد، استانداردها و خط مشی :**در این حوزه، راهبرد، استانداردها و خط مشی داده‌ای سازمان تبیین شده، به تصویب می‌رسد و در نهایت نظارت لازم جهت تحقق آن انجام می‌پذیرد.

۲) **کیفیت داده :**به مباحثی همچون کیفیت، تمامیت و قابلیت استفاده داده‌های سازمان اشاره دارد. کار ارزیابی کیفیت داده‌ها را می‌توان از یک برنامه کاربردی یا از یک واحد سازمانی شروع نمود. در این حوزه، جهت‌گیری اصلی برای کار تعیین می‌شود و داده‌ها مورد نظارت کیفی قرار می‌گیرند. همچنین در این حوزه، وضعیت برنامه‌ها و اقدامات مرتبط با کیفیت داده‌ها ارزیابی شده و ذینفعان، اختیارات تصمیم‌گیری و مسئولیت‌های افراد تعیین می‌گردد.

۳) **امنیت و حریم خصوصی :**مباحثی همچون امنیت، حریم خصوصی، مدیریت دسترسی‌ها، انطباق با قوانین و مقررات و نیازمندی‌های سازمانی و قراردادی و غیره را مورد توجه قرار می‌دهد. ممکن است تمامی انواع داده‌های موجود در یک سازمان در این حوزه مورد توجه نباشند و داده‌های خاص که نیازمند تامل بیشتری هستند، مورد نظر باشند. برای این دسته از داده‌ها، باید ملاحظات لازم در نظر گرفته شده و روش‌ها و فنون لازم برای محافظت از داده‌ها اعمال شود.

۴) **معماری و یکپارچگی :**این حوزه، زمانی مورد توجه است که سیستم جدیدی وارد سازمان شده باشد که نیاز به برقراری ارتباط و یکپارچه‌سازی با سیستم‌های موجود داشته باشد. در این حوزه، باید از تعریف داده‌های سازگار اطمینان حاصل شده و از استانداردها و خط‌مشی‌های امنیتی، اطمینان حاصل شود.

۵) **انباره داده و هوش کسب وکار :**تمرکز این حوزه بر ایجاد انباره‌های داده و دستیابی به هوش کسب و کار در سازمان است.

۶) **همسویی مدیریت :**این حوزه، زمانی مطرح می‌شود که مدیران در اتخاذ تصمیم‌های مدیریت روتین خود به دلیل تاثیرات احتمالی آن بر سایر حوزه‌های عملیاتی سازمان دچار مشکل می‌شوند. ممکن است مدیران بخواهند با دعوت از سایر صاحبنظران، به صورت اجماع و گروهی، تصمیم لازم را اتخاذ نمایند، لیکن در مواردی، حتی مشخص نیست که چه افرادی لازم است تا در جلسات مربوطه، دعوت شوند که این حوزه، می‌تواند در این خصوص راهگشا باشد.

چارچوب حاکمیت داده

چارچوب استقرار حاکمیت داده

اگرچه در مراجع و منابع یک چارچوب و روش ثابت برای استقرار نظام حاکمیت داده و طراحی معماری مطلوب داده پیشنهاد نشده است، نگارنده با بررسی نمونه مراجع معتبر آکادمیک و تجاری موجود، چارچوبی مفهومی برای استقرار نظام حاکمیت داده پیشنهاد کرده است.

در این چارچوب مولفه‌های اصلی نظام حاکمیت داده در پنج ناحیه اصلی **راهبرد، نقش‌ها، نظام‌ها، ابزارها و تطابق** سازماندهی شده است. در هر ناحیه نیز عناصر اصلی تعیین شده است: برای مثال در محور ابزارها، چهار نوع ابزار **حاکمیت داده، معماری داده، مدیریت کیفیت داده و تحلیل و داشبورد** پیشنهاد شده است.

سوال کلیدی درباره پیاده‌سازی نظام حاکمیت داده در سازمان‌ها نحوه تلفیق یا جداسازی آن با سایر نظام‌های مرتبط معماری (معماری کسب‌وکار، معماری نرم‌افزار، معماری امنیت و غیره) است و همچنین چگونگی هم‌افزایی با سایر کلان رویکردها و نظام‌های مدیریتی-فناوری مانند سرویس‌گرایی ، تحول دیجیتالی، سازمان چابک و نظایر اینها.

عناصر تشکیل دهنده چارچوب حاکمیت داده چیست؟

برنامه حاکمیت داده از عناصر مختلفی تشکیل می‌شود؛ از جمله **قوانین، سیاست‌ها، فرآیندها، ساختارهای سازمانی و فناوری**. یکی دیگر از مولفه‌های حاکمیت داده ارائه بیانیه ماموریت، اهداف و ارزیابی میزان موفقیت یک برنامه است. در کنار این موارد، درباره مسئولیت‌ها و عملکرد قسمت‌های مختلف برنامه نیز تصمیم‌گیری خواهد شد.

چهارچوب فعالیت حاکمیت داده باید با همه اعضای داخلی سازمان به اشتراک گذاشته شود. در این صورت همه از نحوه کارکرد آن مطلع خواهند شد و از قبل می‌دانند چه روندی قرار است انجام شود.

همانطور که قبلاً اشاره کردیم فناوری نیز در این فرآیند نقش دارد. نرم‌افزارهای حاکمیت داده باعث می‌شوند برنامه مدیریت داده به‌صورت خودکار پیش رود. این ابزارها همیشه در چارچوب حاکمیت داده استفاده نمی‌شوند، اما به پیشرفت این روند کمک بسیاری می‌کنند. برای مثال:

- **از جریان و برنامه مدیریت داده پشتیبانی می‌کنند؛**
- **با عناصر دیگر همکاری و تعامل دارند؛**
- **سیاست‌های پیشرفته‌تری ارائه می‌دهند؛**
- **روند ارائه اسناد را بهبود می‌بخشند؛**
- **به ساخت کاتالوگ داده و سایر عملیات کمک می‌کنند.**

این ابزارها را می‌توان همراه با ابزارهای کیفیت داده، مدیریت فراداده و مستر داده (MDM) نیز استفاده کرد.

راهنمای اجرای حاکمیت داده

اولین قدم برای اجرای حاکمیت داده تشخیص مسئولان هر بخش از سازمان است. در مرحله بعد، مدیر ارشد طراحی، مدیر اجرایی یا مدیر مربوطه دیگر وظایف خود را انجام می‌دهد. او باید ساختار برنامه را طراحی، گروه حاکمیت داده را هدایت، مسئولان داده را تعیین کند و کمیته حاکمیت را تشکیل دهد. وقتی ساختار اصلی شکل گرفت، تازه اصل ماجرا شروع می‌شود. در این مرحله استانداردها، سیاست‌ها و قوانین استفاده از داده‌ها تنظیم می‌شوند.

همچنین باید بخشی برای نظارت و رسیدگی به انجام کارها طبق سیاست‌های داخلی و خارجی به وجود آید. این بخش ضمانتی برای قرار گرفتن حاکمیت در مسیر درست رسیدن به اهداف تجاری است. علاوه بر این گروه حاکمیت روی رسیدگی به اسناد مربوط به داده‌ها، منابع، محل ذخیره و تامین امنیت آن‌ها تمرکز دارد.

ممکن است اقدامات دیگری نیز در حاکمیت داده صورت گیرد. این اقدامات به شرح زیر هستند:

طراحی مسیر و سازمان بندی داده

طراحی مسیر برای داده به ثبت و سازمان‌بندی آن‌ها کمک می‌کند. با این کار می‌توان داده‌های مختلف را بر اساس عوامل مختلف دسته‌بندی کرد؛ برای مثال، آن‌ها را در گروه‌های اطلاعات شخصی یا داده‌های مهم و حساس قرار داد. مرتب کردن داده‌ها بستگی به نحوه رعایت سیاست‌ها دارد.

واژه نامه تجاری

واژه‌نامه تجاری فهرستی است که شامل اصطلاحات، مفاهیم تجاری و تعاریف آن‌ها می‌شود . به‌طور کلی همه عنصرهای موجود در سازمان در این واژه‌نامه توضیح داده می‌شود. برای مثال، تعریف مشتری فعال چیست؟

به طورکلی جمع‌آوری کلمات و اصطلاحات رایج در یک واژه نامه مشخص به پیشبرد بهتر اهداف تجاری سازمان شما کمک می‌کند.

کاتالوگ داده

کاتالوگ داده شامل فراداده‌های جمع‌آوری شده از سیستم‌های مختلف است. با استفاده از این داده‌ها می‌توان فهرستی مرتب از موجودی و منابع داده‌ها تهیه کرد . فهرست گفته شده پیشینه داده، عملکرد آن در جستجو و ابزارهای تعاملی داده را نشان می‌دهد. در کاتالوگ اطلاعات دیگری را هم نمایش داده می‌شود، از جمله سیاست‌های حاکمیت داده و مکانیسم‌های خودکار.

بهترین روش برای مدیریت عناصر مختلف حاکمیت داده

گاهی حاکمیت داده در مورد نحوه به‌کارگیری و مدیریت داده سختگیری می‌کند. این موضوع برای سازمان‌ها چالش برانگیز است. تحلیلگران داده و کاربران تجاری باید درباره حاکمیت داده آموزش ببینند. آموزش عنصری جدانشدنی در این فرآیند است. این افراد با قوانین استفاده از داده، مقررات مربوط به حفظ حریم خصوصی و مسئولیت‌های خود آشنا می‌شوند.

از مولفه‌های ضروری دیگر در مدیریت عناصر حاکمیت داده می‌توان به برقراری ارتباط مداوم با مدیران اجرایی، مدیران تجاری و مصرف کنندگان سازمان اشاره کرد. حفظ این رابطه برای پیشبرد برنامه حاکمیت داده واجب است و از طریق ارائه گزارش، ایمیل، خبرنامه، کارگاه گروهی و غیره انجام می‌شود.

پس بهترین روش برای مدیریت عناصر حاکمیت داده چیست؟ طبق گزارش‌های منتشر شده بهترین روش برای مدیریت عناصر مختلف حاکمیت داده باید ویژگی‌های زیر را داشته باشد:

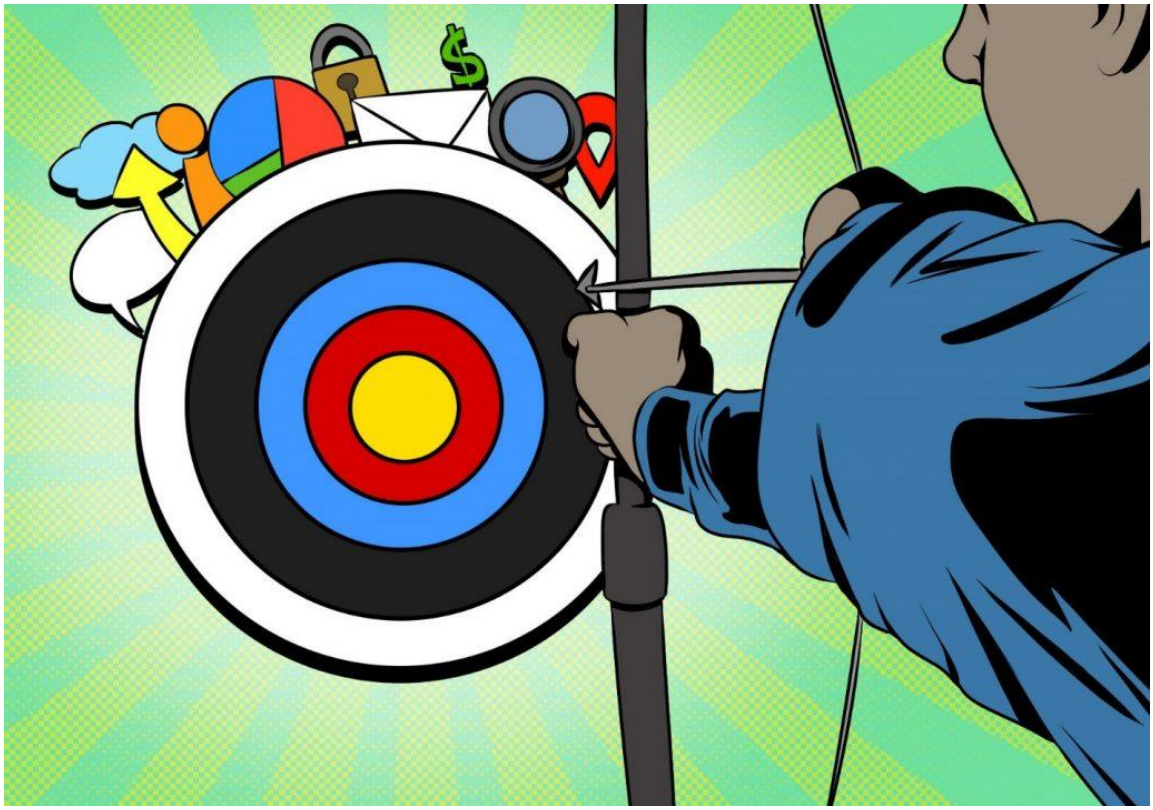
- متمرکز بودن بر ارزش تجاری و نتایج سازمانی؛
- اعتمادسازی و تنظیم حاکمیت داده‌ها بر اساس منابع و مرتب‌سازی داده‌ها؛
- توافق جمعی بر مورد اعتماد بودن داده‌ها و گرفتن تصمیم درست؛
- تصمیم‌گیری شفاف بر مبنای اصول اخلاقی؛
- آموزش پیوسته، ارزیابی و نظارت بر تاثیرگذاری روش‌های آموزشی؛
- مدیریت ریسک و محافظت از امنیت داده‌ها؛
- جا انداختن فرهنگ کار گروهی و همکاری موثر.

اکثر سازمان‌های موفق این موارد را در استراتژی حاکمیت داده خود به کار می‌گیرند.

چالش‌های حاکمیت داده چیست؟

ضرورت حاکمیت داده بر هیچ‌کسی پوشیده نمی‌باشد، اما با این‌وجود بسیاری از سازمان‌ها با وجود مزایای بسیار زیاد آن به خاطر پیچیدگی یا عدم اطمینان از موفقیت برنامه Data Governance اجرا نمی‌کنند.

معمولاً برداشتن اولین قدم‌ها برای اجرای حاکمیت داده با سختی بیشتری همراه است. وجود اختلاف نظر درباره ماهیت داده‌های سازمانی، این روند را سخت‌تر می‌کند. اختلاف نظرها درباره هر چیزی می‌توانند باشند، از جمله محصولات یا مشتریان. بنابراین این مشکلات باید در فرآیند حاکمیت داده حل شوند. مدیران و مسئولان صاحب نظر باید برسر ساختار به توافق برسند؛ این تصمیم‌گیری مسئولیت بزرگی دارد. به همین دلیل کمیته حاکمیت داده به دستورالعمل مشخصی برای حل اختلاف نیاز دارد.



اجرای برنامه Data Governance کاری ساده نمی‌باشد و چالش‌های مهم زیر پیش روی اجرای برنامه حاکمیت داده قرار دارد:

سازماندهی حاکمیت داده

اجرای حاکمیت داده نیازمند فرهنگ‌سازمانی باز است. برای مثال باید بررسی گردد که قوانین Data Governance قابلیت اجرایی در سازمان دارند، حتی اگر این قوانین نیازمند به ایجاد برخی نقش‌ها و تعیین مسئولیت‌ها باشند. درنهایت حاکمیت داده منجر به سیاست‌گذاری جدیدی خواهد شد. این امر نیاز به رویکردی بسیار حساس و دقیق خواهد داشت.

پذیرش و رابطه برقرار کردن با حاکمیت داده

Data Governance نیازمند پذیرش و ارتباط کاری مناسب بین تمامی افراد درگیر با این فرآیند می‌باشد. باید افراد مناسب در مکان مناسب برای اجرای سیاست‌ها با همدیگر در تعامل باشند. مخصوصاً مدیران پروژه نیاز به درک درستی از جنبه‌های فنی و همچنین شناخت کسب‌وکار و ترجیحاً دیدگاه مفهومی فراگیری از سازمان می‌باشند.

بودجه‌بندی و ذینفعان حاکمیت داده

متقاعد کردن ذینفعان سازمان مبنی بر نیاز به برنامه Data Governance و گرفتن بودجه برای اجرای برنامه غالباً دشوار می‌باشد. علاوه بر این تغییر در سازمان دلخواه نمی‌باشد و در برابر آن مقاومت وجود خواهد داشت.

انعطاف‌پذیری و استانداردسازی

کسب‌وکارها برای رفع نیازهایی که خیلی سریع در حال تغییر هستند، باید انعطاف‌پذیر باشند. ایجاد تعادل مناسب بین انعطاف‌پذیری و استانداردهای Data Governance با توجه به نیازمندی‌های کسب‌وکاری از اهمیت حیاتی برخوردار است.

درست نشان دادن ارزش تجاری سازمان

ارائه نمونه اولیه فرآیند حاکمیت داده بسیار چالش برانگیز است. برای ایجاد بهترین وجهه تجاری برای برنامه حاکمیت داده خود باید داده‌ها را مرتب ثبت کنید و اهداف تجاری مورد نظرتان را اولویت‌بندی کنید. در عین حال برای درست نشان دادن ارزش تجاری سازمان خود باید از معیارهای قابل ارزیابی استفاده کنید. این معیارها بیشتر از همه در سنجش میزان پیشرفت کیفیت داده کاربرد دارد. برای مثال، داده‌های مربوط به خطاهای حل شده در هر فصل و درآمدی که با رفع آن خطاها در جیب سازمان می‌ماند. معیارهای دیگری هم برای ارزیابی وجود دارند. این معیارها میزان کیفیت، دقت، نرخ خطاها، کامل بودن و انسجام داده‌ها را ارزیابی می‌کنند.

