

Cassandra Essentials Series Tutorial

Uma visão geral do
Apache Cassandra



Agenda

- O que é Cassandra?
- História
- Arquitetura
- Principais recursos e benefícios
- Quem está usando Cassandra?
- Onde obter Cassandra

Definição de Cassandra

Apache Cassandra TM é livre

Distribuído ...

Alto desempenho ...

extremamente escalável ...

Tolerante a falhas (ou seja, nenhum ponto único de falha) ...

solução de banco de dados pós-relacional. Cassandra pode servir como armazenamento de dados em tempo real (o “sistema de registro”) para aplicações online / transacionais, e, como um banco de dados readintensive para sistemas de business intelligence.



A História de Cassandra

Mesa grande

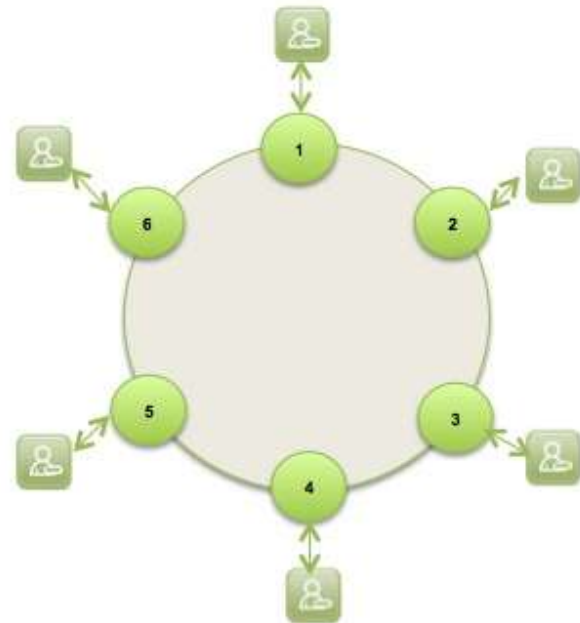


Dínamo



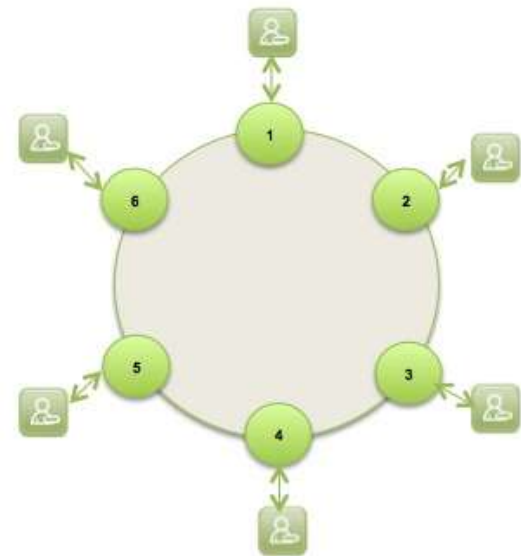
Visão Geral da Arquitetura

- Cassandra foi projetado com o entendimento de que as falhas do sistema / hardware podem acontecer e acontecem
- Peer-to-peer, sistema distribuído
- Todos os nós do mesmo
- Os dados particionados entre todos os nós do cluster
- replicação de dados personalizado para garantir tolerância a falhas
- Leia concepção / Write-em qualquer lugar



