



TWITTER

Frederico dos Santos, Gabriel de Souza, Gabriel Lima de
Souza, Nikolas Louret



SUMÁRIO

01

CONTEXTO

02

ARQUITETURA

03

TECNOLOGIAS

04

HADOOP



01

CONTEXTO

Introdução

- O twitter é uma das maiores redes sociais do mundo
- Arquitetura robusta e altamente distribuída
- 5,8 bilhões de visitas mensais
- Segmentação geográfica: liderança nos EUA.
- Predominância de usuários masculinos, faixa etária de 18 a 34 anos.

Desafios técnicos

- Escala: plataforma com bilhões de tweets e usuários
- Latência: fornecer atualizações em tempo real para milhões de usuários simultâneos
- Consistência: consistência dos dados em um ambiente distribuído.
- Segurança: proteger os dados dos usuários contra ameaças
- Conformidade: Cumprir regulamentos de privacidade de dados em várias regiões geográficas.



02

ARQUITETURA

Arquitetura em Alto Nível

- Camada de Aplicação: interface do usuário do Twitter
- Camada de Lógica de Negócios: recursos como publicação de tweets, pesquisa e tendências
- Camada de Armazenamento: inclui diversos sistemas de armazenamento, como bancos de dados distribuídos, caches e sistemas de armazenamento de objetos.
- Camada de Infraestrutura: data centers em várias regiões geográficas, balanceadores de carga, sistemas de monitoramento e escalonamento automático.



03

Tecnologías

Tecnologias



Manhattan

back-end para tweets,
mensagens diretas



Cache

armazenamento em cache
de usuários, tweets



Kafka

mensagens distribuídas



Blobstore

banco de dados para
armazenar grafos de
relacionamento de
seguidores



SQL

MySQL, PostgreSQL e
Vertica



Hadoop

framework de
computação em nuvem



O4

Hadoop

Características do Hadoop

- Uma biblioteca de software que permite o processamento distribuído de grandes conjuntos de dados em clusters de computadores usando modelos de programação simples.
- Projetado para escalar desde servidores únicos até milhares de máquinas, cada uma oferecendo computação e armazenamento locais.
- Projetado para detectar e tratar falhas na camada de aplicação

Como o Hadoop é utilizado

- Hadoop é a infraestrutura do pipeline de log baseado em Scribe.
- Aborda questões como falta de limitação de taxa para agregadores seletivos.
- Soluciona desafios de garantia de entrega e corrupção de memória.
- Lida com mais de um trilhão de mensagens diariamente.
- Realiza consolidação e cópia seletiva em todos os clusters.
- Demonstra a capacidade de lidar com grande volume de dados de forma eficiente.



OBRIGADO