حمایت از تجزیه و تحلیل خودکفا

هوش تجاری و تجزیه و تحلیل خودکفا برای سازمان‌ها مشکلات تازه‌ای به وجود می‌آورد، زیرا با وجود این فناوری‌ها داده‌های بیشتری در اختیار مصرف‌کنندگان سازمان‌ها قرار می‌گیرد.

فرآیند حاکمیت تضمین می‌کند که داده‌ها به‌صورت دقیق در دسترس مصرف‌کنندگان خودکفا قرار گیرند. مصرف کنندگان شامل تحلیلگران تجاری، مدیران اجرایی یا محققان داده می‌شوند که نباید از داده‌ها سوء استفاده کرده و قوانین حفاظت از حریم شخصی و امنیت داده‌ها را نقض کنند.

نظارت بر کلان داده

راه‌اندازی سیستم کلان داده مشکلات و ملزومات بیشتری برای حاکمیت داده به دنبال دارد. در گذشته برنامه حاکمیت داده بیشتر روی ذخیره مرتب داده‌ها در پایگاه داده مربوطه متمرکز بود . امروزه ساختارهای مختلفی در این حوزه درگیر هستند.

کلان داده امروزی بیشتر شامل ترکیبی از داده‌های سازمان یافته، سازمان نیافته و نیمه سازمان یافته می‌شود. علاوه بر داده‌های گفته شده این سیستم به روز شده دارای انواع مختلفی از پلتفرم‌های داده، مثل سیستم‌های Hadoop، Sparkو NoSQL و فضاهای ابری برای ذخیره داده هستند. مجموعه کلان داده‌ها معمولاً ابتدا به‌صورت خام ذخیره می‌شوند و سپس از فیلترهای مختلف عبور می‌کنند و تحلیل خواهند شد.

حاکمیت داده دریچه‌ای به‌سوی مدیریت داده بهتر

استراتژی مدیریت داده بدون وجود حاکمیت داده آن طور که باید عمل نمی‌کند. برای پیشبرد اهداف سازمانی و تجاری به بهترین نحو باید از حاکمیت داده کمک گرفت.

در این مقاله یاد گرفتیم فرآیند مدیریتی حاکمیت داده چیست و چرا تا این حد در سازمان‌ها اهمیت دارد. تنها با وجود حاکمیت داده می‌توان تجارتی باارزش طبق مقررات و قوانین تصویب‌شده شکل داد. بی‌شک بدون وجود این مقررات نظم هر سازمان برهم می‌ریزد. حاکمیت داده دریچه‌ای به‌سوی ارائه داده‌های باکیفیت‌تر و امن‌تر است.

معرفی ارکان اصلی حاکمیت داده

حاکمیت داده بر اساس بسیاری از جنبه‌های اصلی فرآیند مدیریت داده شکل می‌گیرد. ارکان حاکمیت داده عبارتند از:

- نظارت داده و حفاظت از آن
- کیفیت داده
- مدیریت داده‌های مستر(Master data)
- بررسی کاربرد حاکمیت داده

تجربه‌های مفید و عوامل موفقیت حاکمیت داده

پیاده‌سازی حاکمیت داده خلاق

حاکمیت داده پروژه‌ای نیست که به‌صورت انفجار بزرگ big bang در سازمان پیاده‌سازی گردد. می‌توان گفت Data Governance طرحی جامع در سازمان است که پروژه‌های پیچیده و طولانی‌مدت را باید اجرا نماید. بنابراین همیشه این خطر وجود دارد که ممکن است افراد درگیر باگذشت زمان انگیزه و علاقه خود را از دست بدهند. با توجه به موارد مذکور توصیه می‌شود که پروژه‌ی نمونه اولیه قابل‌کنترل یا کاربردی شروع شود و این رویکرد به‌صورت متناوب ادامه یابد. در این روش، پروژه قابل‌کنترل بوده و می‌توان از تجربه‌های به‌دست‌آمده برای پروژه‌های پیچیده‌تر و گسترش حاکمیت داده در سازمان استفاده کرد.

به‌صورت معمول مراحل پروژه Data Governance عبارت‌اند از:

- تعریف اهداف و درک مزایای حاکمیت داده
- تجزیه‌وتحلیل وضعیت فعلی
- تهیه نقشه راه
- توجیه ذینفعان و بودجه پروژه

- برنامه‌ریزی و تدوین حاکمیت داده
- اجرای برنامه حاکمیت داده
- نظارت و کنترل بر حاکمیت داده

این مراحل باید برای هر برنامه جدید تکرار شوند، همچنین اگر در برنامه قبلی تغییراتی ایجاد گردید نیز باید مراحل مذکور تکرار شوند.

ادامه پیاده‌سازی حاکمیت داده خلاق

قبل از آغاز برنامه حاکمیت داده در سازمان، باید پرسش‌های مربوط به دلایل اجرای Data Governance پاسخ داده شود تا از کارهای غیرضروری و اضافی اجتناب گردد. به همین ترتیب فرآیندهای موجود باید مورد ارزیابی قرار گیرند تا مشخص گردد آیا می‌توان در چارچوب برنامهData Governance ، بجای توسعه غیرضروری فرآیندهای جدید، فرآیندهای موجود را با نیازهای جدید سازگار کرد.

سطوح استراتژیک، تاکتیکی و عملیاتی شرکت، همچنین جنبه‌های سازمانی، کسب‌وکاری و فنی آن پایه و اساس ماتریس شرکت در حاکمیت داده را تشکیل می‌دهند. با این ساختار می‌توان پروژه‌های Data Governance را با مشخصات، فرآیندها، نقش‌ها و وظایف مربوط به هرکدام تعریف کرد. شایان‌ذکر است که سطوح مختلف طرح حاکمیت داده و جنبه‌های ذکرشده و نقش حاکمیت داده در شرکت باید بسیار خاص باشد.

با توجه به مطالب ارائه‌شده برای پیاده‌سازیData Governance ، می‌توان وضعیت فعلی را با وضعیت مورد انتظار به‌صورت ساختاری مقایسه و بررسی نمود .با انجام این کارها می‌توان نقطه ورود به بحث را مشخص نمود، اولویت‌ها را تعیین و نقشه راه را با توجه به اقدامات اصلی طراحی نمود.

نقش‌ها در حاکمیت داده

نقش‌ها در برنامه Data Governance ضروری هستند. امروزه ابزارهایی نرم‌افزاری وجود دارد که الگوهای Data Governanceرا برای مدیریت متا داده (فراداده)، کیفیت داده، مدیریت داده‌های اصلی و یکپارچگی داده‌ها ارائه خواهند داد.

نظریه‌ها درباره نقش‌ها در Data Governance اندکی باهم اختلاف دارند، اما اصلی‌ترین نقش‌های ذکرشده در حاکمیت داده به شرح زیر هستند:

- کمیته استراتژیک حاکمیت داده (کمیته رهبری در سطح استراتژیک)
- کمیته مدیریت حاکمیت داده (سطح تاکتیکی)
- مدیر داده
- مالک داده
- ناظر داده
- کاربران داده

توصیه‌های اجرای حاکمیت داده

نکات زیر در پیاده‌سازی برنامه Data Governance کمک خواهند کرد.

- برنامه Data Governance به‌هیچ‌عنوان بدون حمایت مدیریت ارشد در سازمان اجرا نگردد.
- کار به‌صورت انفجار بزرگ آغاز نشود. Data Governance به‌عنوان روندی مداوم و تکراری است که حاوی پروژه‌های فرعی می‌باشد.
- کار را با پروژه‌های آزمایشی کوچک آغاز کنید و تجربه این موارد را به سراسر شرکت بسط دهید.
- برنامه‌های Data Governance می‌توانند سالیان متمادی اجرا شوند. بااین‌وجود پروژه‌های زیربط نباید بیش از سه ماه طول بکشند.
- اهداف را شفاف و با دقت تنظیم نمایید.
- موفقیت Data Governance در اولویت قرار دارد. دخیل بودن ذینفعان و شفافیت روند اجرای کار مهم است. توصیه می‌شود ارتباطی آزاد و شفاف و به‌دوراز پنهان‌کاری با کلیه ذینفعان برقرار گردد.
- چرخه را مجدداً ابداع نکنید بلکه سعی کنید از الگوهای موجود استفاده نمایید. مدل‌ها و تجربه‌های مفیدی که از قبل در بازار موجود هستند می‌توانند بسیار مفید باشند. ابزارهای نرم‌افزاری، چارچوب‌ها و کتابخانه‌ها یا مشاورین می‌توانند کمک شایانی به سازمان بدهند.
- نقش‌های Data Governance در سازمان را با در نظر گرفتن مسائل سیاسی و حساسیت‌ها به‌درستی مشخص نمایید. مهارت‌های ارتباطی مدیر Data Governance خیلی مهم است.
- فرآیندها و راه‌حل‌های تعیین‌شده را با دقت بررسی کنید و مشخص نمایید چرا به‌اندازه کافی ساده و روشن نیستند.
- بررسی بستر و زمینه‌هایData Governance
- ایجاد ساختارها و مسئولیت‌های واضح و روشن.
- ایجاد روش کامل و خوب برای مستندسازی تجربه‌های مفید سازمان.

ابزارهای زیر در اجرای یک برنامه حاکم بر داده ها کمک می‌کنند:

❖ چارچوب مدیریت داده(DAMA)

چارچوب DAMA زمینه ای جهت شناسایی رشته‌ها و گروه‌های عملکردی را فراهم می‌کند.

❖ ماتریس ۹ زمینه‌ایBARC

BARC ۹-Field Matrix برای تعیین وضعیت فعلی رویکرد سازمان در **مدیریت داده ها** و **استخراج نقشه راه** از آن طراحی شده است.

سه سطح شرکت (استراتژیک، تاکتیکی و عملیاتی) و جنبه‌های سازمانی، تجاری و فنی آن اساس ماتریس را تشکیل می‌دهند . **پروژه‌های مدیریت داده** با ساختار خود می‌توانند با مشخصات موضوعات، فرایندها، نقش‌ها و وظایف مربوطه تهیه شوند.

لازم به ذکر است که پیش بینی سطوح، جنبه‌های سازمانی، تجاری و فنی و همچنین نقش‌های موجود در شرکت باید بسیار مشخص باشد. با این وجود ماتریس برای هر مبحثی در زمینه مدیریت داده مناسب است.

چارچوب DAMA کلیه مباحث مربوط به مدیریت داده را با معیارهای مستند فراهم می‌کند و آنها به یک رشته در ماتریس ۹ زمینه‌ای BARC اختصاص داده می‌شوند. به این ترتیب می‌توان حالت فعلی برای هر زمینه را به صورت ساختاری در برابر حالت هدف مقایسه کرد. با این کار می‌توان دلتا را شناسایی کرد، اولویت‌ها را تعیین کرد و نقشه راهی با اقدامات مشخص بدست آورد.

❖ سیستم عامل های حاکمیت داده

سیستم عامل‌های حاکمیت داده بلوک‌های مختلف عملکردی را برای کیفیت داده ها، مدیریت اصلی داده ها، ادغام داده ها، مدیریت فراداده و حفاظت از داده ها ارائه می‌دهند.

سؤالات اساسی از حکمرانی داده

در این بخش قصد داریم سؤالات اساسی WHO-WHAT-WHEN-WHERE-WHOW-HOW را در مورد حکمرانی داده پاسخ دهیم. در ادامه پاسخ‌های کوتاهی به هر سؤال حکمرانی داده خواهیم داد:

چه کسی با حکمرانی داده درگیر است؟

حکمرانی داده یکی از نگرانی‌های هر فرد یا گروهی است که علاقه‌مند به چگونگی ایجاد، جمع‌آوری، پردازش، دستکاری، ذخیره‌سازی و در دسترس قرار دادن داده‌ها برای استفاده یا از بین بردن است. ما چنین افرادی را سهامداران داده می‌نامیم. غالباً، سهامداران داده با صدور مجوز به گروه‌های مدیریت فناوری اطلاعات و همچنین مدیریت داده، مشخص می‌کنند که آنها پیرامون وظایفی که در بالا ذکر کرده‌ایم، تصمیم بگیرند. اما گاهی اوقات، این فعالیت‌ها نیاز به تصمیماتی دارند که باید توسط گروه‌های ذینفع و آن‌هم طبق یک فرایند توافق شده اتخاذ شوند. این‌زمانی است که Data Governance وارد عمل می‌شود. در چارچوب حکمرانی داده‌های DGI، دفتر حکمرانی داده (DGO) وجود دارد که در آن مدل نقش‌ها و مسئولیت‌های Data Steward در سراسر سازمان توصیف می‌شود.

حکمرانی داده‌ها به چه معناست و چه‌کاری انجام می‌دهد؟

حکمرانی داده‌ها به معنای **"اعمال تصمیم‌گیری و اختیارات درخصوص امور مربوط به داده‌ها"** است.

به‌طور دقیق‌تر، حکمرانی داده "سیستمی مرکب از حقوق تصمیم‌گیری و پاسخگویی برای فرایندهای مربوط به اطلاعات است که بر اساس مدل‌های توافق شده‌ای اجرا می‌شود. حکمرانی داده‌ها توصیف می‌کند چه کسی می‌تواند با چه اطلاعاتی چه اقداماتی را انجام دهد و چه زمانی، در چه شرایطی از چه روش‌هایی استفاده می‌کند".

برنامه‌های حکمرانی داده‌ها بسته به تمرکز (بر انطباق، ادغام داده‌ها، مدیریت کارشناسی ارشد داده و غیره) می‌توانند تفاوت‌های چشمگیری داشته باشند، صرف‌نظر از این مسائل، بااین‌حال، هر برنامه اساساً همان مأموریت سه بخشی را دارد:

۱) برای ایجاد / جمع‌آوری / تنظیم کردن قوانین،
۲) برای حل مسائل،
۳) نظارت و اجرای انطباق ضمن پشتیبانی مداوم از سهامداران داده.

چه زمانی سازمان‌ها به حکمرانی رسمی داده‌ها نیاز دارند؟

سازمان‌ها زمانی که دریکی از چهار موقعیت ذیل قرار می‌گیرند، باید از حکمرانی غیررسمی داده‌ها به حکمرانی رسمی داده‌ها بروند:

۱) سازمان چنان بزرگ می‌شود که مدیریت سنتی قادر به پرداختن به فعالیت‌های عملکردی مرتبط با داده نیست.
۲) سامانه‌های داده سازمان چنان پیچیده می‌شوند که مدیریت سنتی قادر به پرداختن به فعالیت‌های عملکردی مرتبط با داده نیست.
۳) معماران داده سازمان، گروه‌های SOA یا سایر گروه‌ها، نیاز به پشتیبانی از برنامه‌های کاربردی دارند که نگرانی‌ها و انتخاب‌های داده را در نظر بگیرد.
۴) مقررات، انطباق یا الزامات قراردادی، حکمرانی رسمی داده‌ها را می‌طلبد.

برنامه‌های حکمرانی داده‌ها در کجای سازمان قرار دارد؟

این موضوع متفاوت است. آن را می‌توان در ساختارهای سازمانی عملیات تجاری، فناوری اطلاعات، انطباق / حریم خصوصی یا مدیریت داده قرارداد. آنچه مهم است این است که حکمرانی داده‌ها سطح پشتیبانی مناسب مدیریت ارشد سازمان و درگیری مناسب گروه‌های سهام‌دار داده را دریافت کرده است.

چرا از چارچوب حکمرانی داده‌ها استفاده می‌کنید؟

چارچوب‌ها به ما کمک می‌کنند تا نحوه تفکر و برقراری ارتباط درباره مفاهیم پیچیده یا مبهم را تنظیم کنیم. استفاده از چارچوب رسمی می‌تواند به ذی‌نفعان داده از تجارت، فناوری اطلاعات، مدیریت داده، انطباق و سایر رشته‌ها کمک کند تا به‌وضوح اندیشه و هدف برسند. استفاده از چارچوب می‌تواند به مدیریت و کارکنان کمک کند تا تصمیمات خوبی بگیرند – تصمیماتی که باقی می‌مانند. این موضوع می‌تواند به آنها کمک کند تا در مورد چگونگی "تصمیم‌گیری در مورد تصمیم‌گیری" به‌اتفاق نظر برسند . به‌این‌ترتیب، آنها می‌توانند قوانینی با کارایی بیشتر ایجاد کرده، از پیروی از قوانین اطمینان حاصل کنند و با عدم انطباق‌ها، ابهامات و مسائل برخورد کنند.

چگونه سازمان حکمرانی داده‌ها را "انجام" می‌دهد؟

اول، آنها تصمیم می‌گیرند که چه چیزی برای آنها مهم است – برنامه آنها بر چه چیزی تمرکز خواهد کرد. سپس آنها در بیانیه ارزش برای تلاش‌های خود توافق می‌کنند. این موضوع به ایجاد دامنه و همچنین ایجاد اهداف SMART ، اقدامات موفق و معیارها کمک می‌کند. در مرحله بعد، نقشه راهی برای تلاش‌ها تهیه می‌شود و اعضاء از این طریق برای جلب حمایت سهام‌داران استفاده می‌کنند. پس از دستیابی، برنامه‌ای طراحی می‌شود، برنامه مستقرشده، فرایندهای مربوط به حکمرانی داده‌ها انجام می‌گیرند و در مرحله نهایی فرایندهای مربوط به نظارت، اندازه‌گیری و گزارش وضعیت داده‌ها اجرایی می‌گردند.

برنامه‌های حکمرانی داده‌ها با تمرکز خود بر روی مسائل محدود شروع می‌شوند، سپس دامنه خود را برای رفع نگرانی‌های اضافی یا مجموعه اطلاعات بیشتر گسترش می‌دهد؛ بنابراین، تأسیس حکمرانی داده‌ها تمایل به روندی تکراری دارد.

