

عنوان مقاله: گراف دیتابیس: مروری بر Neo4j, ArangoDB و OrientDB [بخش دوم]

نویسنده مقاله: داوود گراوند

تاریخ انتشار: شهریور ماه ۱۳۹۹

منبع: <https://nikamooz.com/graph-databasepart-۰۲>



برای شروع بخش دوم مقاله ابتدا چندین سایت رتبه بندی بررسی و آخرین رتبه ی پایگاه داده های گراف در سال ۲۰۲۰ مورد ارزیابی قرار گرفت که رتبه ها، روند رشد و معیارهای رتبه بندی استفاده شده توسط وبسایت های مورد بررسی در ادامه این بخش آمده است. همچنین نتیجه ی این بررسی نهایتا منجر به انتخاب سه دیتابیس ArangoDB، OrientDB و Neo4j جهت مقایسه ی دقیق تر گردید.

وب سایت DB-Engines

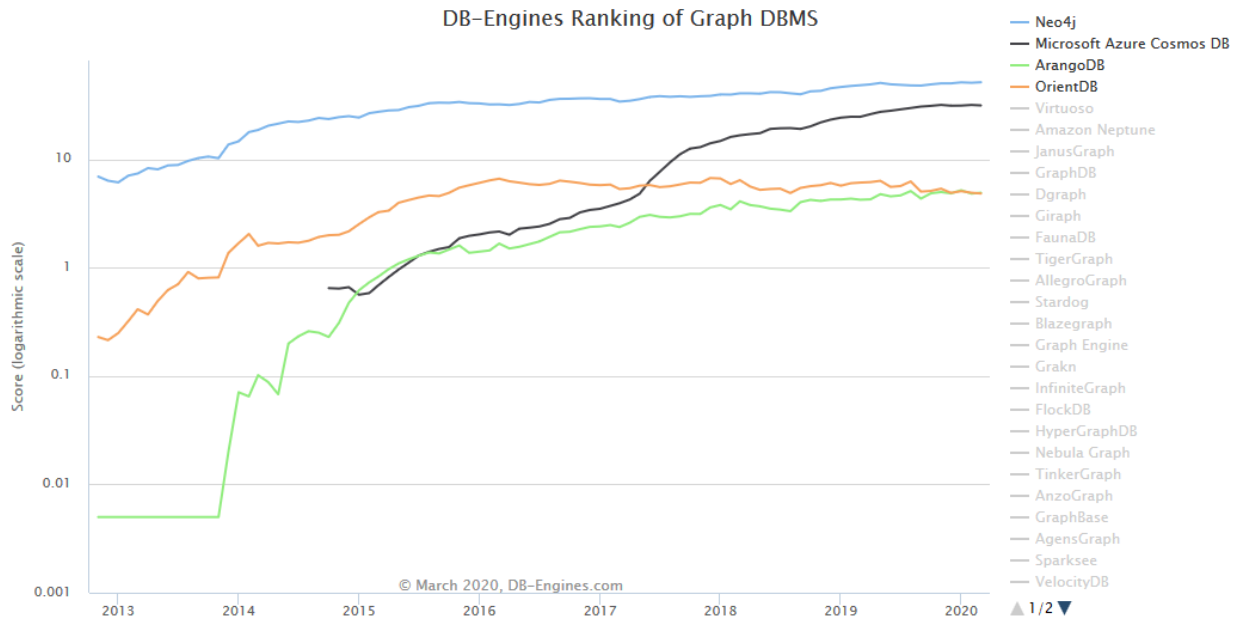
رتبه بندی

include secondary database models

33 systems in ranking, March 2020

Rank			DBMS	Database Model	Score		
Mar 2020	Feb 2020	Mar 2019			Mar 2020	Feb 2020	Mar 2019
1.	1.	1.	Neo4j +	Graph	51.78	+0.57	+3.20