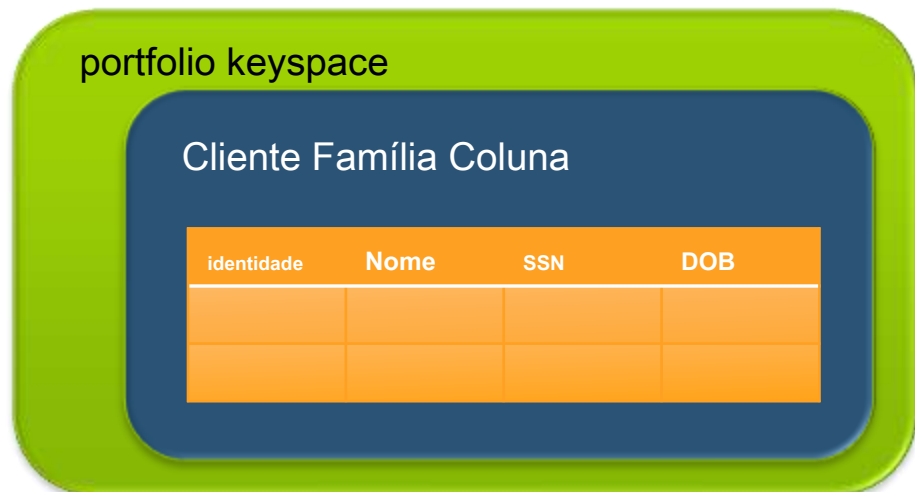
Visão Geral da Arquitetura

- Cada nó comunica com o outro por meio do protocolo Mexerico, que troca informações entre a agrupar cada segundo
- Um log cometer é usado em cada nó para capturar a atividade de gravação.
durabilidade de dados é garantida
- Os dados também escrito para uma estrutura na memória (memtable) e depois para o disco uma vez que a estrutura de memória está cheio (um SStable)



Visão Geral da Arquitetura

- O esquema usado em Cassandra é espelhado depois que o Google Bigtable. Trata-se, uma estrutura de coluna orientada a fileira
- A keyspace é semelhante a um banco de dados no mundo do RDBMS
- Uma família coluna é semelhante a uma mesa de RDBMS, mas é mais flexível / dinâmico
- Uma linha em uma família de colunas é indexado por sua chave. Outras colunas podem ser indexados tão bem



Por Cassandra?

- Gigabyte para escalabilidade petabyte
- ganhos de desempenho linear através adicionar nodes
- Não existe um único ponto de falha
- distribuição fácil de replicação / dados
- Multi-centro de dados e Cloud capaz
- Não há necessidade de camada de cache separado
- a consistência dos dados sintonizáveis
- design de esquema flexível
- Compressão de dados
- linguagem CQL (como SQL)
- Suporte para idiomas principais e plataformas
- Não há necessidade de hardware ou software especial

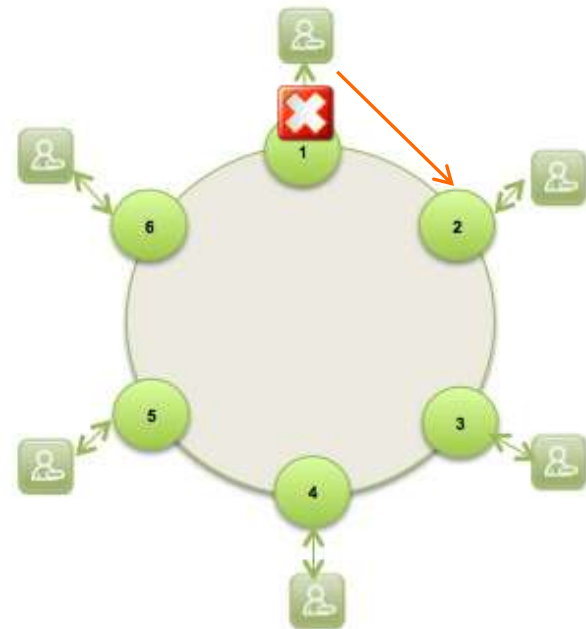
Escalabilidade Big Data

- Capaz de confortavelmente escalar para petabytes
- Novos nós = lineares aumentos de desempenho
- Adicionar novos nós on-line



