

فهرست مطالب

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

بخش ۱: لایه دسته ای ۴۱

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

۶۲	۲-۲-۲. نیاز به یک طرح اجرایی
۶۳	۲-۴. یک مدل داده کامل برای SuperWebAnalytics.com
۶۴	۲-۵. خلاصه

فصل ۳: مدل داده برای کلان داده: تصویر

۶۵	۳-۱. چرا یک چارچوب سریال سازی؟
۶۶	۳-۲. رونق ایاچی
۶۶	۳-۲-۱. گروه
۶۷	۳-۲-۲. لبه‌ها
۶۷	۳-۲-۳. ویژگی‌ها
۶۸	۳-۲-۴. اتصال همه چیز به یکدیگر در اشیاء داده
۶۹	۳-۲-۵. طرح در حال تحول است
۹۷	۳-۳. محدودیت‌های چارچوب‌های سریالی سازی
۷۹	۳-۴. خلاصه

فصل ۴: ذخیره داده در لایه دسته‌ای

۷۲	۴-۱. شرایط ذخیره سازی برای مجموعه داده‌های اصلی
۷۳	۴-۲. انتخاب راه حل ذخیره سازی برای لایه دسته‌ای
۷۴	۴-۲-۱. استفاده از یک فروشگاه کلید / ارزش برای مجموعه داده‌های اصلی
۷۴	۴-۲-۱. سیستم‌های قابل توزیع شده
۷۶	۴-۳. نحوه کار سیستم‌های فایل توزیع شده
۷۸	۴-۴. ذخیره یک مجموعه داده اصلی با سیستم فایل توزیع شده
۸۰	۴-۵. پارتیشن بندی عمودی
۸۱	۴-۶. ماهیت سطح پایین سیستم‌های فایل توزیع شده
۸۳	۴-۷. ذخیره مجموعه داده اصلی استاد SuperWebAnalytics.com در سیستم فایل توزیع شده
۸۴	۴-۸. خلاصه

فصل ۵: ذخیره سازی داده‌ها روی لایه دسته‌ای: تصویر

۸۵	۵-۱. با استفاده از سیستم فایل توزیع شده Hadoop
۸۷	۵-۱-۱. مشکل فایل‌های کوچک
۸۷	۵-۱-۲. به سمت انتزاع سطح بالاتر
۸۸	۵-۲. ذخیره سازی داده‌ها در لایه دسته‌ای با قوطی
۸۹	۵-۲-۱. عملیات پایه Pail
۹۱	۵-۲-۲. مرتب کردن اشیاء در سطل‌ها
۹۳	۵-۲-۳. عملیات دسته‌ای با استفاده از Pail
۹۴	۵-۲-۴. پارتیشن بندی عمودی با Pail
۹۴	۵-۲-۵. فایل‌ها و فرده سازی فایل‌های Pail
۹۶	۵-۲-۶. خلاصه مزایای Pail
۹۶	۵-۳. ذخیره مجموعه داده‌های اصلی SuperWebAnalytics.com
۹۸	۵-۳-۱. یک میله ساختاری برای اشیاء Thrift
۱۰۰	۵-۳-۳. یک قوطی تقسیم شده برای تقسیم بندی عمودی در مجموع داده
۱۰۴	۵-۴. خلاصه

فصل ۶: لایه دسته‌ای

۱۰۷	۶-۱. مثال‌های انگیزشی
۱۰۷	۶-۱-۱. تعداد بازدیدهای صفحه با گذشت زمان
۱۰۷	۶-۱-۲. استنتاج چسبنی
۱۰۸	۶-۱-۳. امتیاز نلود
۱۰۹	۶-۲. محاسبه بر روی لایه دسته
۱۱۲	۶-۳. الگوریتم‌های اعتبار سنجی در مقابل الگوریتم‌های افزایشی
۱۱۳	۶-۳-۱. همگراد
۱۱۴	۶-۳-۲. تحمل خطای انسانی
۱۱۵	۶-۳-۳. کلی بودن الگوریتم‌ها
۱۱۵	۶-۳-۴. انتخاب یک سبک الگوریتم