2.	2.	2.	Microsoft Azure Cosmos DB +	Multi-model i	31.63	-0.32	+6.81
3.	↑ 4.	↑ 4.	ArangoDB +	Multi-model i	4.94	+0.09	+0.68
4.	↓ 3.	↓ 3.	OrientDB	Multi-model i	4.86	-0.08	-1.27
5.	5.	5.	Virtuoso +	Multi-model i	2.86	+0.09	-0.33
6.	6.	↑ 8.	Amazon Neptune	Multi-model i	1.82	-0.14	+0.79
7.	7.	↓ 6.	JanusGraph	Graph	1.79	-0.09	+0.47
8.	↑ 9.	↑ 9.	GraphDB +	Multi-model i	1.18	+0.04	+0.25
9.	↓ 8.	↑ 13.	Dgraph +	Graph	1.08	-0.07	+0.39
10.	↑ 11.	↓ 7.	Giraph	Graph	0.95	-0.03	-0.09

روند رشد

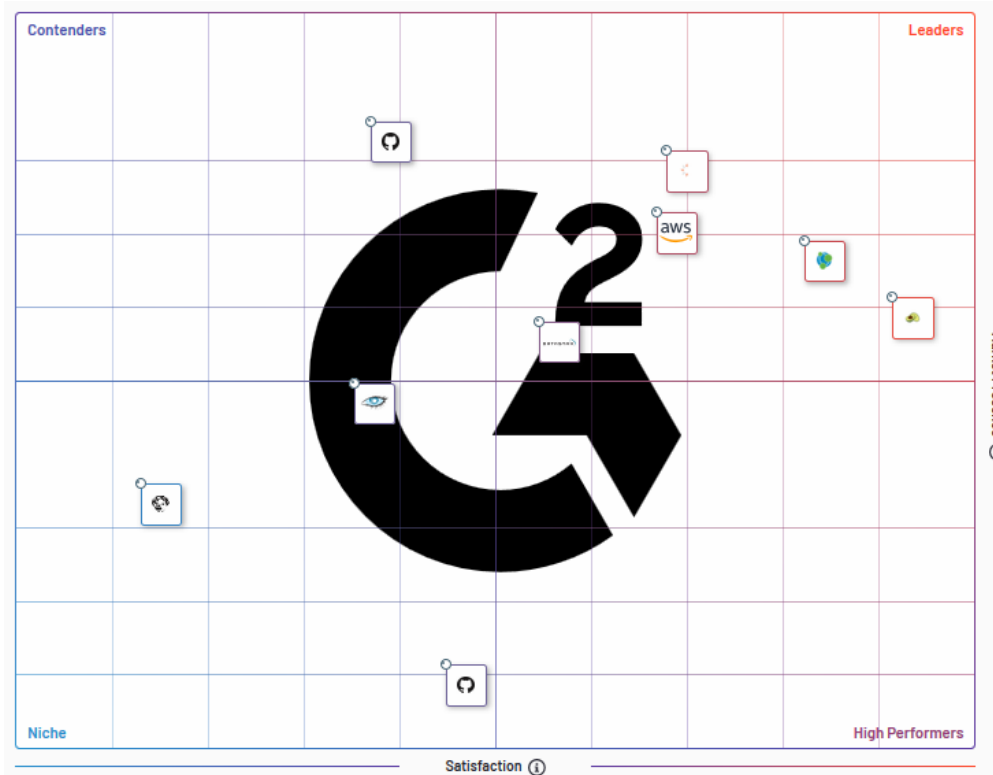


مبنای مقایسه برای رتبه بندی

- تعداد نتایج جستجو در (google,bing,yandex)
- تعدد تکرار در Google Trends
- تعداد سوالات و تعداد تایید کاربران در سایت‌های Stack Overflow و DBA Stack Exchange
- تعداد مشاغل ثبت شده در Indeed و Simply Hired
- تعداد متخصصین و کارشناس در LinkedIn و Upwork
- تعداد تکرارهای Twitter

