

مقدمه‌ای بر کلودرا

www.ketab.ir

نویسندگان

فرساد زمانی بروجنی

عاطفه کریمی

پروین احمدی دول امیری

سرشناسه	:	زمانی بروجنی، فرساد، ۱۳۵۳-
عنوان و نام پدیدآور	:	مقدمه ای بر کلودرا/گردآوری و تالیف فرساد زمانی بروجنی، عاطفه کریمی، پروین احمدی دول امیری.
مشخصات نشر	:	ایلام: دانشگاه آزاد اسلامی، واحد ایلام، ۱۳۹۸.
مشخصات ظاهری	:	۱۲۵ ص: مصور، جدول، نمودار.
شابک	:	۹۷۸-۹۶۴-۱۰۵۸۱۳-۷
وضعیت فهرست نویسی	:	فیبا
یادداشت	:	واژه نامه.
موضوع	:	آپاچی هادوپ
موضوع	:	Apache Hadoop
موضوع	:	داده پردازی—پردازش توزیع شده
موضوع	:	Electronic data processing—Distributed processing
موضوع	:	نرم افزار متن باز
موضوع	:	Open source software
موضوع	:	اس-کی-ال. (زبان برنامه نویسی کامپیوتر)
موضوع	:	SQL (Computer program language)
موضوع	:	سازماندهی فایل ها کامپیوتری
موضوع	:	File organization (Computer science)
موضوع	:	داده های کلان
موضوع	:	Big data
شناسه افزوده	:	کریمی، عاطفه، ۱۳۶۲-
شناسه افزوده	:	احمدی، پروین، ۱۳۶۲-
شناسه افزوده	:	Ilam Branch, Islamic Azad University
رده بندی کنگره	:	QA۷۶/۹
رده بندی دیویی	:	۰۰۵/۷۴
شماره کتابشناسی ملی	:	۵۹۶۱۴۳۶

نام کتاب: مقدمه ای بر کلودرا

نویسندگان: فرساد زمانی، عاطفه کریمی، پروین احمدی دول امیری

ناشر: انتشارات دانشگاه آزاد اسلامی واحد ایلام (تلفکس: ۲۲۲۷۵۳۱- بلوار دانشجو- کدپستی: ۶۹۳۱۱-۳۳۱۴۵)

لیتوگرافی، چاپ و صحافی: چاپخانه دانشگاه آزاد اسلامی

چاپ اول: نیمه دوم سال ۱۳۹۹

شمارگان: ۱۰۰۰

شابک: ۹۷۸-۹۶۴-۱۰۵۸۱۳-۷

قیمت: ۲۸۰۰۰۰ ریال

پیشگفتار

کلودرا ، شرکتی است که در سال ۲۰۰۸ تأسیس شد و نرم افزار های مبتنی بر APACHE HADOOP را توسعه داد. مجموعه ای برای راحتی کاربران و در واقع یک مرکز اطلاعاتی سازمانی که نیاز انبار داده را کاهش داده و به ذخیره سازی داده ها پرداخته است. کلودرا مجوز منبع باز داشته و نیز بدون هزینه می توان از پروژه های منبع باز آن استفاده نمود. استراتژی رشد تجاری از کلودرا به دلیل اهداف تجاری بسیار متفاوت است. همچنین مسیر نرم افزار سنتی را دنبال میکند و توزیع های سفارشی خود را از اکو سیستم که از صنعت استفاده می شود، ارائه می دهد. اکثر سازمانها به دلیل امنیت سختگیرانه داده ها و توزیع از کلودرا بهره می گیرد. توسط کلودرا می توان هادوپ را که بسته ی نرم افزاری بی نظیری است نصب، راه اندازی و مدیریت کرد. هادوپ به تهیه ی نسخه ی پشتیبان از اطلاعات به طور خودکار پرداخته که هر تکه از اطلاعات روی چندین سرور ذخیره شده و در صورت آسیب دیدن یکی از سرور ها ، سرور دیگر میتواند اطلاعات را جایگزین کند. در این کتاب به ساختار جستجوی کلودرا ، تجزیه و تحلیل ، مدیریت پیکر بندی پرداخته شده است. چالش های مدیریت داده مورد بررسی قرار گرفته و همچنین پشتیبانی و ناوبری کلودرا مورد بحث قرار گرفته شده است. و در پایان خداوند منان را شاکریم که به ما توفیق عنایت کرد که خدمتگزارانی در راه علم برای مردمان سرزمینم باشیم و از تمام کسانی که در تهیه ی این کتاب از هیچ کوشش و تلاشی دریغ ننموده اند کمال تشکر را داریم.

