



کلاس دهه‌ها: از تئوری تا کاربرد

مؤلفان:

سید عبدالرضا تاجی

استادیار و عضو هیئت علمی دانشگاه تهران

سجاد شفیعی‌فر

فارغ التحصیل کارشناسی ارشد دانشگاه تهران

تراپی، سید عبدالرضا، ۱۳۶۴.

کلان داده‌ها: ازتثوری تا کاربرد/ مولفان عبدالرضا ترابی، سجاد شفيعی فر.

تهران: جهاددانشگاهی، واحد صنعتی شریف، ۱۳۹۸.

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

97A-96F-6445-46-A

فيا

داده‌های کلان

Big Data

داده‌های کلان -- مدیریت

Big Data-- Management

شفیعی فر، سجاد، ۱۳۶۹.

جهاد دانشگاهی، واحد صنعتی شریف.

QA 

113/50

01.27.20



سازمان جهاد اسلامی صنعتی سرف

عنوان کتاب: کلان داده‌ها: از تئوری تا کاربرد

مولفان: سید عبدالرضا ترابی و سجاد شیعہ فر

ناشر: جهاد دانشگاهی صنعتی شریف

طراح جلد: سید حسن کمالی

نوبت چاپ: اول

تاریخ چاپ: تابستان ۱۳۹۸

تیراژ: ۱۰۰۰ نسخه

قیمت: ۳۰۰۰۰ ریال

کلیه حقوق، برای جهاد دانشگاهی صنعتی شریف محفوظ است.

ISBN: 978-964-6445-46-8

شاک: ۸-۴۶-۶۴۴۵-۹۶۴-۹۷۸

نشانی: تهران، خیابان آزادی، شمال دانشگاه صنعتی شریف، خیابان شهید قاسمی، پلاک ۷۱

تلفن: ۰۱۱-۶۶۰۷۵، شماره: ۰۷۵۲۵۰-۶۶، پست الکترونیک: journal@jdsharif.ac.ir

پیشگفتار مولفان

کلان داده‌ها در سال‌های اخیر یکی از موضوعات جذاب برای پژوهشگران و محققان شده است. مدت‌ها بود که واژه کلان داده به حجم عظیمی از داده‌ها اشاره داشت ولی امروزه با گسترش مفهوم آن، به داده‌های بزرگی گفته می‌شود که با ابزارهای مدیریتی و پایگاه‌های داده معمولی قابل مدیریت نیستند. دولت الکترونیک، تجارت الکترونیک، حوزه بهداشت و امنیت، تجزیه و تحلیل وب و شبکه‌های اجتماعی به نوعی سرویس‌هایی هستند که با موضوع کلان داده عجین شده‌اند. در تمامی این کاربردها پردازش کلان داده‌ها موضوع اول و مهم است. اهمیت این موضوع آنقدر بالا رفته که کرسی‌های درسی دانشگاهی با محوریت این موضوع در سال‌های اخیر شکل گرفته تا به‌صورت مدون و ریشه‌ای به تدریس و پژوهش در این مورد پرداخته شود. اکتساب داده، آماده‌سازی داده، تجزیه و تحلیل داده، رابط پرس و جو و مصورسازی از سرفصل‌های مهم بحث پردازش کلان داده‌هاست.

در این کتاب سعی کرده‌ایم ازبانی ساده ولی تا حد امکان دقیق و شفاف به بررسی موضوع کلان داده‌ها از پله تنوری تا پله کاربردها بپردازیم. برای دستیابی به این هدف یادگیری مفاهیمی چون معماری شناسی کلان داده‌ها، شناسی انواع الگوریتم‌های پردازش کلان داده‌ها، روش‌های جستجو و داده کاوی در کلان داده‌ها، مباحث بیشتری توضیح داده شده است. الگوی MapReduce یکی از مفاهیم مهم در بحث پردازش کلان داده‌هاست که در این کتاب به تفصیل در مورد آن بحث شده است. در انتها به موضوع بسیار مهم امنیت و حریم خصوصی در کلان داده‌ها پرداخته شده است که با توجه به کاربردهای موضوع کلان داده اهمیت بسیار بالایی دارد.