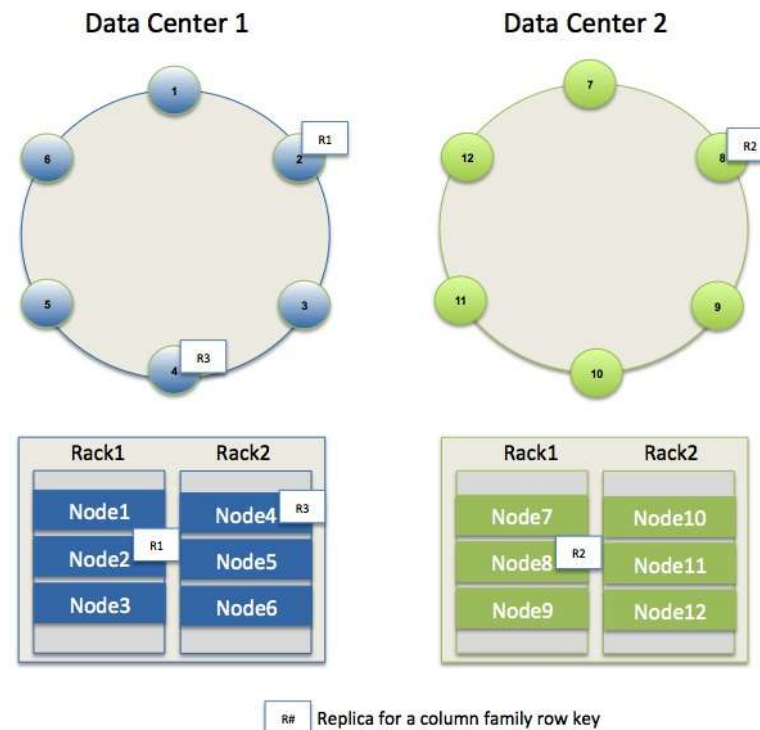
No ponto de falha

- Todos os nós do mesmo
- replicação personalizado proporciona redundância de dados sintonizável
- Leitura / gravação de qualquer nó
- Pode replicar dados entre diferentes racks de data center físico



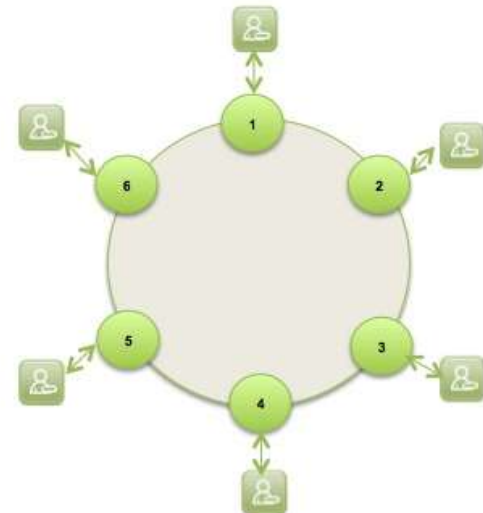
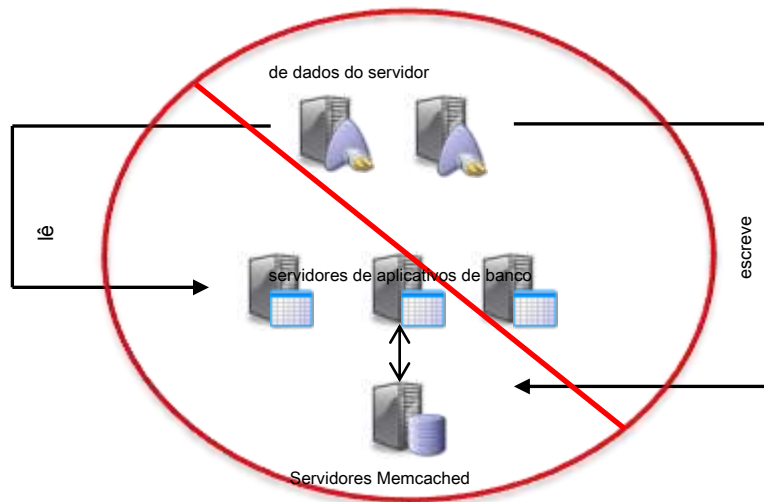
Fácil de replicação / Distribuição de Dados

- Transparente manipulados por Cassandra
- Multi-centro de dados capaz
- Explora todos os benefícios da computação em nuvem
- Capaz de fazer híbrido configuração Nuvem / On-premise



Sem necessidade de software Caching

- Peer-to-peer arquitetura remove necessidade de camada de cache especial e a programação que vai com ele
- O cluster de banco de dados usa a memória de todos os nós participantes para armazenar em cache os dados atribuídos a cada nó
- Não há irregularidades entre um cache de memória e banco de dados são encontrados



Sintonizável consistência de dados

- Escolha entre consistência forte e eventual (Todos com qualquer nó de responder) dependendo da necessidade
- Pode ser feito em uma base per-operação, e para tanto lê e escreve
- Lida com as operações do centro de dados Multi-

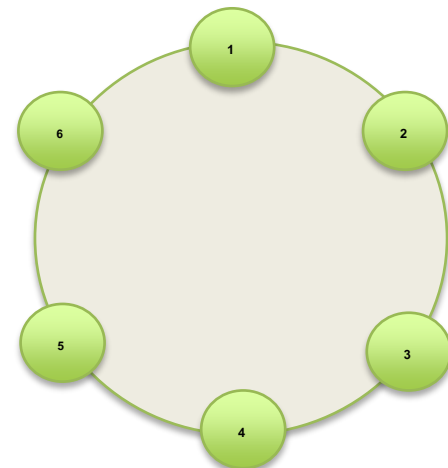


escreve

- Qualquer
- 1
- Quorum
- Local_Quorum
- Each_Quorum
- Todos

lê

- 1
- Quorum
- Local_Quorum
- Each_Quorum
- Todos



esquema flexível

- design de esquema dinâmico permite o armazenamento de dados muito mais flexível do que o RDBMS rígida
- Manipula dados estruturados, semi-estruturado, e não estruturados. Contadores também apoiou
- Sem offline / tempo de inatividade para as mudanças de esquema
- Suporta índices primárias e secundárias