وب سایت g2

رتبه بندی Best Graph Databases

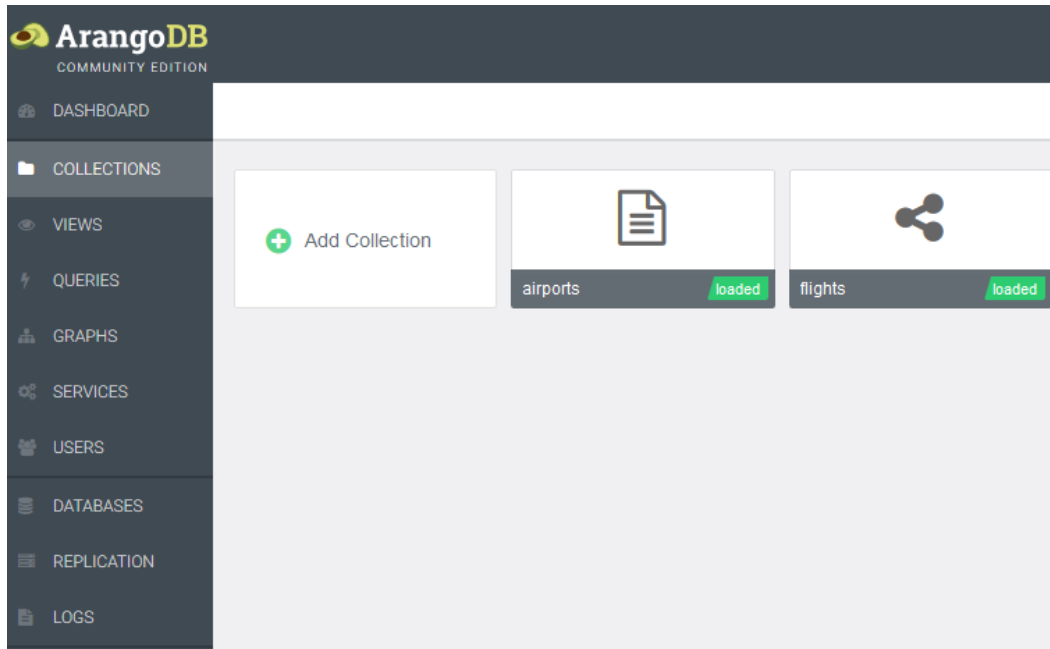


- ArangoDB
- Neo4j
- OrientDB
- Amazon Neptune
- FlockDB
- DataStax
- Cassandra

مبنای مقایسه

G2 محصولات و فروشندگان را بر اساس بررسی‌های جمع آوری شده از جامعه کاربرها ، و همچنین داده‌های جمع آوری شده از منابع آنلاین و شبکه‌های اجتماعی ، رتبه بندی می‌کند. G2 یک الگوریتم منحصر به فرد (۷۳.۰) برای روی داده‌ها اعمال می‌کند و نمرات رضایت مشتری و حضور در بازار را در زمان واقعی محاسبه می‌کند.

