

پایگاه داده‌های نسل آینده

هر آنچه که اهل فن در جهان انبوه داده،
می‌خواهند درباره‌ی پایگاه‌ها بدانند

ترجمه و تألیف:

عمر رضا طباطبایی تبریزی

دانشجوی دکتری تخصصی انفورماتیک پزشکی

تهران، ایران

کارشناس فناوری اطلاعات سلامت



**NAGHOOS
PUBLICATION**

سرسناسه	طباطبایی تبریزی، علیرضا، ۱۳۶۱-
عنوان و نام پدیدآور	پایگاه داده های نسل آینده: هر آنچه که اهل فن در جهان انبوه داده می خواهند درباره ی پایگاه داده ها بدانند/ ترجمه و تالیف علیرضا طباطبایی تبریزی، غزال پارسا.
مشخصات نشر	تهران: ناقوس، ۱۳۹۷.
مشخصات ظاهری	۳۸۴ص: مصور، جدول، نمودار
شابک	978-600-473-126-3
وضعیت فهرست نویسی	فیفا
یادداشت	بخشی از کتاب حاضر از منابع مختلف گردآوری و ترجمه شده است
موضوع	پایگاه های اطلاعاتی
موضوع	Databases :
موضوع	پایگاه های اطلاعاتی رابطه ای
موضوع	Relational databases :
زبان	پارسی، غزال، ۱۳۶۹-
راستی کنگری	QAVP/9/ ط ۲ ۱۳۹۷
رده بندی دیسی	۰۰۵/۷
شماره کتابسناسی	۵۲۳۲۶۳۶



**NAGHOOS
PUBLICATION**

این کتاب را به آدرس زیر مراجعه کنید:

www.naghoospress.ir

انتشارات ناقوس

نام کتاب	پایگاه داده های نسل آینده: هر آنچه که اهل فن در جهان انبوه داده می خواهند درباره ی پایگاه داده ها بدانند
ناشر	انتشارات ناقوس
ترجمه و تالیف	علیرضا طباطبایی تبریزی، غزال پارسا
چاپ اول	۱۳۹۷
تیراژ	۵۰۰ جلد
چاپ و صحافی	نارنجستان
سرپرست صفحه آرا	ثریا صفرزاده
قیمت	۳۶۰۰۰۰ ریال
شابک	۹۷۸-۹۶۴-۳۷۷-۱۲۶-۳
ISBN	978-964-377-126-3

کلیه حقوق برای سرمایه گذار محفوظ است. تکثیر و استفاده از این اثر به صورت حرفه ای یا چاپ مجدد، چاپ افست، پلی کپی، فتوکپی و انواع دیگر چاپ ممنوع است و پیگرد قانونی دارد.

مرکز پخش: انتشارات ناقوس

۱- بلوار کشاورز-خیابان آذر-کوچه پارس-پلاک ۱۰

تلفاکس: ۶۶۴۷۸۹۵۷-۶۶۴۷۸۹۵۸

۲- کتابفروشی الیاس: خیابان انقلاب، نبش ۱۲ فروردین تلفن: ۶۶۴۰۵۰۸۴

مولانا و مگل هر دو معتقدند که در پس همه پدیده های این عالم، وجودی است کامل، آگاه و نامتناهی. هر دو متفکر آن امر نامتناهی را خدا می نامند.

به نام خدا

پیش گفتار

از یک طرف دوران انفجار اطلاعات را گذراندیم و می گذرانیم، بطوریکه در پی رشد نمایی اطلاعات و رخ داد انبوه داده ها (Big Data)، دانایی درباره ی زیرساخت های لازم برای مقابله با تهدیدات و تبدیل نمودن به فرصت ها، امری اجتناب ناپذیر می باشد. از طرف دیگر پیشرفت موفقیت آمیز پایگاه داده رابطه ای در طی بیش از ۳۰ سال ساعه تجاری، گویای پیروزی علوم کامپیوتر و مهندسی نرم افزار برای مدیریت داده است. چنانچه در دنیای پر از تحول فناوری اطلاعات به ندرت یک مفهوم نظری به چنین موفقیت ادامه داری دست یافته و به طور گسترده پیاده سازی می شود. اما پایگاه داده رابطه ای به عنوان یک مفهوم نظری توانست این موفقیت را ثابت نماید. در این کتاب برآن هستیم تا تحولات اخیر، ابزارهایی مانند: هادوپ (Hadoop)، کاساندر (Cassandra)، پایگاه داده های «بدون زبان پرسش» و شاید «نه فقط زبان پرسش» (NoSql) و همچنین بشیر آینده را برای عرصه ی پایگاه داده در متنی ساده در اختیار خوانندگان و پژوهشگران حوزه ی فناوری، شبکه و علوم وابسته قرار دهیم. در این کتاب انواع راه حل ها برای موقعیت های مختلفی که اهل فن با آنها روبرو می شوند، بیان گردیده است.