۱۱۶	۶-۴. مقیاس‌پذیری در لایه دسته
۱۱۷	۶-۵. MapReduce: الگویی برای Big Data computing
۱۱۸	۶-۵-۱. مقیاس‌پذیری
۱۲۱	۶-۵-۲. تحمل خطا
۱۲۱	۶-۵-۳. MapReduce کلیت
۱۲۳	۶-۶. ماهیت سطح پایین MapReduce
۱۲۴	۶-۶-۱. محاسبات چند مرحله غیرطبیعی است
۱۲۴	۶-۶-۲. پیوست‌ها برای اجرای دستی بسیار پیچیده هستند
۱۲۶	۶-۶-۳. اجرای منطقی و بدنی همراه با استحکام
۱۲۷	۶-۷. نمودار لوله: یک روش تفکر در سطح بالاتری است
۱۲۸	۶-۷-۱. مفاهیم نمودارهای لوله
۱۳۳	۶-۷-۲. اجرای نمودارهای لوله‌ای از طریق MapReduce
۱۳۶	۶-۷-۳. جمع‌کننده‌های ترکیبی
۱۳۸	۶-۷-۴. نمونه‌های نمودار لوله‌ای
۱۳۹	۶-۸. خلاصه

فصل ۷. لایه دسته‌ای: تصویر ۱۴۱

۱۴۱	۷-۱. مثال بارز
۱۴۳	۷-۲. اشکالات متداول ابزارهای پردازش داده
۱۴۳	۷-۲-۱. زبان‌های رسمی
۱۴۴	۷-۲-۲. انتزاعات ضعیف سازگار
۱۴۵	۷-۳. مقدمه‌ای برای JCascalog
۱۴۳	۷-۳-۱. مدل داده JCascalog
۱۴۷	۷-۳-۲. ساختار یک پرس‌وجوی JCascalog
۱۴۹	۷-۳-۳. پرس‌وجو از مجموعه داده‌های متعدد
۱۵۱	۷-۳-۴. گروه‌بندی و جمع‌کننده
۱۵۳	۷-۳-۵. پیمایش از مثال یک پرس‌وجو
۱۵۶	۷-۳-۶. عملیات مستندسازی پیش‌فرض

۱۶۳	۷-۴. ترکیب
۱۶۳	۷-۴-۱. ترکیب مواد فرعی
۱۶۵	۷-۴-۲. ایجاد Subqueries بویا
۱۶۸	۷-۴-۳. ماکروها پیش‌بینی‌شده
۱۷۱	۷-۴-۴. ایجاد گزاره‌های ماکروی بویا
۱۷۳	۷-۵. خلاصه

فصل ۸. یک لایه دسته‌ای مثال: معماری و الگوریتم ۱۷۴

۱۷۴	۸-۱. طراحی لایه دسته‌ای SuperWebAnalytics.com
۱۷۵	۸-۱-۱. پرس‌وجوهای پشتیبانی
۱۷۶	۸-۱-۲. نمایش دسته‌ای
۱۷۶	۸-۲. بررسی اجمالی گردش کار
۱۷۹	۸-۳. مصرف داده‌های جدید
۱۸۱	۸-۴. عادی‌سازی URL
۱۸۱	۸-۵. عادی‌سازی شناسه کاربر
۱۸۲	۸-۶. نمایش صفحه‌های اختصاصی
۱۸۹	۸-۷. نمایش دسته‌ای محاسبه
۱۸۹	۸-۷-۱. بازدید از صفحه باگذشت زمان
۱۹۱	۸-۷-۲. بازدیدکنندگان منحصر به فرد باگذشت زمان
۱۹۱	۸-۷-۳. تجزیه و تحلیل نرخ بازگشت
۱۹۳	۸-۸. خلاصه