چقدر به حکمرانی داده نیاز داریم؟

همان مقداری که به شما در رسیدن به اهداف کمک می‌کند. چارچوب حکمرانی داده‌های DGI را می‌توان در برنامه‌های "big-bang" فراگیر استفاده کرد. اما به‌طور خاص برای سازمان‌هایی طراحی‌شده است که قصد دارند حکمرانی را به روشی محدود اعمال کنند و سپس در صورت لزوم مقیاس بندی کنند. تمام ۱۰ مؤلفه حکمرانی داده‌ها که در چارچوب DGI شرح داده‌شده‌اند، در کوچک‌ترین برنامه‌ها و پروژه‌ها وجود دارند. با افزایش تعداد شرکت‌کنندگان یا پیچیدگی سامانه‌های داده، سطح پیچیدگی رشد خواهد کرد.

با استانداردسازی گروه‌های خود در مورد اصطلاحات و مفاهیم توصیف‌شده در چارچوب، شما در حال آموزش کارمندان کسب‌وکار، فناوری اطلاعات و انطباق خود برای برقراری ارتباط با یکدیگر هستید. این امر منجر به درک ارزش دارایی‌های داده سازمان، مدیریت هزینه و پیچیدگی و همچنین اطمینان از انطباق رویکرد "به‌صورت محلی عمل کنید، اما در سطح جهانی فکر کنید" می‌شود.

چگونه ارزیابی می‌کنیم که آیا برای حکمرانی داده آماده هستیم؟

مهم است که قبل از حرکت از وضعیت فعلی خود به رویکرد رسمی‌تر درزمینه حکمرانی و سرپرستی، آمادگی برای Data Governance را ارزیابی کنید. چرا؟ ممکن است دلیل صحیحی وجود داشته باشد که مدل فعلی کفایت می‌کند. به همین ترتیب، ممکن است دلیل موجهی وجود داشته باشد که تغییر می‌تواند برای شرکت، برنامه یا پروژه خاص، یا حتی شغل شخصی فرد مضر باشد.

بیشترین جنبه حکمرانی داده چیست؟

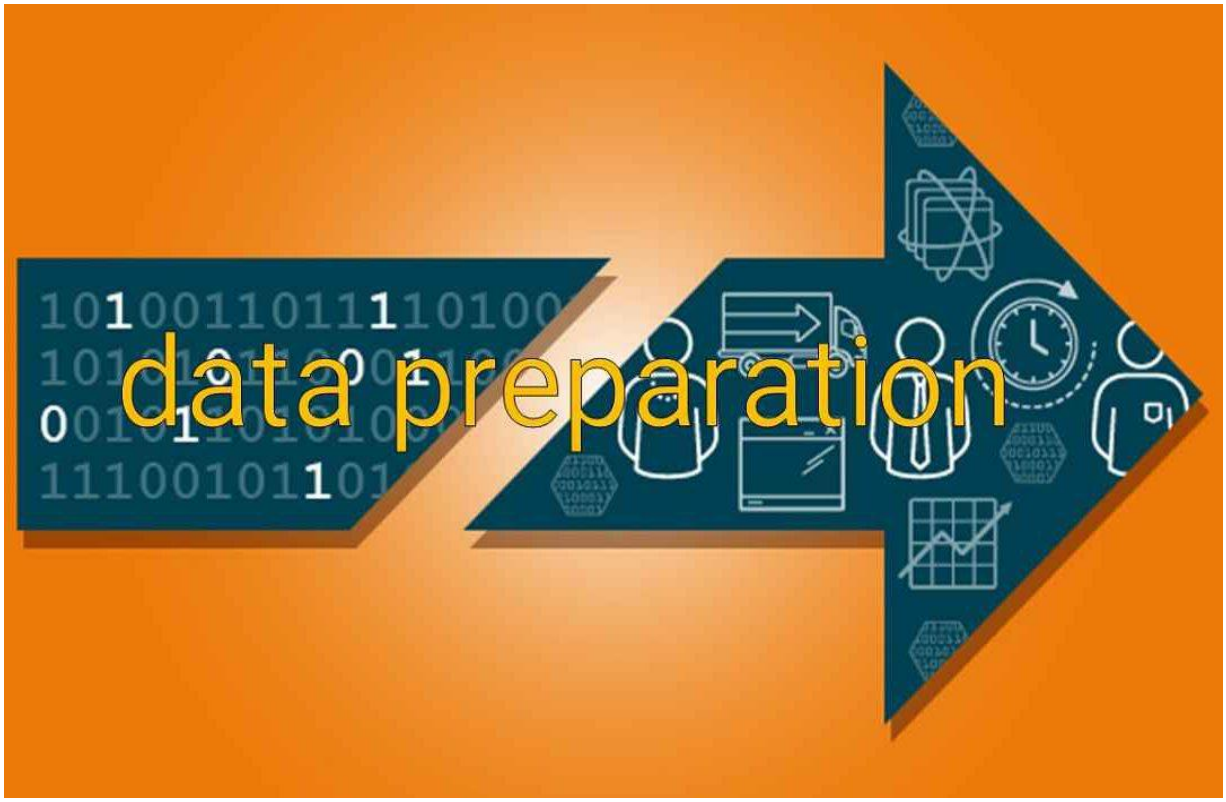
غالباً، کارکنانی که در این سمت‌ها نشسته‌اند زمینه نوشتن یا سخنرانی عمومی ندارند و آنها نیاز به کمک در یادگیری مهارت‌های ارتباطی خاص داده‌ها، ایجاد برنامه‌های ارتباطی و ایجاد قالب‌های متن ایمیل دارند که اطمینان حاصل می‌کند همه ذی‌نفعان سطح مناسب اطلاعات را در زمان و توالی مناسب برای جلوگیری از مسائل سیاسی به‌دست آورده‌اند.

خوشبختانه، آموزش می‌تواند این مهارت‌ها را در اختیار کارکنان باهوش داده قرار دهد. با داشتن توانایی ارتباطی صحیح و آماده، گروه حکمرانی داده سازمان می‌تواند کار مدیریت را انجام دهد و از توانایی خود برای ایجاد همسویی در گروه‌های متنوع ذی‌نفعان داده اطمینان داشته باشد.

data preparation یا آماده سازی داده ها: پالایش داده های خام

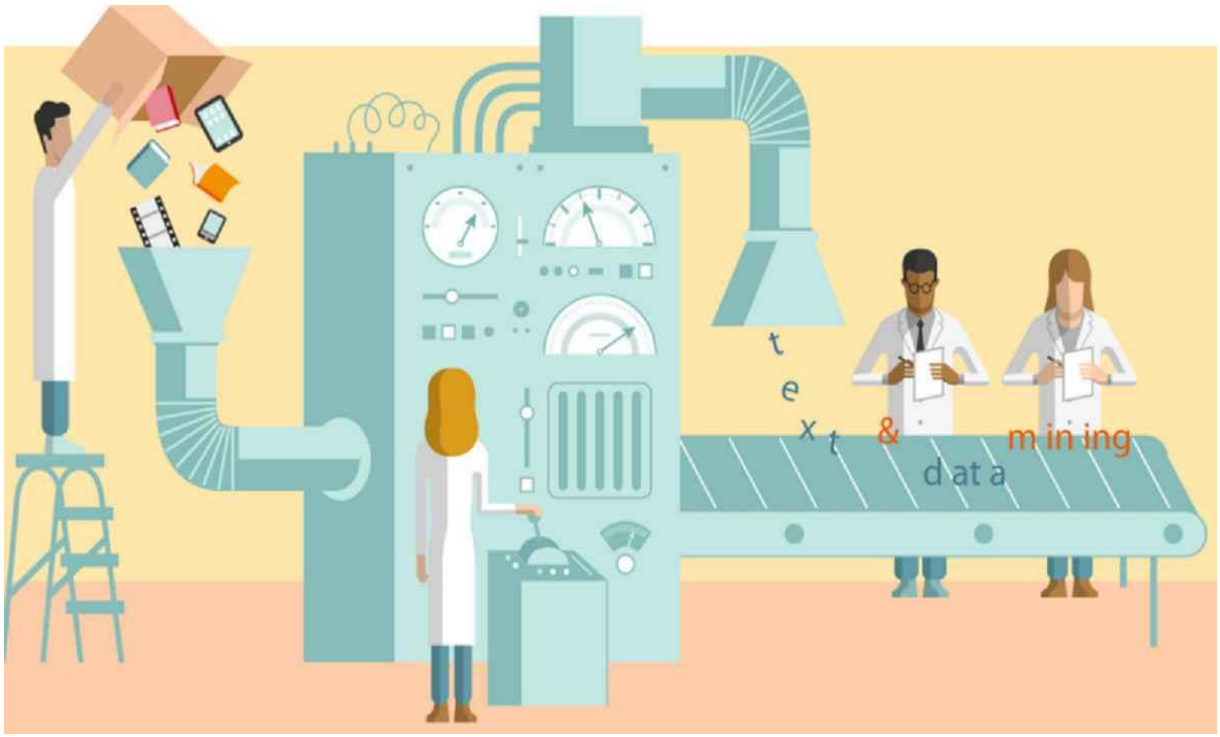
data preparation چیست؟ در این بخش قصد داریم در مورد **آماده سازی داده ها** اطلاعاتی را با شما به اشتراک بگذاریم.

یک رویکرد مدرن برای تهیه داده‌ها مورد نیاز است. دیجیتالی شدن فزاینده فرآیندهای کسب و کار این ضرورت را برای شرکت‌ها ایجاد کرده است که تا آن‌جا که ممکن است کاربران را قادر به دریافت بینش از داده‌ها (دموکراتیک سازی تجزیه و تحلیل) کنند. امروزه بسیاری از شرکت‌ها تهیه داده‌ها را به عنوان کلیدی برای افزایش توانایی آن‌ها در استفاده موثر از داده‌ها به صورت توزیع شده برای بهینه سازی فرآیندهای تجاری یا در وهله اول امکان ایجاد مدل‌های جدید و نوآورانه تجاری می‌دانند. در اقتصاد امروز دستیابی به تهیه داده‌های کارآمد و خوب از اهمیت بالایی برخوردار است. بازارهایی که به طور فزاینده‌ای متزلزل و اشباع می‌شوند، فضای پیچیده‌ای برای تجارت ایجاد می‌کنند که در آن توانایی تمایز با استفاده از قدرت تجزیه و تحلیل حیاتی است. سازمان‌ها تلاش می‌کنند تا تقاضای داده برای تجزیه و تحلیل را دنبال کنند تا بینشی در مورد تغییر شرایط بازار داشته باشند.



آماده سازی داده چیست؟

آماده سازی داده ها یا **data preparation** فرآیند تهیه و ارائه داده‌ها برای کشف داده‌ها، داده کاوی و تجزیه و تحلیل پیشرفته است. هدف از تهیه داده‌ها، پشتیبانی از تحلیل گران کسب و کار و دانشمندان داده با تهیه انواع داده‌ها برای اهداف تحلیلی آن‌ها است. تهیه داده‌ها می‌تواند در بخش‌های تجاری انجام شود و یا به طور متمرکز توسط IT انجام شود. **آماده سازی داده ها** یک زیر دامنه ادغام داده است که می‌تواند با ابزارهای اختصاصی یا ابزارهای سنتی برای ادغام داده‌ها مانند ابزارهای ETL، مجازی سازی داده‌ها یا اتوماسیون انبار داده اجرا شود. BARC برای کسب اطلاعات بیشتر در مورد تفکر فعلی در زمینه تهیه اطلاعات، یک تحقیق مستقل از بیش از ۶۹۵ متخصص BI از طیف وسیعی از صنایع در سراسر جهان انجام داد. نظر سنجی " BARC تهیه داده: پالایش داده‌های خام" یکی از بزرگ‌ترین مطالعات با تمرکز بر شرایط، مزایا و چالش‌های تهیه داده است.



data preparation نیازهای واقعی تجارت را تامین می‌کند و به طور گسترده‌ای مورد استفاده قرار می‌گیرد. مشاغل امروز همانطور که در طول تاریخ با چالش‌های بزرگی روبرو بوده‌اند، هنوز هم چالش برانگیز هستند. آنچه جدید است این است که توانایی استفاده سیستماتیک از داده‌ها به یک مزیت رقابتی تعیین کننده تبدیل شده است. بسیاری از شرکت‌ها این موضوع را تشخیص داده‌اند و در تلاش هستند تا با معرفی یا بهبود **آماده سازی داده ها** بسیاری از مشکلات استفاده از داده‌های خود را برطرف کنند. محرک‌های اصلی پروژه‌ها نشان می‌دهند که هیاهوی ایجاد اطلاعات، که بدون شک وجود دارد، با الزامات ملموس پشتیبانی می‌شود. انتظارات زیاد از مزایای تجزیه و تحلیل و نیاز به چابکی باعث استفاده از **آماده سازی داده ها** می‌شود. سهم شرکت‌هایی که از قبل برای کسب اطلاعات از **آماده سازی داده ها** استفاده می‌کنند به همین ترتیب زیاد است. تقریباً ۷۰٪ پاسخ دهندگان اظهار داشتند که قبلاً از **آماده سازی داده ها** استفاده کرده‌اند.

مزایای آماده سازی داده ها

در این بخش مزایای گسترده‌ای از **data preparation** برای شما بازگو می‌کنیم:

انتظارات سازمان ها از قبل هم بیشتر شده است: زمانی که شرکت‌ها از فناوری روز استفاده می‌کنند، اغلب انتظارات زیادی از این فناوری پیشرفته دارند. در مورد **آماده سازی داده ها** نیز این مورد صادق است. از ابزارها و روش‌های **آماده سازی داده ها** برای مقابله با چالش‌های اساسی استفاده می‌شود و نتایج نظرسنجی ما نشان می‌دهد که آن‌ها در حقیقت مزایایی با خود به همراه دارند. اگر به طور مناسب در سازمان انجام شده باشد؛ آماده سازی داده‌ها فرصتی واقعی برای ارائه داده‌ها برای تجزیه و تحلیل در شکل بهتر و سریع‌تر فراهم می‌کند و بنابراین مزایای فوری برای شرکت ایجاد می‌کند. همچنین شما می‌توانید با ابزارهای داده کاوی تجزیه و تحلیل بهتری داشته باشید.

شرکت‌ها هنوز در حال جستجو هستند!

لازم به ذکر است که بدانید شرکت‌ها هنوز در حال جستجو برای اجرای صحیح هستند. فرمول جادویی **data preparation** در داخل سازمان هنوز یافت نشده است. اگر چه اکنون استفاده از روش‌ها و ابزارهای **آماده سازی داده ها** گسترده شده است، اما تقسیم وظایف مشخصی بین فناوری اطلاعات و کاربران تجاری هنوز به شکل دقیق مشخص نشده است. پاسخ دهندگان به نظر سنجی ما نشان می‌دهند که سازمان‌ها طیف وسیعی از رویکردها را در پیش گرفته‌اند. فناوری اطلاعات نقش مهمی در دو روش محبوب بازی می‌کند. نسبت شرکت‌هایی که کاربران تجاری در آن‌ها به طور فعال در تهیه اطلاعات کار می‌کنند نیز زیاد است .لازم به ذکر است که **اندازه شرکت یا سازمان** در تعیین این که توسط چه کسی و چگونه **آماده سازی داده ها** پیش می‌رود، کمتر از **پیچیدگی فنی نیازها** و مشخصات مهارتی که در هر بخش وجود دارد، تعیین کننده است. بنابراین هیچ نشانه‌ای از یک فرمول جادویی متناسب با همه وجود ندارد؛ به این دلیل است که بازار ابزار آماده سازی داده‌ها به سرعت در حال پیشرفت است.

ارزش داده‌ها

ارزش داده‌ها شناخته شده است اما مهارت‌های مفید در داده‌ها کم است. در مورد **data preparation** برای کشف داده‌ها یا تجزیه و تحلیل پیشرفته، داده کاوی و علم داده، آماده سازی داده‌ها به ویژه از بخش‌های تجاری و دانشمندان داده پشتیبانی می‌کند. در حال حاضر توانایی به دست آوردن بینش ارزشمند از داده‌ها از طریق آماده سازی کارآمد و تا حد زیادی مستقل داده‌ها در بخش‌های تجاری، بسیار ناچیز است. بنابراین برای این که بتوانید استراتژی‌های پیچیده دیجیتال سازی را اجرا کنید، نیاز فوری به آموزش وجود دارد. برای انجام این کار منابع و بودجه باید اختصاص یابد. به گفته شرکت کنندگان در این نظرسنجی، این موارد از بزرگ‌ترین چالش‌ها در زمینه تنظیم ابتکارات **آماده سازی داده ها** هستند. همانند بسیاری از جنبه‌های مدیریت داده، آماده سازی داده نمی‌تواند به صورت گذرا انجام شود. باید به عنوان یک گام ارزشمند در فرآیند ایجاد ارزش از داده‌ها مورد توجه قرار بگیرد. این کار یک پروژه یک بار مصرف نیست که بتوان از خارج از کشور تامین و تکمیل شود بلکه **یک کار مداوم** است که به درجه بالایی از صلاحیت نیاز دارد. بنابراین **آماده سازی داده ها** باید به طور عمیق و گسترده‌ای در سازمان جاسازی شود.

به حداکثر رساندن منافع شرکت‌ها

سوالی که در این بخش مطرح است این است که چگونه شرکت‌ها/سازمان ها می‌توانند منافع را به حداکثر برسانند؟ این سوال در بخش **data preparation** به سه دسته تقسیم می‌شود. در ادامه با ما باشید تا هر کدام را برای شما بررسی کنیم.

