

بر نام خدا

اصول داده‌های بزرگ

مفاهیم، پیشران‌ها و تکنیک‌ها

نویسندگان: توماس ارل، وجید ختاک و پاول بوهرلر

مترجمان:

دکتر حسن رشیدی (دانشیار دانشگاه علامه طباطبایی)
دکتر اسماعیل زینالی (استادیار دانشگاه آزاد اسلامی قزوین)
حامد حیدری (مدرس دانشگاه آزاد اسلامی واحد قزوین)
سعید جلالیان (مدرس دانشگاه آزاد اسلامی واحد قزوین)

اصول داده‌های بزرگ مفاهیم، پیشران‌ها و تکنیک‌ها

نویسندگان: توماس ارل، وجید ختاک و پاول بوهرلر
مترجمان: دکتر حسن رشیدی، دکتر اسماعیل زینالی، حامد حیدری، سعید جلالیان
ناشر: انتشارات آتی‌نگر، وینا
طراحی جلد، صفحه‌آرایی: همتا بیداریان
چاپ اول، ۱۳۹۵
شمارگان: ۵۰۰ نسخه
قیمت: ۱۵۰,۰۰۰ ریال
شابک: ۹۷۸-۶۰۰-۷۶۳۱-۴۳-۰

ISBN: 978-600-7631-43-0

حق چاپ برای انتشارات آتی‌نگر محفوظ است.

نشانی دفتر فروش: خیابان جمالزاده جنوبی، روبه‌روی کوچه رشتچی، پلاک ۱۴۴، واحد ۲

نمابر: ۶۶۵۶۵۳۳۷

تلفن: ۸-۶۶۵۶۵۳۳۶



www.ati-negar.com * info@ati-negar.com

سرشناسه: ارل، توماس، Erl, Thomas

اصول داده‌های بزرگ مفاهیم، پیشران‌ها و تکنیک‌ها/ نویسندگان توماس ارل، وجید ختاک و پاول بوهرلر؛
مترجمان حسن رشیدی... [و دیگران].

تهران: آتی‌نگر، ۱۳۹۵

۲۴۰ ص: (بخشی رنگی) مصور، جدول، نمودار.

ISBN: 978-600-7631-43-0

فیبا.

یادداشت: مترجمان حسن رشیدی، اسماعیل زینالی، حامد حیدری، سعید جلالیان.

یادداشت: عنوان اصلی: techniques & Big data fundamentals : concepts, drivers

موضوع: داده‌های کلان Big Data - داده‌کاوی Data Mining - تصمیم‌گیری - داده‌پردازی - Decision making -- Data processing

شناسه افزوده: ختاک، وجید، Khattak, Wajid

شناسه افزوده: بوهرلر، پاول، Buhler, Paul

شناسه افزوده: رشیدی، حسن، ۱۳۳۹ - مترجم

رده‌بندی کنگره

رده‌بندی دیویی

شماره کتابشناسی ملی

QA۷۶/۹/۵۲۵الف۴ ۱۳۹۵

۰۰۶/۳۱۲

۴۴۶۳۵۸۸



فهرست مطالب

۱۱	سخنی از مترجمان
۱۳	بخش اول: اصول و مبانی داده‌های بزرگ
۱۵	فصل اول: درک داده‌های بزرگ
۱۷	مفاهیم و اصطلاحات
۱۷	مجموعه داده‌ها
۱۸	تحلیل داده‌ها
۱۹	رویکرد تحلیلی داده‌ها
۲۵	هوش کسب‌وکار
۲۵	شاخص‌های کلیدی عملکرد
۲۶	ویژگی‌های داده‌های بزرگ
۲۷	حجم
۲۸	سرعت
۲۸	تنوع
۲۹	صحت
۳۰	ارزش
۳۱	انواع مختلف داده‌ها
۳۲	داده‌های ساخت‌یافته

۳۳ داده‌های بدون ساختار
۳۴ داده‌های نیمه ساخت یافته
۳۴ فراداده
۳۵ سابقه مطالعه موردی
۳۵ تاریخچه
۳۶ زیرساخت فنی و محیط اتوماسیون
۳۷ اهداف و موانع کسب و کار
۳۹ مثال مطالعه موردی
۴۰ شناسایی مشخصات داده‌ها
۴۱ شناسایی انواع داده‌ها

فصل دوم: انگیزه‌های کسب و کار و پیشران‌های به کارگیری داده‌های بزرگ

۴۴ پویایی‌های بازار
۴۶ معماری کسب و کار
۴۹ مدیریت فرآیند کسب و کار
۵۱ فناوری اطلاعات و ارتباطات
۵۱ رویکرد تحلیلی داده‌ها و علوم داده‌ای
۵۱ دیجیتالی شدن
۵۲ فناوری مقرون به صرفه و سخت‌افزار ارزان قیمت
۵۳ رسانه‌های اجتماعی
۵۳ جوامع و دستگاه‌های ابرمتصل
۵۴ رایانش ابری
۵۶ اینترنت همه چیز (IoE)
۵۷ مثال مطالعه موردی