فرساد زمانی بروجنی

عاطفه کریمی

پروین احمدی دول امیری

تابستان ۹۸

فهرست

۵.....	درباره‌ی مقدمه‌ای بر کلودرا
۶.....	مروری بر مستندات
۹.....	مقدمه‌ی CDH
۱۰.....	مرور Apache Hive در CDH
۱۰.....	موارد استفاده‌ی Hive
۱۱.....	اجزای Hive
۱۴.....	مروری بر APACHE IMPALA
۱۴.....	مزایای impala
۱۵.....	Impala چگونه کار می‌کند
۱۶.....	خصوصیات Impalaی اولیه
۱۷.....	مروری بر جستجوی کلودرا
۱۹.....	جستجوی کلودرا چگونه کار می‌کند
۲۱.....	شناخت جستجوی کلودرا
۲۲.....	جستجوی کلودرا و دیگر اجزای کلودرا
۲۶.....	سبک معماری (ساختار) جستجوی کلودرا
۲۸.....	فایل‌های بیکربندی جستجوی کلودرا
۳۰.....	فرآیندها و کارهای جستجوی کلودرا
۳۱.....	نمایه‌سازی نزدیک به زمان واقعی با استفاده از Flume
۳۳.....	نمایه‌سازی NRT با استفاده از API
۳۳.....	مروری بر APACHE SENTRY
۳۴.....	مروری بر APACHE SPARK
۳۵.....	فرمت‌های فایلی و فشرده‌سازی
۳۵.....	استفاده از فایل‌های داده‌های Apache Parquet با CDH
۳۶.....	فشرده‌سازی فایل‌های Parquet
۴۹.....	استفاده از فایل‌های Apache Avro با CDH
۵۰.....	فشرده‌سازی فایل‌های داده‌ی AVRO
۵۴.....	فشرده‌سازی داده
۵۵.....	انواع فشرده‌سازی
۵۷.....	اسناد خارجی

۵۸		مروری بر مدیر کلودرا ۵
۵۹		واژه‌شناسی
۶۳		ساختار
۶۵		مدیریت وضعیت
۶۶		مدیریت پیکربندی
۶۸		پیکربندی سرور و مشتری
۷۱		مدیریت پردازش
۷۲		مدیریت توزیع نرم‌افزاری
۷۴		مدیریت میزبان (هاست)
۷۵		مدیریت منبع
۷۷		مدیریت کاربر
۷۹		سرویس مدیریت کلودرا
۸۲		کنسول ادمین مدیر کلودرا
۸۶		تنظیمات و ورود به کنسول ادمین
۹۰		آغاز و ورودی به کنسول ادمین مدیر کلودرا
۹۱		نمایش اسناد مدیر کلودرا
۹۱		لاگ اوت یا خروج اتوماتیک
۹۲		API مدیر کلودرا
۹۳		بازایی خدمات و خصوصیات میزبان
۹۴		پشتیبان‌گیری و بازایی پیکربندی مدیر کلودرا
۹۵		بازایی پیکربندی مدیر کلودرا
۹۵		استفاده از API مدیر کلودرا برای اتوماتیک‌سازی خوشه
۹۸		توسعه‌ی مدیر کلودرا
۹۹		مروری بر مدیریت داده‌ی ناوبری کلودرا
۹۹		چالش‌های مدیریت داده
۱۰۰		قابلیت‌های مدیریت داده‌ی ناوبری کلودرا
۱۰۱		تجزیه و تحلیل
۱۰۱		حسابرسی
۱۰۱		متادیتا
۱۰۲		سیاست‌ها
۱۰۲		جستجو کردن