شایسته است عنوان داریم که برگردان رگان خاصی قدیمی تر در این کتاب، با بهره گیری از کتاب مفاهیم بنیادی پایگاه داده ها، نوشته ی استاد ارشد دکتر روحانی رانکوهی صورت پذیرفت و برگردان واژگان جدید با نظرسنجی و جستجوی جامع رقم خورد. همچنین باتوجه به اینکه عناوین بسیاری از محصولات تجاری در حوزه ی نرم افزار هم نام شرکت و یا گروه توسعه دهنده می باشد، لذا در متن کتاب حاضر، عناوین محصولات به صورت لاتین و «وین» شرکت های توسعه دهنده با الفبای فارسی ارائه گردیده است.

از آنجایی که مفاهیم ارائه شده در این کتاب از مباحث نوین در حوزه ی پایگاه داده می باشد، صمیمانه خواهشمندیم انتقاد و پیشنهاد خود را با ایمیل A.Tabatabaei@sbmu.ac.ir در میان بگذارید. که البته برچشم خواهیم نهاد...

شایان ذکر است، کلیه عواید فروش این کتاب به جامعه حمایت از بیماران «ام اس» ایران تحقق خواهد یافت.

با احترام

علیرضا طباطبائی تبریزی

غزال پارسا

فهرست مطالب

فصل اول: سه دوره انقلاب پایگاه داده

۴	سیستم‌های پایگاه داده‌ای اولیه
۸	دومین انقلاب پایگاه داده
۸	تئوری رابطه‌ای
۱۰	مدل‌های تراک‌شی
۱۱	اولین پادشاه داده‌های رابطه‌ای
۱۲	جنگ پایگاه داده
۱۳	محاسبات خدمت داده و خدمت گیرنده (کلاینت-سرور)
۱۴	برنامه‌نویسی شریک OODB MS
۱۷	دوران بدون تغییر ماهیت پایگاه داده‌های رابطه‌ای
۱۸	سومین تحول پایگاه داده
۱۸	گوگل و هادوپ
۱۹	سایر وب سایت‌ها در www
۲۰	شبکه ابری خدمات کامپیوتری (ابر)
۲۱	پایگاه داده‌های سندگرا
۲۱	NewSQL
۲۲	انفجار پایگاه داده‌های غیر رابطه‌ای
۲۴	نتیجه گیری: یک محصول ثابت مناسب تمام نیازها نیست
۲۶	ملاحظات برای بررسی بیشتر

فصل دوم: گوگل، انبوه داده (کلان داده) و هادوپ

۲۹	تحول انبوه داده
۳۰	شبکه ابری، تکنولوژی موبایل، شبکه اجتماعی و کلان داده
۳۲	گوگل: پیشگام در کلان داده
۳۲	سخت‌افزار گوگل
۳۴	پشته‌ی نرم‌افزاری گوگل
۳۶	شرحی از MapReduce

۳۸	هadoop: پشته‌ی متن باز گوگل
۳۹	منشاء هadoop
۳۹	قدرت هadoop
۴۰	معماری هadoop
۴۴	HBase
۴۸	Pig
۴۹	اکوسیستم هadoop
۴۹	نتیجه‌گیری
۵۰	ملاحظات برای بررسی بیشتر

فصل سوم: بخش بندی، آمازون و به دنیا آمدن NoSQL

۵۳	پوست اندازی وب به نسبت دوم (وب ۲)
۵۴	چگونه وب ۲.۰ پیروز شد
۵۵	راه حل متن باز
۵۸	فنا شدن بوسیله یک هزار بخش بندی
۵۹	تئوری CAP
۶۱	نهایت سازگاری
۶۱	Amazon's Dynamo
۶۵	درهم سازی سازگار
۶۹	Dynamo و خانواده‌ی ذخیره سازی کلید-مقدار
۶۹	نتیجه‌گیری
۷۰	ملاحظات برای بررسی بیشتر

فصل چهارم: پایگاه داده‌های سندگرا

۷۴	XML و پایگاه داده‌های XML
۷۴	ابزارها و استانداردهای XML
۷۵	پایگاه داده‌های XML
۷۷	حمایت از XML در سیستم‌های رابطه‌ای
۷۷	پایگاه داده‌های استناد JSON
۷۸	JSON و AJAX

۷۹	پایگاه داده‌های JSON
۸۱	مدل‌های داده‌ای در پایگاه داده‌های سندگرا
۸۲	اولین پایگاه داده‌های JSON
۸۳	CouchBase و MemBase
۸۴	MongoDB
۸۶	JOSN, JSON و همه جا JSON
۸۶	نتیجه‌گیری