اساتید و محققان دانشگاه‌ها، دانشجویان مقاطع تحصیلات تکمیلی بخصوص رشته مهندسی کامپیوتر، هوش مصنوعی، مهندسی برق-مخابرات و پردیس‌های نوآوری می‌توانند مخاطبان اصلی این کتاب باشند. در انتها بر خود لازم می‌بینیم که از حمایت‌های مرکز پژوهشی رادیومتری فرکانس پایین دانشگاه تهران در نگارش و تکمیل این کتاب و در اختیار قرار دادن بستر مناسب در راستای این چنین تحقیقاتی سپاسگزاری کنیم. از سرکار خانم آرزو مکرمی مدیر انتقادات جهاد دانشگاهی صنعتی شریف به دلیل هماهنگی و پیگیری‌های مجدانه و جناب آقای محسن حسن‌کمالی به دلیل طراحی جلد تشکر می‌نماییم.

تمام دانش‌پژوهان و کارشناسان می‌توانند جهت بهبود کار از طریق آدرس الکترونیکی ab.reza.torabi@ut.ac.ir مکاتبه کرده و ما را از نظرات و پیشنهادات سازنده خود مطلع سازند.

دکتر سید عبدالرضا ترابی

استادیار و عضو هیئت علمی دانشگاه تهران

سجاد شفیعی

فارغ التحصیل کارشناسی ارشد دانشگاه تهران

۱۵	فصل اول: مقدمه
۱۵	۱-۱- کاربردهای کلان داده‌ها
۱۹	۲-۱- کلان داده‌ها به‌عنوان سرویس
۲۲	۳-۱- پردازش کلان داده‌ها
۲۳	۱-۳-۱- گام‌های پردازش
۲۴	۲-۳-۱- چالش‌ها
۲۶	۳-۳-۱- استمرار عمل‌ها
۲۷	۴-۱- سیستم کلان داده‌ها
۲۸	۵-۱- فناوری ابزارها
۲۹	۱-۵-۱- فناوری سخت‌افزار
۳۰	۲-۵-۱- فناوری نرم‌افزاری
۳۲	۶-۱- ابزار کلان داده‌ها
۳۸	۷-۱- موضوعات برای مطالعه بیشتر
۴۱	۸-۱- منابع
۴۵	فصل دوم: معماری کلان داده‌ها
۴۶	۱-۲- فضای کلان داده‌ها
۵۰	۲-۲- سطح بعدی داده‌ها
۵۱	۳-۲- ویژگی‌های داده
۵۶	۱-۳-۲- دریاچه‌ی داده
۵۷	۲-۳-۲- تولید داده
۵۸	۳-۳-۲- تحقیق و توسعه داده
۵۹	۴-۳-۲- برداشت ارزش از داده‌ها
۶۱	۴-۲- استفاده از سه مؤلفه عملکردی در عمل
۶۱	۵-۲- تحقیق و توسعه‌ی داده زمینه ظهور نوآوری‌ها
۶۶	۶-۲- ساخت معماری داده
۶۶	۱-۶-۲- پلتفرم داده یا دریاچه داده
۷۰	۲-۶-۲- تولید داده
۷۱	۳-۶-۲- تحقیق و توسعه داده یا پلتفرم کشف