portfolio keyspace

Cliente Família Coluna

identidade	Nome	SSN	DOB

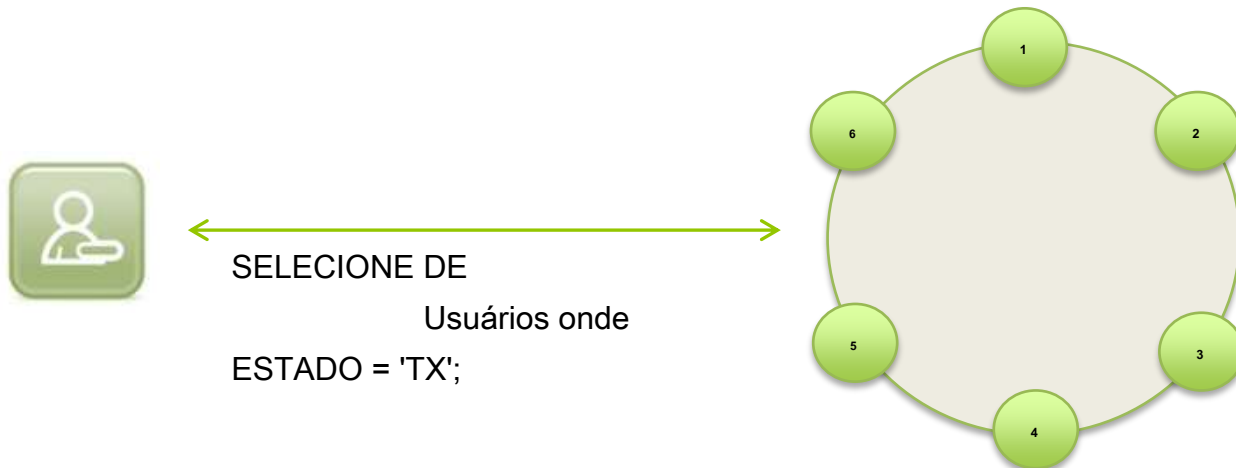
Compressão de dados

- Usa o algoritmo de compressão de dados Snappy do Google
- Comprime os dados num nível familiar per coluna
- Os testes internos no DataStax mostrar-se para 80% + de compressão de dados não processados
- Nenhuma penalidade de desempenho (e alguns aumentos de desempenho global devido à menor física I / O)!



CQL Idioma

- Muito semelhante ao RDBMS sintaxe SQL
- Criar objetos via DDL (por exemplo, criar ...)
- comandos núcleo DML suportados: INSERT, UPDATE, DELETE
- dados de consulta com SELECIONAR



Quem está usando Cassandra?

