

داده های کلان

برآوردی:

دکتر عباس خدایی

مهندس مریم نیکوپور

تابستان ۱۳۹۴



سازمان اسناد و کتابخانه ملی
جمهوری اسلامی ایران

سرشناسه	: سقایی، عباس، ۱۳۵۱ -
عنوان و نام پدیدآور	: داده‌های کلان/ تألیف و ترجمه عباس سقایی، مریم نیکوپور.
مشخصات نشر	: تهران: عباس سقایی، ۱۳۹۴.
مشخصات ظاهری	: ۱۷۵ ص.: مصور (بخشی رنگی)، جدول، نمودار.
شابک	: ۱۲۰۰۰۰ ریال ۸-۴۱۶۷-۰۴-۹۷۸-۶۰۰-۰۵/۷۴
وضعیت فهرست نویسی	: فیا
یادداشت	: بخش اعظم کتب برگرفته از کتاب "Big data : related technologies, challenges and future prospects, 2014" تألیف مین چن... [و دیگران] است.
یادداشت	: واژه نامه.
یادداشت	: کتابنامه: ص. ۴۴-۴۵.
عنوان دیگر	: فناوری‌ها و چالش‌های مرتبط و چشم‌انداز آینده
عنوان دیگر	: داده‌های بزرگ: فناوریهای مرتبط، چالشها و چشم‌اندازهای آینده.
موضوع	: داده‌های کلان
موضوع	: داده‌های کلان
موضوع	: داده‌های اطلاعاتی -- مدیریت
شناسه افزوده	: نیکوپور، مریم، ۱۳۶۶ -
شناسه افزوده	: میر، ۱۹۰۰ -
شناسه افزوده	: Chen, L. (Lin)
رده بندی کنگره	: ۰۵۷۶/۰۰۵۷۴ ۱۳۹۴
رده بندی دیویی	: ۰۰۵/۷۴
شماره کتابشناسی ملی	: ۴۱۱۲۶۹۲

عنوان کتاب: داده‌های کلان

گردآورندگان: عباس سقایی، مریم نیکوپور

ناشر: عباس سقایی

طرح جلد: خمدی گرافیک

نوبت چاپ: اول

تیراژ: ۱۰۰۰ نسخه

شابک: ۹۷۸-۶۰۰-۰۴-۴۱۶۷-۸

قیمت: ۱۲۰۰۰۰ ریال

مقدمه مؤلفین

خداوند متعال را سپاسگزاریم که توفیق تدوین این مجموعه را برایمان فراهم نمود. کتابی که پیش رو دارید به شرح یکی از مفاهیم نوین داده کاوی به نام "داده های کلان" یا "Big Data" پرداخته است. کتاب حاضر می تواند به عنوان راهنما برای دانشجویان مهندسی صنایع، مهندسی فن آوری اطلاعات، مهندسی کامپیوتر، مدیریت صنعتی، مدیریت بازرگانی، مدیریت اجرایی و ... در مقطع کارشناسی و تحصیلات تکمیلی استفاده شود. به علاوه این کتاب مورد کاربرد عموم محققین، مدیران، کارشناسان و تحلیلگران حوزه های مختلف کسب و کار است. تحلیلگران کسب و کار افرادی هستند که به معرفی، مدیریت و نظارت بر فرایندهای کسب و کار سازمان می پردازند. این افراد عموماً متخصصان فن آوری اطلاعات و یا برنامه ریزانی هستند که فرایند های کسب و کار را درون سیستم های مدیریت فرایند کسب و کار به اجرا در می آورند.

کتاب حاضر در هشت بخش تنظیم شده است. بخش اول به مفاهیم مقدماتی داده های کلان، بخش دوم به معرفی فن آوری های رایسته و بخش سوم به شرح تولید و دریافت داده های کلان اختصاص دارد. در بخش چهارم، دهم و دوازدهم داده های کلان و در بخش پنجم تحلیل داده های کلان ارائه شده است. در بخش ششم به برنامه های داده های کلان، در بخش هفتم به مسائل حل نشده و چشم انداز و در بخش آخر به بیان دوره نامه پرداخته ایم.

در خاتمه لازم می دانیم از تمامی بزرگواریانی که به نحوی در تهیه و تألیف این کتاب یاری نمودند، کمال تشکر و قدردانی را به عمل آوریم. امید است خوانندگان عزیز با ارائه نظرات و پیشنهادات سازنده خود ما را مورد لطف و عنایت خویش قرار دهند.