همکاری بین بخش‌های IT و بازرگانی

یک رویکرد امیدوار کننده در زمینه **آماده سازی داده ها** برای تجزیه و تحلیل، تقسیم کار در مرزهای وزارت است. شرکت‌هایی که بیشترین سود و بیشترین رضایت را از **data preparation** دارند گزارش می‌دهند که :**آماده سازی داده ها** را به یک وظیفه مشترک بین بخش‌های فناوری اطلاعات و بازرگانی/کسب و کار اصلی تبدیل کرده‌اند، جایی که کاربران با حس قوی از تجارت قادر به تهیه داده‌ها به طور مستقل هستند، اما با پشتیبانی از کارشناسان فنی که مطابقت با استانداردها را تضمین می‌کنند. این امر باعث می‌شود تا در صورت بروز مشکلات تجاری سریعاً پاسخ داده شود. سپس راه حل‌های توسعه یافته توسط متخصصان فنی خودکار می‌شوند و در دسترس مخاطبان گسترده‌تری قرار می‌گیرند. برای افزایش دستیابی به اهداف، نه تنها کاربران تجاری و فناوری اطلاعات باید همکاری کنند بلکه مدیریت باید منابع مناسب برای آموزش (به ویژه برای کاربران تجاری) و ابزارهای مناسب را فراهم کند و همچنین موقعیت خود را به عنوان یک درایور برای پیشرفت در زمینه تجزیه و تحلیل تقویت کند.

حاکمیت داده‌ها

حاکمیت داده‌ها از استانداردها و قواعدی برای ایجاد ارزش کارآمد از داده‌ها استفاده می‌کند. پیش نیاز ایجاد ارزش از داده‌ها و بنابراین وظیفه حاکمیت داده‌ها؛ **ایجاد مسئولیت‌های مشخص، تعریف اهداف و ساختارهای شفاف، ارائه تعاریف استاندارد و اطمینان از کیفیت داده‌ها و امنیت داده‌ها است.** به نظر می‌رسد نیاز به اقدام برای اجرای مطلوب این نکات بسیار زیاد است. این مسئله برای **آماده سازی داده ها** همیشه یک کار ساده نیست چون اجرای موفقیت آمیز آن اغلب به تعدیل سازمان نیاز دارد. در زمینه تهیه داده‌ها این امر در درجه اول یافتن تعادل مناسب بین ثبات مرکزی و سطح مطلوب انعطاف پذیری غیر متمرکز در آماده سازی داده‌ها است.

ابزار مناسب

استفاده از اکسل Excel گسترده است و به نظر می‌رسد امروزه اولین ابزار برای تهیه داده‌ها باشد. این واقعیت که اکسل از تهیه داده‌های پیچیده پشتیبانی نمی‌کند باید برای همه روشن باشد و همچنین باید توضیح بدهد که چرا برنامه‌های کاربردی عمدتاً ساده و با استفاده از **آماده سازی داده ها** اجرا می‌شوند. صفحات گسترده برای عملکردهای پیشرفته یکپارچه سازی و همچنین توانایی نقشه برداری از فرآیندهای خودکار با عملکرد بالا و پایدار برای **data preparation** فاقد قابلیت هستند. سوال اصلی در این جا است که آیا اکسل به دلیل محدودیت عملکرد مانع توسعه پتانسیل آماده سازی داده‌ها می‌شود یا این که صرفاً مواردی که به ابزارهای ویژه‌ای برای تهیه داده‌ها نیاز دارند، کمبود دارد؟

سه گام راه تلفیق داده + مراحل یکپارچه سازی داده ها

سه گام راه تلفیق داده را با هم در این بخش مرور می‌کنیم. در حقیقت سازمان‌ها برای به دست آوردن ارزش از داده ها باید داده‌ها را بدون در نظر گرفتن محل سکونت، جمع کنند. این موضوع در اصل به معنی اتصال منابع داده، تاباندن نور به داده‌های تاریک، پردازش و تمیز کردن داده‌ها در زمان واقعی و خودکار سازی محیط‌های تجزیه و تحلیل است. سوال بی جواب در این بخش این است که از کجا باید شروع کنید؟ به گفته اعضای انجمن متخصصان فناوری اطلاعات، تحلیل گران صنعت و کارشناسان فناوری، پاسخ در ساخت یک نقشه راه برای ادغام داده‌ها نهفته است. در ادامه **سه مرحله تلفیق داده** را با هم مرور می‌کنیم.

مرحله اول

مرحله اول از **سه گام راه تلفیق داده** در حقیقت قرار دادن زمینه در اطراف داده‌ها است. مشاغل صرف نظر از اندازه، داده‌های فراوانی دارند. بیشتر آن‌ها در اصل تاریک و در سیلوهای مختلف یا مخازنی مانند صفحات گسترده، انبارهای داده، پایگاه داده‌های غیر رابطه‌ای و غیره نشسته‌اند. در اصل اولین قدم در **سه گام برای تلفیق داده** ، درک آنچه شما دارید است.

رییس StarCLO و نویسنده کتاب Driving Digital می‌گوید: چیزی که سازمان‌ها به آن در اولین قدم نیاز دارند این است که باید با یک تمرین کشف داده شروع کنند. این مورد شامل بررسی حجم، عرض، سرعت و الزامات ادغام منابع مختلف داده است. ساکولیک رییسStarCLO می گوید: این کار به شناسایی دارندگان داده، انتخاب ابزارهای ادغام داده‌ها و یادگیری

مهارت‌های پردازش داده کمک می‌کند. شما در طی این فرآیند می‌توانید در داده‌های تاریک یا بدون ساختار سلطنت کنید. کین مک گلادری، معمار امنیت در Ascent Solutions می‌گوید: سعی کنید که تعداد مکان‌هایی را که داده‌های بدون ساختار در آن‌ها ذخیره می‌شود، محدود کنید.

سازمان شما در اصل با داشتن تعمدی در مورد ذخیره سازی اطلاعات شما و ممنوعیت ذخیره یا پردازش داده‌های غیر عمومی در سرویس‌های پرخطر یا غیرمجاز، با خطرات قانونی و نظارتی روبرو خواهد شد .**سه مرحله تلفیق داده** را جدی بگیرید تا شاهد نتایج آن باشید. علاوه بر این با تابش نور به داده‌های تاریک، شرکت‌ها می‌توانند به این درک برسند که ارزش در کجا است؟ اسکات نلسون مدیر عامل CTO در Reuleaux Technology می‌گوید: در اصل هر سیستم، دنیای واقعی پیچیده‌ای است و شما باید در اسرع وقت یاد بگیرید که داده‌ها چه اهمیتی دارند. در این راستا جیسون جیمز، مدیرعامل CIO می‌گوید : کشف و فهرست بندی داده‌ها، زمینه را برای آن فراهم می‌کند. همچنین او اضافه می‌کند که قبل از شروع نقشه راه، ادغام داده‌ها باید وضعیت فعلی داده‌ها و نتیجه مطلوب را درک کنید. در حقیقت نقشه جایی که هستید و کجا می‌خواهید بروید است.

مرحله دوم

در مرحله دوم از **سه گام راه تلفیق داده** شما باید در اصل اهداف و نقش‌ها را تعریف کنید. در این مرحله شما باید مانند هر نقشه راه دیگری، اهداف خود را تعیین کنید. تریستان، رییس جامعه در CTO می‌گوید : اهداف خود را زودتر مشخص کنید. مشکل تجاری را که در حال حل آن هستید، شناسایی کنید. آیا شما مقدار زیادی داده می‌گیرید و آن را قابل فهم می‌کنید؟ آیا این مسئله به مشتریان شما می‌گوید که چگونه از آن استفاده کنند؟ این فاکتورها، موارد مهمی در هر تصمیمی هستند. غیر از **سه مرحله تلفیق داده** این موارد را در تصمیم گیری‌های شخصی خود هم رعایت کنید. IDG اینفلوئنسر معروف در این زمینه می‌گوید که مشتری نقشی اساسی در شکل دادن به داستان ارزش داده شما دارد. فرانک کاتیتا، مدیر عامل و بنیانگذار آزمایشگاه تصمیمات Health Tech می‌گوید: از بیرون شروع کنید! اغلب اوقات نقشه راه بدون ورود مشتری تهیه می‌شود و منعکس کننده داده‌های بازار محور نیست، به ویژه در دنیایی مصرف کننده! پیتر مدیر ارشد فناوری و نوآوری‌های Oroca با این مسئله موافق است. او می‌گوید، خیلی اوقات ما با تعداد زیادی داده، بینش گیج کننده و هیچ داستان مشخصی روبرو هستیم. نیکول پیشنهاد می‌کند، سوالاتی مانند:

- چه کسی سهامدار داده را مدیریت می‌کند؟
- چه داستانی را با داده‌ها در تلاش هستیم، تا بگوییم؟
- چه کسی مسئول توسعه و تفصیل پرونده تجاری است؟

در مرحله بعدی از سه گام برای تلفیق داده به چارچوبی که چرخه عمر داده‌ها را تعریف می‌کند، بپردازید. چاک بروکس، مشاور بین المللی در این زمینه می‌گوید: یک چارچوب مدیریت ریسک از بهترین روش‌ها برای سازمان شما استفاده می‌کند تا از داده‌هایی که برای شما بیشترین ارزش دارد، محافظت کند. طراحی برای **امنیت داده‌ها** همراه با مدیریت بنیادی ریسک سایبری، بخشی جدایی ناپذیر از چارچوب کلی مدیریت ریسک شرکت ERM برای جلو ماندن از تهدیدات است. جک گلد، تحلیل گر اصلی و بنیانگذار شرکت گلد می‌گوید: همچنین به داده‌ها به عنوان یک چرخه کامل زندگی فکر کنید؛ مسیری از کسب تا تجزیه و تحلیل کامل و چشم نواز!

داده‌ها را یتیم نکنید: در اصل باید مطمئن شوید که در تصویر کل داده، ادغام شده است. اگر فقط ۱۰ تا ۱۵درصد از اطلاعات خود را تجزیه و تحلیل می‌کنید، ۸۵ تا ۹۰٪ بینش کسب و کار شما محقق نمی‌شود. در ادامه با گام بعدی و گام آخر ما از سه مرحله تلفیق داده همراه باشید.

مرحله سوم

گام آخر از **سه گام راه تلفیق داده** به شما می‌گوید تا راه حل‌های مناسب داده را پیدا کنید. زمانی که سازمان شما، داستان ارزش داده خود را درک کرد، زمان آن رسیده است که ادغام را عملی کنید .استراتژی‌ها و راه حل‌هایی را که به بهترین وجه، مناسب تجارت شما هستند، کاوش کنید. بن شین، معاون و مشاور داده کاوی در DOMO می‌گوید: من تیم‌ها را تشویق می‌کنم که به این فکر کنند، چگونه تغییر مناسب خود را ایجاد کنند ! در حقیقت مهم‌ترین کاری که شما می‌توانید انجام دهید، ساختن سیستمی است که بتواند هم شرایط متغیر و هم فناوری در حال پیشرفت را پیش بینی کرده و به آن پاسخ دهد . **سه گام برای تلفیق داده** شاید به نظر برسد ساده است اما زمان بر است و نیاز به کار گروهی قوی دارد. بنابراین باید اطمینان حاصل کنید که شما یک فرهنگ سازنده و نگرشی که موجب تغییر می‌شود، را ایجاد می‌کنید.

شما همیشه راهی خواهید داشت که از چشم انداز در حال پیشرفت، جلوتر باشید. از دیدگاه راه حل‌ها، دی لیبرو مدیر ارشد استراتژی در حقیقت توسعه یک استراتژی مدیریت داده را اصلی را توصیه می‌کند. او می‌گوید MDM جایی است که لاستیک جاده را برای حفظ داده‌های سازمانی به عنوان یک دارایی پیدا می‌کند. همانطور که گفته شد، MDMدر طرح‌ها و شکل‌ها و طعم‌های مختلف وجود دارد؛ آنچه برای یک سازمان مناسب است، ممکن است برای سازمان دیگری مناسب نباشد. یک دسته نکات ریز و مهم هستند که شما باید در **سه مرحله تلفیق داده** رعایت کنید. علاوه بر این موارد، استیو توت، مدیر ارشد در Optivمی‌گوید :یک نقشه راه برای ادغام داده‌ها باید فناوری‌های نوظهور را در نظر بگیرد .سیستم عامل‌های یکپارچه سازی داده‌های نسل بعدی به سازمان‌ها این امکان را می‌دهد تا انواع داده‌ها را به طور خودکار کشف کنند و به صورت پویا از کنترهای دسترسی گرانول برای مصرف کنندگان پایین دست داده استفاده کنند .کاتیتا دیگر مدیر ارشد با این نظریه موافق است: اطمینان حاصل کنید که نقشه راه جدا از فناوری اساسی موجود ساخته شده است؛ در عوض باید از یک نقشه راه زیرساخت داده جدید هدایت شود در مقابل زیرساختی که ما با آن گیر کرده‌ایم .سرانجام باید بگیم که در این **سه گام راه تلفیق داده** دقیقاً جایی که بیشتر از همه به آن نیاز دارید، می‌توانید کمک بگیرید.

نتیجه

در صورت عدم تنظیم صحیح آن، ایجاد یک نقشه راه برای ادغام داده‌ها می‌تواند مشکل باشد و زمان زیادی از دست می‌رود. فایده استفاده از کمک خارجی برای چنین پروژه‌ای این است که به تیم‌های داخلی شما این امکان را می‌دهد تا این طرز فکر را ایجاد کنند که دیگران چگونه نقشه راهی برای ادغام داده‌ها ایجاد کرده‌اند؟ چه چیزی مفید بود؟ چه عواملی باعث موفقیت شد؟ از انجام چه چیزهایی باید خودداری کنیم؟

کلان داده (Big Data) چیست؟

کلان داده اصطلاحی است که حجم زیاد داده های بزرگ را توصیف می کند. کلان داده چه به صورت ساختاری و چه غیر ساختاری، به طور روزمره یک تجارت را می تواند اشباع کند. اما مقدار داده مهم نیست، این که سازمان ها با این داده های مهم چه کاری انجام می دهند، مسئله اصلی می باشد. داده های کلان را می توان برای بینش هایی که منجر به تصمیم گیری بهتر و حرکت های استراتژیک تجاری می شوند، تجزیه و تحلیل کرد.

استفاده از کلان داده این روزها توسط شرکت ها برای پیشی گرفتن از همتایان خود رایج شده است. در بیشتر صنعت ها، رقبای قدیمی و تازه واردان به طور یکسان از استراتژی های حاصل از داده های تجزیه و تحلیل شده برای رقابت، نوآوری و ارزش پیدا کردن استفاده می کنند.

کلان داده به سازمان ها کمک می کند تا فرصت های جدید رشد کردن را پیدا کنند و دسته بندی های کاملاً جدیدی از شرکت ها را ایجاد کنند که می توانند داده های صنعتی خود را ترکیب، تجزیه و تحلیل کنند. این شرکت ها اطلاعات کافی در مورد محصولات و خدمات، خریداران و تأمین کنندگان، ترجیحات مصرف کنندگان دارند که می توانند جذب، تجزیه و تحلیل شوند.

در حالی که اصطلاح **"کلان داده"** نسبتاً جدید است، اما خودِ عمل جمع آوری و ذخیره اطلاعات و داده های عظیم برای تجزیه و تحلیل نهایی قدیمی است. این مفهوم در اوایل دهه ۲۰۰۰ وقتیDoug Laney ، تحلیلگر صنعتی، تعریف اصلی و جریان کلان داده را به عنوان **سه V (Volume, Velocity, Variety)** بیان کرد، تبدیل به بحث روز شد:

Volume حجم :سازمان ها داده ها را از منابع مختلف، از جمله تراکنش های تجاری، رسانه های اجتماعی، اطلاعات از داده های سنسور ها و یا به شکل ماشین به ماشین جمع آوری می کنند. در گذشته، ذخیره آن یک مشکل بود اما فناوری های جدید مانند Hadoop بار روی دوش کسب و کار ها را کاهش داده اند. نام **"کلان داده"** خود به بسیار بزرگ بودن اندازه داده مربوط می شود. اندازه داده ها در تعیین مقدار داده ها بسیار مهم است. همچنین، این که آیا داده خاصی می تواند به عنوان یک داده بزرگ در نظر گرفته شود یا خیر، به حجم داده بستگی دارد. از این رو، **"حجم "**یکی از ویژگی هایی است که باید در هنگام پرداختن به **"کلان داده"** مورد توجه قرار گیرد.

Velocity سرعت :جریان داده ها امروزه با سرعتی بی سابقه در حال انجام می باشد و باید به موقع با آن ها برخورد شود. بر چسب هایRFID ، سنسور ها و اندازه گیری هوشمند نیاز به مقابله با تورنت های داده ها را در زمان تقریباً real time ایجاد می کنند. اصطلاح "**Velocity"** به سرعت تولید داده ها اشاره دارد. این که سرعت تولید و پردازش داده ها برای پاسخگویی به خواسته ها چگونه باشد، پتانسیل واقعی را در داده ها تعیین می کند.

Big Data Velocity با سرعتی که داده ها از منابعی مانند فرآیند های تجاری، گزارش ها، شبکه ها و سایت های رسانه های اجتماعی، سنسور ها، دستگاه های تلفن همراه و ... سرازیر می کنند، سرو کار دارد. جریان داده ها گسترده و مداوم است.

