اصول ذخیره و پردازش

داده‌های حجیم (Big Data)

مؤلف:

مهندس محمد مهدیان

کارشناس‌ارشد مهندسی فن‌آوری اطلاعات گرایش شبکه‌های کامپیوتری

دانشگاه صنعتی امیرکبیر (پلی تکنیک تهران)

NAGHOOS
PUBLICATION

۱۳۹۹

سرشناسه : مهدیان , محمد , ۱۳۶۴ -
عنوان و نام پدیدآور : اصول ذخیره و پردازش داده های حجیم(Big Data)/ مولف محمد مهدیان
مشخصات نشر : تهران: ناقوس,۱۳۹۹
مشخصات ظاهری : ۲۱۳ص .: جدول, نمودار(رنگی).
شابک : ۴۵۰,۰۰۰ریال: 978-600-473-271-0
وضعیت فهرست‌نویسی : فیپا
موضوع : داده های کلان
موضوع : Big Data
موضوع : داده های کلان -- مدیریت
موضوع : Big Data--Management
موضوع : پایگاه های اطلاعاتی-- مدیریت
موضوع : Database Management
رده بندی کنگره : QA۷۶/۹
رده بندی دیویی : ۰۰۵/۷۴
شماره کتابشناسی ملی : ۶۱۲۳۵۸۸

NAGHOOS PUBLICATION

برای خرید Online به آدرس زیر مراجعه کنید:
www.naghoospress.ir

انتشارات ناقوس

چاپ اول

نام کتاب : اصول ذخیره و پردازش داده های حجیم(Big Data)
ناشر : انتشارات ناقوس
مولف : مهندس محمد مهدیان
چاپ اول : ۱۳۹۹
تیراژ : ۵۰۰نسخه
ناظر چاپ : یاسمن تجملی محدث
چاپ و صحافی : حیدری
قیمت : ۴۵۰,۰۰۰ ریال
شابک : ۹۷۸ -۶۰۰-۴۷۳-۲۷۱-۰
ISBN : 978-600-473-271-0

انتشارات ناقوس
۱-بلوار کشاورز-خیابان ۱۶آذر-کوچه پارسی-پلاک ۱۰
تلفن و فاکس : ۶۶۴۷۸۹۵۱-۶۶۴۷۸۹۵۳

فهرست مطالب

مقدمه

امروزه منابع زیادی درحال تولید داده هستند و دیوایس های خیلی زیادی به اینترنت متصل‌اند که توسط این منابع و دیوایس ها، حجم دیتای زیادی با سرعت بالا در حال تولید شدن است. مثلا اطلاعات شرکت مخابرات و شبکه‌های اجتماعی و بازار بورس، اطلاعات ژنتیکی و... در Google روزانه ۱۵۰۰۰ پنتا بایت داده درحال ذخیره و ۱۰۰ پنتا بایت پردازش اطلاعات داریم[25] این مقادیر آن قدر بالا هست که در یک کامپیوتر عادی نمی‌توانیم نه این حجم ذخیره سازی و نه این حجم پردازشی را داشته باشیم و نیاز به مجموعه ای از کامپیوترها داریم که با هم کار کنند تا این حجم داده را ذخیره و پردازش کنند.

مفهوم داده‌های حجیم (Big Data) زمانی مطرح شد که داده‌هایمان را نتوانستیم ذخیره و پردازش کنیم و با عنایت به اینکه داده‌های حجیم (Big Data) فقط دیتا هستند که تنها حجم و یک سری مشخصاتشان تغییر کرده است و با توجه به محدودیت‌هایی که از لحاظ Resource داریم نمی‌توانیم با استفاده از روش های سنتی، دیتا را ذخیره و تحلیل کنیم پس به یک راه و یک بستر جدید برای ذخیره و تحلیل این دیتا احتیاج داریم. طبیعتا این را در بعضی شرکت ها مثل Google که داده‌های بیشتری داشتند بیشتر احساس شد.

در رابطه با داده‌های حجیم سه سوال اصلی پیش روی ما قرار دارد و در این کتاب سعی می‌کنیم آن‌ها را بررسی کنیم که این سوالات به شرح زیر می باشند:

۱-چگونه داده‌های حجیم را ذخیره (Store) کنیم؟ ذخیره سازی داده‌ها اولین مسئله ای است که باید آن را بررسی کنیم که در این رابطه Distributed File System و NoSQL databaseها را معرفی و بررسی می‌کنیم.

۲-چگونه داده‌های حجیم را پردازش (Process) کنیم؟ حال بسته به اینکه داده‌های ما از چه جنسی باشند Platformهای مختلفی را معرفی می‌کنیم. مثلا اگر داده‌ها ما ثابت غیر قابل تغییر باشند و اگر همه داده‌ها را یکجا داشته باشیم از Batch Processing Platform استفاده می‌کنیم و اگر داده‌های Stream (جاری) داشته باشیم از Streaming Processing Platform استفاده می‌کنیم و اگر داده‌ها ساختاریافته باشند می‌توانیم از Platformهایی مثل Spark SQL یا ... استفاده کنیم و اگر داده‌ها از جنس گراف باشند از Graph Platform استفاده می‌کنیم.