

مقدمه‌ی برکلان داده‌ها؛ مفاهیم و کاربردها

اثر: کریش کریشنان



ترجمہ و تالیف:

بہار عسگری

مصطفیٰ قبائی آرائی

بہارہ رحیمی

سرشناسه	: کریشان، کریش Krishnan, Krish
عنوان و نام پدیدآور	: مقدمه‌یی بر کلان داده‌ها؛ مفاهیم و کاربردها/ اثر کریش کریشان : ترجمه و تألیف بهار عسگری، مصطفی قبائی آرائی، بهار رحیمی.
مشخصات نشر	: کرج: چکامه باران، ۱۳۹۵.
مشخصات ظاهری	: ۲۵۲ ص: مصور(رنگی)، جدول، نمودار.
شابک	: ۹۷۸-۶۰۰-۸۰۲۶-۲۲-۸
وضعیت فهرست نویسی	: فیبا
یادداشت	: عنوان اصلی: Big data.
یادداشت	: واژه‌نامه.
یادداشت	: کتابنامه.
موضوع	: داده‌های کلان
موضوع	: Big data
موضوع	: پایگاه‌های اطلاعاتی - مدیریت
موضوع	: Database management
شناسه افزوده	: رج - بهار، ۱۳۶۸ - مترجم
شناسه افزوده	: گری، آر، ۱۳۶۷ - مترجم
شناسه افزوده	: قبائی آئی، مصطفی، ۱۳۶۴ - مترجم
رده بندی کنگره	: ۴۱۲.۱۵۰۰۵/۷۶۱
رده بندی دیویی	: ۰۰۷۴
شماره کتابشناسی ملی	: ۲۲۸۴۴۶۶



مقدمه‌یی بر کلان داده‌ها؛ مفاهیم و کاربردها

مؤلفان:	بهار عسگری، مصطفی قبائی آرائی، بهار رحیمی
چاپ:	اول - تابستان ۱۳۹۵
شمارگان:	۱۰۰۰ نسخه
شابک:	۹۷۸-۶۰۰-۸۰۲۶-۲۲-۸ قیمت: ۲۰۰۰ تومان
ویرایش:	حمیدرضا جدیدی طراح جلد و صفحه آرای حسین فلاح
کرج، بلوار شهید مؤذن، ابتدای شهرک شهیدان جهازی ها، پلاک ۳۱	
۰۲۶-۳۴۴۹۵۵۸۴-۳۴۴۹۵۶۰۲	
Chakamebaran.com - chakamebaran@yahoo.com	

حقوق چاپ برای ناشر محفوظ است

پیشگفتار

عصر حاضر زمان پیشرفت فناوری و تکنولوژی است. امروزه همه‌ی جوامع در پی یافتن روش‌های نوینی در راستای افزایش بازده و راندمان سازمان‌ها و شرکت‌های دولتی و خصوصی خود هستند. با توجه به افزایش و گسترش روزافزون حجم انتقال اطلاعات، مسأله کلان داده‌ها، ورود و ذخیره سازی حجم زیادی از اقلام اطلاعاتی داده‌ها و مدیریت آن، سرعت، دقت و امنیت اطلاعات از اهمیت بسیار زیادی برخوردار شده و توجه بسیاری از محققان و متخصصین در حوزه فناوری اطلاعات را به خود معطوف کرده است. کلان داده‌ها معمولاً اشاره به حجم‌های عظیمی از داده‌ها دارد که توسط سازمان‌های بزرگی مانند گوگل یا ناسا ذخیره و تحلیل و مورد استفاده قرار می‌گیرد. با ایجاد یک دید مناسب در خصوص کلان داده و اهمیت آن، جامعه‌ای را تصور کنید که در آن جمعیت به‌طور نمایان در حال افزایش است، اما خدمات و زیرساخت‌های عمومی آن نتواند پاسخگوی رشد جمعیت باشد و از عهده مدیریت آن برآید. چنین شرایطی در حوزه داده در حال وقوع است و کلان داده به‌فرازی بزرگ و حجیم هستند که با ابزارهای مدیریتی و پایگاه‌های داده سنتی و معمولی قابل مدیریت نیستند. مشکلات اصلی در کار با این نوع داده‌ها مربوط به برداشت و جمع‌آوری، ذخیره‌سازی، جستجو، اشتراک‌گذاری، تحلیل و نمایش آن‌ها است. حوزه کلان داده‌ها، به این دلیل هر روز جذابیت و مقبولیت بیشتری پیدا می‌کند که با استفاده از تحلیل حجم‌های بیشتری از داده‌ها، می‌توان تحلیل‌های عمیق‌تر و پیشرفته‌تری را برای مقاصد مختلف، از جمله مقاصد تجاری، پزشکی و امنیتی، انجام داد و نتایج مناسب‌تری را دریافت کرد. بیشتر تحلیل‌های مورد نیاز در پردازش داده‌های عظیم، توسط دانش‌اندیشان در علوم‌ی مانند هواشناسی، ژنتیک، شبیه‌سازی‌های پیچیده فیزیک، تحقیقات زیست‌شناسی، محیطی، جست‌وجوی اینترنت، تحلیل‌های اقتصادی و مالی و تجاری مورد استفاده قرار می‌گیرد.