<http://www.datastax.com/cassandrausers#all>



Get Cassandra

Developer Center

Contact Us



CONTACT US

First Name *

Last Name *

Email *

Company *

Company Size *

State *

Phone *

Comments *

Cassandra Users

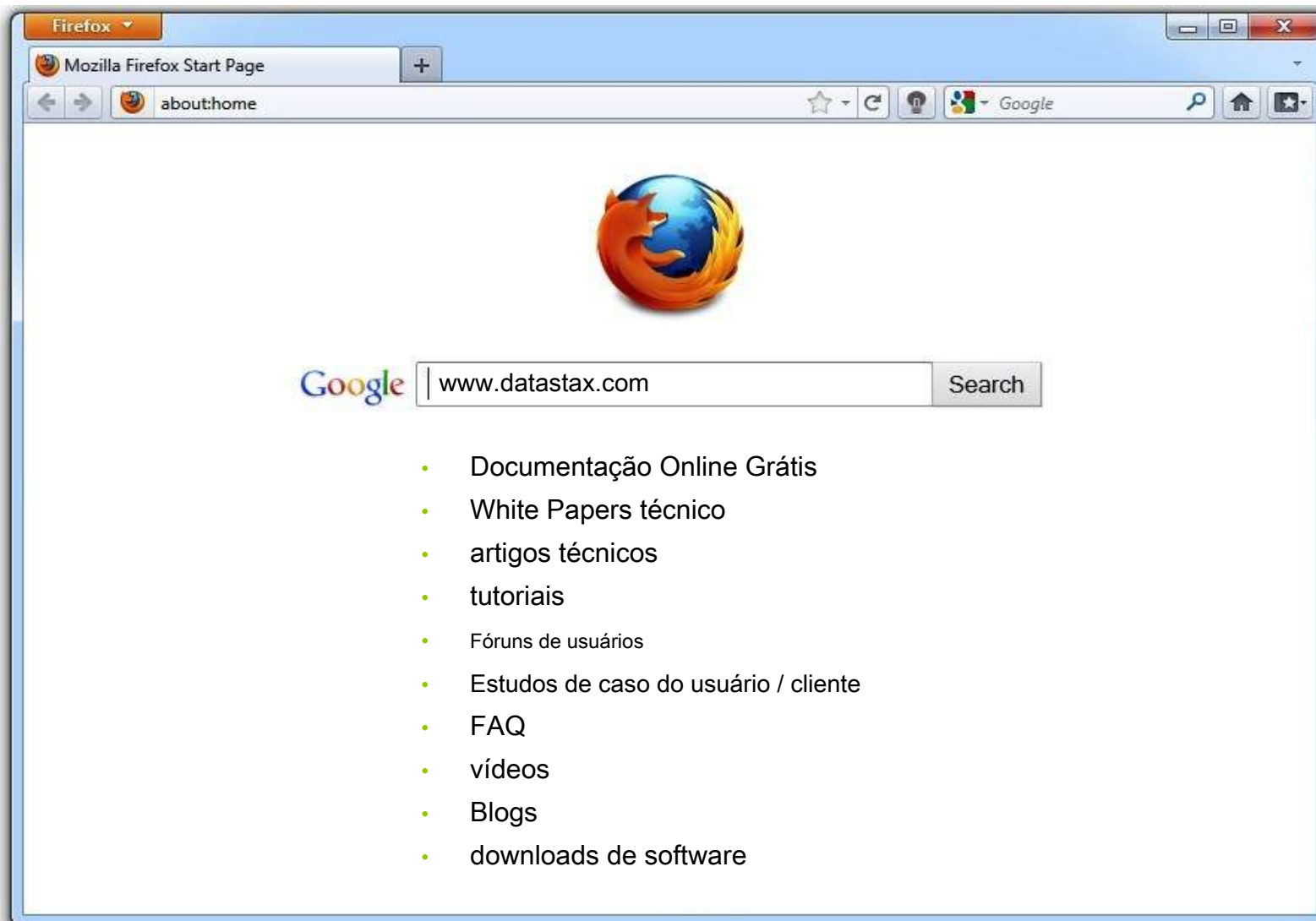
As the Cassandra community continues to grow, the amount of use cases grows as well. Here's a list of companies using Cassandra and how Cassandra is being utilized. If you would like to be added, send an e-mail out to info@datastax.com with your use case.

Company	Cassandra Use Case
	<ul style="list-style-type: none">A9 uses Cassandra as a storage solution to a dataset for their E-commerce Product and Visual Search technologies
	<ul style="list-style-type: none">Accenture created a Cassandra Beginner's Guide
	<ul style="list-style-type: none">Adform uses Cassandra to help power their digital advertising platform
	<ul style="list-style-type: none">Why Adku Chose Cassandra over Hbase
	<ul style="list-style-type: none">AdXpose™ Analytics helps advertisers and platforms verify and optimize billions of campaign data points captured in real time using Cassandra

Onde obter Cassandra?

- Ir para www.datastax.com
- DataStax torna livres instaladores de início inteligentes disponíveis para Cassandra que incluem:
 - A versão mais up-to-date Cassandra que é qualidade da produção
 - A versão do DataStax OpsCenter, que é uma ferramenta de gestão visual, baseada em browser para gerenciar e monitorar Cassandra
 - Drivers e conectores para linguagens de desenvolvimento populares
 - Mesmo banco de dados e aplicação
 - assistência automática de configuração para garantir o desempenho ideal e configuração tanto para implementações independentes ou de fragmentação
 - Guia de Introdução

Onde posso saber mais?



Cassandra Essentials Series Tutorial

Uma visão geral do
Apache Cassandra

Obrigado...!

