

جهانی از داده

پا داده هایی به وسعت جهان

مؤسسه تحقیقاتی گارتنر، دارایی‌های اطلاعاتی با حجم بالا، شتاب بالا و تنوع بالا را که نیازمند اشکال جدید پردازشی است و تصمیم‌گیری پیشرفته، کشف بینش و بهینه‌سازی فرآیند را امکان‌پذیر می‌سازد، "کلان داده" (Data Big) می‌نامد.

داده‌های کلان، داده‌هایی هستند که اندازه آنها فراتر از حدی است که با نرم افزارهای عادی امکانپذیر باشد و معمولاً به مجموعه‌ای از داده‌ها گفته می‌شود که اندازه آن‌ها فراتر از حدی است که بتوان عملیات دریافت آن‌ها را با نرم‌افزارهای معمول و در زمانی معقول انجام داد و بتوان مواردی چون دقیق سازی، مدیریت و پردازش را روی این داده‌ها پیاده سازی کرد. مهمترین ویژگی داده‌های بزرگ این است که مفهوم مقدار یا اندازه در این داده‌ها به‌طور مستمر در حال تغییر است و به‌مرور بزرگ‌تر می‌شود. کلان داده‌ها مجموعه‌ای از تکنیک‌ها و تاکتیک‌هایی است که شکل جدیدی از یکپارچگی را طلب می‌کند. چنانچه یکپارچگی حاصل شود، این امیدواری وجود دارد که ارزش‌های بزرگی را که در مجموعه‌های بزرگ، پیچیده و متنوع داده پنهان شده‌اند، آشکار سازند.

سرچشمه و بستر تولید داده‌های کلان معمولاً شبکه و رسانه‌های اجتماعی، اینترنت اشیا، توسعه دسترسی نسل‌های جدید تلفن همراه و ثابت، توسعه خدمات نوین الکترونیکی و توسعه کاربردهای فناوری اطلاعات و ارتباطات در بخش‌های مختلف از جمله صنعت بانکداری است. رشد داده‌ها نیز همچون سایر پدیده‌ها در حوزه فناوری، چالش‌ها و فرصت‌هایی برای استفاده کنندگان خود ایجاد می‌کند از این بین می‌توان به افزایش در میزان داده، افزایش سرعت تولید داده‌های ورودی و خروجی و افزایش محدوده، تنوع و منابع داده‌ها اشاره کرد.

۵ ویژگی کلان داده‌ها

حجم بالا:

کمیت داده‌هایی که جمع‌آوری می‌شود اهمیت زیادی دارند. اندازه داده‌ها مشخص می‌کند که آیا این داده‌ها پتانسیل کلان داده تلقی شدن را دارند یا خیر؟

تنوع داده‌ها:

یکی دیگر از ویژگی‌های داده‌های بزرگ تنوع داده‌ها است؛ اینکه داده‌ها مربوط به چه حوزه‌ای است و چه مفهومی را پشتیبانی می‌کند. اینکه چقدر نیاز به تجزیه و تحلیل داده‌ها وجود دارد، در تعیین آن‌ها به عنوان داده عظیم نقش دارد.

سرعت تولید داده:

منظور از سرعت تولید داده: میزان داده تولیدشده در یک بازه زمانی یا به عبارت بهتر، سرعت ایجاد داده‌های جدید است. موضوع سرعت تولید اطلاعات در سامانه‌های رایانه‌ای به سرعت در حال افزایش است. آمارها نشان می‌دهد در سال ۲۰۱۰ سرعت تولید اطلاعات به حدی رسید که در هر دو روز، بیش از کل داده‌هایی که تا سال ۲۰۰۳ تولیدشده بود، داده تولید می‌شد. پیش بینی می‌شود در سال ۲۰۲۰ سرعت تولیدات داده‌ای پنجاه برابر داده‌های تولیدشده در سال ۲۰۱۱ باشد. حجم اطلاعاتی که تا سال ۲۰۰۳ توسط انسان ایجاد شد تنها ۵ اگزابایت است، اما امروزه این حجم از اطلاعات کمتر از دو روز ایجاد می‌شود.

تغییر داده‌ها:

تغییر داده‌های ورودی، یکی از اصلی‌ترین مشکلات مرتبط با داده‌های عظیم است. این مفهوم به ناپایداری محیط‌هایی که داده در آن تولید می‌شود برمی‌گردد. هر چه ناپایداری در محیط بیشتر باشد، مدیریت و تحلیل داده‌ها سخت‌تر خواهد شد.

پیچیدگی استفاده از داده‌ها:

مدیریت داده‌ها فرآیند بسیار پیچیده‌ای است، به‌ویژه وقتی که حجم زیادی از داده‌های پیچیده از منابع در دسترس باشد.

کاربردهای داده های بزرگ

داده های بزرگ نحوه کار سازمان ها و افراد را تحت تاثیر قرار می دهد. این داده ها فرهنگی را در سازمان ها ایجاد می کند که از طریق آن کسب و کارها و مدیران فناوری اطلاعات را به سمت استفاده از تمامی ارزش های پنهان در داده ها سوق می دهد. ادراک این ارزش ها به کارکنان سازمان ها این امکان را می دهد که با بینش وسیع تری تصمیم گیری کنند، نزدیکی بیشتری با مشتریان داشته باشند، فعالیت های خود را بهینه کنند، با تهدیدات مقابله کنند و در نهایت سرمایه های خود را روی منبع جدیدی از سود سرشار پنهان در داده ها متمرکز سازند. سازمان ها برای رسیدن به این مرحله نیازمند معماری جدید، ابزارهای نو، فعالیت ها و تلاش های مستمری هستند تا بتوانند از مزیت های چهارچوب های مبتنی بر داده های بزرگ بهره مند شوند.

تحلیل داده های عظیم

با ورود به عصر فناوری اطلاعات و ارتباطات و آغاز استفاده از داده ها و اطلاعات به عنوان سرمایه های اصلی در حرکت علمی، اقتصادی، اجتماعی و فرهنگی جوامع، سازمان ها و شرکت های مختلف و توسعه مشارکت افراد در جهان اینترنت و ارتباطات شبکه ای در دنیا، دغدغه ای بروز پیدا کرد که از جنس همین داده هایی بود که همه روز و با سرعت وصف ناپذیری در دنیا و در عرصه های مختلفی که فناوری اطلاعات ورود پیدا کرده بود، تولید می شود و دغدغه این بود که چگونه حجم بزرگ و متنوع داده ها و اطلاعات را می توان با توجه به ساختارهایی که در فضای فناوری اطلاعات وجود دارد، مدیریت، کنترل و پردازش کرد و از آن در جهت بهبود ساختارها و سودآوری بیشتر بهره جست؟ این موضوع باعث شده است که پژوهشگران و دانشمندان به دنبال ایجاد ساختارها، متدلوژی ها، روش ها و رویکردهای جدیدی برای مدیریت، کنترل و پردازش این حجم از داده ها باشند که این تلاش ها در ذیل سایه **DataB g** مطرح شده است.

ترجمه و تالیف : شیوا حسینی - منبع: dzone.com