فصل پنجم: پایگاه داده‌های مبتنی بر گراف

۹۲	گراف چیست
۹۳	مدل‌های DBMS برای گرافها
۹۶	RDF و SPARQL
۹۷	گراف‌های ویژگی و Neo4j
۱۰۰	Gremlin
۱۰۱	نخیره سازی پایگاه داده گراف محور
۱۰۲	موتورهای محاسباتی گراف
۱۰۳	نتیجه‌گیری

فصل ششم: پایگاه داده‌های ستونی

۱۰۷	طرح‌های پایگاه داده‌های تحلیلی
۱۰۹	قالب ستونی به عنوان جایگزین
۱۱۱	فشرده‌سازی در معماری ستونی
۱۱۲	جریمه و هزینه عملیات نوشتن در معماری ستونی
۱۱۳	Sybase IQ, C-Store, and Vertica
۱۱۴	معماری پایگاه داده‌های ستونی
۱۱۶	پرتوها
۱۱۷	فن‌آوری ستونی در دیگر پایگاه داده‌ها
۱۱۸	نتیجه‌گیری
۱۱۹	ملاحظات برای بررسی بیشتر

فصل هفتم: پایان دیسک؟

۱۲۳	SSD و پایگاه داده‌های مقیم در حافظه
۱۲۳	پایان دیسک؟
۱۲۴	دیسک‌های حالت جامد
۱۲۶	نگاهی اقتصادی به دیسک
۱۲۷	پایگاه داده‌ها، با قابلیت بهره‌گیری از SSD
۱۲۸	پایگاه داده‌های مقیم در حافظه
۱۳۰	Times Ten
۱۳۱	Redis
۱۳۴	SAPHANA
۱۳۶	VoltDB
۱۳۹	Oracle 12c «پایگاه داده، مقیم در حافظه»
۱۴۰	پشته داده‌ی تحلیلی بر کلی Spark و
۱۴۱	BDAS شامل چند مولفه‌ی اصلی است
۱۴۲	معماری Spark
۱۴۴	نتیجه‌گیری
۱۴۴	ملاحظات برای بررسی بیشتر

فصل هشتم: الگوهای پایگاه داده توزیع شده

۱۴۷	پایگاه داده رابطه‌ای توزیع شده
۱۴۸	محاوره جهت یکسان‌سازی
۱۵۰	عدم اشتراک‌گذاری و دیسک به اشتراک گذاشته شده
۱۵۳	پایگاه داده توزیع شده غیر رابطه‌ای
۱۵۴	بخش‌بندی و محاوره جهت یکسان‌سازی در MongoDB
۱۵۵	بخش‌بندی
۱۵۶	مکانیزم‌های بخش‌بندی
۱۵۸	متعادل نمودن خوشه
۱۵۸	محاوره جهت یکسان‌سازی
۱۶۰	ملاحظات نوشتن و ترجیحات خواندن

۱۶۰	HBase
۱۶۱	جداول، مناطق و سرورهای منطقه‌ای
۱۶۳	قراردادن در نهانگاه و محل قرارگیری داده
۱۶۶	تقسیم سرور منطقه‌ای، برقراری تعادل و شکست
۱۶۶	نسخه‌های المثنی منطقه‌ای
۱۶۷	Cassandra
۱۶۸	درهم سازی سازگار
۱۷۲	حفظ تیب تقسیم بندی
۱۷۲	السی ها
۱۷۴	خبررسان ها
۱۷۴	خلاصه

فصل نهم: مدل‌های سازگاری

۱۷۷	انواع سازگاری
۱۷۸	MVCC و AICID
۱۷۸	شماره‌های توالی تراکنش جهانی
۱۸۱	تثبیت دو فازه
۱۸۲	سایر سطوح سازگاری
۱۸۳	سازگاری در MongoDB
۱۸۴	مجموعه‌ای از دربرگیرندگان نسخه المثنی و سازگاری نهایی (شماره ۱)
۱۸۴	سازگاری HBase
۱۸۷	فاکتور محاوره جهت یکسان سازی
۱۸۷	سازگاری در نوشتن
۱۸۷	سازگاری در خواندن
۱۸۹	تعامل بین سطوح سازگاری
۱۹۰	دست به دست کردن پنهانی و اصلاح خواندن
۱۹۱	مهر زمانی و دانه‌ای بودن
۱۹۲	ساعت‌های بردار
۱۹۴	تراکنش‌های سبک وزن
۱۹۸	نتیجه‌گیری