Variety تنوع :داده ها در انواع قالب ها وجود دارند، که شامل داده های ساختار یافته، داده های عددی در پایگاه های داده سنتی تا اسناد متنی بدون ساختار مثل ایمیل، ویدئو، صدا، داده های مربوط به سهام و تراکنشات مالی می باشند. **"تنوع"** به منابع نا همگن و ماهیت داده ها اعم از ساختاری و غیر ساختاری اشاره دارد. در طی روز های گذشته، صفحات گسترده و پایگاه داده تنها منابع داده ای بودند که توسط اکثر برنامه ها مورد توجه قرار گرفتند. اکنون، داده هایی به صورت ایمیل، عکس، فیلم، دستگاه نظارت، PDF، صدا و ... نیز در برنامه های تجزیه و تحلیل در نظر گرفته شده اند. این تنوع داده های غیر ساختاری مسائل خاصی را برای ذخیره سازی، استخراج، تجزیه و تحلیل داده ها به وجود می آورد.

چرا Big Data (کلان داده) مهم است؟

اهمیت داده های کلان به میزان داده های یک شرکت بر نمی گردد بلکه، به نحوه استفاده یک شرکت از داده های جمع آوری شده است. هر شرکتی از داده ها به روش خود استفاده می کند. هر چه شرکت ها از کارآیی داده های خود استفاده انجام دهند، توانایی بیشتری برای رشد دارند .این شرکت ها می توانند داده ها را از هر منبعی گرفته و آن ها را تجزیه و تحلیل کنند تا پاسخی پیدا کنند که این موارد را تحت تاثیر قرار دهد:

۱. **صرفه جویی در هزینه :**برخی از ابزار های Big Data مانند Hadoop و Cloud-based Analytics می توانند کاهش هزینه ای را برای تجارت به ارمغان بیاورند؛ وقتی که مقدار زیادی داده ذخیره می شود، این ابزار ها به شناسایی روش های کارآمد تر تجارت کمک می کنند.
۲. **کاهش های زمانی :**سرعت بالای ابزاری مانند Hadoop و تجزیه و تحلیل حافظه می تواند به راحتی منابع جدید داده را شناسایی کند که به صاحبان کسب و کار ها در تجزیه و تحلیل سریع داده ها و تصمیم گیری سریع بر اساس آموخته ها کمک می کند.
۳. **شرایط بازار را درک کنید :**با تجزیه و تحلیل کلان داده می توانید درک بهتری از شرایط فعلی بازار داشته باشید. به عنوان مثال، با تجزیه و تحلیل رفتار های خرید مشتریان، یک شرکت می تواند محصولی را که بیشترین فروش را دارد بیابد و مطابق این روند محصولات بعدی خود را تولید کند و با این کار می تواند از رقبای خود پیشی بگیرد.
۴. **اعتبار آنلاین را کنترل کنید :**ابزار های کلان داده می توانند تجزیه و تحلیل احساسات را انجام دهند. بنابراین، می توانید درباره اینکه چه کسی درباره شرکت شما چه چیزی می گوید، بازخورد بگیرید. اگر می خواهید حضور آنلاین مشاغل خود را رصد کرده و بهبود ببخشید، ابزار های کلان داده می توانند به همه این مسائل کمک کنند.
۵. **استفاده از تجزیه و تحلیل کلان داده ها برای تقویت جذب و نگهداری مشتری :**مشتری مهم ترین سرمایه ای است که هر کسب و کاری دارد و تماما به آن بستگی دارد. هیچ تجارت واحدی وجود ندارد که بتواند ادعای موفقیت کند، بدون این که ابتدا نیاز به ایجاد یک پایگاه مشتری ثابت داشته باشد. با این حال، حتی با داشتن مشتری، یک تجارت نمی تواند از رقابت بالایی که با آن روبرو است، چشم پوشی کند. اگر یک کسب و کار در زمینه این که بفهمد مشتری ها به دنبال چه چیزی هستند، عمل کند، پس تنها محصولات کسل کننده و ضعیف ارائه خواهند شد و در پایان به از دست دادن مشتری ختم خواهد شد و این یک اثر کلی نامطلوب بر موفقیت کسب و کار شما ایجاد می کند. استفاده از کلان داده ها به مشاغل اجازه می دهد تا الگو ها و روند های مختلف مربوط به مشتری را بهتر درک کنند. بررسی رفتار مشتری باعث ایجاد وفاداری آن مشتری برای شما می شود.
۶. **استفاده از تجزیه و تحلیل کلان داده ها برای حل مشکل تبلیغ کنندگان و ارائه بینش بازاریابی :**تجزیه و تحلیل کلان داده ها می تواند به تغییر تمام عملیات بیزینسی کمک کند. این شامل توانایی مطابقت پیدا کردن با انتظارات مشتری، تغییر خط تولید شرکت و البته اطمینان حاصل کردن از قدرتمند بودن در حیطه بازاریابی است.
۷. **Big Data Analytics به عنوان محرک نو آوری و توسعه محصول :**یکی دیگر از مزایای بزرگ کلان داده ها، توانایی کمک به شرکت ها در زمینه نو آوری و توسعه مجدد محصولات خود است.

سوالات متداول درخصوص کلان داده های سازمان

مزایای کلان داده چیست؟

- کسب و کار ها می توانند هنگام تصمیم گیری از سورس های بیرونی استفاده کنند
- خدمات مشتری بهبود یافته است
- شناسایی زود هنگام ریسک های پیش رو برای محصولات یا خدمات، در صورت وجود
- کارایی عملیاتی بهتر

کلان داده (Big Data) چه استفاده ای برای سازمان ها دارد؟

- کلان داده به سازمان ها کمک می کند تا فرصت های جدید رشد کردن را پیدا کنند و دسته بندی های کاملاً جدیدی از شرکت ها را ایجاد کنند که می توانند داده های صنعتی خود را ترکیب، تجزیه و تحلیل کنند.

اهمیت کلان داده در چیست؟

- اهمیت داده های کلان به میزان داده های یک شرکت بر نمی گردد، بلکه به نحوه استفاده یک شرکت از داده های جمع آوری شده است .

بهترین نمونه های کلان داده

بهترین نمونه های کلان داده ها را می توان هم در بخش دولتی و هم در بخش خصوصی یافت. از تبلیغات هدفمند، آموزش و پرورش و صنایع گسترده (بهداشت و درمان، بیمه، تولید یا بانکداری)، تا سناریو های زندگی واقعی، در مهمانی ها یا انواع سرگرمی. در سال ۲۰۲۰، در هر ثانیه ۱,۷ مگابایت داده برای هر فرد روی کره زمین تولید شده است، پتانسیل رشد سازمانی داده محور در بخش مهمان نوازی بسیار زیاد است.

کلان داده ها می توانند در برخی از حوزه هایی که کسی فکرش را هم حتی نمی کند، باعث ایجاد مزایا شود.

➢ داده های کلان در صنعت آموزش

در زیر برخی از زمینه های صنعت آموزش و پرورش وجود دارد که با تغییرات انگیزه داده های بزرگ تغییر شکل داده اند:

- برنامه های یادگیری سفارشی سازی شده و پویا
- تنظیم مجدد مطالب دوره ها
- سیستم های درجه بندی
- پیش بینی شغلی

➢ داده های کلان در صنعت بیمه

صنعت بیمه نه تنها برای اشخاص بلکه شرکت های تجاری نیز دارای اهمیت است. دلیل اینکه بیمه جایگاه قابل توجهی دارد این است که در زمان سختی ها و بلاتکلیفی ها از افراد پشتیبانی می کند. داده های جمع آوری شده از این منابع فرمت های مختلفی دارند و با سرعت بسیار زیادی تغییر می کنند.

جمع آوری اطلاعات

از آن جا که داده های بزرگ به جمع آوری داده ها از منابع مختلف اشاره دارد، این ویژگی یک مورد حیاتی را برای صنعت بیمه ایجاد می کند. به عنوان مثال: هنگامی که مشتری قصد خرید یک بیمه اتومبیل را دارد، شرکت ها می توانند اطلاعاتی را بدست آورند، که می توانند از آن ها سطح ایمنی راننده یعنی سوابق رانندگی گذشته وی را متوجه شوند. بر این اساس آن ها می توانند هزینه بیمه اتومبیل را نیز به طور موثر محاسبه کنند.

کسب بصیرت مشتری

تعیین تجربه مشتری و تبدیل مشتری به کانون جذب یک شرکت برای سازمان ها از اهمیت بالایی برخوردار است.

تشخیص تقلب

کلاهبرداری در بیمه یک اتفاق رایج است. پرونده استفاده از کلان داده برای کاهش تقلب بسیار موثر است.

نقشه تهدید

هنگامی که یک آژانس، بیمه ای را می فروشد، آن ها می خواهند از همه احتمالات نا مطلوب کار با مشتری خود آگاه شوند و آن ها را وادار به اثبات ادعا کنند.

➢ کلان داده در صنعت دولتی

همراه با بسیاری از زمینه های دیگر، داده های کلان در دولت می توانند به شکل محلی، ملی و جهانی تاثیر بسزایی داشته باشند. امروزه با وجود بسیاری از مسائل پیچیده که مطرح می باشد، دولت ها تلاش کرده اند تا تمام اطلاعاتی را که دریافت می کنند، معنی دهند و تصمیمات حیاتی را در رابطه با آن ها بگیرند که این مسئله میلیون ها نفر را تحت تأثیر قرار می دهد. دولت ها و هر کشوری، تقریباً روزانه با حجم بسیار زیادی از داده ها روبرو می شوند. از این رو، آن ها باید سوابق و بانک های اطلاعاتی مختلفی را در مورد شهروندان ثبت کنند. مطالعه و تجزیه و تحلیل مناسب این داده ها به روش های بی پایان به دولت ها کمک می کند. تعداد کمی از آن ها عبارتند از:

- طرح های رفاهی
- امنیت سایبری

➢ کلان داده در بخش بانکداری

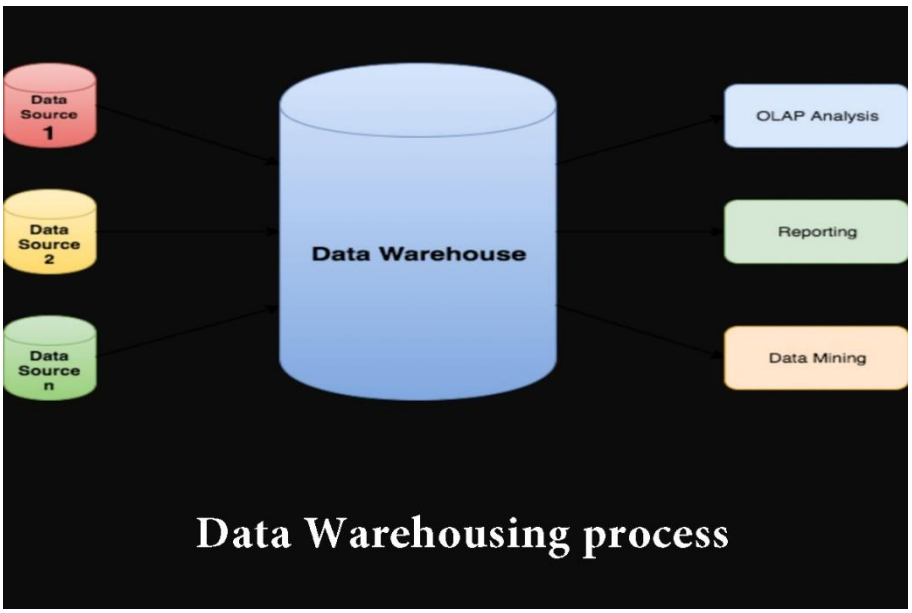
میزان داده ها در بخش های بانکی هر ثانیه سر به فلک می کشد. طبق پیش بینیGDC ، تخمین زده می شود که این داده ها تا سال ۲۰۲۱ تا ۷۰۰ درصد رشد کنند!

مطالعه و تجزیه و تحلیل داده های بزرگ می تواند به تشخیص موارد زیر کمک کند:

- سواستفاده از کارت های اعتباری
- واضح شدن کسب و کار
- تغییر آمار مشتری
- پول شویی
- کاهش ریسک

انبار داده یا Data Warehouse ، مزایا ، معایب و تفاوت آن با پایگاه داده

انبار داده چی است و چه کارآیی دارد؟ آیا همان پایگاه داده است یا این دوتا باهم تفاوت دارند؟ اصلاً استفاده از این تکنولوژی مزایا و معایبی هم دارد؟ اگر دنبال جواب این سوال‌ها و یک توضیح کامل و جامع درباره **Data Warehouse** هستین تا آخر این بخش را مطالعه نمایید.



انبار داده چیست؟

انبار داده Data Warehouse یا **DW** or **DWH** به سیستمی میگویند که داده ها را از منابع مختلف برای ارائه پیش بینی‌های **منطقی** در کسب و کار جمع آوری و مدیریت می کند. یک **Data Warehouse** معمولاً برای تولید گزارش و **تجزیه و تحلیل** داده های تجاری از منابع مختلف استفاده میکند. بهتر است بدانید که **DW** هسته اصلی **BI** به حساب می آید که خود **Business Intelligence** یا **هوش تجاری** هم برای تجزیه و تحلیل و گزارش دهی ساخته شده است.

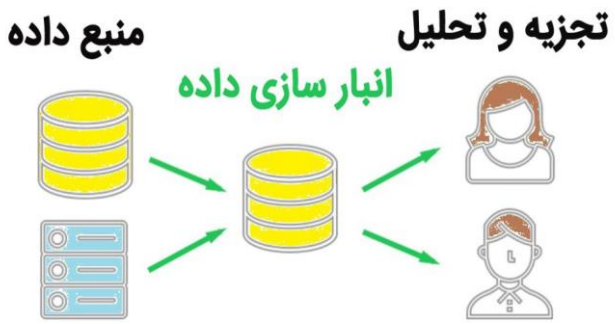
انبار داده یا Data Warehouse یک **پایگاه داده رابطه‌ای** هست که داده های فعلی و گذشته را در یک مکان واحد جمع آوری میکند و هدف اصلی آن پوشش **گزارش گیری** و رسیدگی به نیازهای **تحلیلی** یک سازمان می باشد.

ضمناً یک پروسه است که برای تبدیل **داده** به **اطلاعات** و به موقع تحویل دادن آن به کاربران شکل گرفته تا بلکه تغییری در رفتار آینده شرکت یا سازمان به وجود آید!

سیستم Data Warehouse با اسامی زیر هم شناخته میشود:

- سیستم پشتیبانی تصمیم گیری Decision Support System
- سیستم اطلاعات اجرایی Executive Information System
- سیستم اطلاعات مدیریت Management Information System
- راه حل هوشمندی کسب و کار Business Intelligence Solution
- برنامه تحلیلی Analytic Application
- پایگاه داده تحلیلی Data Warehouse

از **DW** معمولاً برای ارتباط داده های تجاری گسترده استفاده میشود تا بینش اجرایی بیشتری نسبت به **عملکرد** شرکت‌ها ارائه بشود.



نحوه عملکرد DW

Data Warehouse به عنوان یک مخزن مرکزی، که در آن اطلاعات از یک یا چند منبع داده به دست می آید، کار میکند. داده ها از طریق سیستم تراکنش و همچنین پایگاه های داده مرتبط، به یک **انبار داده** راه پیدا می‌کنند. داده ها ممکن است به حالت های زیر باشند:

- ساختار یافته، Structured
- نیمه ساختار یافتهSemi-structured
- داده های بدون ساختارUnstructured

داده ها پردازش، تبدیل و آماده میشوند تا کاربران بتوانند از طریق ابزارهای اطلاعاتی کسب و کار ، زبان پرس و جوی ساختار یافته یا Structured Query Language یا SQL و صفحات گسترده Spreadsheets به این **داده های پردازش شده** در انبار داده دسترسی پیدا کنند. یک Data Warehouse ، اطلاعاتی که از منابع مختلف هستند را با یک **پایگاه داده جامع** ادغام میکند.

حالا با ادغام همه این اطلاعات در یک مکان، سازمان می تواند مشتری‌هایش را بهتر **تحلیل** کند.. این امر کمک می کند که مطمئن بشوید همه اطلاعات موجود را در نظر گرفته اید، همچنین انبار داده باعث میشود تا **داده کاوی Data Mining** امکان‌پذیر شود و همانطور که می‌دانید **داده کاوی** می‌تواند به فروش و سود بیشتر منجر بشود.

انبار داده چه تفاوتی با پایگاه داده Database دارد؟

انبار داده از یک طرح متفاوت نسبت به **پایگاه داده** استفاده می‌کند. وظیفه اصلی پایگاه داده ، پشتیبانی از **تراکنش های آنلاین** و **پردازش کوئری** است که به آن سیستم پردازش تراکنش آنلاین یا **OLTP** میگوییم و بیشتر عملیات روزمره یک سازمان را پوشش میدهد. از طرف دیگه انبار داده به کاربر، خدماتی در نقش **تحلیل گر** و تصمیم گیرنده ارائه میدهد که می‌تواند داده ها را از چندتا منبع، تجزیه و تحلیل کند و درمورد **اختلافات** در طرح ذخیره سازی با استفاده از فرایند **ETL** بحث کند. البته سیستم ها در این مدل می‌توانند داده ها را در قالب‌های مختلفی برای هماهنگی با نیازهای متفاوت کاربران سازماندهی کنند و ارائه دهند که به آن سیستم پردازش تحلیلی آنلاین یا **OLAP** هم میگوییم.

چند مورد از تفاوت‌های انبار داده و پایگاه داده به صورت خلاصه:

- کابران پایگاه داده، کارمند دفتری و مسئولان هستند در حالی که کاربران انبار داده مدیران و تصمیم گیرندگانند.
- مقدار داده یک پایگاه داده کمتر از انبار داده است، در انبار داده این مقدار بین **چند گیگ تا چند ترابایت** است در حالی که در پایگاه داده بین **چند مگابایت تا چند گیگابایت** می باشد.
- پایگاه داده برای مدل‌های **OLTP** بهینه شده در حالی که انبار داده برای **پردازش تحلیلی آنلاین** طراحی شده.