فصل ۹. یک لایه دسته‌ای مثال: پیاده‌سازی ۱۹۵

۱۹۵	۹-۱. نقطه شروع
۱۹۶	۹-۲. تهیه گردش کار

۱۹۷	۹-۳. مصرف داده‌های جدید
۲۰۱	۹-۴. عادی‌سازی URL
۲۰۲	۹-۵. نرمال‌سازی شناسه کاربر
۲۰۹	۹-۶. نمایش صفحه‌های اختصاصی
۲۰۹	۹-۷. نمای دسته‌ای محاسبه
۲۰۹	۹-۷-۱. بازدید از صفحه باگذشت زمان
۲۱۱	۹-۷-۲. منحصربه‌فرد در طول زمان
۲۱۳	۹-۷-۳. تجزیه و تحلیل نرخ پرش
۲۱۶	۴-۸. خلاصه

بخش ۲: لایه سرویس ۲۱۷

فصل ۱۰: خدمت لایه ۲۱۸

۲۲۰	۱۰-۱. معیارهای عملکرد برای لایه سرویس
۲۲۲	۱۰-۲. راه‌حل لایه خدمت به مشکل عادی‌سازی / نرمال‌سازی
۲۲۴	۱۰-۳. مورد نیاز برای یک بانک اطلاعاتی لایه سرویس
۲۲۶	۱۰-۴. طراحی یک لایه سرویس دهنده برای SuperWebAnalytics.com
۲۲۶	۱۰-۴-۱. بازدید از صفحه باگذشت زمان
۲۲۷	۱۰-۴-۲. منحصربه‌فرد در طول زمان
۲۲۸	۱۰-۴-۳. تجزیه و تحلیل نرخ پرش
۲۲۸	۱۰-۵. تضاد با یک راه‌حل کاملاً افزایشی
۲۲۹	۱۰-۵-۱. راه‌حل کاملاً افزایشی منحصربه‌فرد در طول زمان
۲۳۵	۱۰-۵-۲. مقایسه با راه‌حل معماری لامبدا
۲۳۶	۱۰-۶. خلاصه

فصل ۱۱: لایه خدمت: تصویر ۲۳۷

۲۳۷	۱۱-۱. میانی ElephantDB
۲۳۸	۱۱-۱-۱. ایجاد مشاهده در ElephantDB
۲۳۸	۱۱-۱-۲. مشاهده سرویس در ElephantDB
۲۳۸	۱۱-۱-۳. استفاده از ElephantDB
۲۴۱	۱۱-۲. ساخت لایه سرویس دهنده برای ۲۰۰ SuperWebAnalytics.com
۲۴۱	۱۱-۲-۱. بازدید از صفحه باگذشت زمان
۲۴۴	۱۱-۲-۲. منحصربه‌فرد باگذشت زمان
۲۴۵	۱۱-۱-۳. تجزیه و تحلیل نرخ پرش

۱۱-۳. خلاصه ۲۴۶

بخش ۳: لایه سریع ۲۴۷

فصل ۱۲: نمایش زمان واقعی ۲۴۸

۲۵۰	۱۲-۱. محاسبه نماهای بلادرنگ
۲۵۱	۱۲-۲. ذخیره نماهای در زمان واقعی
۲۵۲	۱۲-۲-۱. دقت نهایی
۲۵۳	۱۲-۲-۲. مقدار حالت ذخیره شده در لایه سرعت
۲۵۳	۱۲-۳. چالش‌های محاسبات افزایشی
۲۵۴	۱۲-۳-۱. اعتبار قضیه CAP
۲۵۶	۱۲-۳-۲. تعامل پیچیده بین قضیه CAP و الگوریتم‌های افزایشی
۲۵۷	۱۲-۴. ناهم‌زمان در مقابل به‌روزرسانی‌های هم‌زمان
۲۵۹	۱۲-۵. نمایان شده در زمان واقعی
۲۶۱	۱۲-۶. خلاصه

