

بررسی مقایسه ای پایگاه های داده

N o S Q L

توسط

افشین عباسپور
فاطمه سلطان محسنی

نام استاد

آقای دکتر شجری

نام تدریس یار

مهندس مرتضی سرگلزایی جوان

فهرست

۲ معرفی NoSQL
۳ انواع پایگاه داده NOSql
۴ پایگاه های key/value
۴ پایگاه های داده ای استنادی
۵ پایگاه های داده ای توزیع شده
۶ پایگاه داده گرافی
۸ بحث و نتیجه گیری
۹ منابع

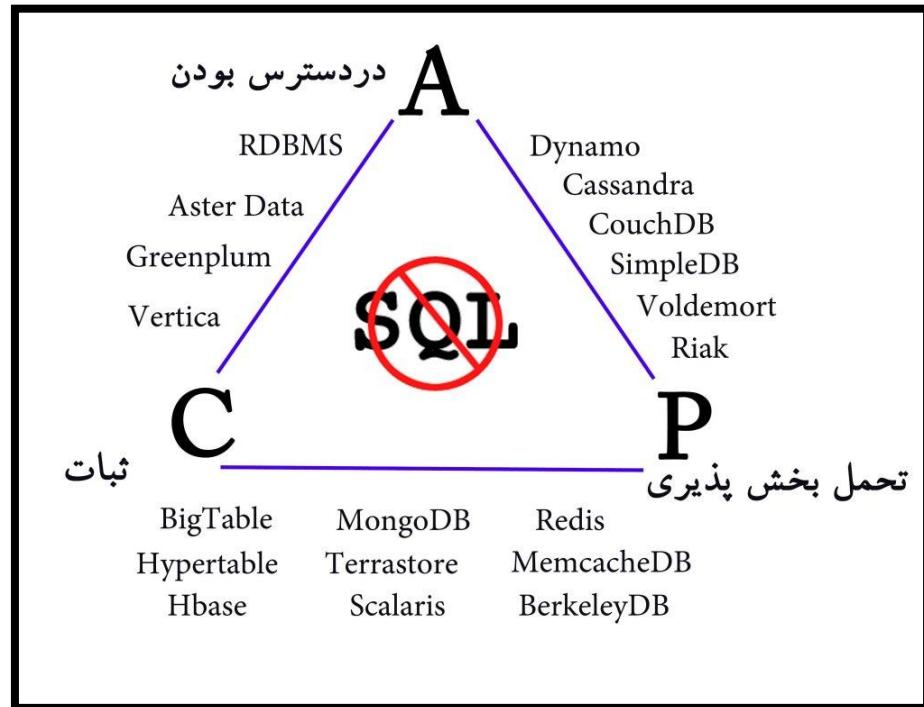
NoSQL معرفی

مقدار داده‌ای که تا آوریل ۲۰۱۱ توسط کتابخانه کنگره آمریکا جمع آوری شده است حدود ۲۳۵ تریابیت می‌باشد تنها در عرض چند سال حجم داده‌های دیجیتالی در دنیای مابه صورت نمایی رشد کرده است و این روند به طور روز افزون ادامه دارد BigData در آینده‌ای نه چندان دور به بحث سود آور و چالش پذیر در حوزه فناوری اطلاعات تبدیل خواهد شد.

در واقع به مجموعه داده‌هایی می‌گویند که هانداره آنها از آنچه ابزارهای ذخیره سازی معمول می‌توانند مدیریت کنند فراتر می‌رود و همزمان با پیشرفت فناوری رشد می‌کند و به روند رشد و زندگی خود ادامه می‌دهد.

نیازهای جدید ذخیره اطلاعات امروزی، پایگاه‌های داده‌ای با معاری جدیدی را طلب می‌کند مدل سنتی پایگاه داده‌های رابطه-ای با همه مزایا و معایب آن دیگر پاسخگوی داده‌های عظیم تجاری و برنامه‌های تحت وب با رشد روز افرونو تغییرات فراوان نمی‌باشد به همین دلیل پایگاه داده جدیدی با معماری کاملاً متفاوت با نام NoSql پا به عرضه وجود گذاشت و علاوه بر رفع برخی از نواقص مدل رابطه‌ای برای سیستم‌های پایگاه داده قابل گسترش و تزیع شده امکانات جدیدی را برای سیستم‌ها معرفی نمود که از جمله این امکانات دسترسی پذیری بالا high availability و بخش پذیری consistency و ثبات partitioning می‌باشد. مواردی که اهمیت و کاربرد NoSql را در پردازش ابری به خوبینمایان می‌سازد. بر اساس همین اهمیت و بخاطر OpenSource بودن ایده اولیه NoSql نمونه‌های بسیاری از پایگاه داده‌ای NoSql با اهداف و ساختارهای متفاوت نسبت به هم ایجاد گردید که از جمله آنها می‌توان به Hbase ، Redis ، BigTable ، Mongo DB ، CouchDB ، Cassandra ، vertica Simple DB و... اشاره کرد.

