تکلیف چهارم

امنیت در پایگاه های داده

حسین زارعی

**کلان داده چیست؟**منظور از داده های کلان یا Big Data چیست؟ Big data اصطلاحی است برای مجموعه داده های حجیم که بزرگ ، متنوع ، با ساختار پیچیده و با دشواریهایی برای ذخیره سازی ، تحلیل و تصویرسازی (نمایش) ، پردازشهای بیشتر یا نتایج میباشد. پروسه تحقیق بر روی داده های حجیم جهت آشکارسازی الگوهای مخفی و راز همبستگی ها ، تجزیه و تحلیل big data نامیده میشود.

این اطلاعات مفید برای سازمانها وشرکتها در جهت کسب بینش غنی تر و عمیق تر و موفقیت در رقابت کمک میکند. به همین دلیل اجراهای big data نیاز دارند تا در صورت امکان ، تحلیل شوند و به طور دقیق اجرا شوند. این گزارش خلاصه ای از محتوای big data ، هدف ، مثالها ، روشها ، منافع و چالش ها ارائه میدهد و در خصوص نگرانیهای حفظ حریم خصوصی در big data بحث میکند.

Big data و تجزیه و تحلیلش در مرکز علوم و تجارتهای مدرن هستند. این داده ها از تراکنشهای online ، email ها ، ویدئوها ، صوتها ، کلیک کردن ها ، log ها و ارسالها ، درخواستهای جستجو ، یادداشتهای درست ، تعاملات شبکه های اجتماعی ، داده های علمی ، سنسورها و تلفنهای همراه و برنامه های کاربردی آنها تولید میشوند. آنها بر روی پایگاه داده ها که به شکل حجیم رشد میکنند ، ذخیره میشوند و ضبط ، شکل دهی ، ذخیره سازی ، مدیریت ، به اشتراک گذاری ، تحلیل و نمایش آنها از طریق ابزارهای نوعی نرم افزار پایگاه داده ها ، دشوار میشود.

پنج اگزا بایت (10 به توان 18 بایت) دیتا تا سال 2003 به وسیله انسان به وجود آمده است. امروزه این مقدار اطلاعات در دو روز تولید میشود. در سال 2012 داده های دنیای دیجیتال به 2.72 زتا بایت ( 10 به توان 21 بایت) توسعه پیدا کرد. پیش بینی میشود که این مقدار هر دو سال ، دو برابر شود وبه حدود 8 زتا بایت داده در سال 2015 میرسیم.

در گذشته پروسه آشکارسازی ژنهای بشر تقریباً 10 سال وقت میگرفت ، هم اکنون بیشتر از یک هفته نمیشود. داده های چند رسانه ای وزن بزرگی در ترافیک اینترنت دارند و انتظار میرود به %70 در سال 2013 افزایش یابد. فقط google بیش از یک میلیون سرور در دور جهان دارد. 6 میلیارد اشتراک موبایل در دنیا وجود دارد و هر روز 10 میلیارد متن پیام (message) ارسال میشود. در سال 2020 ، 50 میلیارد دستگاه به شبکه ها و اینترنت متصل خواهند بود.

در سال 2012 مواجهه بشر با big data به عنوان یک پروژه جهانی انجام شد با مرکزیت مجموعه ای بلادرنگ که مقدار بزرگی از داده ها را نمایش میدهد و تحلیل میکند. بر اساس این پروژه خیلی آمارها نتیجه گرفته شد. Facebook ، نهصد و نود و پنج میلیون حساب کاربری فعال ماهانه با 70 زبان دارد ، 140 میلیارد تصویر آپلود شده است ، 125 میلیارد اتصالات دوستها و روزانه 30 میلیارد قطعه های با محتوا و 2.7 میلیارد like و comment ارسال میشود.

هر دقیقه 48 ساعت ویدئو آپلود میشود و هر روز 4 میلیارد view در YouTube اجرا میشود. Google خدمات بسیاری را پشتیبانی میکند از جمله 7.2 میلیارد page را هر روز مونیتور میکند و همچنین 20 پتا بایت (10 به توان 15 بایت) داده را روزانه به 66 زبان ترجمه میکند.