

Big Data

بیگ دیتا در حقیقت بیانگر حجم زیادی از داده می باشد که آنالیز و دسته بندی آن ها به وسیله زیرساخت ها و دیتابیس های رایج مشکل می باشد. سازمان های بزرگ با انبوه اطلاعات متنوعی چون ایمیل، عکس، ویدیو، گزارشات روزانه، نوشته و اطلاعات روزانه مواجه هستند که می بایست به صورت هوشمندانه ای طبقه بندی گردند تا به سرعت و در زمان کوتاهی قابل بررسی و دستیابی باشند.

برای نمونه یک فروشگاه اینترنتی را در نظر بگیرید که بتواند با بررسی کلیک بازدیدکنندگان بر روی لینک های مختلف وب سایت، به علایق و سلیقه مشتریان پی ببرد و بدین سان کسب و کار خود را بهبود بخشد. یا دولت یک کشور می تواند با رصد شبکه های اجتماعی به مقابله و پیشگیری از ناهنجاری های جامعه بپردازد. نمونه بارز آن موتور جستجوی گوگل است که به دسته بندی و اولویت بندی علایق کاربران در کشورهای مختلف می پردازد و کاربران را در دستیابی به اطلاعات مورد نظرشان راهنمایی می نماید. همچنین شهرداری یکی از شهرهای بزرگ آمریکا توانسته با جمع آوری داده ها و اطلاعات در بیش از ۳۰ نقطه جغرافیایی شهری و آنالیز این اطلاعات، به منظور کاهش جرم و جنایت و نیز آرایه خدمات بهتر شهری استفاده کند.

در واقع بیگ دیتا به مجموعه اطلاعات ایجاد شده و ذخیره شده به همراه تجزیه و تحلیل آن گویند که از نظر **سرعت^۱، حجم^۲ و تنوع^۳** قابل بررسی باشد.

بیگ دیتا متشکل از دو بخش اساسی و پایه می باشد:

¹ Velocity

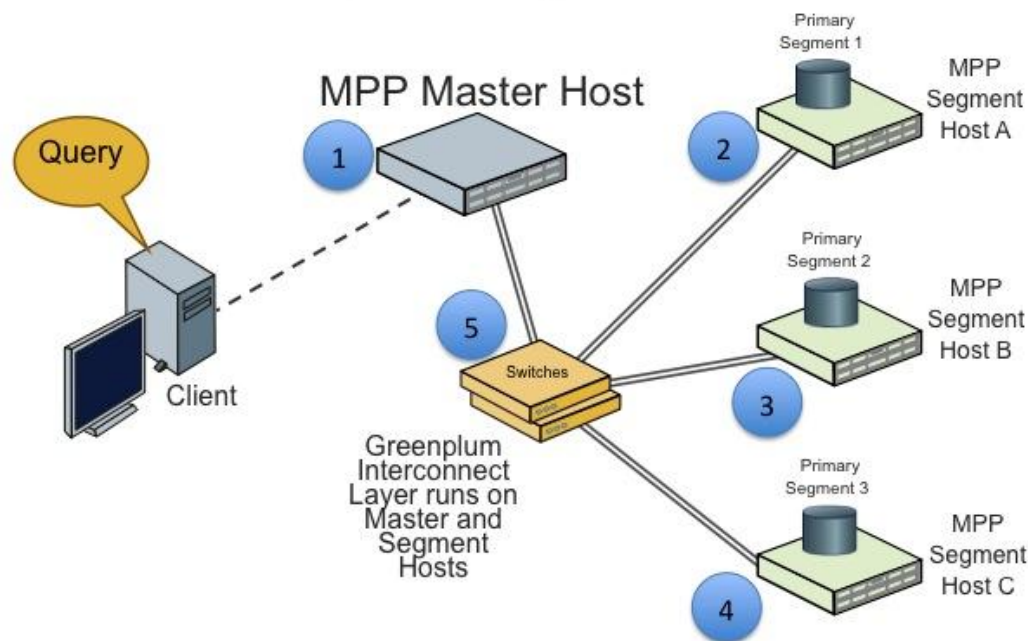
² Volume

³ Variety

۱- سیستم های پردازشی که قابلیت ثبت و ضبط اطلاعات را دارند. از نمونه های بارز چنین سیستم های پردازشی، دیتابیس های NoSQL است که قابلیت پاسخگویی به درخواست های هم زمان را با کمترین زمان تاخیر دارند. از ویژگی های چنین دیتابیس می توان به کدنویسی کمتر، سرعت بالا و افزایش حجم ذخیره اطلاعات نسبت به دیتابیس های رایج اشاره کرد.

۲- سیستم های آنالیز کننده اطلاعات هستند که می توانند همراه با سیستم های پردازشی ارائه شده باشند و یا به صورت نرم افزاری جداگانه مورد استفاده قرار گیرند. در مفهوم بیگ دیتا، سیستم های آنالیز کننده بر پایه مدل برنامه نویسی MapReduce و معماری پردازش موازی به صورت انبوه (MPP^۴) پیاده سازی می شوند.