عباس سقایی

مریم نیکوپور

abbas.saghaei@gmail.com

nikoopour.mn@gmail.com

پیشگفتار

تغییرات سریع تکنولوژی در دهه اخیر و افزایش چشمگیر نقش اطلاعات، نحوه مدیریت و رهبری را دگرگون ساخته و اداره سازمان ها به روش های سنتی تقریباً ناشدنی است. محیط امروز، سازمانها و مدیران را با چالش های جدیدی روبرو می سازد و هر روز پویاتر از قبل می شود. موسساتی که خواهان افزایش سهم خود در بازار و کسب منافع بیشتری هستند باید خود را با تغییرات محیط تطبیق دهند.

سازمانها جهت اتخاذ تصمیمات راهبردی تجاری به داده ها متکی اند. اما "داده های کلان" چالش های خاص خود را دارد. بی بایست پیش از آنکه اثر واقعی خود را بر جای بگذارند، بررسی شوند. داده های کلان به مجموعه داده هایی اطلاق می شود که مدیریت، کنترل و پردازش آنها فراتر از توانایی ابزارهای نرم افزاری در یک زمان مورد انتظار است. یکی از مهم ترین مسائل مرتبط با داده های کلان، مشکل بودن کار با آنها به وسیله پایگاه های داده های اطلاعاتی و دسته های نرم افزاری تصویرنگاری داده ها و نرم افزارهای آماری رومیزی است. چرا که این داده ها برای پردازش شدن در یک زمان معقول به نرم افزارهای به شدت موازی شده با قابلیت اجرا روی ده ها، صدها یا هزاران سرور نیاز دارند.

امروزه به دلیل تولید و جمع آوری داده ها از محبوبیت بزرگی از تجهیزات و ابزارهای مختلف (مانند گوشی های موبایل، حسگرهای محیطی، دوربین ها، میکروفون ها، دستگاه های تشخیص امواج رادیویی، شبکه های حسگر بی سیم و غیره) حجم داده های ذخیره شده در پایگاه ها و انبار های داده با سرعت خیره کننده ای در حال افزایش است. نکته جالب توجه در این زمینه آن است که ۹۰ درصد داده هایی که اکنون در اختیار ما قرار دارد، تنها در ۲ سال اخیر تولید شده اند! به عنوان یک بزرگ هیچ گاه مانند اکنون داده های عظیمی نداشته اند و یا حداقل به فکر جمع آوری و استفاده از آن نبوده اند. به همین دلیل، تعجبی ندارد که ابزارهای ما در زمینه مدیریت چنین داده های کلانی پیشرفت چندانی نکرده اند.

فهرست مطالب

صفحه

عنوان

۱	بخش اول: مفاهیم مقدماتی داده های کلان
۱	۱-۱) طلوع عصر داده های کلان
۴	۲-۱) شکل گیری داده های کلان
۶	۳-۱) تعریف و مشخصات داده های کلان
۸	۴-۱) ابزارهای مدیریت داده های کلان
۱۱	۵-۱) ارزش داده های کلان
۱۳	۶-۱) دروغ داده های کلان
۱۴	۷-۱) قانون داده های کلان
۱۵	منابع بخش اول
۱۷	بخش دوم: فن آوری های وابسته
۱۷	۱-۲) رایانش ابری
۱۷	۱-۱-۲) اصول مقدماتی رایانش ابری
۱۸	۲-۱-۲) رابطه میان رایانش ابری و داده های کلان
۲۰	۲-۲) اینترنت اشیا
۲۰	۱-۲-۲) مقدمات اینترنت اشیا
۲۱	۲-۲-۲) رابطه میان اینترنت اشیا و داده های کلان
۲۲	۳-۲) مرکز داده ها
۲۳	۴-۲) هادوب
۲۳	۱-۴-۲) اصل مقدماتی هادوب
۲۶	۲-۴-۲) رابطه میان هادوب و داده های کلان
۲۸	۵-۲) ده منبع بزرگ داده های کلان
۳۰	منابع بخش دوم