فصل سوم: به کارگیری داده‌های بزرگ و ملاحظات برنامه‌ریزی

۶۱ پیش‌نیازهای سازمان
۶۱ تهیه داده
۶۱ حریم خصوصی

۶۲	امنیت
۶۳	اصالت
۶۳	پشتیبانی بلادرنگ محدود
۶۴	چالش‌های عملکرد مجزا
۶۵	نیازهای حاکمیتی مجزا
۶۵	روش مجزا
۶۶	ابرها
۶۷	دوره عمر رویکرد تحلیلی داده بزرگ
۶۸	ارزیابی مورد کسب‌وکار
۷۰	شناسایی داده
۷۱	اکتساب و پالایش داده‌ها
۷۳	استخراج داده‌ها
۷۵	اعتبارسنجی داده‌ها و پاکسازی داده‌ها
۷۷	تجمیع و بازنمایی داده‌ها
۸۰	تحلیل داده‌ها
۸۱	مصورسازی داده‌ها
۸۳	استفاده از نتایج تحلیل
۸۴	مثال مطالعه موردی
۸۶	دوره عمر رویکرد تحلیلی داده بزرگ
۸۶	ارزیابی مورد کسب‌وکار
۸۷	شناسایی داده‌ها
۸۷	اکتساب و پالایش داده‌ها
۸۸	استخراج داده‌ها
۸۸	اعتبارسنجی و پاکسازی داده‌ها
۸۸	تجمیع و بازنمایی داده‌ها
۸۸	تحلیل داده‌ها
۸۹	مصورسازی داده‌ها
۸۹	استفاده از نتایج تحلیل

فصل چهارم: فناوری‌های سازمانی و هوش کسب‌وکار داده‌های بزرگ

۹۱

- ۹۲ پردازش تراکنش آنلاین (OLTP)
- ۹۳ پردازش تحلیلی آنلاین (OLAP)
- ۹۴ استخراج، تبدیل، بارگذاری (ETL)
- ۹۵ انبار داده
- ۹۵ بازارهای داده
- ۹۶ هوش کسب‌وکار سنتی
- ۹۷ گزارش‌های موردی
- ۹۷ داشبوردها
- ۹۸ هوش کسب‌وکار داده بزرگ
- ۱۰۰ مصورسازی داده‌های سنتی
- ۱۰۰ مصورسازی داده‌ها برای داده‌های بزرگ
- ۱۰۱ مثال مطالعه موردی
- ۱۰۱ فناوری سازمانی
- ۱۰۲ هوش کسب‌وکار داده بزرگ

بخش دوم: ذخیره‌سازی و رویکرد تحلیلی داده‌های بزرگ

۱۰۵

فصل پنجم: مفاهیم ذخیره‌سازی داده‌های بزرگ

۱۰۷

- ۱۰۸ خوشه‌ها
- ۱۰۹ سیستم‌های فایل و سیستم‌های فایل توزیع‌شده
- ۱۱۱ NoSQL
- ۱۱۱ بخش‌بندی
- ۱۱۳ کپی‌برداری
- ۱۱۴ اصلی - فرعی
- ۱۱۶ هم‌تا به هم‌تا
- ۱۱۸ بخش‌بندی و کپی‌برداری
- ۱۱۹ بخش‌بندی و کپی‌برداری اصلی-فرعی
- ۱۲۰ ترکیب بخش‌بندی و کپی‌برداری هم‌تا به هم‌تا

۱۲۱	قضیه CAP
۱۲۴	ACID
۱۲۸	BASE
۱۳۱	مطالعه موردی

۱۳۳ فصل ششم: مفاهیم پردازش داده‌های بزرگ

۱۳۴	پردازش در حالت بلادرنگ
۱۳۴	پردازش داده موازی
۱۳۵	پردازش داده‌های توزیع شده
۱۳۵	هادوپ
۱۳۶	پردازش حجم کار
۱۳۶	دسته‌ای
۱۳۷	تراکنشی
۱۳۷	خوشه
۱۳۹	پردازش در حالت دسته‌ای
۱۳۹	پردازش دسته‌ای با MapReduce
۱۴۰	وظایف نگاشت و کاهش
۱۴۷	یک مثال ساده از MapReduce
۱۴۷	درک الگوریتم‌های MapReduce
۱۵۰	پردازش در حالت بلادرنگ
۱۵۱	سرعت سازگاری حجم (SCV)
۱۵۴	پردازش جریان رویداد
۱۵۴	پردازش رویداد پیچیده
۱۵۴	پردازش داده‌های بزرگ بلادرنگ و SCV
۱۵۵	پردازش بلادرنگ داده‌های بزرگ و MapReduce
۱۵۶	مطالعه موردی
۱۵۶	پردازش حجم کار
۱۵۶	پردازش در حالت دسته‌ای
۱۵۷	پردازش بلادرنگ