انواع Data Warehouse

ما سه تا مدل اصلی برای انبار داده ها داریم که عبارتند از:

۱) انبار داده های سازمانی Enterprise Data Warehouse

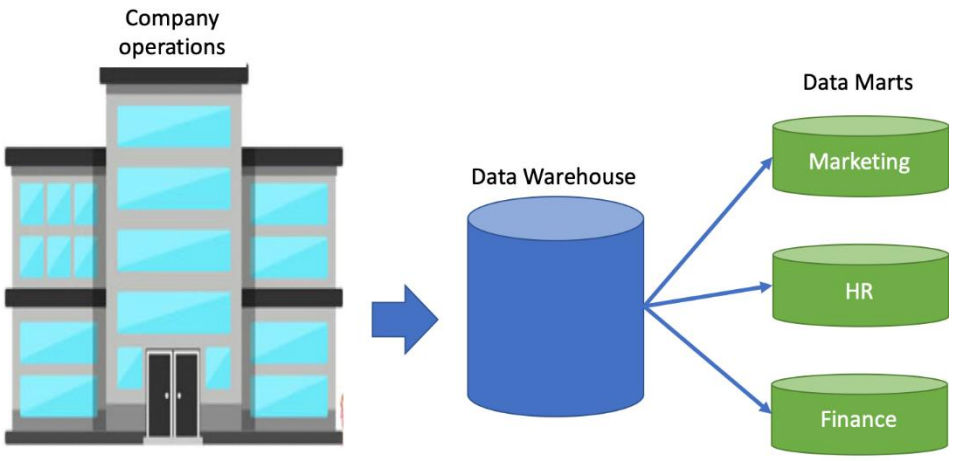
Enterprise Data Warehouse یک انبار متمرکز است که خدمات پشتیبانی و تصمیم گیری را در سراسر شرکت ارائه میدهد. این مدل همچنین یک روش یکپارچه برای سازماندهی و نمایش داده ها ارائه میدهد و امکان **طبقه بندی** داده ها را با توجه به موضوع و امکان دسترسی به داده ، مطابق با این بخش‌ها فراهم میکند.

۲) انبار داده های عملیاتی Operational Data Store

انبار داده های عملیاتی که **ODS** هم نامیده میشود، چیزی نیست به جز داده های مورد نیاز برای وقتی که نه Data warehouseو نه سیستم OLTP از نیازهای سازمان پشتیبانی نمی‌کنند. در **ODS** انبار داده بلافاصله تجدید میشود، برای همین هم بیشتر برای فعالیت‌های روزمره مثل ذخیره سازی سوابق کارمندها مناسب می باشد.

۳) بازار داده Data Mart

بازار داده یا Data Mart برای یک خط ویژه از کسب و کار ، مثل فروش یا امور مالی طراحی شده است. در یک **بازار داده** مستقل، داده ها می‌توانند مستقیماً از خود منابع جمع آوری شوند.



تکامل DW در کاربرد سازمانی

از خیلی سال پیش، سازمان ها شروع کرده بودند خیلی ساده از **انبار سازی** داده ها استفاده کنند؛ با این حال با گذشت زمان، این استفاده پیچیده‌تر و پیچیده‌تر شد:

انبار داده عملیاتی آفلاین

در این مرحله، داده ها فقط از یک سیستم عامل روی یک سرور دیگر **کپی** میشوند. در این روش بارگیری، پردازش و گزارش داده های کپی شده بر **عملکرد سیستم عملیاتی** تأثیری نمیگذارد.

انبار داده آفلاین

داده های موجود در انبار داده به طور منظم **آپدیت** میشوند و داده های داخل DW برای تحقق اهداف انبار داده ، نقشه برداری و تبدیل میشوند.

انبار داده لحظه ای

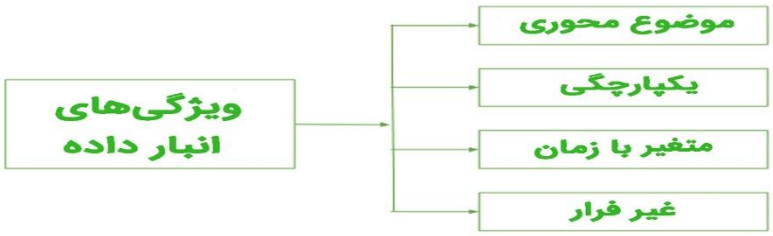
در این مرحله، هر وقت تراکنشی در پایگاه داده انجام بشود، انبار داده **به روز** میشود. مثلاً سیستم رزرواسیون هواپیمایی یا راه آهن.

انبار داده یکپارچه

در این مدل، داده ها را از بخش‌های مختلف کسب و کار گردآوری می‌کنند و برای همین هم کاربران می‌توانند **اطلاعاتی** که از سیستم‌های دیگه نیاز دارند را ببینند.

ویژگی‌های DW

ما در DW یک سری ویژگی‌ها داریم که **داده ها** رو تعریف می‌کنند. حالا این ویژگی ها شامل چه مواردی میشوند؟



موضوع محوری

هر انباری، داده های مربوط به یک موضوع خاص را در خودش نگه میدارد و این داده ها را به منظور **استخراج نتایج** و **مفاهیم کلی** به یک شکل خاصی سازماندهی میکند. پس این طور **سرعت جستجو** افزایش پیدا میکند. گردآوری اشیای مورد نیاز، **موضوع محوری** نامیده میشود.

یکپارچگی/اجتماع

در سیستم‌های مختلف، داده ها ممکن است از جنبه‌های متفاوتی مثل: قراردادهای نامگذاری، خصوصیات فیزیکی داده و اندازه گیری متغیرها و غیره با هم نامتناسب باشند. این **ناسازگاری ها** باید حذف بشوند و داده های یک انبار داده ، یکپارچه بشوند. حالا چون ما میدانیم منابع داده متفاوت است، باید قبل از ذخیره سازی آن ها در **DW** از تکنیک‌های پاکسازی داده و یکپارچه سازی استفاده کنیم.

متغیر با زمان

سیستم‌های عملیاتی بخاطر **پشتیبانی** از عملیات هر روزه، مقادیر فعلی را نشان میدهند اما DW نشان دهنده داده هایی با مقیاس زمانی طولانی مدت‌تری می باشد. یعنی این که انبار داده شامل **داده های تاریخی** است .این انبار برای **داده کاوی** و **پیش بینی** استفاده میشود، مثلاً اگر کاربری دنبال یک الگوی خرید باشد باید داده هایی که مربوط به خریدهای الان و گذشته است را دنبال کند.

غیر فرّار

داده های داخل انبار ، خواندنی هستند! یعنی چی؟ یعنی اینکه لازم نیست انها را **تغییر** بدهید یا **آپدیت** کنید و نیازی به ایجاد و دسترسی **انحصاری** به داده ها نداریم، فقط به دو تا فعالیت کلیدی احتیاج داریم: باز کردن داده و دسترسی به داده.

جمع بندی

داده های Data Warehouse در سطح‌های متفاوتی **جمع بندی** میشوند. مثلاً کاربر DW اول به واحدهای فروش کلی یک محصول در یک منطقه نگاه می‌کند و سپس به آمار آن منطقه نگاهی می اندازد و در نهایت ممکن است فروشگاه‌های تکی را در منطقه خاصی بررسی کند. پس معمولاً تحلیل از **سطح‌های بالاتر** آغاز میشود و برای این که جزئیات را پیدا کنند به **سطح‌های پایینی** منتقل میشود.

مزایا و معایب Data Warehouse

مزایا

- به کاربران مشاغل اجازه میدهد تا به **سرعت** به داده های مهم بعضی منابع در همه جا دسترسی پیدا کنند.
- اطلاعات کاملی راجع به فعالیت‌های مختلف عملکردی ارائه میدهد و از گزارش موقت و پرس و جو **پشتیبانی** میکند.
- به ادغام خیلی از منابع داده کمک میکند تا **استرس** رو در سیستم تولید کاهش بدهند.
- به **کاهش زمان** کل گردش مالی برای تجزیه و تحلیل و گزارش دهی کمک میکند.
- تجدید ساختار و ادغام را برای گزارش و **تجزیه و تحلیل** آسان میکند.
- امکان دسترسی به داده های منابع موجود در یک مکان رو میدهد و موجب صرفه جویی در وقت کاربر برای **بازیابی** داده از چندتا منبع میشود.
- داده های تاریخی را ذخیره میکند و باعث میشود کاربر، با استفاده از اطلاعات دوره‌های مختلف، **پیش بینی دقیق‌تری** برای آینده داشته باشد.

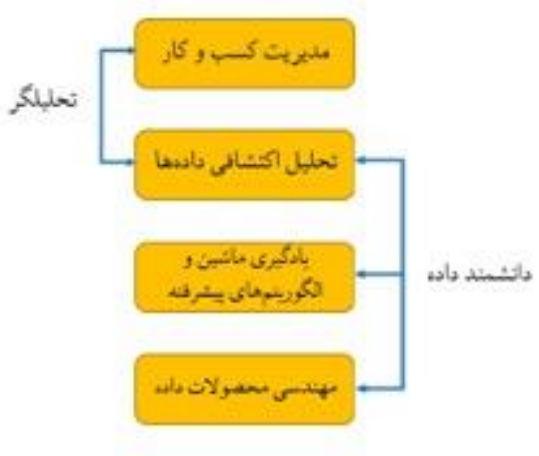
معایب

- گزینه ایده‌آلی برای داده های **بدون ساختار** نیست.
- ایجاد و **پیاده سازی** آن زمان بر و گیج کننده است.
- میتواند سریعاً **منسوخ** بشود.
- ایجاد تغییر در انواع داده، دامنه، طرح واره منابع داده، فهرست‌ها و نمایش داده ها **کار سختی** است.
- ممکن است آسان به نظر برسد ولی برای **کاربران سطح متوسط** خیلی پیچیده است.
- علیرغم بهترین تلاش‌ها در مدیریت پروژه، دامنه انبار داری داده ها همیشه **افزایش** پیدا میکند.
- گاهی وقت‌ها کاربران انبار داده، قوانین مختلف **کسب و کار** را برای خودشان توسعه میدهند!!
- سازمان باید **انرژی** زیادی را برای آموزش و اجرای اهداف صرف کند.

علم داده یا Data Science چیست؟

از مباحث مهم برای یادگیری هوش تجاری و درک بهتر آن Data Science می باشد. علم داده ترکیبی از ابزارهای مختلف، الگوریتم ها و اصول machine learning با **هدف کشف الگو های پنهان از داده های خام** است. اما چگونه این تفاوت با آن که سال هاست که به صورت آماری انجام می دهند، متفاوت است؟

پاسخ آن در تفاوت بین توضیح و پیش بینی نهفته است.



همانطور که از تصویر بالا مشاهده می کنید، یک تحلیلگر داده معمولاً با پردازش تاریخچه داده ها، آن چه اتفاق می افتد را توضیح می دهد. از طرف دیگر، Data Scientist نه تنها برای کشف بینش از آن، تحلیل اکتشافی انجام می دهد، بلکه از الگوریتم های مختلف پیشرفته machine learning برای شناسایی وقوع یک رویداد خاص در آینده استفاده می کند. یک تحلیل گر داده، دیتاها را از زوایای مختلف بررسی می کند، گاهی اوقات نیز حتی زاویه هایی که قبلاً شناخته نشده اند!

بنابراین، Data Science در درجه اول برای **تصمیم گیری و پیش بینی** با استفاده از "تجزیه و تحلیل عامل سببی پیش بینی"، "تجزیه و تحلیل تجویزی" (علم پیش بینی به علاوه تصمیم گیری) و "machine learning" استفاده می شود.

- **تجزیه و تحلیل عامل سببی پیش بینی** – اگر مدلی می خواهید که بتواند احتمالات یک رویداد خاص را در آینده پیش بینی کند، باید از تحلیل های علّی پیش بینی استفاده کنید. فرض کنید، اگر پول خود را به صورت اعتباری تأمین می کنید، پس احتمال این که مشتریان سر وقت پرداخت های اعتباری خود را انجام دهند، باعث نگرانی شما می شود. در اینجا، می توانید مدلی بسازید که بتواند تجزیه و تحلیل پیش بینی کننده تاریخ پرداخت مشتری را انجام دهد تا پیش بینی کند آیا پرداخت های آینده به موقع خواهند بود یا خیر.

- **تجزیه و تحلیل تجویزی** – اگر مدلی می خواهید که خود، هوش تصمیم گیری داشته باشد و توانایی اصلاح آن را با توجه به پارامتر های داینامیک داشته باشد، مطمئناً به تجزیه و تحلیل تجویزی برای آن نیاز دارید. این زمینه ی نسبتاً جدید، همه چیز به نحو ارائه مشاوره بستگی دارد. به عبارت دیگر، این نه تنها طیف وسیعی از اقدامات تجویز شده و نتایج مرتبط با آن را پیش بینی می کند، بلکه پیشنهاداتی نیز برای شما دارد.

بهترین مثال در این زمینه اتومبیل خودران Google است .این خودرو از داده های جمع آوری شده توسط وسایل نقلیه دیگر می تواند برای آموزش اتومبیل های خودران خود استفاده کند. می تواند الگوریتم هایی را بر روی این داده ها اجرا کند تا هوش خود را به خوبی به نمایش بگذارد. این امر باعث می شود اتومبیل شما بتواند تصمیماتی مانند زمان چرخش، انتخاب مسیر، زمان کاهش سرعت را اتخاذ کند.

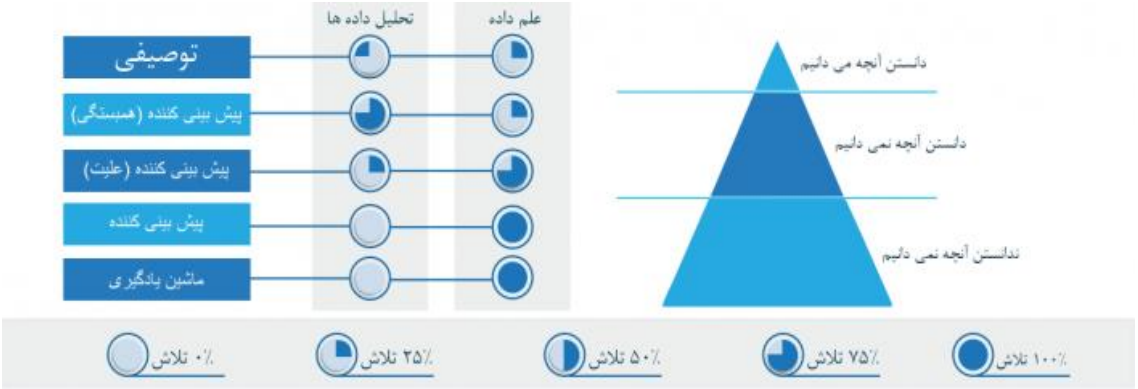
- **Machine learning برای پیش بینی** – اگر داده های معاملاتی یک شرکت مالی را دارید و باید مدلی برای تعیین روند آینده بسازید، الگوریتم های Machine learning بهترین گزینه هستند. این الگوی یادگیری تحت نظارت قرار می گیرد. دلیل این که نظارت نامیده می شود این است که شما قبلاً داده هایی داشته اید که بر اساس آن می توانید ماشین های خود را آموزش دهید. به عنوان مثال، یک مدل کشف تقلب می تواند با استفاده از سوابق تاریخی خرید های تقلبی آموزش ببیند.

- **Machine learning برای کشف الگو** – اگر پارامتر هایی ندارید که بتوانید بر اساس آن ها پیش بینی کنید، باید الگو های پنهان را در مجموعه داده ها پیدا کنید تا بتوانید پیش بینی های معنی داری داشته باشید. این چیزی نیست جز مدل بدون نظارت، زیرا هیچ بر چسب از پیش تعیین شده ای برای این گروه بندی ها ندارید. متداول ترین الگوریتم مورد استفاده برای کشف الگو، "خوشه بندی" است.

بگذارید بگوییم شما در یک شرکت تلفنی مشغول به کار هستید و باید با قرار دادن آنتن در یک منطقه، شبکه ای ایجاد کنید. حال می توانید از روش خوشه بندی برای یافتن مکان های آنتن استفاده کنید تا اطمینان حاصل شود که قدرت سیگنال مطلوب را برای همه کاربران دریافت می کنید.

حال بیایید ببینیم که نسبت رویکرد های فوق الذکر برای تجزیه و تحلیل داده ها و همچنین Data Science چه تفاوتی دارند. همان طور که در تصویر زیر مشاهده می کنید، تجزیه و تحلیل داده شامل تجزیه و تحلیل توصیفی و پیش بینی حدودی است. از طرف دیگر، Data Science بیشتر در مورد تجزیه و تحلیل علّی پیش بینی و Machine learning است.