۷۲	۲-۶-۴- یکپارچه کردن
۷۴	۲-۶-۵- ویژگی‌های معماری
۷۶	۲-۷- منابع
۷۷	فصل سوم: الگوریتم‌های پردازش کلان داده‌ها
۷۷	۳-۱- چالش‌ها با کلان داده‌ها
۷۹	۳-۲- سیستم چند هسته‌ای در مقایسه با سیستم‌های توزیع شده
۷۹	۳-۳- الگوریتم‌های توزیع شده
۸۰	۳-۴- گشت تصادفی
۸۱	۳-۵- جدول درهم توزیع شده
۸۱	۳-۶- پردازش دسته‌ای هم‌زمان
۸۲	۳-۷- الگوی MapReduce
۸۵	۳-۸- معرفی ابزارهای هدوب
۸۵	۳-۸-۱- سیستم فایل توزیع شده هدوب
۸۷	۳-۸-۲- گره نام
۸۸	۳-۸-۳- گره داده‌ها
۹۰	۳-۸-۴- پلتفرم محاسباتی MapReduce
۹۴	۳-۸-۵- یارن
۹۶	۳-۸-۶- تقسیم‌کننده‌ها و ادغام‌کننده‌ها
۹۷	۳-۸-۷- خواننده‌های ورودی و نویسنده‌های خروجی
۹۹	۳-۸-۸- جمع‌بندی از MapReduce
۱۰۶	۳-۹- روش‌های پیشرفته MapReduce
۱۰۶	۳-۹-۱- جریان هدوب
۱۰۷	۳-۹-۲- نهان‌سازی توزیع شده
۱۰۷	۳-۹-۳- خروجی چندگانه
۱۰۸	۳-۹-۴- MapReduce تکراری
۱۰۸	۳-۹-۵- یادگیری ماشین با MapReduce
۱۱۰	۳-۱۰- نتیجه‌گیری
۱۱۲	۳-۱۱- منابع
۱۱۳	فصل چهارم: جستجو و داده‌کاوی در کلان داده‌ها
۱۱۶	۴-۱- خوشه‌بندی K-Means

۱۱۸	۲-۴- شبکه‌های اجتماعی-تشخیص انجمن
۱۲۰	۳-۴- خوشه‌بندی شبکه‌های اجتماعی-تشخیص توپولوژی
۱۲۰	۴-۳-۱- معیارهای اندازه‌گیری در شبکه‌های اجتماعی
۱۲۴	۴-۴- الگوریتم خوشه‌بندی برای پیدا کردن توپولوژی شبکه
۱۲۴	۴-۴-۱- تعیین وزن یال
۱۲۵	۴-۴-۲- خوشه‌بندی
۱۲۵	۴-۴-۳- ارزیابی آزمایشی
۱۲۷	۴-۵- خلاصه‌ای از شبکه‌های اجتماعی
۱۲۹	۴-۵-۱- احساس کلوی متن
۱۳۰	۴-۵-۲- ارزیابی احساسات
۱۳۱	۴-۵-۱-۱- گیری فعال
۱۳۲	۴-۵-۴- <i>Concept drift</i>
۱۳۳	۴-۵-۵- استفاده از احساسات از وب
۱۳۵	۴-۵-۶- تعمیم محلی
۱۳۵	۴-۵-۷- بهروزرسانی محلی
۱۳۶	۴-۵-۸- هسته در فضای ویژگی
۱۳۷	۴-۵-۹- بازخورد کاربر در بهروزرسانی همسایگی
۱۴۱	۴-۶- ابزارهای آنالیز و داده‌کاوی کلان داده‌ها
۱۴۳	۴-۷- منابع
۱۴۵	فصل پنجم: امنیت و حریم خصوصی در کلان داده‌ها
۱۴۵	۵-۱- ملزومات طراحی
۱۴۹	۵-۲- امنیت
۱۵۰	۵-۲-۱- امنیت کلان داده‌ها در IT
۱۵۰	۵-۲-۲- امنیت کلان داده‌ها در زیرساخت‌های حیاتی
۱۵۲	۵-۲-۳- مجرمانه بودن
۱۵۴	۵-۲-۴- یکپارچگی
۱۵۴	۵-۲-۵- در دسترس بودن
۱۵۵	۵-۲-۶- حریم خصوصی
۱۵۶	۵-۲-۷- حریم خصوصی آنلاین
۱۵۷	۵-۲-۸- حریم خصوصی آفلاین
۱۵۷	۵-۲-۹- حریم خصوصی داده‌های آرشیو شده بسیار قدیمی

- ۱۵۸ ۳-۵- ظهور روندها
- ۱۵۸ ۱-۳-۵- روند تا به اینجا
- ۱۵۸ ۲-۳-۵- چشم انداز
- ۱۵۸ ۳-۳-۵- چالش ها
- ۱۶۰ ۴-۵- منابع

www.ketab.ir