کتاب حاضر در شش فصل تدوین گردیده است. در فصل‌های اول و دوم به معرفی کلان داده‌ها و چگونگی استفاده از آنها پرداخته شده، و مروری بر کار با کلان داده‌ها مطرح شده است. در فصل سوم معماری پردازش کلان داده‌ها تشریح شده است و در فصل چهارم از تکنولوژی‌های کلان داده‌ها ارائه شده است. در فصل پنجم این کتاب NoSQL توصیف شده و مدیریت کلان داده‌ها با استفاده از آن بیان شده است و نهایتاً در فصل ششم کتاب پیش‌رو، با معرفی چند نمونه مطالعاتی در حوزه کلان داده‌ها، ارزش تجاری کلان داده‌ها در بعضی حوزه‌ها، بررسی شده است.

سخن آخر:

در اینجا لازم می‌دانیم از تمامی استادان، فرهیختگان و پژوهشگران و دانشجویان گرامی که مشوق ما در تدوین این کتاب بوده‌اند، کمال تشکر و قدردانی کنیم. از اینکه توانسته‌ایم این کتاب را با صرف وقت زیاد و فشرده عرضه کنیم، بسیار خرسندیم. با امید اینکه کتاب مورد نظر مورد قبول

علاقه مندان قرار گیرد. از خوانندگان گرامی خواهشمندیم، پیشنهادهای و انتقادات و دیدگاه های خود را در مورد مطالب این کتاب با آدرسهای اینترنتی Bahar_Asgari88@yahoo.com و Mostafaghobaye@yahoo.com و Bahar.Rahimi68@gmail.com یا از طریق ناشر در میان بگذارند و ما را مرهون سپاس خود کنند.

بهار عسگری

مصطفی قبائی آرائی

بهاره رحیمی

www.ketab.ir

فهرست مطالب

فصل اول

- ۱۷ مقدمه ای بر کلان داده
- ۱۸ ۱-۱ مقدمه
- ۱۹ ۲-۱ کلان داده ها
- ۲۳ ۳-۱ تعریف کلان داده ها
- ۲۳ ۴-۱ چرا کلان داده ها و چرا الان؟
- ۲۵ ۵-۱ توسعه و تکامل فناوری های مربوط به کلان داده ها
- ۲۸ ۶-۱ مثال از کلان داده
- ۳۵ ۷-۱ تحلیل و یکپارچه سازی
- ۳۸ ۸-۱ انواع داده های سراف
- ۳۹ ۹-۱ خلاصه

فصل دوم

- ۴۱ کار با کلان داده
- ۴۳ ۱-۲ مقدمه
- ۴۳ ۲-۲ انفجار داده ها
- ۴۴ ۳-۲ تغییر مدل کسب و کار
- ۴۷ ۱-۳-۲ حجم داده ها
- ۵۷ ۲-۳-۲ سرعت داده ها
- ۶۱ ۳-۳-۲ تنوع داده ها
- ۶۳ ۴-۲ خلاصه

فصل سوم

- ۶۵ معماری های پردازش کلان داده ها
- ۶۶ ۱-۳ مقدمه
- ۶۷ ۲-۳ بازنگری در پردازش داده ها
- ۶۸ ۳-۳ تکنیک های پردازش داده ها

۷۱	۳-۴ چالش‌های زیرساخت‌های پردازش داده‌ها
۷۱	۳-۴-۱ ذخیره‌سازی
۱۳	۳-۴-۲ انتقال
۷۳	۳-۴-۳ پردازش
۷۴	۳-۴-۴ سرعت (راندمان)
۷۵	۳-۵ معماری‌های همه‌چیز-مشترک و هیچ‌چیز-مشترک
۷۵	۳-۵-۱ معماری همه-مشترک
۷۷	۳-۵-۲ معماری هیچ-مشترک
۷۸	۳-۵-۳ OLTP در مقابل انبارش داده‌ها
۷۸	۳-۶ پردازش کلان داده‌ها
۸۴	۳-۶-۱ زیرساخت توزیع داده‌ها
۸۵	۳-۶-۲ پردازش داده‌های توضیح داده‌ها
۸۷	۳-۷ مطالعه کلان داده‌های یک سازمان
۹۱	۳-۸ خلاصه
	فصل ۴
۹۳	توصیف فناوری‌های کلان داده
۹۵	۴-۱ مقدمه
۹۷	۴-۲ پردازش داده‌های توزیع شده
۱۰۲	۴-۳ ملزومات پردازش کلان داده‌ها
۱۰۳	۴-۴ فناوری‌های پردازش برای کلان داده‌ها
۱۰۴	۴-۴-۱ سیستم فایل گوگل
۱۰۸	۴-۵ Hadoop
۱۱۰	۴-۵-۱ اجزای اصلی Hadoop
۱۱۹	۴-۵-۱-۱ ردیاب کار و ردیاب وظیفه
۱۲۲	۴-۵-۲ MapReduce
۱۳۶	۴-۵-۳ Zookeeper
۱۴۲	۴-۵-۴ Pig