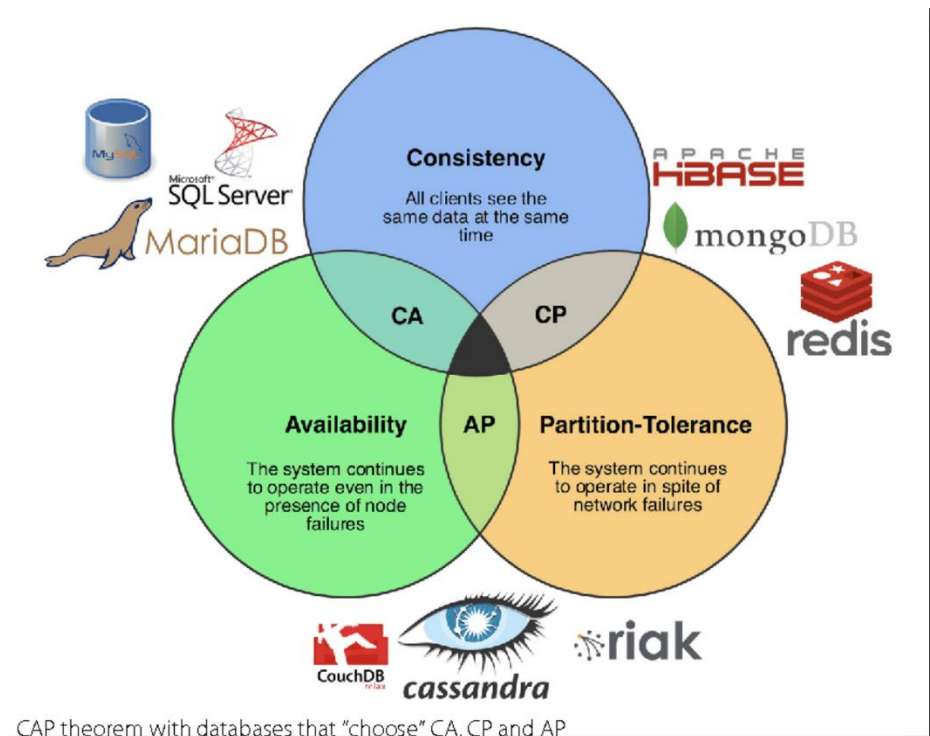
ArangoDB



ArangoDB یک سیستم پایگاه داده چند مدلی است که توسط ArangoDB GmbH ساخته شده است. در این دیتابیس داده‌ها را می‌توان به صورت جفت‌های کلید-مقدار، اسناد یا گراف ذخیره کرد و همه این موارد فقط با یک زبان پرس و جو به نام AQL قابل دسترسی هستند.

ArangoDB برای تمام مدل‌های داده از همان زبان اصلی و همان پرس و جو استفاده می‌کند. اگر محصول جدیدی تولید شود و یا ایده‌های جدید ایجاد شود، مدل‌ها به طور مکرر نیاز به تغییر دارند و ArangoDB امکان به روز کردن محصول و مدل را برای ما فراهم می‌کند. ترکیب مدل‌های مختلف داده در یک پرس و جو، داده‌ها را پیچیده تر می‌کند. استفاده و انتخاب ArangoDB به دلیل امکان چند مدل بودن، مقیاس پذیری ساده، کارایی مناسب، افزایش انعطاف پذیری، تحمل خطا، حافظه ذخیره سازی و هزینه کمتر نسبت به سایر بانک‌های اطلاعاتی منطقی است و این موارد را می‌توان از مزایای آن برشمرد. همچنین این بانک اطلاعاتی در نسخه ۳.۴ خود از ویژگی Search Engine خود رونمایی کرد که مورد اقبال عمومی و رشد این نرم افزار در رتبه بندی‌ها شد. این نرم افزار دارای دو نسخه است: نسخه‌های Community و Enterprise

معماری خوشه ArangoDB یک Master / Master CP (سازگار و پارتیشن) است، یعنی اینکه با استفاده از تئوری CAP در گروه CP قرار می‌گیرد. قضیه CAP بیانگر این موضوع است که در سیستم‌های توزیع شده شما فقط امکان فراهم کردن دو گزینه از سه گزینه Consistency, Availability و Partition tolerance را خواهید داشت و گزینه باقیمانده قربانی فراهم کردن سایر گزینه‌ها خواهد شد.



منظور از "CP" این است که علاوه بر پارتیشن تولرنس شبکه، ArangoDB ترجیح می‌دهد به جای در دسترس بودن، ثبات داخلی خود را حفظ کند، انتخاب Consistency و Partition tolerance به معنی آن است که هر درخواستی برای دریافت اطلاعات، دقیقاً و حتماً باید آخرین نسخه موجود را به عنوان پاسخ ارسال نماید. از طرفی، این آخرین نسخه موجود از هر اطلاعاتی بر روی سیستم کاملاً مشخص و واضح است. یعنی اگر دو درخواست یکسان از دو مکان مختلف به سیستم ارسال شود، سیستم جواب یکسان و مشخصی را به هر دو خواهد داد. نکته دیگر اینکه با افزایش Consistency، مدت تاخیر (Latency Time) نیز افزایش می‌یابد. هر میزان که Consistency مورد نیاز خود را افزایش دهید، سیستم دچار رکورد شده و سرعت پاسخ‌دهی کاهش می‌یابد. با حفظ Consistency تمامی کلاینت‌ها دید یکسانی از پایگاه داده خواهند داشت بدون توجه به اینکه به کدام نود وصل شده‌اند و نکته بعد اینکه کلاستر به پاسخ‌دهی خود ادامه می‌دهد حتی اگر یک ماشین خراب و از سیستم خارج شود. ArangoDB بعنوان یک پایگاه داده چند مدلی توزیع شده طراحی شده است.

