

---

---

# کلان داده

## اصول و بهترین اقدامات

### سیستم‌های داده‌ای بلادرنگ مقیاس پذیر

---

---

**مؤلفین:**

مارز، ناتان

با جیمز وارن

**مترجمین:**

دکتر جواد وحیدی – دانشگاه علم و صنعت ایران

دکتر رمضان عباس نژادورزی

مهندس زهرا علیجان نژادبایی



فن‌آوری نوین

---

---

سرشناسه	: مارز، ناتان <b>Marz, Nathan</b>
عنوان و نام پدیدآور	: کلان داده: اصول و بهترین اقدامات سیستم‌های داده‌ای بلادرنگ مقیاس پذیر / مولفین ناتان مارز، جیمز وارن؛ مترجمین جواد وحیدی، رمضان عباس‌زادورزی، زهرا علیجان‌زادبایی.
مشخصات نشر	: بابل: فناوری نوین، ۱۳۹۹.
مشخصات ظاهری	: ۳۴۹ ص: مصور(بخشی رنگی)، جدول(بخشی رنگی)، نمودار(بخشی رنگی).
شابک	: 0-20-7393-622-978
وضعیت فهرست نویسی	: فیبا
یادداشت	: عنوان اصلی: <b>Big data : principles and best practices of scalable real-time data systems.</b>
یادداشت	: ترجمه دیگری از کتاب حاضر با عنوان "داده‌های کلان: اصول و بهترین تجارب سیستم‌های داده‌ای بلادرنگ مقیاس پذیر" توسط ارسطو فیبا دریافت کرده است.
عنوان دیگر	: اصول و بهترین اقدامات سیستم‌های داده‌ای بلادرنگ مقیاس پذیر.
عنوان دیگر	: داده‌های کلان: اصول و بهترین تجارب سیستم‌های داده‌ای بلادرنگ مقیاس پذیر.
موضوع	: داده‌های کلان
موضوع	: <b>Big data</b>
موضوع	: پایگاه‌های اطلاعاتی -- مدیریت
موضوع	: <b>Database management</b>
موضوع	: پایگاه‌های اطلاعاتی -- طراحی
موضوع	: <b>Database design</b>
موضوع	: داده کاوی
موضوع	: <b>Data mining</b>
شناسه افزوده	: وارن، جیمز
شناسه افزوده	: <b>Warren, James</b>
شناسه افزوده	: وحیدی، جواد، ۱۳۴۸ - مترجم
شناسه افزوده	: <b>Vahidi, Javad</b>
شناسه افزوده	: عباس‌زاد ورزی، رمضان، ۱۳۴۸ - مترجم
شناسه افزوده	: علیجان‌زاد بایی، زهرا، ۱۳۶۶ - مترجم
زده بندی کنگره	: ۷۶۱۹۹۸
زده بندی دیویی	: ۶۵۸/۴۰۳۸
شماره کتابشناسی ملی	: ۷۳۹۱۵۹۶
وضعیت رکورد	: فیبا

تلفن: ۰۱۱-۳۲۲۵۶۶۸۷

بابل، کد پستی ۷۳۴۴۸-۴۷۱۶۷

فن آوری نوین

کلان داده اصول و بهترین اقدامات سیستم‌های داده‌ای بلادرنگ مقیاس پذیر  
ترجمه: جواد وحیدی، رمضان عباس‌زادورزی، زهرا علیجان‌زادبایی.

نوبت چاپ: چاپ اول

سال چاپ: تابستان ۹۹

شمارگان: ۲۰۰

نام چاپخانه و صحافی: دفتر فنی سورنا

حروف چینی و تایپ: فناوری نوین

شابک: 978-622-7393-20-0

نشانی ناشر: بابل، چهارراه نواب، کاظم بیگی، جنب مسجد منصور کاظم بیگی، طبقه اول

طراح جلد: کانون آکبھی و تبلیغات آبان (احمد فرجی)