هر یک از این سیستم‌ها مزایای خاص و منحصر به فرد خود را دارا می‌باشند. شکل زیر دسته‌بندی کلی این پایگاه‌های داده‌ای را بر اساس سه اصل دسترسی پذیری، بخش پذیری و صحت نشان می‌دهد.



تعدد ابزارها و سیستمها ی مبتنی بر NOSql انتخاب یک روش مناسب برای کاربر خاص رامشکل می کند در این مقاله بعد از توضیح کلی در مورد توانایی های NOSql و معرفی انواع پایگاه داده های مبتنی بر آن به مقایسه برخی از این سیستمها می پردازیم.

انواع پایگاه داده NOSql

همانطور که گفته شد معماری سیستمها ی مبتنی بر NOSql بخاطر تفاوت در نوع عملکرد و کاربرد آنها متفاوت است. انواع پایگاه داده NOSql از نظر معماری به چند دسته مهم تقسیم می شود.

۱. پایگاه های key/value
۲. پایگاه های داده ای اسنادی
۳. پایگاه های داده ای توزیع شده
۴. پایگاه های داده ای گرافی

در ادامه به معرفی اجمالی هر یک از این انواع می پردازیم

پایگاه های key/value

این سرویس در واقع یک سرویس بازیابی اطلاعات مشابه HashTable می باشد که راهی بهینه برای ذخیره داده های ساده اما در مقیاسی بسیار وسیع می باشد . در این روش داده ها تنها با کلیدهای اولیه و برونق ارتباطات پیچیده نگه داری می شوند.

تاخیر کم ، توزیع پذیری و جایگزین پذیری بالا از مزایای این نوع پایگاه داده ای NoSql می باشد . هر چند که در این نمونه پایگاه Shema ، NoSql برای محتوى ذخیره شده وجود ندارد.

این می تواند به عنوان یک نقطه ضعف برای آن مطرح شود اما همین شکل ساده و استفاده از ساختمان HashTable در سرعت مراجعه به اطلاعات نقش بسزایی دارد و در کاربرهایی که جستجو اطلاعات شده در اولویت اهمیت می باشد پایگاه های داده ای Cashe میتوانند بهتر و مفید تر عمل کنند . از نمونه های پایگاه داده NoSql که با این روش Redis ، Voldemort ، Oracle ...



پایگاه های داده ای اسنادی

این نوع پایگاه داده در اصل نوعی از پایگاه داده key/value مقداری است که داده ها و مقادیر در قالب اسناد ذخیره می شوند که هر سند در این روش حاوی مقادیر داده ای دلخواه به همراه MetaData های آن سند می باشد . در این روش نوع داده ای مقادیر اسناد می تواند متفاوت و متغیر باشد و هیچ محدودیت خاصی در اندازه متن یا تعداد المانهای آن وجود ندارد. هر تغییری که روی اسناد صورت گیرد می تواند در بانک اطلاعاتی ذخیره شود و یا تغییر کند.

نمونه ای از این نوع پایگاه داده ای DB Mongo می باشد که داده ها را با فرمت Json که برای برنامه نویسان جawa اسکریپت آشنا است به عنوان سند ذخیره خواهد کرد. علاوه بر این در Mongo DB پرس و

جوها به شکل توابع معمولی جاوا اسکریپت است. به همین دلیل این پایگاه داده ای مورد توجه بسیاری از تولید کنندگان وب واقع شده است

اما بهترین نمونه و نماینده پایگاه داده استنادی CouchDB می باشد که کتابخانه های مورد نیاز برای همه زبانها ای برنامه نویسی مهم را ارائه داده و مورد توجه بسیاری از توسعه دهندها قرار گرفته است. ضمن اینکه شرکت Cloudant سرویس میزبانی این نوع پایگاه داده ای را در ابر ارائه می دهد و استفاده از این پایگاه داده به صورت سرویس روز به روز ساده تر و کم هزینه تر شده است. از نمونه های دیگر این نوع پایگاه داده ((استنادی)) می توان به Redis اشاره کرد که برای کاربردهایی که به سرعت تغییر می کنند مناسب است.