۱۶۰ دستگاه‌های ذخیره‌سازی روی دیسک
۱۶۱ سیستم‌های فایل توزیع‌شده
۱۶۲ پایگاه‌های داده سیستم‌های مدیریت پایگاه داده رابطه‌ای
۱۶۵ پایگاه‌های داده NoSQL
۱۷۷ پایگاه داده NewSQL
۱۷۸ دستگاه‌های ذخیره‌سازی درون حافظه‌ای
۱۸۱ شبکه‌های داده درون حافظه
۱۹۰ پایگاه‌های داده درون حافظه‌ای
۱۹۴ مطالعه موردی

۱۹۷ تحلیل کمی
۱۹۸ تحلیل کیفی
۱۹۸ داده‌کاوی
۱۹۹ تحلیل آماری
۱۹۹ آزمون A/B
۲۰۱ همبستگی
۲۰۳ رگرسیون
۲۰۵ یادگیری ماشین
۲۰۶ دسته‌بندی (یادگیری ماشین تحت نظارت)
۲۰۷ خوشه‌بندی (یادگیری ماشین بدون نظارت)
۲۰۸ تشخیص نقاط پرت
۲۰۹ فیلترگذاری
۲۱۰ تحلیل معنایی
۲۱۱ پردازش زبان طبیعی
۲۱۱ رویکرد تحلیلی متن
۲۱۳ رویکرد تحلیلی احساسات
۲۱۳ تحلیل مصور

۲۱۴	نقشه‌های حرارتی
۲۱۵	نمودارهای سری زمانی
۲۱۷	گراف‌های شبکه
۲۱۸	نقشه‌نگاری داده‌های مکانی
۲۲۰	مطالعه موردی
۲۲۰	همبستگی
۲۲۰	رگرسیون
۲۲۰	نمودار سری‌های زمانی
۲۲۱	خوشه‌بندی
۲۲۱	دسته‌بندی

۲۲۳ پیوست: نتیجه‌گیری مطالعه موردی

۲۲۵ واژه‌نامه



سخنی از مترجمان

نویسندگان در این کتاب راجع به داده‌های بزرگ سخن گفته‌اند. داده‌های بزرگ اصطلاحی است برای مجموعه داده‌های حجیم که بزرگ، متنوع، با ساختار پیچیده و با دشواری‌هایی برای ذخیره‌سازی، تحلیل و مصورسازی است. فرآیند پژوهش روی داده‌های بزرگ برای آشکارسازی الگوهای پنهان، تجزیه و تحلیل داده‌های بزرگ نامیده می‌شود. این اطلاعات مفید به سازمان‌ها و شرکت‌ها در جهت کسب بینش غنی‌تر و عمیق‌تر و موفقیت در رقابت کمک می‌کند. کتاب حاضر در دو بخش ارائه شده است که در بخش اول مفاهیم پایه و اصول داده‌های بزرگ ارائه می‌شود. بخش دوم در خصوص ذخیره و تحلیل داده‌های بزرگ است.

دانشجویان مقاطع مختلف رشته‌های مهندسی، علوم کامپیوتر، فناوری اطلاعات و آمار که به دنبال رویکردهایی نو در فناوری اطلاعات برای حل مسائل و مشکلات هستند، می‌توانند از مفاهیم و تکنیک‌های تشریح شده در این کتاب استفاده کنند. ضمن این‌که در ترجمه کتاب، اغلب اسامی و نام‌های اختصاصی برای مفاهیم، اصطلاحات، تجهیزات، سخت‌افزارها و نرم‌افزارها به همان حالت انگلیسی خود آورده شده تا دانشجویان بیشتر با آن‌ها آشنا شوند.

امید است ترجمه این کتاب از نویسندگانی که در پیشبرد و ارتقای علم در حوزه فناوری اطلاعات نقش برجسته‌ای داشته‌اند، گامی رو به جلو در جامعه دانشگاهی و حرفه‌ای کشورمان محسوب شود. بی‌شک پیشنهادها و انتقادهای خوانندگان کتاب برای افزایش هرچه بیشتر سطح کیفی این ترجمه در چاپ‌های بعدی مؤثر خواهد بود.

گروه مترجمان

