بسمه تعالی

**Apache Cassandra**

کاساندرا یک دیتابیس متن باز است که اولین بار فیس بوک آن را تولید کرد و بعدا توسط کپانی آپاچی توسعه یافت ، از دیگر ویژگی های این دتابیس میتوان به موارد زیل اشاره کرد :

1 . توضیع شده : به این معنی که داده ها میتوانند در سرور های مختلف ذخیره شوند و کاساندرا میتواند آن ها را مدیریت کند .

2 . مقیاس بزرگ داده : این دیتابیس میتواند مقادیر بسیار زیاد داده راکنترل کند ، از این رو برای مواردی که حجم ترکنش بسیار زیاد است میتوان از این دیتابیس استفاده کرد .

3 . در کاسانردا اگر نقطه ای از سیسیتم (سرور یا بخشی از آن) از بین برور و یا crash کند کل سیستم آسیب نمیبیند . یکی از دلایل آن کپی شدن داده ها در سرور های مختلف و همچنین مکانیزم های بازگردانی داده ها از روی لاگ های آن میباشد (Commit Log ) .

می توانیم پایگاه های داده را به سه دسته تقسیم کنیم :

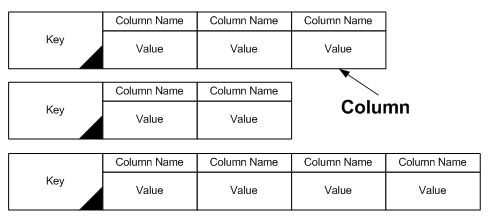
* RDBMS
* OLAP
* NoSQL

در این صورت پایکاه داده کاساندرا جزء دسته سوم محسوب میشود . از تفاوت های مهمی که ای دسته با نوع رابطه ای دارد میتوان به موارد زیر اشاره کرد (توجه داشته باشید بدلیل ملموس بودن نوع پایگاه داده رابطه ای ):

1 . در این نوع پایگاه داده تعریفی به نام جدول مانند پایگاه داده رابطه ای وجود ندارد و قهرا زمانی که جدولی وجود نداشته باشد رابطه ای نیز بین آنها وجود ندارد ؛ ا زاین رو به این نوع پایگاه داده ها غیر رابطه ای هم گفته میشود .

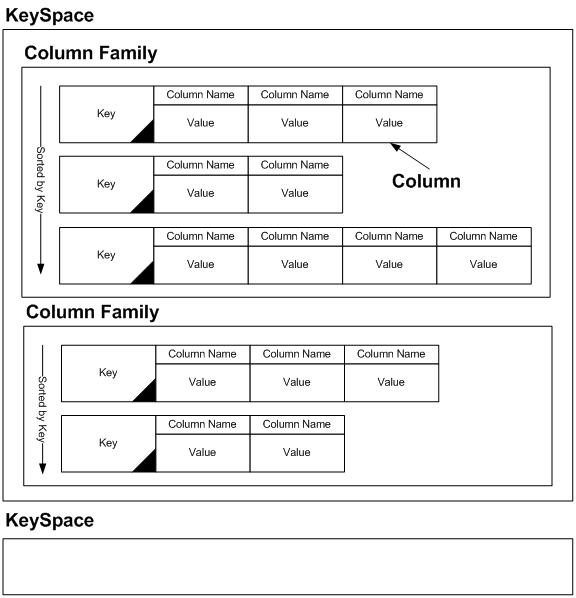
در پایگاه داده رابطه ای ابتدا یک شما تعریف شده ، سپس جداول در آن تبدل میشوند و در آن ستون ها مشخص میشود و رکورد ها باتوجه به ستون ها مقدار دهی میشوند ؛ اما یکی از مشکلات پایگاه داده رابطه ای زمانی مشخص میشود که رکورد ها مقادیری برای برخی ستون ها نداشته باشند ، از این رو مقدار آن ستون - نال - میشود ، ان اتفاق در حجم داده بالا اصلا مناسب نیست ، در صورتیکه در پایگاه داده - نواس کیو ال – ساختار جدول وجود ندارد و ساختار ها بر اساس ستون ها میباشد و از این رو وقتی رکوردی نیازی به ستونی نداشته باشد به کلی آن را حذف میکند . شکل زیر ساختار ستونی را در پایگاه داده ای غیر رابطه ای نشان میدهد .

همانطور که درشکل مشخص است در هر سه رکورد ستون ها متفاوت هستند و فقط دو ستون اول آنها یکسان است .