۱۴۶	HBase ۵-۵-۴
۱۵۲	Hive ۶-۵-۴
۱۵۹	Chukwa ۷-۵-۴
۱۶۰	Fluma ۸-۵-۴
۱۶۰	Oozie ۹-۵-۴
۱۶۰	HCatalog ۱۰-۵-۴
۱۶۱	Sqoop ۱۱-۵-۴
۱۶۴	Sqoop2 ۱۱-۵-۴

فصل پنجم

۱۶۷	مدیریت داده ها، NoSQL برای کلان داده ها
۱۶۸	۱-۵ مقدمه
۱۷۰	NoSQL ۲-۵
۱۷۱	۲-۵ اتنوری CAP
۱۷۳	۲-۵ زوج کلید مقدار: ولدمورت
۱۷۳	۳-۵ ذخیره ساز ستون خانواده: کاساندرا
۱۸۶	۴-۵ پایگاه داده اسناد Riak
۱۸۸	۵-۲-۵ پایگاه داده گرافی
۲۰۳	NoSQL ۶-۲-۵ خلاصه
۲۰۴	۳-۵ پردازش متنی ETL
۲۰۵	۴-۵ خلاصه
۲۰۵	۵-۵ مطالعه بیشتر

فصل ششم

۲۰۹	ارزش تجاری محرک کلان داده ها
۲۱۱	۱-۶ مقدمه
۲۱۲	۲-۶ نمونه ی مطالعاتی: داده های سنسور
۲۱۲	Vestas ۱-۲-۶
۲۱۳	۲-۶ تولید برق از باد

- ۲۱۶ ۳-۲-۶ تبدیل آب و هوا به سرمایه
- ۲۱۷ ۴-۲-۶ فائق آمدن بر چالش های کلان داده ها
- ۲۱۸ ۵-۲-۶ حفظ تولید انرژی در مرکز داده های آن
- ۲۱۹ ۳-۶ نمونه مطالعاتی ۲: داده های جریان دار
- ۲۱۹ ۱-۳-۶ نظارت و امنیت: TerraEchos
- ۲۲۰ ۲-۳-۶ منفعت
- ۲۲۱ ۳-۳-۶ ترکیب فیبرهای نوری پیشرفته با جریان دار data بی درنگ
- ۲۲۴ ۴-۶ نمو: مطالعاتی ۳: نسخه نویسی صحیح: بهبود پیامدهای بیمار با تحلیل های کلان داده ها
- ۲۲۴ ۱-۴-۶ هدایت تجاری
- ۲۲۴ ۲-۴-۶ چالش ها
- ۲۲۴ ۳-۴-۶ بازبینی: ارائه پیشنهادهایی جدید به پزشکان در هدایت مراقبت از بیمار
- ۲۲۵ ۴-۴-۶ چالش ها: ترکیب سیستم های انبار داده های قدیمی با کلان داده ها
- ۲۲۶ ۵-۴-۶ راه حل: آماده شدن برای تحلیل های کلان داده ها
- ۲۲۷ ۶-۴-۶ نتایج: حذف "تله های داده ها"
- ۲۲۹ ۸-۴-۶ دربارۀ Aurora:
- ۵-۶ نمونه مطالعاتی ۴: دانشگاه انتاریو، موسسه فناوری مهار داده های کلیدی جهت فراهم سازی مراقبت قابل پیش بینی از بیمار
- ۲۳۰ ۱-۵-۶ بازبینی
- ۲۳۲ ۲-۵-۶ منافع تجاری
- ۲۳۲ ۳-۵-۶ استفاده بهتر از منابع داده ها
- ۲۳۳ مراقبت از سلامت هوشمندانه تر
- ۲۳۳ ۴-۵-۶ اجزاء راه حل
- ۲۳۵ ۵-۵-۶ ترکیب دانش انسان و فناوری
- ۲۳۷ ۶-۵-۶ گسترده سازی تأثیر Artemis
- ۲۳۸ ۶-۶ نمونه مطالعاتی ۵: راه حل مشتری مایکروسافت SQL سرور
- ۲۳۸ ۱-۶-۶ برتری راه حل
- ۲۳۹ ۲-۶-۶ نیازهای تجارت

۲۴۰	۶-۶-۳ راه حل
۲۴۱	۶-۶-۴ منافع
۲۴۱	۶-۶-۵ بهره وری سرعت و قطع هزینه ها
۲۴۳	۶-۷ نمونه ی مطالعاتی ۶: یکپارچه سازی (ادغام) داده های مشتری محور
۲۴۳	۶-۷-۱ بازبینی
۲۴۹	۶-۷-۲ طراحی راه حل
۲۴۹	۶-۷-۳ تمند سازی فرصت Cross-Sell و Upsell بهتر
۲۵۲	۶-۷ خلاصه