مزایا و ویژگی‌ها

مهمترین مزایا و ویژگی‌های ArangoDB عبارتند از:

- مدیریت چندین مدل داده با یک زبان اصلی و پرس و جو (SQL)؛
- HTTP API برای مدیریت پایگاه داده؛
- معماری چندگانه (اینستنس یکتا، کلاستر و سرویس های متفاوت)
- ادغام با چارچوب JavaScript Foxx
- Sharding
- قابل استفاده در فضای ابری Azure Cloud;

• **تلفیق:** استفاده از حداقل کامپوننت‌ها که منجر به کاهش پیچیدگی می‌شود. این به این معنی است که هزینه کل مالکیت کاهش می‌یابد، انعطاف پذیری افزایش و نیازهای فنی با هم تلفیق و یکی می‌شوند.

• **مقیاس پذیری:** به راحتی می‌توانید کارایی و ذخیره سازی را با مقیاس پذیری افزایش دهید، ArangoDB هم به صورت عمودی و هم به صورت افقی قابل مقیاس پذیری است، در واقع اگر عملکرد ما نیاز به کاهش داشته باشد، می‌توانیم به راحتی سیستم back-end را down scale کنیم، که این کار منجر به صرفه جویی در سخت افزار و نیازهای عملیاتی می‌شود.

• قابلیت استفاده به عنوان search engine

• کاهش پیچیدگی عملیاتی با استفاده از Polyglot Persistence

○ اصطلاح Polyglot Persistence یک نوع معماری نرم افزاری است که این مفهوم را بیان میکند

که برای هر کار، دیتا مدل مناسب همان کار انتخاب شود، یعنی برای یک اپلیکیشن می‌شود از چندین دیتابیس مختلف استفاده کرد. برای مثال می‌شود session های کاربرها را در دیتابیس از نوع key-value ذخیره کرد و تراکنش‌ها را در oracle و کامنت‌های کاربران را در mongodb. از آنجایی که بانک اطلاعاتی ArangoDB چند مدلی است این امکان را به ما می‌دهد که مفهوم Polyglot را بدون پیچیدگی پیاده کنیم و سازگاری داده ها را بدون خطا روی یک سیستم پیاده کنیم.

• **سازگاری داده‌ها:** مدل‌های مختلف داده را با پشتیبانی از ACID مدیریت می‌کند.

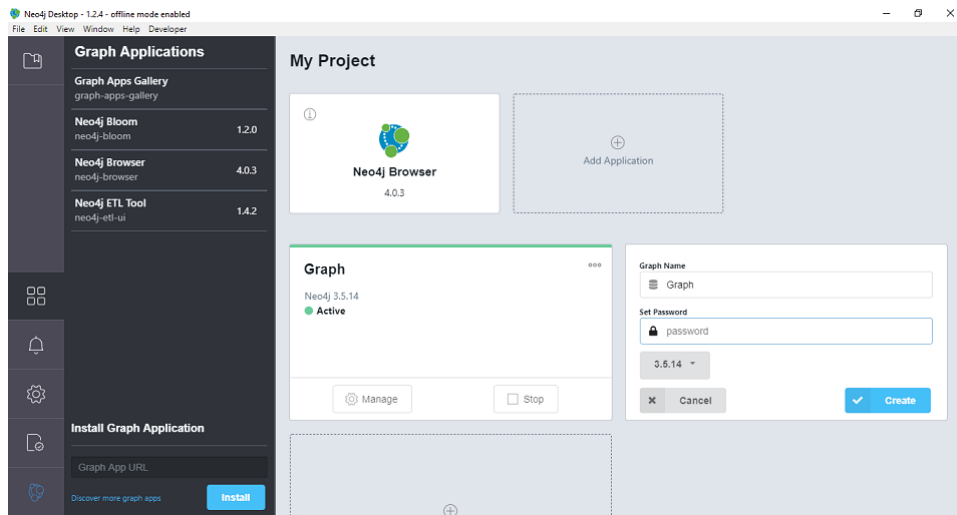
• **تحمل خطا:** مولتی مدل بودن، معماری ماژولار، اجرای مدل‌های داده‌ای متفاوت و استفاده مناسب از کلاسترها منجر به افزایش تحمل خطا در این دیتابیس می‌شود.