- بخش سوم: تولید و دریافت داده های کلان ۳۱
- ۳۱-۱-۳ تولید داده های کلان ۳۱
- ۳۱-۱-۳ داده های شرکتی ۳۲
- ۳۱-۲-۳ داده های اینترنت اشیا ۳۲
- ۳۱-۳-۳ داده های اینترنتی ۳۴
- ۳۱-۴-۳ داده های زیست-بزشکی ۳۴
- ۳۱-۵-۳ داده ها در زمینه های دیگر ۳۶
- ۳۲-۲-۳ کسب داده های کلان ۳۷
- ۳۲-۲-۳ جمع آوری داده ها ۳۷
- ۳۲-۲-۳ انتقال داده ها ۴۱
- ۳۲-۲-۳ پیش پردازش داده ها ۴۳
- ۳۲-۲-۳ یکپارچگی ۴۳
- ۳۲-۳-۳ تصفیه (بالایش) ۴۴
- ۳۲-۳-۳ کاهش داده های زائد ۴۵
- ۳-۳-۱۰ روش درک بهتر داده های کلان ۴۶
- ۳-۳-۱ درک اهدافتان! ۴۶
- ۳-۳-۲ ایجاد یک نقشه راه! ۴۷
- ۳-۳-۳ داده هایتان را کشف کنید! ۴۷
- ۳-۳-۴ کشف کنید چه داده هایی را ندارید! ۴۸
- ۳-۳-۵ درک گزینه های تکنولوژی! ۴۸
- ۳-۳-۶ طرح امنیت در زمینه داده های کلان! ۴۹
- ۳-۳-۷ طراحی داده با استراتژی حاکمیت! ۴۹
- ۳-۳-۸ طرح داده نظارت! ۴۹
- ۳-۳-۹ مکرراً فرضیات خود را آزمایش کنید! ۵۰
- ۳-۳-۱۰ بهترین روشها و الگوهای اعمال نفوذ را مطالعه کنید! ۵۰
- منابع بخش سوم ۵۱

- بخش چهارم: ذخیره داده های کلان ۵۵
- ۱-۴) سیستم ذخیره داده های کلان ۵۵
- ۲-۴) سیستم ذخیره سازی توزیع شده ۵۷
- ۳-۴) مکانیسم ذخیره سازی برای داده های کلان ۶۱
- ۱-۳-۴) فن آوری پایگاه داده ۶۲
- ۱-۱-۳-۴) پایگاه داده های کلید-ارزش ۶۳
- ۲-۱-۳-۴) پایگاه داده های ستون-محور ۶۵
- ۱-۳-۴) پایگاه داده های اسناد ۶۸
- ۲-۱-۳-۴) پایگاه ذخیره سازی جدولی جهانی نیممل ۷۰
- ۲-۳-۴) عوامل طراحی ۷۱
- ۳-۳-۴) مدل برنامه نویسی داده ۷۲
- ۱-۳-۳-۴) نگاشت کاهش ۷۲
- ۲-۳-۳-۴) دریافت ۷۳
- ۳-۳-۳-۴) جفت-گونه ها ۷۴
- ۴-۳-۲-۴) پریگل ۷۵
- منابع بخش چهارم ۷۷
- بخش پنجم: تحلیل داده های کلان ۸۰
- ۱-۵) تحلیل سنتی داده ها ۸۰
- ۲-۵) روش های تحلیل داده های کلان ۸۳
- ۳-۵) معماری تحلیل داده های کلان ۸۵
- ۱-۳-۵) زمان حقیقی در برابر تحلیل آفلاین ۸۵
- ۲-۳-۵) تحلیل در سطوح مختلف ۸۶
- ۳-۳-۵) تحلیل با پیچیدگی های مختلف ۸۶
- ۴-۵) ابزارهای داده کاوی و تحلیل داده ها ۸۷
- ۵-۵) کاوش داده های کلان ۹۰

۹۰	۱-۵-۵ تعریف جریان داده و انواع آن
۹۲	۲-۵-۵ داده کاوی موازی و توزیعی
۹۴	۳-۵-۵ فضای طراحی موازی
۹۵	۱-۳-۵ سیستم های اشتراک حافظه در مقابل ماشین های حافظه توزیع شده
۹۵	۲-۳-۵ موازی سازی داده ها در مقابل موازی سازی وظایف
۹۶	۱-۳-۵ توازن بار کاری به شیوه های پویا و ایستا
۹۶	۴-۳-۵ ساختار های داده ای افقی و عمودی
۹۷	۴-۵-۵ روش های کس موازی و توزیعی
۹۷	۵-۵-۵ چالش ها در کامپیوتر موازی و توزیعی
۹۷	۱-۵-۵ ابعاد بالای داده ها
۹۸	۲-۵-۵ اندازه بزرگ
۹۸	۳-۵-۵ موقعیت مکانی داده ها
۹۹	۴-۵-۵ نوع داده ای
۹۹	۵-۵-۵ انحراف داده ها
۱۰۰	۶-۵-۵ توزیع پویای بار کاری
۱۰۰	۷-۵-۵ روش های افزایشی
۱۰۰	۸-۵-۵ کاوش از چندین جدول، ساختار داده ها و شیماهای شاخص گذاری
۱۰۱	۶-۵-۵ سامانه‌ی مدیریت پایگاه داده و سیستم های مدیریت فایل موازی
۱۰۲	منابع بخش پنجم
۱۰۳	بخش ششم: برنامه های داده های کلان
۱۰۳	۱-۶ تکامل برنامه ها
۱۰۵	۲-۶ زمینه های تحلیل داده های کلان
۱۰۶	۱-۲-۶ تحلیل داده های دارای ساختار
۱۰۶	۲-۲-۶ تحلیل داده های متنی
۱۰۸	۳-۲-۶ تحلیل داده های وب

