

پایگاه داده آپاچی کاساندرای چیست؟



cassandra



پایگاه داده آپاچی کاساندرای چیست؟



<https://iracode.com/what-is-apache-cassandra-database/>



۰۹۱۲۴۸۰۴۵۳۵

پایگاه داده آپاچی کاساندرای چیست؟

آپاچی کاساندرای یک سیستم مدیریت پایگاه داده توزیع شده در زبان برنامه نویسی آزاد و متن باز است که برای مدیریت حجم زیادی از داده ها در بسیاری از سرورهای کالای طراحی شده است و دسترسی بالا را بدون هیچ نقطه ضعفی فراهم می کند. کاساندرای پشتیبانی قوی برای خوشه هایی که چندین مرکز داده را پوشش می دهند.

کاساندرای همچنین ارزش بالایی برای عملکرد قائل است. در سال ۲۰۱۲، محققان دانشگاه تورنتو که سیستم های NOSQL را مطالعه می کردند، به این نتیجه رسیدند که "از نظر مقیاس پذیری، برنده واضحی در تمام آزمایش های ما وجود دارد. کاساندرای به بالاترین توان عملیاتی برای حداکثر تعداد گره ها در تمام آزمایش ها دست می یابد.

تاریخچه آپاچی کاساندرای

Avinash Lakshman (یکی از نویسندگان آمازونی دینامو) Prashant Malik در ابتدا کاساندرای را در فیسبوک توسعه دادند تا ویژگی جستجوی اینباکس فیسبوک را در اختیار کاربران قرار دهند. فیسبوک در ژوئیه ۲۰۰۸ کاساندرای را به عنوان یک پروژه متن باز در کد گوگل منتشر کرد. در مارس ۲۰۰۹ تبدیل به یک پروژه پرورش دهنده آپاچی شد. در ۱۷ فوریه ۲۰۱۰ به یک پروژه سطح بالا فارغ التحصیل شد.

توسعه دهندگان فیسبوک، پایگاه داده خود را به نام کاساندرای، پیامبر افسانه ای تروجان، با اشارات کلاسیک به نفرین یک غیبگو نامگذاری کردند.



ویژگی های اصلی آپاچی کاساندرای

ویژگی های آپاچی کاساندرای

- 1 تمرکز زدایی
- 2 پشتیبانی از تکرار و تکرار چند مرکز داده
- 3 مقیاس پذیری
- 4 تحمل اشتباه
- 5 سازگاری قابل تنظیم
- 6 پشتیبانی از MapReduce
- 7 زبان پرس و جو
- 8 مدل داده
- 9 خوشه بندی
- 10 مدیریت و نظارت

WWW.IRACODE.COM

تمرکز زدایی

هر گره در خوشه نقش یکسانی دارد. هیچ نقطه شکستی وجود ندارد. داده ها در سراسر خوشه توزیع می شوند (بنابراین هر گره حاوی داده های متفاوتی است)، اما هیچ اربابی وجود ندارد زیرا هر گره می تواند هر درخواستی را ارائه دهد.

پشتیبانی از تکرار و تکرار چند مرکز داده

استراتژی های تکرار قابل پیکر بندی هستند. آپاچی کاساندریا به عنوان یک سیستم توزیع شده، برای استقرار تعداد زیادی از گره ها در مراکز داده متعدد طراحی شده است. ویژگی های کلیدی معماری توزیع شده کاساندریا به طور خاص برای استقرار چند مرکز داده، برای افزونگی، برای شکست و بازیابی فاجعه مناسب است.

مقیاس پذیری

با اضافه شدن ماشین های جدید، توان عملیاتی خواندن و نوشتن هر دو به صورت خطی افزایش می یابد، بدون اینکه وقفه ای در برنامه ها ایجاد شود.

تحمل اشتباه

داده ها به طور خودکار برای تحمل خطا در گره های متعدد تکرار می شوند. تکرار در چندین مرکز داده پشتیبانی می شود. گره های شکست خورده را می توان با هیچ زمان نزولی جایگزین کرد.

سازگاری قابل تنظیم

نوشته ها و خواندن ها سطح قابل تنظیمی از سازگاری را ارائه می دهند، در تمام طول مسیر از "نوشته ها هرگز شکست نمی خورند" گرفته تا "مسدود کردن برای قابل خواندن بودن تمام کپی ها"، با سطح حد نصاب در وسط.

پشتیبانی از MapReduce

کاساندریا یکپارچگی Hadoop را با پشتیبانی MapReduce دارد. همچنین پشتیبانی از Apache Pig و Apache Hive نیز وجود دارد.

زبان پرس و جو

کاساندریا زبان پرس و جوی کاساندریا (CQL) را معرفی کرد. CQL یک رابط ساده برای دسترسی به کاساندریا، به عنوان جایگزینی برای زبان پرس و جوی ساختار یافته سنتی (SQL) است. CQL یک لایه انتزاعی اضافه می کند که جزئیات پیاده سازی این ساختار را پنهان می کند و نحوه های بومی را برای مجموعه ها و دیگر کدگذاری های رایج فراهم می کند.