پایگاه های داده ای توزیع شده

برای مدیریت میلیونها بروز رسانی داده ای از افراد مختلف در یک شبکه اجتماعی بزرگ با سریعترین روش و ارزانترین راهکار باید با ایده های پایگاه های توزیع شده NoSQL آشنا شویم . این موضوع وقتی از درجه اهمیت بالایی برخوردار می گردد که قرار باشد حجم عظیمی از داده ها روی شبکه بزرگی با سرورهای متصل به هم قرار گیرد.

در این روش هر ماشین باید در شبکه p2p داخل خود عملکرد با ثباتی داشته باشد و هر ماشین دقیقا حاوی داده های یکسانی باشد.

Cassandra نمونه ابتدایی این نوع پایگاه داده ای می باشد که عملکرد نسبتاً پیچیده تری نسبت به نمونه های دیگر پایگاه داده ای NoSQL دارد. در این پایگاه داده می توان سرعت همگام سازی ستونها و مجموعه هایی از مقادیر را تا یین کنیم و با توجه به شرایط بستر شبکه ای موجود برای هماهنگی و مدیریت کلاسترها ای این پایگاه داده تنظیمات مختلفی را اعمال کنیم که تنظیمات این پایگاه داده ای نیز نسبت به انواع دیگر پایگاه داده ای NoSQL برای توسعه دهندها جهت نصب و راه اندازی Cassandra نسبتاً پیچیده تر می باشد .

Cassandra تحت پلتفرم جاوا توسعه داده شده و با سیستمهای تحت جاوا بهتر کار می کند. از نمونه های دیگر پایگاه داده توزیع شده Riak می باشد . این نمونه پایگاه داده ای NoSQL مشابه Cassandra می باشد با این تفاوت که از پیچیدگی ها و سردرگمی های کاربران آن به شدت کاسته است و برای مواردی که دسترسی پذیری و مقیاس پذیری در محیط توزیع شده مورد نیاز می باشد گزینه خوبی برای ذخیره حجم زیادی از داده ها می باشد این نوع پایگاه داده ای در برابر خطای مقاوم است و قابلیت جستجوی FullText ایندکس کردن، پرس و جو با استفاده از Riak Search Server و... را دارا می باشد.

پایگاه داده گرافی

این پایگاه داده امکان پر کردن انبار داده ای با نودهای مختلف و برقراری ارتباط میان آن ها را برای ایجاد معانی و مفاهیم پیچیده تر و یا اشیاء ایجاد می کند. که پایه این نوع پایگاه داده ای با مجموعه ای از الگوریتمها ی محاسباتی مرتبط با گراف عرضه شده است که برای مثال امکان تعیین کوتاهترین مسیر را فراهم می سازد. از نمونه های بارز این نوع پایگاه داده NoSQL میتوان Neo4j را نام برد که مهمترین نوع پایگاه داده ای گرافی می باشد. هرچند که توسعه این پایگاه داده ای در ابتدای راه قرار دارد و یک نمونه بسیار جوان از این دسته است ولی هر روز قابلیت های بیشتری به آن اضافه می گردد که افزایش سرعت کارکرد با استفاده از روشهای Caching و استفاده از یک زبان پرس و جوی جدید شبیه xsI از جمله این قابلیت های اضافه شده جدید به این پایگاه داده می باشد.

در ادامه به بررسی مقایسه ای برخی از این پایگاههای داده ای NoSQL خواهیم پرداخت

نام DataBase	زبان برنامه نویسی	پروتوكول	بهترین کاربرد	توضیحات
MongoDB	C++	اختصاصی (BSON)	کاربرهایی که به سرعت عملکرد بالا برای داده های عظیم احتیاج دارند و داده ها بشدت تغییر میکنند	محدودیت تا حجم ۲/۵ گیگابایت برای سیستمهای ۳۲ بیتی و اشغال فضای حدود ۱۹۲ مگابایت از حافظه می باشد حفظ برخی از خصوصیات Sql مانند Query و Index از شاخصه های آن می باشد
Redis	C++	اختصاصی (Telnet)	برای داده هایی که به سرعت تغییر می کنند و حجم مشخصی دارند که در حافظه جای می گیرند	کارکرد بسیار سریع از شاخصه های آن می باشد کارکرد اصلی در حافظه و اعمال نا همزمان بر روی دیسک