۱۱۰	۴-۲-۶) تحلیل داده های چند رسانه ای
۱۱۲	۵-۲-۶) تحلیل داده های شبکه
۱۱۵	۶-۲-۶) تحلیل ترافیک سیار
۱۱۷	۳-۶) برنامه های کاربردی کلیدی
۱۱۷	۱-۳-۶) کاربرد داده های کلان در شرکت ها
۱۱۸	۲-۳-۶) برنامه های کاربردی داده های کلان مبتنی بر اینترنت اشیا
۱۱۹	۳-۳-۶) کاربرد داده های کلان مبتنی بر شبکه های اجتماعی آنلاین
۱۲۳	۴-۳-۶) کاربرد بهداشت و درمان و پزشکی داده های کلان
۱۲۴	۵-۳-۶) هوش جمعی
۱۲۵	۶-۳-۶) شبکه های دشمن
۱۲۸	منابع بخش ششم
۱۳۲	بخش هفتم: مسائل حل نشده و چشم انداز
۱۳۲	۱-۷) مسائل حل نشده
۱۳۲	۱-۱-۷) تحقیقات نظری
۱۳۳	۲-۱-۷) توسعه فن آوری
۱۳۶	۳-۱-۷) معانی کاربردی
۱۳۷	۴-۱-۷) امنیت داده ها
۱۳۹	۲-۷) چشم انداز
۱۴۳	۳-۷) توسعه نرم افزار در آینده
۱۴۴	۴-۷) توسعه آینده الگوریتم
۱۴۶	۵-۷) بایدها و نبایدهای داده های کلان
۱۴۶	۱-۵-۷) همه واحدهای کسب و کار را در استراتژی داده های کلان دخالت دهید
۱۴۶	۲-۵-۷) مدل های تحویل داده های کلان را ارزیابی کنید
۱۴۶	۳-۵-۷) منابع سنتی داده ها را به عنوان بخشی از استراتژی داده های کلان تصور کنید
۱۴۶	۴-۵-۷) برای فراداده های با ثبات برنامه ریزی کنید

۱۴۷	۵-۵-۷) داده های خود را توزیع کنید.....
۱۴۷	۶-۵-۷) برای تحلیل داده های کلان، تنها به یک رویکرد تکیه نکنید.....
۱۴۷	۷-۵-۷) با آمادگی به سراغ داده های کلان بروید.....
۱۴۷	۸-۵-۷) فراموش نکنید داده ها باید یکپارچه شوند.....
۱۴۸	۹-۵-۷) داده ها را با امنیت اداره کنید.....
۱۴۸	۱۰-۵-۷) از تاراج عملکرد داده ها غافل نباشید.....
۱۴۸	۶-۷) چالش های داده های کلان.....
۱۵۱	۷-۷) رویکردهای آینده.....
۱۵۲	منابع بخش هفتم.....
۱۵۳	بخش هشتم: واژه نامه.....

«بخش اول»

- شکل (۱-۱): رشد نمایی کاربران بر خط و انفجار بانک های داده..... ۲
- شکل (۲-۱): روند افزایش ابعاد داده..... ۴
- شکل (۳-۱): در هرم داده، داده های کلان با ترابایت، پتابایت و حتی اگزابایت تعیین می شود..... ۷
- شکل (۴-۱): شرح افزایش مداوم داده های کلان روند افزایش ابعاد داده..... ۹
- شکل (۵-۱): خصوصیات چهارگانه داده های کلان..... ۱۱
- شکل (۶-۱): فرصت های کلیدی برای سناریوهای مشتریان در بانکداری با استفاده از داده های کلان..... ۱۳

«بخش دوم»

- شکل (۱-۲): مباحث کلیدی رانانش ابر..... ۱۹
- شکل (۲-۲): نمایش معماری اینترنت اشیا..... ۲۰

«بخش سوم»

- شکل (۱-۳): نمایش منبع اصلی داده حسگرها..... ۳۳

«بخش ششم»

- شکل (۱-۴): توانمندسازی فن آوری ها برای داده های کلان شبکه های اجتماعی..... ۱۲۰
- شکل (۲-۴): رابطه میان توثیت های مربوط به قیمت برنج و تورم قیمت غذا..... ۱۲۲

«بخش هفتم»

- شکل (۱-۷): تکنولوژی های کلیدی در حوزه داده های کلان..... ۱۳۵