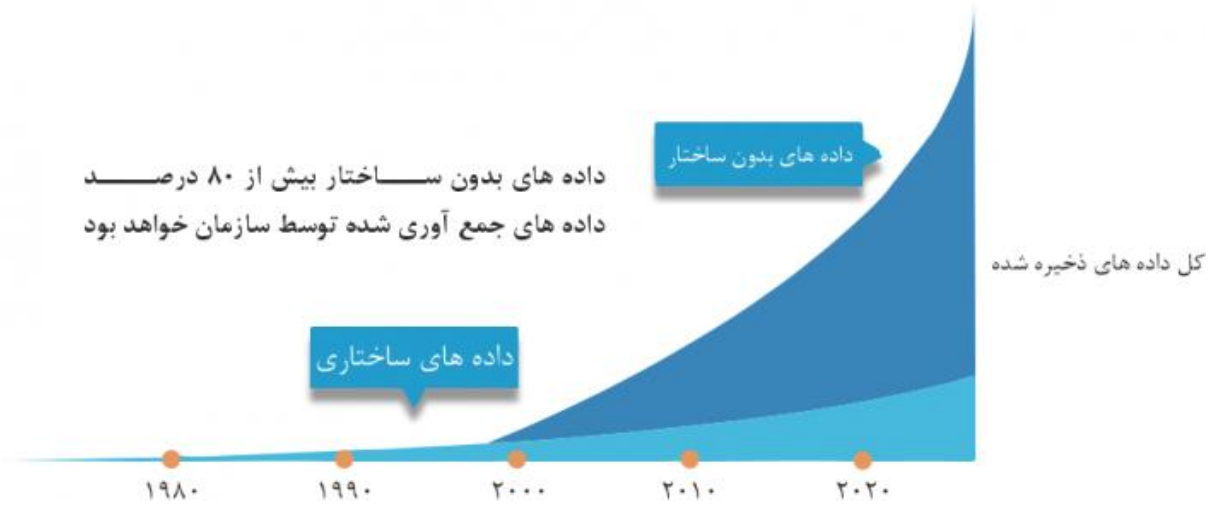
اکنون که می دانید Data Science دقیقاً چیست، حال وقتش است که چرایی نیاز ما به آن را بیشتر بشناسید.



چرا علم داده؟

به طور سنتی، داده هایی که ما در اختیار داشتیم بیشتر دارای ساختار و اندازه کوچک بودند که با استفاده از ابزار های ساده BI قابل تحلیل می باشند. بر خلاف داده ها در سیستم های سنتی که بیشتر ساختار داشتند، امروزه بیشتر داده ها بدون ساختار یا به اصطلاح نیمه ساختاری هستند. بیایید نگاهی به روند داده ها در تصویر زیر بیاندازیم که نشان می دهد تا سال ۲۰۲۰، بیش از ۸۰٪ داده ها بدون ساختار بودند.

این داده ها از منابع مختلف مانند لاگ های مربوط به پرونده های مالی، پرونده های متنی، فرم های چند رسانه ای، حسگر ها و ابزار ها تولید می شوند. ابزار های ساده BI قادر به پردازش این **حجم عظیم و تنوع داده** نیستند. به همین دلیل است که برای پردازش، تجزیه و تحلیل و ترسیم بینش معنا دار از آن، به ابزار ها و الگوریتم های تحلیلی پیچیده و پیشرفته تری نیاز داریم.



حال اگر بتوانید از اطلاعات موجود مانند سابقه مرور گذشته مشتری، سابقه خرید، سن و درآمد مشتری، نیاز های دقیق مشتریان خود را درک کنید، بدون شک شما همه این داده ها را زودتر نیز بدست آورده اید، اما اکنون با حجم گسترده و تنوع داده ها، می توانید مدل ها را به طور موثرتری آموزش دهید و محصول را با دقت بیشتری به مشتریان خود توصیه کنید. آیا شگفت آور نیست؟ آیا این مسئله تجارت بیشتری را برای سازمان شما ایجاد نمی کند؟

بیایید سناریویی متفاوت برای درک نقش Data Science در تصمیم گیری در نظر بگیریم. حال اگر ماشین شما از هوش لازم برای رسیدن شما به خانه برخوردار باشد، چطور؟ اتومبیل های خودران، داده های زنده حسگرها از جمله رادارها، دوربین ها و لیزرها را برای ایجاد نقشه از محیط اطراف خود جمع می کنند. بر اساس این داده ها، تصمیماتی مانند زمان افزایش سرعت، زمان کاهش سرعت، زمان سبقت، دور زدن اتخاذ می شود که همه این موارد با استفاده از الگوریتم های پیشرفته machine learning امکان پذیر می شود.

بیایید ببینیم چگونه Data Science می تواند در تجزیه و تحلیل های پیش بینی شده استفاده شود. بیایید این بار پیش بینی کردن هوا را به عنوان مثال در نظر بگیریم. داده های کشتی ها، هواپیما ها، رادار ها، ماهواره ها را می توان جمع آوری و برای ساخت مدل تجزیه و تحلیل از آن ها استفاده کرد. این مدل ها نه تنها آب و هوا را پیش بینی می کنند بلکه به پیش بینی وقوع هر گونه بلایای طبیعی نیز کمک می کنند .این به شما کمک می کند که از قبل اقدامات مناسبی را تعبیه کنید و جان بسیاری از افراد را از پیش نجات دهید. بیایید نگاهی به نکات زیر بیندازیم تا تمام زمینه هایی را که Data Science در حال ساخت آن است را ببینیم.

۱. **مسافرت**
   - قیمت گذاری هوشمند
   - پرواز امروز را پیش بینی کنید
۲. **بازاریابی**
   - گران فروشی
   - فروش متقابل
   - پیش بینی ارزش طول عمر مشتری
   - ریزش
۳. **مراقبت های بهداشتی**
   - پیش بینی بیماری
   - اثربخشی دارو
۴. **شبکه های اجتماعی**
   - تجزیه و تحلیل احساسات
   - بازاریابی دیجیتال
۵. **فروش**
   - پیشنهاد تخفیف
   - پیش بینی تقاضا
۶. **اتوماسیون**
   - اتومبیل های خودران
   - هواپیمای بدون خلبان، پهپاد
۷. **اعتبار و بیمه**
   - ادعای پیش بینی
   - کشف تقلب و ریسک

تفاوت BI و Data Science

اکنون اجازه بدهید که درباره BI بحث کنیم. حتما تا به حال اصطلاح Business Intelligence یا همان هوش تجاری به گوش شما خورده است. غالبا Data Science با BI اشتباه گرفته می شود. تضادهای مختصر و واضحی بین این دو وجود دارد که به شما در درک بهتر هر دو کمک می کند. موارد پایین را با دقت نگاه کنید:

BI چیست؟

هوش تجاری (BI) اساساً **داده های قبلی موجود را تجزیه و تحلیل می کند** تا بینش حال و آینده برای توصیف روند کسب و کار پیدا کند. در اینجا BI به شما این امکان را می دهد تا داده ها را از منابع خارجی و داخلی تهیه کنید، آن ها را آماده کنید، queryها را روی آن ها پیاده سازی کنید و داشبورد ایجاد کنید تا به سوالاتی مانند تجزیه و تحلیل درآمد سه ماهه یا مشکلات تجاری پاسخ دهید BI .می تواند تأثیر وقایع خاص را در آینده نزدیک برای شما ارزیابی کند.

Data Science چیست؟

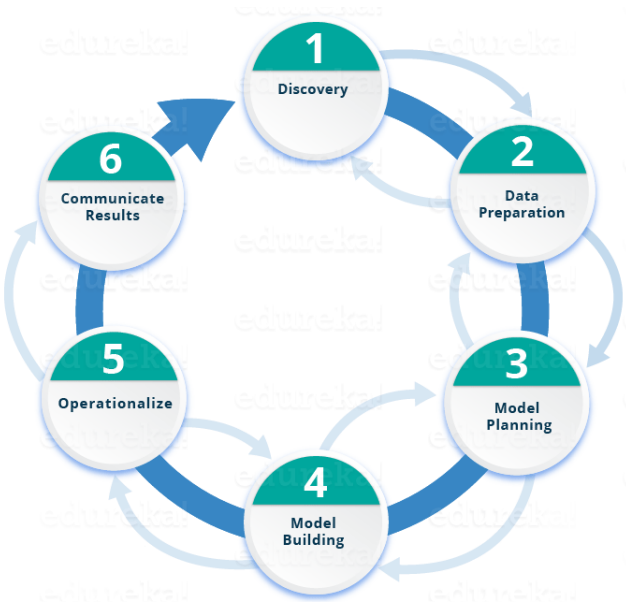
Data Science یک **رویکرد آینده نگرانه تر** است؛ روشی اکتشافی با تمرکز بر تجزیه و تحلیل داده های گذشته یا فعلی و پیش بینی نتایج آینده با هدف تصمیم گیری آگاهانه است. این سوالات راجع به وقایعی که با "چه" و "چگونه" شروع می شوند، پاسخ می دهد.

بیایید نگاهی به برخی از ویژگی های متضاد آن ها بیندازیم.

ویژگی ها	هوش تجاری BI	علم داده Data Science
رویکرد	آمار و بصری سازی	آمار ، یادگیری ماشین ، تجزیه و تحلیل نمودار ، برنامه نویسی عصبی(NLP)
تمرکز	گذشته و حال	حال و آینده
ابزارها	Pentaho, Microsoft BI, QlikView, R	RapidMiner, BigML, Weka, R
منابع داده	ساختارمند معمولاً (Microsoft SQL)	ساختار یافته و ساختار نیافته (logs ، فضای ابری , SQL, NoSQL، متن)

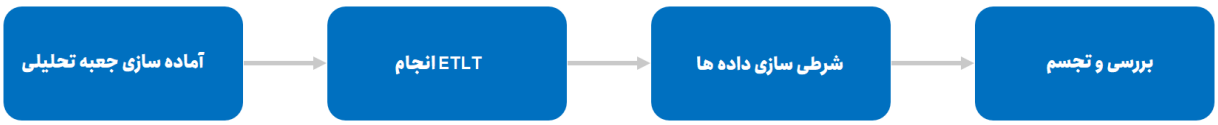
همه این توضیحات تا به حال راجع به خود Data Science بود، حالا بیایید چرخه زندگی Data Science را با هم بهتر درک کنیم. یک اشتباه متداول در پروژه هایData Science ، عجله در جمع آوری و تجزیه و تحلیل داده ها، بدون درک نیازها و یا حتی طرح صحیح برای مشکل تجاری مورد نظر است. بنابراین، برای شما بسیار مهم است که برای اطمینان از عملکرد پیوسته و روان پیش رفتن پروژه، تمام مراحل را در طول چرخه عمر Data Science دنبال کنید.

چرخه زندگیData Science



فاز ۱ – کشف: قبل از شروع پروژه، درک مشخصات مختلف، الزامات، اولویت ها و بودجه مورد نیاز مهم است. شما باید توانایی پرسیدن سوالات صحیح را داشته باشید. در اینجا، شما ارزیابی می کنید که آیا منابع مورد نیاز از نظر افراد، فناوری، زمان و داده ها برای پشتیبانی از پروژه مد نظر خود را دارید یا خیر. در این مرحله، شما همچنین باید مسئله مشاغل را تنظیم کرده و فرضیه های اولیه را برای آزمایش فرموله کنید.

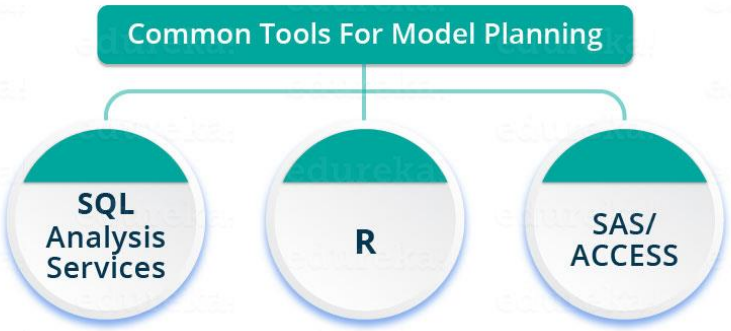
فاز ۲ – آماده سازی داده ها: در این مرحله شما به یک Sand Box تحلیلی نیاز دارید که در آن بتوانید تجزیه و تحلیل را برای کل مدت پروژه انجام دهید .قبل از مدل سازی، باید داده ها را کشف، پیش پردازش و شرایط آن ها را تنظیم کنید. بعلاوه، شما ETLT (استخراج، تبدیل، بارگذاری و تبدیل) را برای ورود داده ها به Sand Box انجام خواهید داد. بیایید نگاهی به جریان تجزیه و تحلیل آماری در زیر بیندازیم.



برای خالص تر کردن، تبدیل و مجسم سازی داده ها می توانید از R استفاده کنید .این به شما کمک می کند تا با دید بهتر و بزرگ تر مسئله را بررسی کنید و بین متغیر ها رابطه برقرار کنید. پس از پاک کردن و آماده سازی داده ها، وقت آن رسیده است که تجزیه و تحلیل اکتشافی را روی آن انجام دهید. بیایید ببینیم که چگونه می توانید به آن دست پیدا کنید.

فاز ۳ – برنامه ریزی مدل: در اینجا، روش ها و تکنیک های ترسیم روابط بین متغیر ها را تعیین خواهید کرد. این روابط اساس الگوریتم هایی را ایجاد می کند که در مرحله بعدی پیاده سازی می کنید. شما از تجزیه و تحلیل داده های اکتشافی (EDA) با استفاده از فرمول های آماری مختلف و ابزار تجسم استفاده خواهید کرد.

بیایید نگاهی به ابزار های مختلف برنامه ریزی مدل بیندازیم.



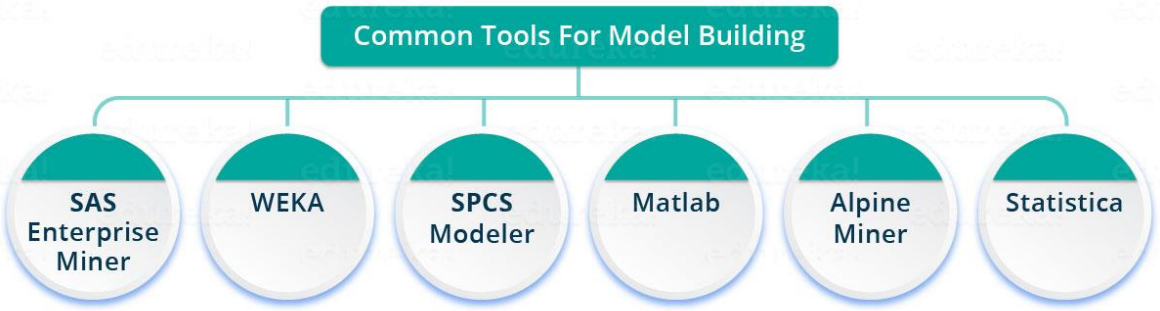
۱. R دارای مجموعه کاملی از قابلیت های مدل سازی است و فضای خوبی برای ساخت مدل های تفسیری فراهم می کند.
۲. سرویس های SQL Analysis می توانند با استفاده از توابع داده کاوی متداول و مدل های اساسی پیش بینی، تجزیه و تحلیل database را انجام دهند.
۳. SAS / ACCESS می تواند برای دسترسی به داده ها از Hadoop استفاده کند و برای ایجاد نمودار های جریان، مدل تکرار پذیر و قابل استفاده مجدد بسازد.

اگر چه ابزار های زیادی در بازار وجود دارد اما R پر کاربرد ترین ابزار است.

اکنون که اطلاعاتی درباره ماهیت داده های خود پیدا کرده اید و تصمیم گرفته اید که چه الگوریتم هایی مورد استفاده قرار گیرند، در مرحله بعدی، شما الگوریتم را اعمال کرده و یک مدل ایجاد می کنید.

فاز ۴ – ساخت مدل: در این مرحله، مجموعه داده هایی را برای اهداف آموزشی و آزمایشی ایجاد خواهید کرد. در اینجا باید بررسی کنید که آیا ابزار های موجود شما برای اجرای مدل ها کافی می باشند یا به محیط مستحکم تری نیاز دارند (مانند پردازش سریع و موازی). برای ساخت مدل، روش های مختلف یادگیری مانند طبقه بندی، تداعی و خوشه بندی را تجزیه و تحلیل خواهید کرد.

شما می توانید از طریق ابزار های زیر به ساخت مدل بپردازید.



فاز ۵ — عملیاتی سازی: در این مرحله، شما گزارش های نهایی، جلسات توجیهی، کد و اسناد فنی را ارائه می دهید. علاوه بر این، گاهی اوقات یک پروژه آزمایشی نیز در یک محیط تولید در real-time اجرا می کنید. با این کار قبل از استقرار کامل، تصویری واضح از عملکرد و سایر محدودیت های مربوطه در مقیاس کوچک به شما ارائه می شود.

فاز ۶ – نتایج را اعلام کنید: اکنون ارزیابی اینکه آیا توانسته اید به هدفی که در مرحله اول برنامه ریزی کرده اید، برسید مهم است. بنابراین، در آخرین مرحله، شما تمام یافته های کلیدی را شناسایی می کنید، با سهامدارانتان ارتباط بر قرار می کنید و بر اساس معیار های تدوین شده در فاز ۱، موفقیت یا شکست نتایج پروژه را بررسی و تعیین می کنید.

مطالعه موردی: پیشگیری از دیابت

"اگر بتوانیم وقوع دیابت را پیش بینی کنیم و قبل از آن اقدامات مناسبی برای جلوگیری از آن انجام دهیم، چه می کنیم؟"

در این مورد، ما وقوع دیابت را با استفاده از کل چرخه زندگی که قبلاً بحث کردیم پیش بینی خواهیم کرد. اجازه دهید مرحله به مرحله بررسی کنیم.

مرحله ۱:

- در ابتدا، ما داده ها را بر اساس سابقه پزشکی بیمار همانطور که در فاز ۱ بحث شد، جمع آوری خواهیم کرد. می توانید به نمونه داده های زیر مراجعه کنید.

;npreg;glu;bp;skin;bmi;ped;age,income
1;6;148;72;35;33.6;0.627;50
2;1;85;66;29;26.6;0.351;31
3;1;89;80;23;28.1;0.167;21
4;3;78;50;32;31;0.248;26
5;2;197;70;45;30.5;0.158;53
6;5;166;72;19;25.8;0.587;51
7;0;118;84;47;45.8;0.551;31
8;1;103;30;38;43.3;0.183;33
9;3;126;88;41;39.3;0.704;27
10;9;119;80;35;29;0.263;29
11;1;97;66;15;23.2;0.487;22
12;5;109;75;26;36;0.546;60
13;3;88;58;11;24.8;0.267;22
14;10;122;78;31;27.6;0.512;45
15;4;97;60;33;24;0.966;33
16;9;102;76;37;32.9;0.665;46
17;2;90;68;42;38.2;0.503;27
18;4;111;72;47;37.1;1.39;56
19;3;180;64;25;34;0.271;26
20;7;106;92;18;39;0.235;48
21;9;171;110;24;45.4;0.721;54

- همانطور که می بینید، ما ویژگی های مختلفی داریم که در زیر ذکر شده است.