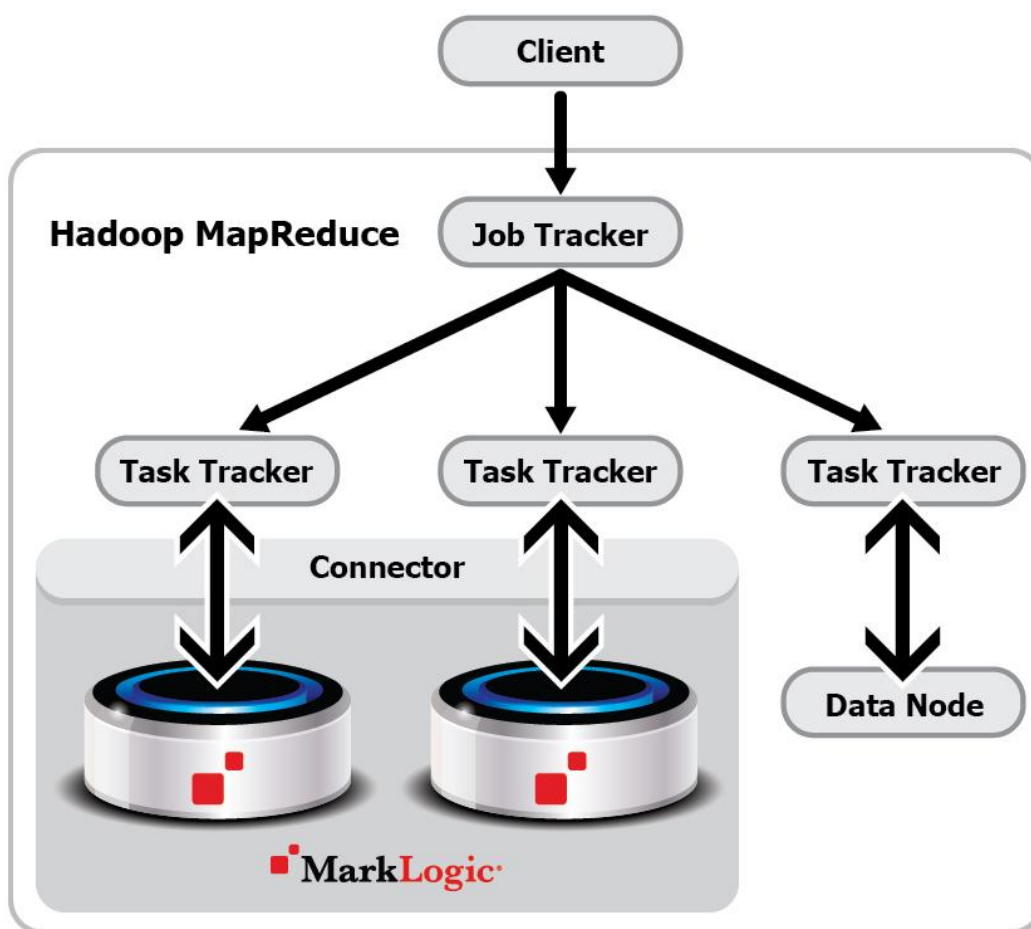
Greenplum Physical Architecture



⁴ Massively Parallel Processing

استفاده از چنین تکنولوژی هایی در حقیقت پاسخی به محدودیت های موجود در ارتباطات دیتابیس های رایج و پردازش بر روی تک سرور بوده است که توانسته ضمن حل این مشکلات، متدها جدیدی را برای آنالیز اطلاعات به کار گیرد.

همانطور که گفته شد، برخی سیستم های آنالیز کننده همراه دیتابیس NoSQL ارایه می شوند در غیر اینصورت می توان اطلاعات دیتابیس NoSQL را در نرم افزار آنالیز کننده جداگانه ای مانند Hadoop که از مدل برنامه نویسی MapReduce استفاده می کند، وارد نمود.



در واقع Hadoop یک تکنولوژی نرم افزاری رایگان^۵ است که مختص پردازش حجم انبوهی از اطلاعات بر روی ذخیره ساز ها و سرورهای کلاستر شده، طراحی گردیده است. این نرم افزار شامل سیستم گسترده فایل با نام HDFS^۶ می باشد.

پردازش ابری^۷:

همانطور که می دانیم سیستم های پردازش ابری در یک دهه گذشته توانسته تحولی شگرف در زیرساخت های فناوری اطلاعات و ارتباطات به وجود آورد به نحوی که از یک طرف به کاهش هزینه ها منجر شده و از طرف دیگر کارایی سیستم ها را بالا برده است. به همین دلیل می توان گفت که طراحی سیستم های بیگ دیتا برای استفاده از مزایای پردازش ابری بوده است. زیرا برای استفاده از مزایای چنین زیرساخت هایی طراحی شده اند و می توانند حجم انبوهی از اطلاعات را از طریق سرورهای کلاستر شده جمع آوری و پردازش کنند. برای درک بهتر مطلب می توانید نگاهی به ساختار و کارکرد MPP بیندازید که چگونه اطلاعات بخش^۸ های مختلف را به واسطه سرور میزبان^۹ جمع آوری می نماید.