Cassandra	Java	Thrift		
کاربرهایی که نوشتن در آنها بیشتر از خواندن اتفاق می افتد و برنامه کاربردی مربوطه از جاوا استفاده می کند				
قابلیت ارائه بهترین راهکار مبتنی بر BigTable برای سیستمها توزیع شده پیچیدگی و ایجاد سر درگمی بسیار برای نصب و پیکربندی ثبات پایگاه داده و سادگی در استفاده طراحی قابل اعتماد، صحت سنجی سمت Server برای اسناد، قابلیت های اعتبار سنجی از ویژگیهای آن می باشد	جمع آوریداده هایی که به ندرت تغییر می کنند و مناسب برای استفاده متمادی از Query های مشخص	Http/Rest	Erlong	CouchDB
قابلیت http با پشتیبانی از Xml Protobuf. و باینری به همراه درگاه Thrift با عملکرد بسیار بالا از مزایای آن می باشد	کاربردهایی که برای حجم عظیمی از داده ها نیاز به استفاده از مدل داده ای BigTable می باشد	HTTP/Rest Thrift	Java	HBase
سرعت بسیار بالا با استفاده از دسترسی داده با کلید ، رابط کاربری خوب برای مدیریت کلاسترها تحت وب و قابلیت محافظت در مقابل تکرار داده و ذخیره وقت داده ها در حافظه مانند MemCached و معادل بودن همه نودها در کلاستر از مزایای ویژه این دیتابیس است	کاربردهایی که نیازمند دسترسی داده با تاخیر بسیار کم و همگام سازی و دسترسی پذیری بالا هستند	پروتوكل همانند Memcached	C/Erlang	Membase

بحث و نتیجه گیری

موارد ذکر شده تنها چند نمونه از پایگاه داده NOSql راچ میباشد . نمونه های بسیار زیاد دیگری نیز وجود دارد که همه در مراحل اولیه توسعه خود بسر می برند . ضمن اینکه با توسعه بیشتر و کشف قابلیتها و فرصتهای جدید شرکتهای دیگری نیز همواره به این موج اضافه خواهند شد . در چنین شرایطی انتخاب پایگاه داده ای متناسب با نیاز های ما بسیار پیچیده به نظر میرسد . زیرا هر یک از این پایگاه های داده ای ویژگی خاص خود را دارا می باشند . سوال دیگر اینکه آیا با اینهمه تاکید بر روی SQL عمر پایگاه داده ای رابطه ای که سالها به عنوان بهترین راه حل معرفی شده به پایان رسیده و کاربردی ندارد ؟

در جواب باید به این نکته دقیق کنیم که میتوان از چند پایگاه داده بطور همزمان برای یک کاربرد خاص استفاده نمود . همچنان میتوان از پایگاه های داده ای SQL در کنار سیستمهای رابطه ای بطور همزمان استفاده کرد . مثلاً "در یک برنامه اتوماسیون تحت وب که از پایگاه داده رابطه ای استفاده میکند برای ذخیره LOG های حجمی که کارایی سیستم پایگاه داده ای را کم میکند از Cassandra و برای دسترسی سریع کاربرای به اطلاعات خاصی که باید با تأخیر کم در دسترس باشند از Membase و ... استفاده کرد و از مزایای هر یک به درستی بهره جست .

با این حال با وجود قابلیتهای فراوان برای SQL به نظر میرسد که عمر پایگاه داده سنتی رابطه ای به این زودی به پایان نمیرسد و همچنان میتوان از بسیاری از فواید این مدل بهره جست .

منابع

- [١] Nishtha Jatana , Assistant professor, Department of Computer Science and Engineering, “**A Survey and Comparison of Relational and Non-Relational Database**” , International Journal of Engineering Research & Technology (IJERT) , August - ٢٠١٢
- [٢] Rick Cattell , “**Scalable SQL and NoSQL Data Stores**” , IEEE Paper December ٢٠١٠.
- [٣] Rick Cattell , “**High Performance Scalable Data Stores**” IEEE Paper April ٢٧, ٢٠١٠.
- [٤] Chad DeLoatch and Scott Blindt , “**NoSQL Databases: Scalable Cloud and Enterprise Solutions**” , University of Illinois at Urbana Champaign
- [٥] NoSQL Conferences in WebSite Site : <http://nosql-database.org>
- [٦] <http://www.thoughtworks.com/articles/nosql-comparison>
- [٧] Shashank-Tiwari , “**Professional NoSQL**” Wrox Programmer to Programmer Ebook