این پایگاه داده در سال ۲۰۱۲ ایجاد شد اما Community آن هنوز خیلی گسترده نیست، همچنین برای استفاده از آن به دلیل پیچیدگی‌های نرم افزاری ممکن است به یک کارشناس پایگاه داده نیاز داشته باشد.

نصب ArangoDB به دو صورت اتوماتیک و دستی انجام می‌شود و بعد از نصب در آدرس <http://۱۲۷.۰.۰.۱:۸۵۲۹/> و با نام کاربری root در دسترس است. طی بررسی صورت گرفته، عمده مشتریان این

بانک اطلاعاتی در حوزه مالی و بانکی و هواپیمایی فعال بودند. و روی سیستم عامل‌ها و کلاینت‌های مختلفی قابل نصب و در دسترس می‌باشد.

NeoFz



NeoFz یک پایگاه داده گرافیکی منبع باز است که در جاوا پیاده سازی شده است. توسعه دهندگان NeoFz را به عنوان یک بانک اطلاعاتی کاملاً تراکنشی با یک موتور جاوا قوی توصیف می‌کنند که می‌توانیم به جای استفاده از جداول در آن، ساختارها را به شکل گراف ذخیره کنیم.

NeoFz محبوب ترین و پرکاربردترین پایگاه داده گرافیکی در سراسر جهان به حساب می‌آید که در صنایع مختلفی از قبیل بهداشت، دولت، تولید خودرو، منطقه نظامی و سایر مناطق مورد استفاده قرار می‌گیرد و مرجع اصلی در این زمینه محسوب می‌شود.

این نرم افزار در سال ۲۰۰۷ ایجاد شد و به سه نسخه گسترده تقسیم می‌شود: یک نسخه عمومی که یک نسخه رایگان است، یک Enterprise Edition، که در آن امکان آزمایش یک نسخه کامل تر را به مدت ۳۰ روز فراهم میکند و نمونه سوم آن یک نسخه دولتی است، که مانند یک نسخه به روز شده از نسخه Enterprise است. این نسخه روی خدمات دولتی متمرکز است. تفاوت‌های اصلی بین دو نسخه اصلی NeoFz عبارتند از: وجود پشتیبان گیری آنلاین، عملکرد بهتر و بالاتر حافظه پنهان، نظارت سیستمی دقیق، مدیریت قوی قفل‌های دیتابیس و قابلیت مقیاس پذیری بیشتر بانک اطلاعاتی از دیگر مزایای نسخه Enterprise محسوب می‌شود.

سناریوهای معمول استفاده از NeoFz عبارتند از: سیستم‌های پیشنهاد دهنده برخط، شناسایی و مدیریت دسترسی، عملیات شبکه و فناوری اطلاعات، کشف تقلب، پولشویی / فرار مالیاتی، نمودار دانش، تجزیه و تحلیل گرافیکی و الگوریتم‌ها، هوش مصنوعی که دارای نمودار گرافیکی هستند، خانه‌های هوشمند و IoT.

در حال حاضر بیشتر بانکهای اطلاعاتی از طریق یک سرور که از طریق یک کتابخانه محلی در دسترس قرار می‌گیرد، اجرا می‌شوند. NeoFz می‌تواند در حالت تعبیه شده و همچنین در حالت سرور اجرا شود. حالت Embedded مستقیماً با دیسک کار میکند و تمام اطلاعات را روی دیسک ذخیره می‌کند.