فصل دهم: ذخیره سازی و مدل های داده

۲۰۳	مدل های داده
۲۰۴	مروری بر مدل رابطه ای داده ها
۲۰۶	ذخیره سازی کلید-مقدار
۲۰۷	انواع داده محاوره شده جهت یکسان سازی، بصورت همگرا
۲۱۰	مدل های داده در HBase و BigTable
۲۱۰	ساختار خانوادگی ستون
۲۱۲	نسخه ها
۲۱۳	Cassandra
۲۱۶	مجموعه های Cassandra
۲۱۶	مدل داده های JSON
۲۱۷	JSON دوتایی (BJSON)
۲۱۸	ذخیره سازی
۲۱۸	مدل ذخیره سازی رابطه ای معمور
۲۲۱	درخت ادغام مبتنی بر ساختار ثبت (SM)
۲۲۳	SSTable و فیلترهای Bloom
۲۲۵	بروزرسانی ها و نشانگر مزار (سنگ مزار)
۲۲۵	تراکم
۲۲۵	شاخص گذاری ثانویه
۲۲۶	خودتان انجام دهید- شاخص گذاری ثانویه
۲۲۸	شاخص های جهانی و محلی
۲۲۹	پایگاه داده NoSQL
۲۳۰	نتیجه گیری

فصل یازدهم: زبان ها و رابط های برنامه نویسی

۲۳۶	SQL
۲۳۷	NoSQL APIs
۲۳۸	Riak
۲۴۱	Hbase

۲۴۴	MongDB
۲۴۶	زبان پرسش Cassandra (CQL)
۲۵۰	MapReduce
۲۵۲	Pig
۲۵۴	گراف جهت‌دار غیر مدور (DAG)
۲۵۵	آپشار
۲۵۵	Spark
۲۵۷	بازگشت SQL
۲۵۸	Hive
۲۵۹	Impala
۲۶۰	SQL مربوط به Spark
۲۶۰	NIQL مربوط به Couchbase
۲۶۴	Apache Drill
۲۶۶	سایر SQLها روی NoSQL
۲۶۷	نتیجه‌گیری
۲۶۷	ملاحظات برای بررسی بیشتر

فصل دوازدهم: پایگاه داده‌های آینده

۲۷۱	بازبینی انقلاب
۲۷۳	ضد انقلاب‌ها
۲۷۴	آیا دور باطل زده‌ایم؟
۲۷۵	فقدان انتخاب
۲۷۷	مدل‌های سازگاری
۲۷۸	شیما
۲۸۰	زبان‌های پایگاه داده
۲۸۲	نخیره سازی
۲۸۴	دیدگاهی برای پایگاه داده همگرا
۲۸۵	در این میان، بازگشت به اوراکل
۲۸۶	پشتیبانی JSON در اوراکل
۲۸۹	دسترسی JSON از طریق اوراکل REST

۲۹۱	دسترسی REST به جداول اوراکل
۲۹۳	گراف اوراکل
۲۹۳	بخش بندی اوراکل
۲۹۵	اوراکل به عنوان پایگاه داده ترکیبی
۲۹۶	سایر پایگاه های داده های همگرا
۲۹۷	فناوری های پایگاه داده تفاق افکن
۲۹۷	فناوری های ذخیره سازی
۲۹۹	زنچیر بلوک
۳۰۰	محاسبه کوانتوم
۳۰۲	نتیجه گیری
۳۰۴	ملاحظات برای بررسی بیشتر

فصل سیزدهم ویژگی های محصول SQLSERVER شرکت

مایکروسافت

۳۰۷	چند سکویی بودن SQL Server 2017
۳۰۹	ویژگی های جدید توسعه پایگاه داده در SQL Server 2017
۳۱۲	ویژگی های جدید تنظیم و کارایی در SQL Server 2017
۳۱۶	ویژگی جدید حوزه امنیت در SQL Server 2017
۳۱۶	ویژگی های جدید دسترسی پذیری بالا (HA) در SQL Server 2017
۳۱۸	ویژگی های جدید یادگیری ماشین در SQL Server 2017

ضمیمه ۱

۳۲۳	Aerospike
۳۲۴	Cassandra
۳۲۵	CouchBase
۳۲۵	DynamoDB
۳۲۶	HBase

۳۲۷	MarkLogic
۳۲۷	MongoDB
۳۲۸	Neo4j
۳۲۸	NuoDB
۳۲۹	Oracle RDBMS
۳۲۹	Redis
۳۳۰	Riak
۳۳۰	SAP HANA
۳۳۱	Times Ten
۳۳۱	Vertica
۳۳۲	VoltDB

ضمیمه ۲

۳۳۳	واژه‌نامه
-----	-----------

ضمیمه ۳

۳۶۱	مکاتبات با نویسنده مرجع اصلی
-----	------------------------------