در زیر نمونه ای از ایجاد فضای شامل یک خانواده ستونی در CQL آمده است:

```
CREATE KEYSPACE MyKeySpace
;{ WITH REPLICATION = { 'class' : 'SimpleStrategy', 'replication_factor' : 3
;USE MyKeySpace
CREATE COLUMNFAMILY MyColumns (id text, Last text, First text,
;((PRIMARY KEY(id
;('INSERT INTO MyColumns (id, Last, First) VALUES ('1', 'Doe', 'John
;SELECT * FROM MyColumns
```

خروجی:

id | first | last-----+-----+----- 1 | John | Doe
سازگار نیست، به این معنی که درج ها و به روز رسانی ها در جدول که بر همان ردیف که تقریباً همزمان پردازش می شوند تاثیر می گذارد، ممکن است به روش های متناقض بر ستون های غیر کلیدی تاثیر بگذارد... یک به روز رسانی ممکن است بر یک ستون تاثیر بگذارد در حالی که به روز رسانی دیگر بر ستون دیگر تاثیر می گذارد، که منجر به مجموعه ای از مقادیر در داخل ردیف می شود که هرگز مشخص یا در نظر گرفته نشده اند.

مدل داده

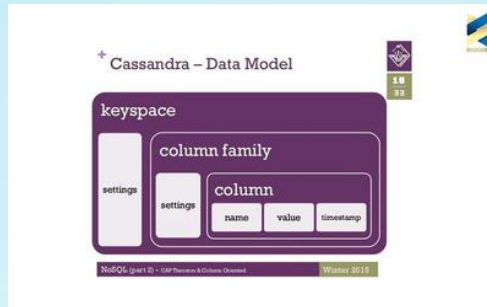
کاساندریا در اصل ترکیبی بین یک کلید - مقدار و یک سیستم مدیریت پایگاه داده ستون گرا (یا جدولی) است. مدل داده ای آن یک ذخیره ردیفی پارتیشن بندی شده با ثبات قابل تنظیم است. ردیف ها در جدول ها سازماندهی می شوند؛ اولین جز کلید اصلی جدول، کلید پارتیشن است؛ در داخل یک پارتیشن، ردیف ها توسط ستون های باقی مانده کلید دسته بندی می شوند. ستون های دیگر ممکن است به طور جداگانه از کلید اصلی نمایه شوند.

جداول ممکن است در زمان اجرا بدون مسدود کردن به روز رسانی ها و پرس و جوها ایجاد، حذف و تغییر داده شوند.

آپاچی کاساندریا نمی تواند به هم بپیوندد یا به هم بپیوندد. در عوض، کاساندریا بر هنجار زدایی از طریق ویژگی هایی مانند مجموعه ها تاکید می کند.

یک خانواده ستونی (که از SQL3 به بعد "جدول" نامیده می شود) شبیه به یک جدول در RDBMS است. خانواده های ستونی شامل ردیف ها و ستون ها هستند. هر ردیف به طور منحصر به فردی توسط یک کلید ردیف شناسایی می شود. هر سطر دارای چندین ستون است که هر کدام از آن ها یک نام، مقدار و یک تایم لپس دارند. برخلاف یک جدول در RDBMS، ردیف های مختلف در یک خانواده ستونی نیازی به اشتراک گذاری مجموعه ستون های یک سان ندارند، و یک ستون ممکن است در هر زمان به یک یا چند ردیف اضافه شود.

هر کلید در کاساندریا متناظر با یک مقدار است که یک شی است. هر کلید مقادیری به عنوان ستون دارد و ستون ها در مجموعه هایی به نام خانواده ستونی با هم گروه بندی می شوند. بنابراین، هر کلید یک ردیف از تعداد متغیر عناصر را مشخص می کند. این خانواده های ستونی را می توان به صورت جدول در نظر گرفت. یک جدول در کاساندریا یک نقشه چند بعدی توزیع شده است که توسط یک کلید نمایه شده است. علاوه بر این، اپلیکیشن ها می توانند ترتیب ستون ها را در خانواده Super Column یا Simple Column مشخص کنند.



خوشه بندی

هنگامی که کلاستر برای آپاچی کاساندریا طراحی می شود، یک نکته مهم انتخاب پارتیشن بندی مناسب است. دو افراز وجود دارد:

- مرتب سازی انتخابی (OPP): این مرتب سازی زوج های مقدار کلید را به روشی طبیعی توزیع می کند به طوری که کلیدهای مشابه از هم دور نباشند. مزیت این است که باید به گره های کمتری دسترسی داشت. اشکال توزیع نامتوازن جفت های مقدار کلید است.
- مرتب سازی تصادفی (RP): این پارتیشن به طور تصادفی جفت های مقدار کلید را در شبکه توزیع می کند که منجر به تعادل بار خوبی می شود. در مقایسه با OPP، برای به دست آوردن تعدادی کلید باید به گره های بیشتری دسترسی داشت.

مدیریت و نظارت

کاساندریا یک سیستم مبتنی بر جاوا است که می تواند از طریق ضمیمه های مدیریت جاوا (JMX) مدیریت و نظارت شود. برای مثال، برنامه نود سازگار با JMX می تواند برای مدیریت یک خوشه کاساندریا (اضافه کردن گره ها به یک حلقه، تخلیه گره ها، از بین بردن گره ها، و غیره) استفاده شود. Nodetool همچنین تعدادی دستور برای بازگرداندن معیارهای کاساندریا مربوط به استفاده از دیسک، تاخیر، فشردگی، جمع آوری زباله و غیره ارائه می دهد. معیارهای اضافی از طریق ابزارهای JMX مانند JConsole و از طریق گزارشگران معیارهای قابل اتصال برای ابزارهای نظارت خارجی در دسترس هستند، که با نسخه ۲/۰/۲ کاساندریا در دسترس قرار گرفت.