NeoFJ تعبیه شده برای دستگاه‌های سخت افزاری، برنامه‌های دسکتاپ و برنامه‌های جاسازی شده در سرورها ایده آل است. برای پیاده سازی پایگاه داده اجرای NeoFJ در حالت سرور معمول ترین و بهترین انتخاب است. باید به این نکته اشاره کرد که در هر سرور یک embedded instance از NeoFJ وجود دارد. NeoFJ مزایای رقابتی زیادی دارد، که این نرم افزار را به یکی از پرکاربردترین آنها در این زمینه تبدیل می‌کند. در زیر، ویژگی‌های اصلی NeoFJ را شرح می‌دهیم:

مزایا و ویژگی‌ها

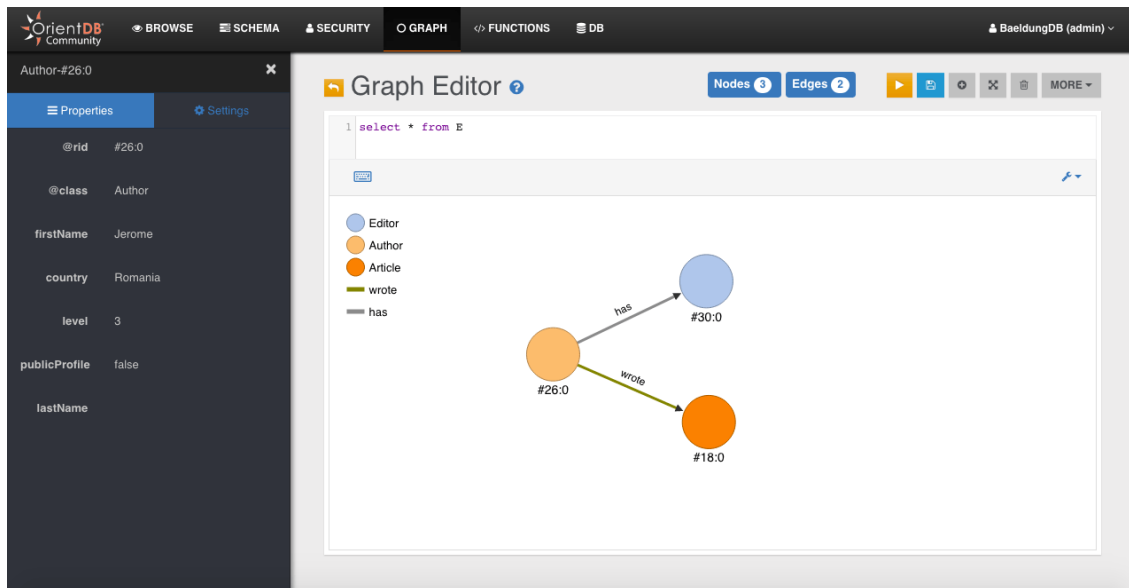
- شمای انعطاف پذیر
- مدل داده گراف
- مقیاس پذیری و قابلیت اطمینان؛
- زبان پرس و جوی Cypher
- HTTP API برای مدیریت پایگاه داده؛
- پشتیبانی از indexها با استفاده از Apache Lucene
- پشتیبانی از درایورهای مانند Java، Spring، Scala، JavaScript
- Online backups
- پشتیبانی از Cloud
- خروجی گرفتن از نتایج پرس و جو در قالب JSON و XLS؛
- فعال ترین community گراف در جهان؛
- کارایی بالا به دلیل استفاده از حافظه گرافیکی محلی و پردازش محلی
- یادگیری و استفاده آسان
- بارگیری آسان داده‌ها در نرم افزار؛
- دیتا مدل دوستدار تخته برای ساده سازی چرخه توسعه

NeoFJ از sharding پشتیبانی نمی‌کند و در نسخه Community که نسخه رایگان آن است محدودیت‌هایی در تعداد نودها، ارتباطات و خصوصیات وجود دارد.

برخی از مشتری‌های NeoFJ عبارتند از :

- eBay
- Adob
- Microsoft
- IBM
- NBC News

OrientDB



OrientDB یک سیستم مدیریت پایگاه داده NoSQL منبع باز چند مدل است که از مدل‌های داده اسناد، گراف، کلید / مقدار و اشیاء پشتیبانی می‌کند. این پایگاه داده که با زبان جاوا توسعه داده شده بود در سال ۲۰۱۱ توسط Orient Technologies به بازار عرضه شد. Orientdb یک دیتابیس تراکنشی است و از معماری توزیع شده و replication پشتیبانی می‌کند. تغییرات پایگاه داده می‌تواند در جاوا، SQL و یا با Gremlin انجام شود. ذخیره سازی داده‌های فیزیکی می‌تواند در حافظه و یا دیسک انجام شود.

