

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

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

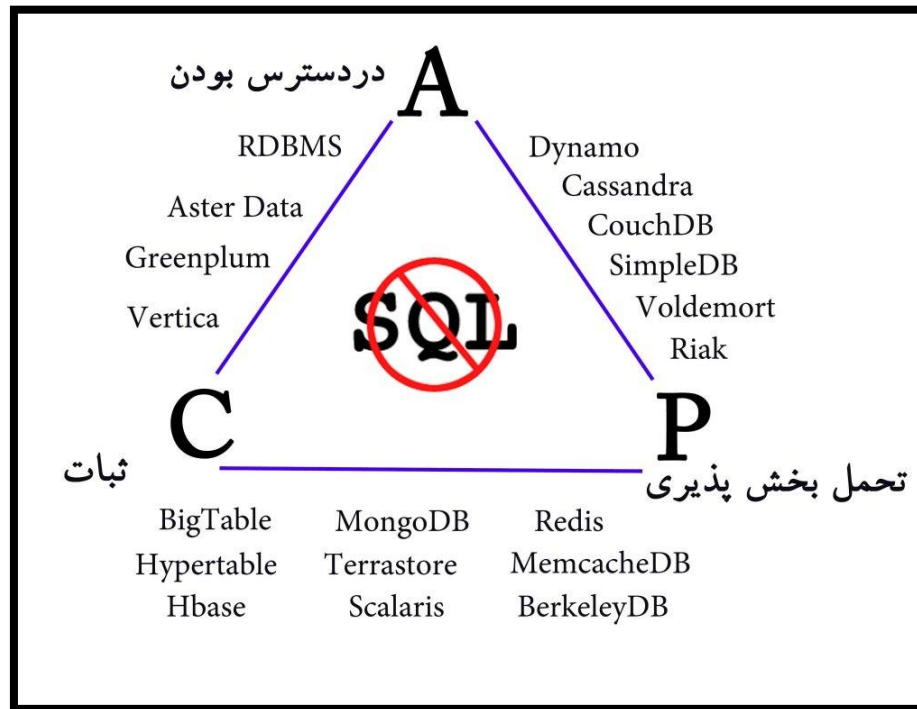
معرفی NoSQL

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

BigData در واقع به مجموعه داده هایی می گویند که اندازه آنها از آنچه ابزارهای ذخیره سازی معمول می توانند مدیریت کنند فراتر می رود و همزمان با پیشرفت فناوری رشد می کند و به روند رشد و زندگی خود ادامه می دهد.

نیازهای جدید ذخیره اطلاعات امروزی ، پایگاه های داده ای با معماری جدیدی را طلب می کند مدل سنتی پایگاه داده های رابطه -ای با همه مزایا و معایب آن دیگر پاسخگوی داده های عظیم تجاری و برنامه های تحت وب با رشد روز افزون تغییرات فراوان نمی باشد به همین دلیل پایگاه داده جدیدی با معماری کاملاً متفاوت با نام NoSql پا به عرصه وجود گذاشت و علاوه بر رفع برخی از نواقص مدل رابطه ای برای سیستمهای پایگاه داده قابل گسترش و توزیع شده امکانات جدیدی را برای سیستمها معرفی نمود که از جمله این امکانات دسترسی پذیری بالا high availability و بخش پذیری partitioning و ثبات consistency می باشد. مواردی که اهمیت و کاربرد NoSql را در پردازش ابری به خوبی نمایان می سازد . بر اساس همین اهمیت و بخاطر OpenSource بودن ایده اولیه NoSql نمونه های بسیاری از پایگاه داده ای NoSql با اهداف و ساختارهای متفاوت نسبت به هم ایجاد گردید که از جمله آنها می توان به Cassandra ، CouchDB ، Mongo DB ، BigTable ، Redis ، Hbase ، vertica ، Simple DB و... اشاره کرد .

هر یک از این سیستمها مزایای خاص و منحصر به فرد خود را دارا می باشند . شکل زیر دسته بندی کلی این پایگاه های داده ای را بر اساس سه اصل دسترسی پذیری ، بخش پذیری و صحت نشان می دهد.



تعدد ابزارها و سیستمهای مبتنی بر NoSql انتخاب یک روش مناسب برای کاربر خاص را مشکل 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 سیستمهای 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 گرافیکی

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 برون 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 محتوی 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 Cashe شده در اولویت اهمیت می باشد پایگاه های داده ای key/value می توانند بهتر و مفید تر عمل کنند. از نمونه های پایگاه داده NoSql که با این روش Oracle، Voldmort، Redis ...



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 Mongo DB می باشد که داده ها را با فرمت Json که برای برنامه نویسان جاوا 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 اسنادی CouchDB می باشد که کتابخانه های مورد نیاز برای همه زبانهای برنامه نویسی مهم را ارائه داده و مورد توجه بسیاری از توسعه دهندگان قرار گرفته است. ضمن اینکه شرکت Cloudbant 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 ((اسنادی)) 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 دقیقاً 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 بین کنیم و با توجه به شرایط بستر شبکه ای موجود برای هماهنگی و مدیریت کلاسترهای 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 دهندگان 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 خطا 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 الگوریتمهای محاسباتی مرتبط با گراف عرضه شده است که برای مثال امکان تعیین کوتاهترین مسیر را فراهم میسازد. 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 جو ی جدید شبیه xsl 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	پروتوکل	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 ((شبیه 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 پیکر بندی
CouchDB	Erlong	Http/Rest	جمع آوریداده هایی که به ندرت تغییر می کنند و مناسب برای استفاده متمادی از Query های مشخص	ثبات پایگاه داده و سادگی در استفاده طراحی قابل اعتماد ، صحت سنجی سمت Server برای اسناد، قابلیت های اعتبار سنجیاز ویژگیهای آن می باشد
HBase	Java	HTTP/Rest Thrift	کاربردهایی که برای حجم عظیمی از داده ها نیاز به استفاده از مدل داده ای BigTable می باشد	رابطه http با پشتیبانی از Xml، Protobuf، و باینری به همراه درگاه Thrift با عملکرد بسیار بالا از مزایای آن می باشد
Membase	C/Erlang	پروتوکل همانند Memcached	کاربردهایی که نیازمند دسترسی داده با تاخیر بسیار کم و همگام سازی و دسترسی پذیری بالا هستند	سرعت بسیار بالا با استفاده از دسترسی داده با کلید ، رابط کاربری خوب برای مدیریت کلاسترهای 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

بحث و نتیجه گیری

موارد ذکر شده تنها چند نمونه از پایگاه داده NoSql رایج 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 نیازهای 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 NoSQL 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 دقت کنیم که میتوان از چند پایگاه داده بطور همزمان برای یک کاربرد خاص استفاده نمود. همچنین میتوان از پایگاه های داده ای NoSQL 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 حجیم که کارایی سیستم پایگاه داده ای را کم میکند از Cassandra و برای دسترسی سریع کاربرای به اطلاعات خاصی که باید با تاخیر 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 NoSQL 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.

[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