⁵ Open Source

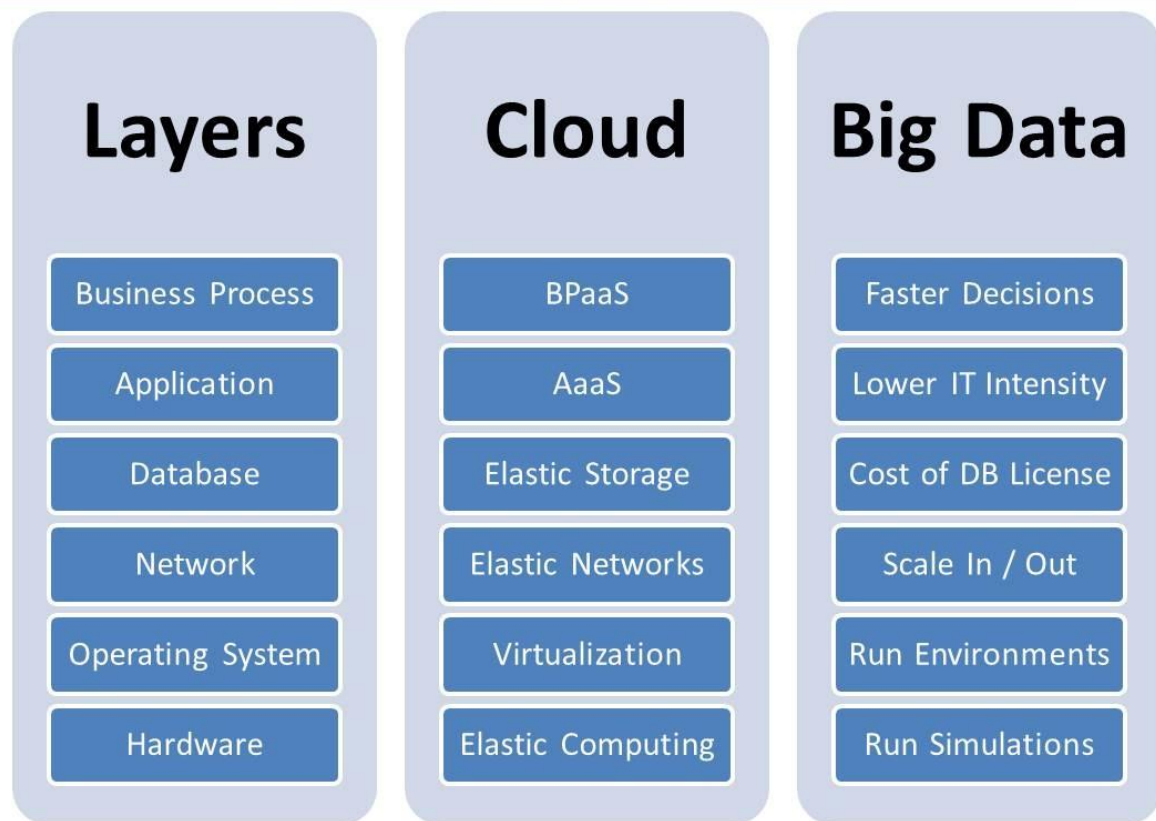
⁶ Hadoop Distributed File System

⁷ Cloud Computing

⁸ Segment

⁹ Master Node

Big Data & Cloud



دیتابیس های بر پایه مفهوم بیگ دیتا مانند NoSQL می توانند به راحتی بر بستری از سرورهای کلاستر شده پیاده سازی و اجرا شوند و این بدان دلیل است که برخلاف دیتابیس های معمولی که اطلاعات در جداول جداگانه ذخیره می شوند، در اینجا اطلاعات به صورت مجتمع ذخیره می گردند.

برای نمونه می توان به دیتابیس MangoDB اشاره نمود که بر پایه تکنولوژی NoSQL بنا نهاده شده است. این دیتابیس توسط یک سری رابط های اینترفیس ¹⁰ API قابل اتصال به نرم افزار Hadoop می باشد تا اطلاعات

¹⁰ Application Programming Interface

ثبت شده توسط پایگاه داده ی MongoDB به وسیله سیستم Hadoop و دیتابیس های MPP بررسی و آنالیز شوند.

پس نتیجه می گیریم سیستم های NoSQL به همراه MPP Databases و Hadoop با هم ما را به دنیای بیگ دیتا وارد می کنند.

تالیف و ترجمه توسط: محمدرضا نیکوکلام

Sources:

mongodb.com

dwarehouse.wordpress.com

watalon.com