Orientdb مشابه سایر سیستم‌ها از لیست مجاورت برای پردازش کوئری‌های محلی استفاده می‌کند اما برخلاف سایر سیستم‌ها از قابلیت‌های پایگاه داده اسناد قابلیت‌های Object-Orientation برای ذخیره سازی نودهای فیزیکی استفاده می‌کند. Orientdb از اسکیمای‌های مختلفی پشتیبانی می‌کند، Schema-Full، Schema-Less و Schema-Hybrid همچنین از زبان SQL استفاده می‌کند و replication و sharding در اکثر موارد قابل استفاده است. Orientdb ایمنی را با استفاده از تایید هویت، رمز عبور و رمزگذاری داده‌های محرمانه فراهم می‌کند. نسخه Community Edition رایگان (مجوز آپاچی ۲) است اما از ویژگی‌هایی مانند مقیاس پذیری افقی، fault tolerance و خیلی ویژگی‌های دیگر پشتیبانی نمی‌کند. در نسخه Enterprise ویژگی‌های بسیاری از قبیل پروفایلر کوئری، پیکربندی کلاستر توزیع شده، ضبط متریک، نظارت آنلاین، ابزارهایی برای Migration و تنظیمات هشدار وجود دارد.

موارد استفاده OrientDB شناسایی تقلب، عملیات شبکه IT، جستجوی گراف، موتورهای پیشنهاددهنده، مدیریت داده‌ها، مدیریت شناسایی هویت و تجزیه و تحلیل‌های پزشکی قانونی است. معماری OrientDB در حالت توزیع شده استفاده می‌شود، OrientDB را می‌توان در سرورهای مختلف توزیع کرد و از روش‌های متنوعی برای دستیابی به حداکثر عملکرد، مقیاس پذیری و توانمندی آن استفاده کرد. OrientDB از پروژه Hazelcast Open Source برای کشف خودکار گره، ذخیره پیکربندی خوشه در زمان اجرا و همگام سازی عملیات خاص بین گره‌ها استفاده می‌کند.

مزایا و ویژگی‌ها

در زیر، ویژگی‌ها و مزایای اصلی OrientDB را ارائه می‌دهیم:

- پشتیبانی از زبان SQL
- پشتیبانی از فن آوری‌های وب مانند HTTP، پروتکل RESTful، کتابخانه‌های JSON؛
- توزیع شده و پشتیبانی از Multi-Master Replication
- قابل استفاده کلاد
- قابلیت تغییر دیتابیس با استفاده از جاوا
- پشتیبانی از روابط در عین حال که می‌توان استاد را با قابلیت‌های خودشان ذخیره کرد.
- معماری Multi-master
- مقیاس پذیری و قابلیت اطمینان افقی؛
- نصب سریع
- نسخه رایگان با استفاده از مجوز Apache ۲

OrientDB محدودیت‌هایی مانند عدم داشتن ابزار ورود داده دارد و برخی کاربران نیز ایراداتی را در ویرایشگر آن مشاهده کرده‌اند.

در نهایت اما برای انتخاب و استفاده از یک دیتابیس گراف در یک پروژه واقعی باید موارد زیادی را در نظر گرفت، هزینه، زمان، نیروی انسانی، حجم دیتا، ویژگی‌ها و کاربرد مورد نیاز و در کنار همه‌ی اینها توجه به نسخه‌ی از نرم افزار که ما توانایی تهیه آن را داریم.

در مقاله پایانی ما سعی می‌کنیم به شکل فنی تری به بیان ویژگی‌هایی بپردازیم که در انتخاب شما تاثیر گذار هستند و به این موضوع اشاره خواهیم کرد که هر یک از دیتابیس‌های فوق آن ویژگی‌ها را با چه کیفیتی پیاده سازی کرده‌اند.