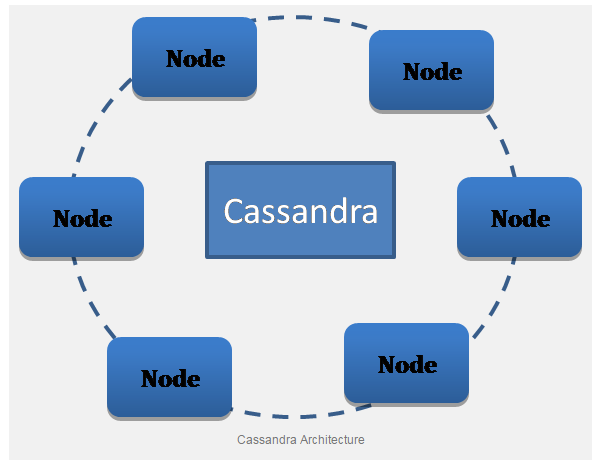
چند رکورد که ستون های مشترک دارند را - ستون های فامیل - میگویند .

بالا ترین سطح داده درپایگاه داده غیر رابطه ای - فضای کلید - فضای کلید نام دارد که میتواند شامل یک یا چند ستون فامیل باشد . برای فهم بهتر به شکل زیر توجه کنید .



**اجزاء پایگاه داده کاساندرا**

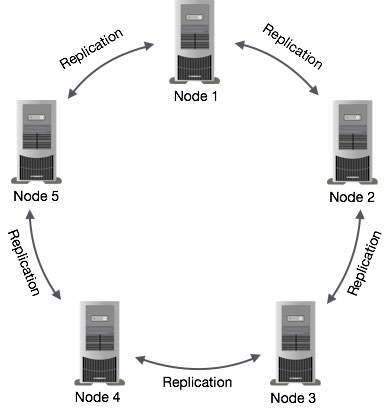
* 1 . node : محلی که داده ها در آن ذخیره میشوند (کامپیوتر، سرور،...)
* 2. Data Center : مجموعه نود های وابسته به یکدیگر
* 3.Cluster : مجموعه ای از یک یا چند دیتاسنتر



همانطور که در شکل مشاهده میکنید نود هایی که با یکدیگر رابطه دارند یک کلاستر را تشکیل میدهند . همچنین نود ها داده های خود را در اختیار نود های دیگر قرارمیدهند ؛ یکی از دلایل این کار امنیت سیستم میباشد به این شکل که اگر یکی از نود ها به هر دلیلی از بین رفت یا با مشکل مواجه شد اطلاعات از بین نمیرود .این عمل که - رپلیکیشن - نام دارد با دومشخصه ی موجود در - فضای کید - انجام میشود:

* Replication Factor
* Replica Placement Strategy

که مورد اول مکانیزم برای دریافتت کپی داده ها و مورد دوم برای قرار دادن داده ها در حلقه ی نود ها (کلاستر) میباشد .



**روند انجام عملیات نوشتن برروی پایگاه داده کاساندرا**

1.در اولین قدم باید یک – Session – به یکی از نود ها بزنیم یا به عبارت دیگه باید یه مسیر ارتباطی بین Client و Node در بستر شبکه اینجاد کنیم (شبکه داخلی یا شبکه اینترنت) ؛ اینکار مستلزم اطلاع از Node IP و پورتی است که کاساندرا آنرا مشخص کرده است . کاساندرا بصورت پیش فرض پورت 9042 رابه خود اختصاص میدهد .

2. زمانی که یک ارتباط به نود ایجاد شد میتوانیم عملیات خود را روی دیتابیس انجام دهیم .در اینجا به سه مورد ازروش های آن اشاره میکنیم .

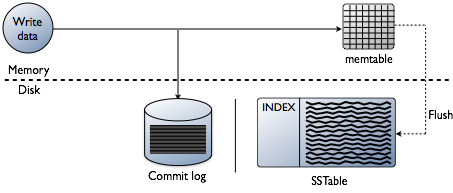
* CQL
* CQLSH
* SQL

3. زمانی که با یکی از روش های فوق عملیاتی را روی کاساندرا انجام دادیم ابتدا –Log – عملیات ذخیزه میشود(Commit Log) . اگر داده ای به هر دلیلی از بین برود از این طریق میتوان آن را بازیابی کرد ؛ همچنین این استراژی به ما کمک میکند که بتوانیم روند اتفاقات مهم را کنترل کنیم .

4. بعد از ذخیره شدن در لاگ داده ها روی مموری نوشته میشوند . (MemTable)

5. بعد از نوشته شدن داده روی مموری ، سپس مموری فلاش میشود و داده های روی دیسک قرار میگیرند . (SSTable)

روند بالا را میتوان در شکل زیر مشاهده کرد .



* نکته : همچنین میتوان از CQLSH (شل اسکریپت) برای Query نوشتن استفاده کرد .
* نکته :در دیتابیس NewSQL مشکلات NoSQL رفع شده است و بهتر است در مورد آن نیز تحقیق انجام شود .

باتشکر