Apache Cassandra

"manage massive amounts of data, fast, without losing sleep."

Fernando Fagonde
UNIPAMPA
SABD
Bagé - RS

Summary

- What is Cassandra?
  - NoSQL*;
    - CAP Theorem;
  - Proven;
  - Fault Tolerant;
  - Performant;
  - Decentralized;
  - Scalable;
  - Durable;
  - Elastic;
  - CQL.
- Installation.
- Next Steps.

What is Cassandra?

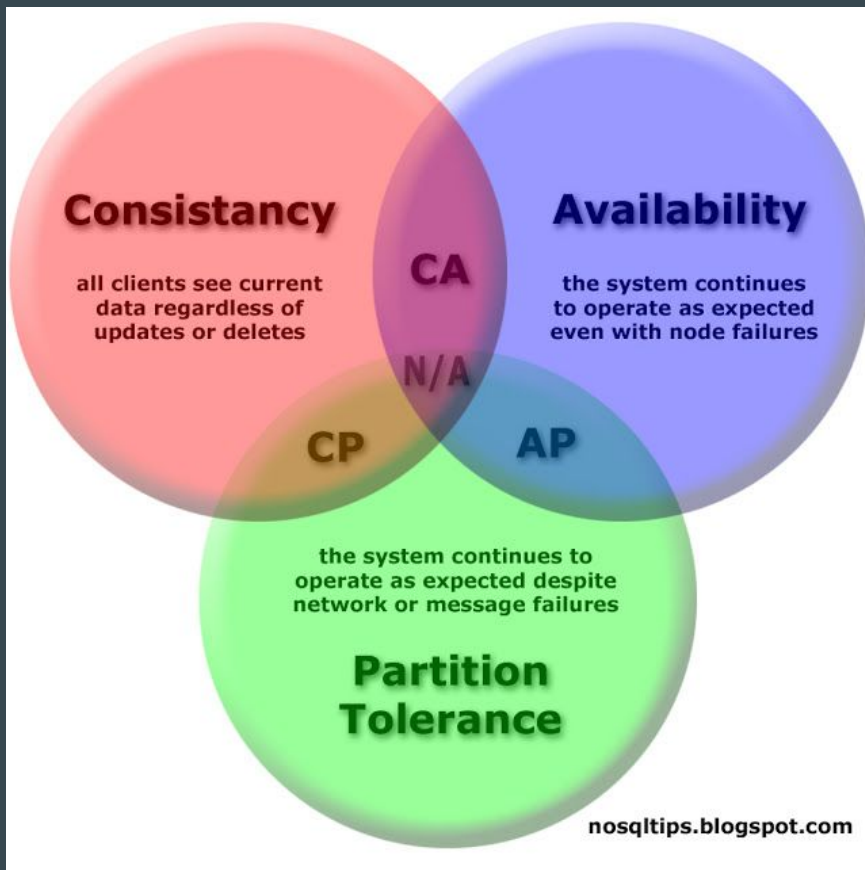# What is Cassandra?

- NoSQL*

*NOSQL

- NoSQL é um termo usado para descrever bancos de dados não relacionais de alto desempenho.
- Os bancos de dados NoSQL usam diversos modelos de dados, incluindo documentos, gráficos, chave-valor e colunares.
- Bancos de dados NoSQL são amplamente reconhecidos pela facilidade de desenvolvimento, desempenho escalável, alta disponibilidade e resiliência.

CAP Theorem

Dr. Eric Brewer

CAP Theorem

"Cassandra are always weakly consistent, favoring high availability and response time (latency), ie PA / L."

fonte:http://blog.caelum.com.br/nosql-do-teorema-cap-para-paccl/

Proven

- Cassandra is in use at enterprises that have large, active data sets:
  - CERN;
  - eBay;
  - Github;
  - GoDaddy;
  - Instagram;
  - Netflix;
  - Reddit;
  - The Weather Channel;
  - +1500.

fontes:

http://cassandra.apache.org/

Fault Tolerant

- Data is automatically replicated to multiple nodes for fault-tolerance;
- Replication across multiple data centers is supported;
- Failed nodes can be replaced with no downtime.

fontes:

http://cassandra.apache.org/

Performant

- Cassandra Consistently outperforms popular NoSQL alternatives in benchmarks and real applications, primarily because of fundamental architectural choices;

fontes:

http://cassandra.apache.org/

Decentralized

- There are no single points of failure.
- There are no network bottlenecks.
- Every node in the cluster is identical.

fontes:

http://cassandra.apache.org/

Scalable

- Some of the largest production deployments include:
  - Apple's, with over 75,000 nodes storing over 10 PB of data;
  - Netflix (2,500 nodes, 420 TB, over 1 trillion requests per day);
  - Chinese search engine Easou (270 nodes, 300 TB, over 800 million requests per day);
  - eBay (over 100 nodes, 250 TB).

fontes:

http://cassandra.apache.org/

Durable

- Cassandra is suitable for applications that can't afford to lose data, even when an antire data center goes down;

fontes:

http://cassandra.apache.org/

Elastic

- Read and write throughput both increase linearly as new machines are added, with no downtime or interruption to applications.

fontes:

http://cassandra.apache.org/

CQL

- **Cassandra Query Language (CQL)**

Installation

- Debian like:
  - echo "deb http://www.apache.org/dist/cassandra/debian 310x main" | sudo tee -a /etc/apt/sources.list.d/cassandra.sources.list
  - sudo apt-get install cassandra

- Docker
  - sudo docker run --name instance-name -d cassandra:3.10 (stable)

fontes:

https://hub.docker.com/_/cassandra/

http://cassandra.apache.org/download/

Next Steps

- Testar com uma aplicação.

@fernandofagonde