تهران، خ اردیبهشت، نبش وحید نظری، پلاک ۱۴۲ تلفکس: ۶۶۴۰۰۲۲۰-۶۶۴۰۰۱۴۴

## فهرست مطالب

### فصل ۱: یک الگوی جدید برای کلان داده ..... ۱۴

۱-۱. نحوه ساخت این کتاب .....	۱۵
۱-۲. مقیاس گذاری با یک بانک اطلاعاتی سنتی .....	۱۵
۱-۲-۱. مقیاس گذاری با یک صف .....	۱۶
۱-۲-۲. مقیاس گذاری با خرد کردن بانک اطلاعات .....	۱۷
۱-۲-۳. آغاز مسائل مربوط به تحمل خطا .....	۱۸
۱-۲-۴. مسائل مربوط به انحراف .....	۱۸
۱-۲-۵. چه اشتباهی رخ داده است؟ .....	۱۸
۱-۲-۶. چگونه تکنیک های کلان داده کمک خواهد کرد؟ .....	۱۹
۱-۳. NoSQL یک پاناسرا نیست .....	۱۹
۱-۴. اصول اول .....	۲۰
۱-۵. خواص کلان داده .....	۲۱
۱-۵-۱. استحکام و تحمل خطا .....	۲۱
۱-۵-۲. خواندن و به روز رسانی با زمان تأخیر کم .....	۲۱
۱-۵-۳. مقیاس پذیری .....	۲۲
۱-۵-۴. قابلیت توسعه .....	۲۲
۱-۵-۵. پرس و جوهای Ad hoc .....	۲۲
۱-۵-۶. حداقل نگهداری .....	۲۲
۱-۵-۷. اشکال زدایی .....	۲۲
۱-۶. مشکلات معماری کاملاً افزایشی .....	۲۳
۱-۶-۱. پیچیدگی عملیاتی .....	۲۴
۱-۶-۲. پیچیدگی شدید دستیابی به قوام نهایی .....	۲۵
۱-۶-۳. عدم تحمل خطای انسانی .....	۲۵
۱-۶-۴. راه حل کاملاً افزایشی در مقابل راه حل معماری لامبدا .....	۲۸
۱-۷. معماری لامبدا .....	۲۹
۱-۷-۱. لایه دسته ای .....	۳۱
۱-۷-۲. لایه سرویس .....	۳۲
۱-۷-۳. پیچیدگی عملی لایه های دسته ای و خدماتی تقریباً تمام خصوصیات را بر آورده می کنند .....	۳۳
۱-۷-۴. لایه سرعت .....	۳۴
۱-۸. روندهای اخیر در فناوری .....	۳۷
۱-۸-۱. پردازنده ها سریع تر نمی شوند .....	۳۷
۱-۸-۲. ابرهای الاستیک .....	۳۷
۱-۸-۳. اکوسیستم فعال منبع باز برای کلان داده ها .....	۳۸
۱-۹. برنامه مثال: SuperWebAnalytics.com .....	۳۹
۱-۱۰. خلاصه .....	۴۰

### بخش ۱: لایه دسته ای ..... ۴۱

### فصل ۲: مدل داده برای کلان داده ..... ۴۲

۲-۱. خصوصیات داده ها .....	۴۳
۲-۱-۱. داده خام است .....	۴۶
۲-۱-۲. داده تغییرناپذیر است .....	۴۹
۲-۱-۳. داده همیشه صحیح .....	۵۲
۲-۲. مدل مبتنی بر واقعیت برای نمایش داده ها .....	۵۳
۲-۲-۱. مثال ها و خصوصیات آن ها .....	۵۴
۲-۲-۲. مزایای مدل واقعیت محور .....	۵۶
۲-۳. نمودارهای نمودار .....	۶۰
۲-۳-۱. عناصر شمای یک نمودار .....	۶۰