فصل ۱۳: نمایش زمان واقعی: تصویر ۲۶۲

۲۶۲	۱۳-۱. مدل داده کاساندر
۲۶۴	۱۳-۲. با استفاده از کاساندر

۲۶۶	۱۳-۲-۱. کاساندرای پیشرفته
۲۶۷	۱۳-۳. خلاصه
۲۶۸	فصل ۱۴: پردازش صف و جریان
۲۶۹	۱۴-۱. صف
۲۶۹	۱۴-۱-۱. سرورهای صف تک مصرفی
۲۷۱	۱۴-۱-۲. صف‌های چند مصرف‌کننده‌ای
۲۷۲	۱۴-۲. پردازش جریان
۲۷۳	۱۴-۲-۱. صف و کارگران
۲۷۴	۱۴-۲-۲. مشکلات در صف کار و کارگران
۲۷۵	۱۴-۳. پردازش جریان سطح یک و یک بار
۲۷۵	۱۴-۳-۱. مدل طوفان
۲۸۱	۱۴-۳-۲. ضمانت پردازش پیام
۲۸۳	۱۴-۴. لایه سرعت SuperWebAnalytics.com
۲۸۵	۱۴-۴-۱. ساختار توپولوژی
۲۸۶	۱۴-۵. خلاصه
۲۸۸	فصل ۱۵: صف‌بندی و پردازش جریان: تصویر
۲۸۸	۱۵-۱. تعریف توپولوژی با طوفان آب‌جی
۲۹۰	۱۵-۲. خوشه‌های طوفان آب‌جی و استقرار
۲۹۴	۱۵-۳. ضمانت پردازش پیام
۲۹۶	۱۵-۴. اجرای لایه سرعت منحصربه‌فرد SuperWebAnalytics.com
۳۰۰	۱۵-۵. خلاصه
۳۰۱	فصل ۱۶: پردازش جریان میکرو دسته‌ای
۳۰۱	۱۶-۱. دستیابی به معانی دقیقاً یک‌بار
۳۰۲	۱۶-۱-۱. پردازش سفارشی شدید
۳۰۲	۱۶-۱-۲. پردازش جریان میکرو دسته‌ای
۳۰۴	۱۶-۱-۳. توپولوژی پردازش میکرو دسته‌ای
۳۰۷	۱۶-۲. مفاهیم اصلی پردازش جریان میکرو دسته‌ای
۳۰۸	۱۶-۳. گسترش نمودارهای لوله برای پردازش میکرو دسته
۳۰۹	۱۶-۴. به پایان رساندن لایه سرعت برای SuperWebAnalytics.com
۳۰۹	۱۶-۴-۱. بازدید از صفحه باگذشت زمان
۳۱۰	۱۶-۴-۲. تجزیه و تحلیل نرخ برش
۳۱۵	۱۶-۵. نگاهی دیگر به مثال تحلیلی نرخ انفجار
۳۱۶	۱۶-۶. خلاصه
۳۱۷	فصل ۱۷: پردازش جریان میکرو دسته‌ای: تصویر
۳۱۷	۱۷-۱. با استفاده از Trident
۳۲۱	۱۷-۲. به پایان رساندن لایه سرعت SuperWebAnalytics.com
۳۲۱	۱۷-۲-۱. بازدید از صفحه باگذشت زمان
۳۲۴	۱۷-۲-۲. تجزیه و تحلیل نرخ برش
۳۳۰	۱۷-۳. پردازش میکرو دسته‌ای کاملاً تحمل‌پذیر، در حافظه
۳۳۲	۱۷-۴. خلاصه
۳۳۳	فصل ۱۸: معماری لامبدا در عمق
۳۳۳	۱۸-۱. تعریف سیستم‌های داده
۳۳۴	۱۸-۲. دسته‌ها و لایه‌های سرور
۳۳۵	۱۸-۲-۱. پردازش دسته‌ای الزامی
۳۴۲	۱۸-۲-۲. اندازه‌گیری و بهینه‌سازی استفاده از منابع لایه‌ای دسته‌ای
۳۴۶	۱۸-۳. لایه سرعت
۳۴۷	۱۸-۴. لایه پرس‌وجو
۳۴۸	۱۸-۵. خلاصه