

دریاچه داده : معماری جدید برای زیرساخت تحلیل و پردازش اطلاعات

در دنیای Big Data، مدتیست که عبارت دریاچه داده به یک اصطلاح فراگیر در حوزه معماری‌های پیشنهادی برای پردازش داده‌ها تبدیل شده است.

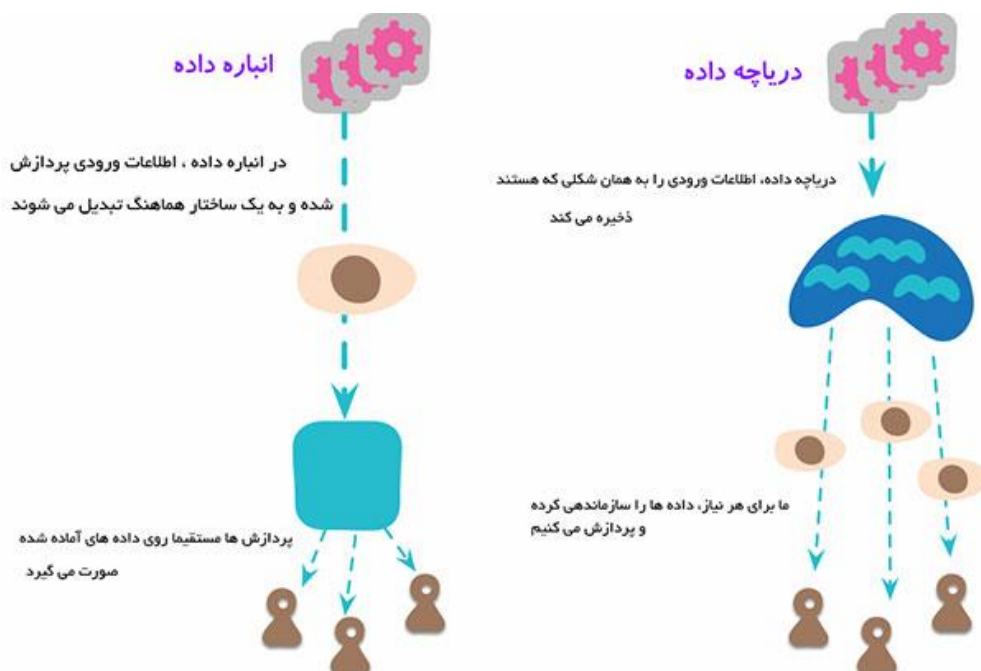
دریاچه داده از دید مفهومی، یک مکان مشترک برای ذخیره انواع داده‌های یک سازمان به صورت خام و پردازش نشده است.

برخلاف انباره‌های داده که به صورت خاص برای مقاصد تحلیل و گزارش‌گیری‌های مدیریتی طراحی می‌شوند و شکل داده‌های ورودی را به قالب مورد نیاز خود تبدیل و ذخیره می‌کنند، دریاچه‌های داده، اطلاعات و داده‌ها را به صورت خام ذخیره می‌کنند تا در آینده، ابزارهای پردازشی و تحلیل‌گران، خود تصمیم بگیرند که چه تبدیلی و چه پردازشی را روی آنها انجام بدهند.

از طرفی، انباره‌های داده دارای قالب و ساختار از پیش تعیین شده هستند و این موضوع، انواع داده‌های نوینی که روزانه در سازمانها تولید می‌شود را برای ذخیره به منظور مقاصد تحلیلی به چالش دچار می‌کند چون یکسره نیاز به تیمی برای مدیریت و طراحی ساختار و تبدیل داده‌ها خواهیم بود. در صورتیکه در دریاچه داده، داده‌ها به همان شکلی که هستند (مثلا داده‌های خام شبکه، حسگرها، متون استخراج شده از شبکه‌های اجتماعی و ...)، ذخیره می‌شوند و نیاز مداوم به پالایش ندارند.

نکته دیگری که در مورد انباره‌های داده وجود دارد این است که در این سامانه‌ها، اطلاعات به صورت تجمعی ذخیره می‌شوند مثلا فروش ماهیانه و فروش هفتگی یک محصول اما فروش ساعتی یا لحظه‌ای آن دیگر ذخیره نمی‌شود و از بین می‌رود که دریاچه داده سازمانی، این نقیصه را هم با حفظ داده‌های اصلی، مرتفع می‌سازد.

مارتین فاولر به عنوان یکی از پیشروهای مهندسی نرم افزار و مباحث مدیریت داده در مقاله‌ای که در زمینه دریاچه داده در سایت خود منتشر کرده است این تفاوت را به خوبی با شکل زیر نمایش داده است:





پردازش اطلاعات

رابطه‌ای که میان دریاچه داده و سایر داده‌ها و نرم افزارهای یک سازمان می‌توان پیشنهاد کرد این است که به عنوان یک محیط جنبی و کاملاً مجزا از داده‌های عملیاتی، به ذخیره داده‌ها و اطلاعات در قالب‌های باز یا بانک‌های اطلاعاتی No-SQL پردازد و تحلیل گران از روی این داده‌ها، ساختارهای تحلیلی خود را طراحی و اجرا کنند. دریاچه داده هم مثل بسیاری دیگر از فناوری‌های حوزه کلان داده تا به بلوغ رسیدن و رفع اشکالات و نهادینه شدن در سازمانها، راه درازی را در پیش دارد.