۶۲	..... نیاز به یک طرح اجرایی	۲-۳-۲
۶۳	..... SuperWebAnalytics.com یک مدل داده کامل برای	۲-۴
۶۴	..... خلاصه	۲-۵
<b>۶۵</b>	<b>..... فصل ۳: مدل داده برای کلان داده: تصویر</b>	
۶۵	..... چرا یک چارچوب سریال سازی؟	۳-۱
۶۶	..... رونق ایچی	۳-۲
۶۶	..... گره	۳-۲-۱
۶۷	..... لبه‌ها	۳-۲-۲
۶۷	..... ویژگی‌ها	۳-۲-۳
۶۸	..... اتصال همه چیز به یکدیگر در اشیاء داده	۳-۲-۴
۶۹	..... طرح در حال تحول است	۳-۲-۵
۹۷	..... محدودیت‌های چارچوب‌های سریالی سازی	۳-۳
۷۹	..... خلاصه	۳-۴
<b>۷۲</b>	<b>..... فصل ۴: ذخیره داده در لایه دسته‌ای</b>	
۷۲	..... شرایط ذخیره سازی برای مجموعه داده‌های اصلی	۴-۱
۷۳	..... انتخاب راه حل ذخیره سازی برای لایه دسته‌ای	۴-۲
۷۴	..... استفاده از یک فروشگاه کلید / ارزش برای مجموعه داده‌های اصلی	۴-۲-۱
۷۴	..... سیستم‌های فایل توزیع شده	۴-۲-۱
۷۶	..... نحوه کار سیستم‌های فایل توزیع شده	۴-۳
۷۸	..... ذخیره یک مجموعه داده اصلی با سیستم فایل توزیع شده	۴-۴
۸۰	..... پارتیشن بندی عمودی	۴-۵
۸۱	..... ماهیت سطح پایین سیستم‌های فایل توزیع شده	۴-۶
۸۳	..... ذخیره مجموعه داده اصلی اسناد SuperWebAnalytics.com در سیستم فایل توزیع شده	۴-۷
۸۴	..... خلاصه	۴-۸
<b>۸۵</b>	<b>..... فصل ۵: ذخیره سازی داده‌ها روی لایه دسته‌ای: تصویر</b>	
۸۵	..... با استفاده از سیستم فایل توزیع شده Hadoop	۵-۱
۸۷	..... مشکل فایل‌های کوچک	۵-۱-۱
۸۷	..... به سمت انتزاع سطح بالاتر	۵-۱-۲
۸۸	..... ذخیره سازی داده‌ها در لایه دسته‌ای با قوطی	۵-۲
۸۹	..... عملیات پایه Pail	۵-۲-۱
۹۱	..... مرتب کردن اشیاء در سطرها	۵-۲-۲
۹۳	..... عملیات دسته‌ای با استفاده از Pail	۵-۲-۳
۹۴	..... پارتیشن بندی عمودی با Pail	۵-۲-۴
۹۴	..... قالب‌ها و فشرده سازی فایل‌های Pail	۵-۲-۵
۹۶	..... خلاصه مزایای Pail	۵-۲-۶
۹۶	..... ذخیره مجموعه داده‌های اصلی SuperWebAnalytics.com	۵-۳
۹۸	..... یک میله ساختاری برای اشیاء Thrift	۵-۳-۱
۱۰۰	..... یک قوطی تقسیم شده برای تقسیم بندی عمودی در مجموع داده	۵-۳-۳
۱۰۴	..... خلاصه	۵-۴
<b>۱۰۶</b>	<b>..... فصل ۶: لایه دسته‌ای</b>	
۱۰۷	..... مثال‌های انگیزشی	۶-۱
۱۰۷	..... تعداد بازدیدهای صفحه با گذشت زمان	۶-۱-۱
۱۰۷	..... استنتاج جنسیتی	۶-۱-۲
۱۰۸	..... امتیاز نفوذ	۶-۱-۳
۱۰۹	..... محاسبه بر روی لایه دسته	۶-۲
۱۱۲	..... الگوریتم‌های اعتبار سنجی در مقابل الگوریتم‌های افزایشی	۶-۳
۱۱۳	..... عملکرد	۶-۳-۱
۱۱۴	..... تحمل خطای انسانی	۶-۳-۲
۱۱۵	..... کلی بودن الگوریتم‌ها	۶-۳-۳
۱۱۵	..... انتخاب یک سبک الگوریتم	۶-۳-۴