ویژگی ها:

- ✓ npreg تعداد دفعات بارداری
- ✓ glucose غلظت گلوکز پلاسما
- ✓ bp فشار خون
- ✓ skin ضخامت پوست بند سه سر
- ✓ bmi شاخص توده بدن
- ✓ ped عملکرد شجره نامه دیابت
- ✓ age سن
- ✓ income درآمد

مرحله ۲:

- اکنون، پس از دستیابی به داده ها، باید دیتا را برای تجزیه و تحلیل مرتب و آماده کنیم.
- این داده ها دارای تناقضات زیادی مانند مقادیر از دست رفته، ستون های خالی، مقادیر ناگهانی و قالب داده نادرست است که باید مرتب شوند.
- در اینجا، ما داده ها را در یک جدول واحد تحت ویژگی های مختلف سازماندهی کرده ایم که ساختار سازی آن ها بهتر به نظر می رسد.

- بیایید نگاهی به نمونه داده های زیر بیندازیم.

	npreg	glu	bp	skin	bmi	ped	age	income
1	6	148	72	35	33.6	0.627	50	
2	1	85	66	29	26.6	0.351	31	
3	1	89	6600	23	28.1	0.167	21	
4	3	78	50	32	31	0.248	26	
5	2	197	70	45	30.5	0.158	53	
6	5	166	72	19	25.8	0.587	51	
7	0	118	84	47	45.8	0.551	31	
8	one	103	30	38	43.3	0.183	33	
9	3	126	88	41	39.3	0.704	27	
10	9	119	80	35	29	0.263	29	
11	1	97	66	15	23.2	0.487	22	
12	5	109	75	26	36	0.546	60	
13	3	88	58	11	24.8	0.267	22	
14	10	122	78	31	27.6	0.512	45	
15	4		60	33	24	0.966	33	
16	9	102	76	37	32.9	0.665	46	
17	2	90	68	42	38.2	0.503	27	
18	4	111	72	47	37.1	1.39	56	
19	3	180	64	25	34	0.271	26	
20	7	106	92	18		0.235	48	
21	9	171	110	24	45.4	0.721	54	

این داده ها ناسازگاری زیادی دارند.

۱. در ستونnpreg ، ”one“با حروف نوشته شده است، در حالی که باید به شکل عددی مانند ۱ باشد.
۲. در ستون bp یکی از مقادیر ۶۶۰۰ است که حداقل (برای انسان) غیر ممکن است; زیرا bp نمی تواند به چنین مقدار عظیمی برسد.
۳. همان طور که مشاهده می کنید ستون درآمد خالی است و همچنین در پیش بینی دیابت معنی ندارد. بنابراین وجود آن در اینجا زائد است و باید از جدول حذف شود.

بنابراین، ما با حذف اشتباهات، پر کردن مقادیر صفر و درست کردن نوع داده، این داده ها را مرتب و پیش پردازش می کنیم. اگر به یاد داشته باشید، این مرحله دوم ما است که پیش از پردازش داده انجام می شود.

در آخر، داده های سالم را همانطور که در زیر نشان داده شده است، بدست می آوریم که می توانند برای تجزیه و تحلیل استفاده شوند.

	npreg	glu	bp	skin	bmi	ped	age
1	6	148	72	35	33.6	0.627	50
2	1	85	66	29	26.6	0.351	31
3	1	89	80	23	28.1	0.167	21
4	3	78	50	32	31	0.248	26
5	2	197	70	45	30.5	0.158	53
6	5	166	72	19	25.8	0.587	51
7	0	118	84	47	45.8	0.551	31
8	1	103	30	38	43.3	0.183	33
9	3	126	88	41	39.3	0.704	27
10	9	119	80	35	29	0.263	29
11	1	97	66	15	23.2	0.487	22
12	5	109	75	26	36	0.546	60
13	3	88	58	11	24.8	0.267	22
14	10	122	78	31	27.6	0.512	45
15	4	97	60	33	24	0.966	33
16	9	102	76	37	32.9	0.665	46
17	2	90	68	42	38.2	0.503	27
18	4	111	72	47	37.1	1.39	56
19	3	180	64	25	34	0.271	26
20	7	106	92	18	39	0.235	48
21	9	171	110	24	45.4	0.721	54

مرحله ۳:

حال اجازه دهید مقداری تجزیه و تحلیل انجام دهیم، همانطور که قبلاً در فاز ۳ بحث شد.

- ابتدا داده ها را بارگذاری می کنیم و توابع آماری مختلفی را روی آن اعمال می کنیم. به عنوان مثال، زبان R تابعی مانند describe دارد که تعداد مقادیر از دست رفته و مقادیر منحصر به فرد را به ما ارائه می دهد. همچنین می توانیم از تابع summary استفاده کنیم که اطلاعات آماری مانند مقادیر میانگین، دامنه، حداقل و حداکثر را به ما می دهد.

- سپس، ما از تکنیک های بصری مانند هیستوگرام(Histograms) ، نمودار های خطی(line graphs) ، نمودار های جعبه ای (box plots) استفاده می کنیم تا ایده مناسبی از توزیع داده ها بدست آوریم.

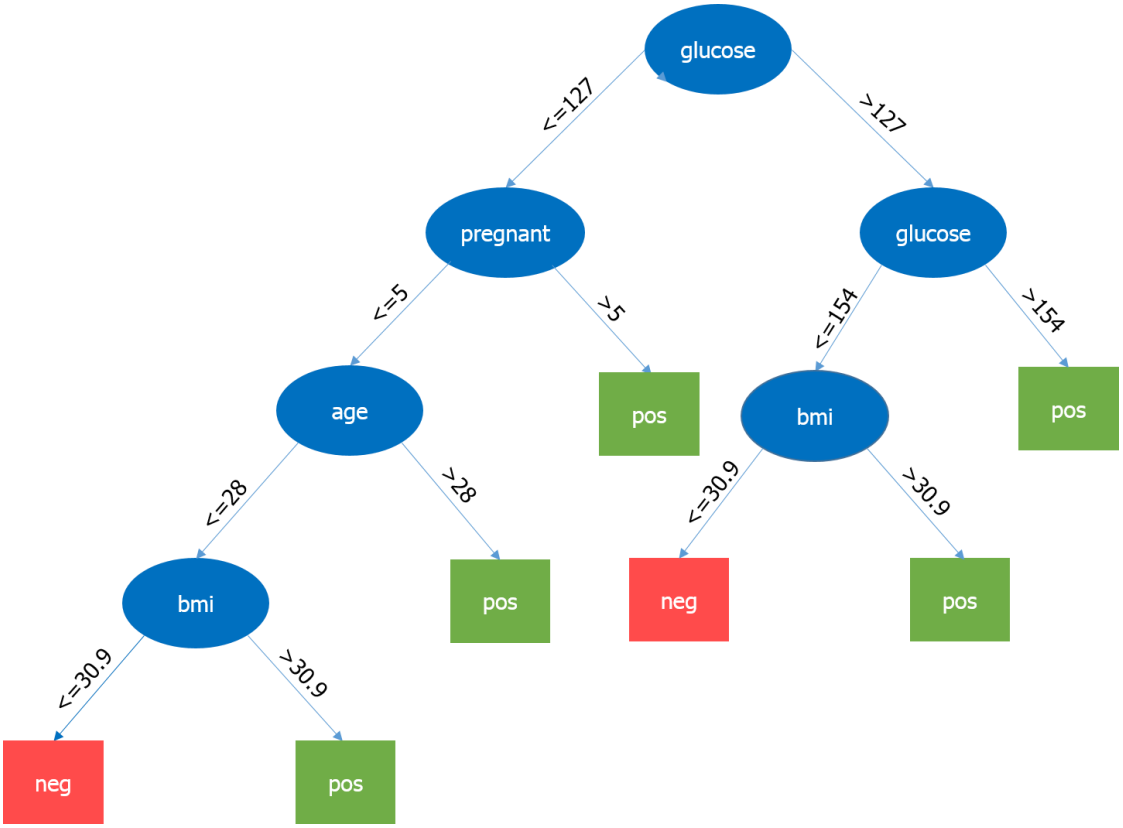
مرحله ۴:

اکنون، بر اساس بینش های حاصل از مرحله قبلی، بهترین گزینه برای این نوع مشکلات درخت تصمیم (decision tree) است:

- از آن جا که، ما در حال حاضر ویژگی های اصلی برای تجزیه و تحلیل مانند npreg، bmiو ... را در اختیار داریم، بنابراین ما از روش یادگیری نظارت شده برای ساختن یک مدل در اینجا استفاده خواهیم کرد.

- بعلاوه، ما به ویژه از درخت تصمیم استفاده کرده ایم زیرا همه ویژگی ها را یک جا مورد توجه قرار می دهد، مانند خصوصیاتی که رابطه خطی و هم چنین رابطه غیر خطی دارند. در این مورد، ما یک رابطه خطی بین npreg و سن داریم، در حالی که رابطه غیر خطی بین npreg و ped هم وجود دارد.
- مدل های درخت تصمیم نیز بسیار قوی هستند زیرا می توانیم از ترکیبات مختلف ویژگی ها برای ساختن درخت تصمیم های مختلف استفاده کنیم و در نهایت یکی را که حداکثر کارایی را دارا می باشد، پیاده سازی کنیم.

بیایید نگاهی به درخت تصمیم خود بیندازیم.



در این جا، مهم ترین پارامتر سطح گلوکز است، بنابراین ریشه ما یک گره است. اکنون، گره فعلی و مقدار آن پارامتر مهم بعدی را تعیین می کند. این کار ادامه می یابد تا زمانی که از نظر مثبت یا منفی نتیجه بگیریم Pos .به این معنی است که تمایل به دیابت مثبت است و منفی به معنای منفی بودن دیابت می باشد.

مرحله ۵:

در این مرحله، ما یک پروژه آزمایشی کوچک را برای بررسی مناسب بودن نتایج خود اجرا خواهیم کرد. همچنین در صورت وجود به دنبال محدودیت های عملکردی خواهیم بود. اگر نتایج دقیق نباشند، باید مدل را دوباره طراحی و بازسازی کنیم.

مرحله ۶:

هنگامی که پروژه را با موفقیت اجرا کردیم، خروجی را برای گسترش کامل به اشتراک خواهیم گذاشت.

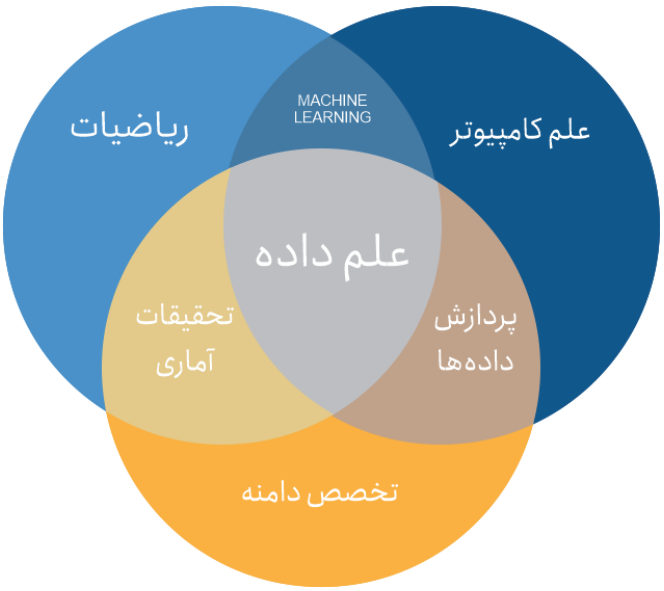
تحلیلگر داده Data Scientist کیست؟

تعاریف مختلفی در مورد Data Scientist ها (دانشمند داده) وجود دارد. به عبارتی دیگر، Data Scientist کسی است که هنر Data Science را تمرین می کند. اصطلاح "Data Scientist" پس از در نظر گرفتن این واقعیت که دانشمند داده، اطلاعات زیادی را از زمینه ها و برنامه های علمی اعم از آمار یا ریاضیات به دست می آورد، ابداع شده است.

Data Scientist چه کاری انجام می دهد؟

Data Scientist ها کسانی هستند که کاربرد علم داده را نشان می دهند و با تخصص قوی خود در برخی از رشته های علمی، مشکلات پیچیده مرتبط به داده را حل می کنند. آن ها با چندین عنصر مرتبط با ریاضیات، آمار، علوم کامپیوتر و غیره کار می کنند (اگر چه ممکن است در همه این زمینه ها متخصص نباشند). آن ها از آخرین فناوری ها در یافتن راه حل و نتیجه گیری برای رشد و توسعه سازمان بسیار آگاه هستند و از آن ها استفاده می کنند. Data Scientist داده ها را به شکل بسیار مفید تری در مقایسه با داده های خام موجود از فرم های ساختار یافته و غیر ساختاری ارائه می دهند.

دانشمند دیتا بودن، خیلی سخت تر از چیزی که به نظر می رسد، است. بنابراین، بیایید ببینیم برای دانشمند بودن چه چیزی نیاز داریم. یک دانشمند داده به مهارت هایی که در زیر نشان داده شده است نیاز دارد.



همان طور که در تصویر بالا مشاهده می کنید، شما باید مهارت های سخت و مختلفی را کسب کنید. برای تجزیه و تحلیل و تجسم داده ها، باید در آمار و ریاضیات مهارت کافی داشته باشید. نیازی به گفتن نیست که Machine Learning قلب علم داده را تشکیل می دهد و از شما می خواهد در آن مهارت داشته باشید. همچنین، شما باید درک درستی از دامنه ای که در آن کار می کنید داشته باشید تا مشکلات تجاری را به وضوح درک کنید .وظیفه شما به اینجا ختم نمی شود. شما باید بتوانید الگوریتم های مختلفی را که نیاز به مهارت های کدگذاری خوبی دارند، پیاده سازی کنید. سر انجام، هنگامی که تصمیمات اساسی خاصی را اتخاذ کردید، مهم است که آن ها را به سهام داران تان تحویل دهید. پس ارتباطات اجتماعی نیز جزو اساسی ترین نیاز ها می باشد.

داده کاوی چیست؟

داده‌کاوی، استخراج دانش و افزایش آگاهی از داده‌هایی است که در حالت کلی معنا و مفهوم خاصی برای کاربر ندارد. به بیان دیگر هدف داده‌کاوی این است که از میان داده‌هایی با حجم و وسعت زیاد، پیچیده و پیشرفته بتواند اطلاعات مفیدی را به دست آورد و آن را در اختیار ما قرار دهد.

تفاوت علم داده با داده‌کاوی چیست؟

علم داده بسیار گسترده بوده و شامل تمام موارد مدل‌سازی، ریاضیات، آمار، آنالیز داده‌ها و حتی داده‌کاوی است، اما داده‌کاوی زیرمجموعه‌ای از این علم بزرگ است که هدف آن استخراج اطلاعات مفید از بین داده‌های عظیم است. در واقع علم داده یا data science یک **علم و رشته** است و داده‌کاوی **تکنیکی** است که با کمک علم داده، اطلاعات موردنیاز را استخراج می‌کند.

نتیجه و خروجی علم داده می‌تواند بنا بر نیازی که وجود دارد، متفاوت باشد. در واقع هدف از استفاده علم داده، **ساخت محصولاتی داده‌محور برای یک سازمان است** که استفاده از آن بتواند میزان فروش، سود خالص و تعداد مشتریان وفادار را افزایش دهد و ایرادات موجود در روند کاری سازمان را شناسایی و برطرف کند. اما خروجی مورد انتظار از داده‌کاوی، **الگوریتم‌هایی است که می‌توان با کمک آن، داده‌های قابل استفاده و مفید را استخراج کرد.** در واقع با وجود تفاوت‌های زیادی که علم داده و داده‌کاوی با یکدیگر دارند، در نهایت **مکمل** هم هستند و داده‌کاوی **زیرمجموعه‌ای** از این علم گسترده و بی‌نظیر محسوب می‌شود.

منابع:

https://virakam.com/%D۸۰%AD%D۸۰%A۷%DA%A۹%D۹۰%۸۵۰%۹DB۸۰%C%D۸۰%AA-%D۸۰%AF%D۸۰%A۸۰%۷D۸۰%AF%D۸۷۰%۹-%DA۸۰%۸۶۰%DB۸۰%C%D۸۰%۸B۰%۳D۸۰%AA/

https://nafisbi.com/%D۸۰%AD%D۸۰%A۷۰%DA%A۹۰%D۹۰%۸۵۰%DB۸۰%۸C%D۸۰%AA-%D۸۰%AF%D۸۰%A۷۰%D۸۰%AF%D۹۰%۸۷/

https://nor.co.ir/blog/data-governance/

https://peivast.com/p/۱۰۶۸۰۰

https://www.sahab.ir/events/event-۱۰/

http://hesfa.ir/%D۸۰%AD%D۸۰%A۷۰%D۹۰%۸۳۰%D۹۰%۸۵۰%D۹۰%۸A%D۸۰%AA-%D۸۰%AF%D۸۰%A۷۰%D۸۰%AF%D۹۰%۸۷/

https://ponisha.ir/blog/%D۸۰%B۹۰%D۹۰%۸۴۰%D۹۰%۸۵-%D۸۰%AF%D۸۰%A۷۰%D۸۰%AF%D۹۰%۸۷/