Recapitular o que foi passado até agora.

Usar Slides dos desafios do big data e agora focar em real time analytics

Real time Big Data Analytics

“Usuários finais podem aproveitar das capacidades analíticas cada vez mais sofisticadas através de real-time analytics embarcado em ferramentas de descoberta de dados e aplicações com capacidades de processamento em tempo real e sem necessidade de desenvolvedores intervirem.”

W. Roy Schulte, Analista do Gartner

Real time Big Data Analytics – Desafios

Escalabilidade – Bancos de dados tradicionais não escalam bem.

Schema de dados flexível – Dificuldades de manipulação de fontes de dados heterogêneas e dados não estruturados.

Tempo de ingestão e preparo dos dados – Sistemas tradicionais operam com janelas de tempo pré-definidas em dia-1 ou n-horas de atraso.

Casos de uso – Realtime Big Data Analytics

Na Black Friday de 2016, a Sprint presenciou uma baixa significativa no acesso ao sprint.com. Anteriormente, teria levado horas para investigar manualmente os dados para identificar a causa raiz e consertar o problema. Ao invés disso, o time de TI foi alertado imediatamente, identificando o problema e rapidamente corrigindo a falha.

Esse tipo de agilidade é crucial para a empresa vendar mais celulares.

3 bilhões de eventos por dia de logs, bancos de dados, e-mails, logs de sistema, mensagem de texto e APIs internas e externas.

<https://www.elastic.co/use-cases/sprint>

ELK Stack

O que é :

Uma coleção de três produtos open source – Elasticsearch, Logstash e Kibana – mantidos pela Elastic.

Elasticsearch : NoSQL database baseado no Apache Lucene (motor de busca)

Logstash : ferramenta de pipeline de dados com conexão a vários tipos de fontes, executando transformações e exportando dados para vários destinos.

Kibana : Camada de visualização que funciona sobre o Elasticsearch

<https://logz.io/wp-content/uploads/2018/08/image21-1024x328.png>

Produtos adicionais – Canvas

Ferramenta de visualização com capacidade de ajuste fino via CSS e criação de novos componentes

<https://www.elastic.co/assets/blte82ce6dcd493cdb4/business-analytics-canvas-animation.gif>

ELK - Benefícios

* Open Source,
* Comunidade ativa,
* Altamente escalável e distribuído,
* Schemaless data model,
* APIs de fácil acesso

Utilização

Social media Analytics

<https://www>.elastic.co/assets/blte771400b1f9a444b/business-analytics-sentiment-analysis.png

Serviços Financeiros

<https://www>.elastic.co/assets/bltf73b414d490af1f0/business-analytics-financial-services.png

Varejo

<https://www>.elastic.co/assets/bltea8c5544a2e677e3/business-analytics-retail-ops.png

Web Analytics

<https://www.elastic.co/assets/blt9da3b921073c9dae/business-analytics-web-analytics.png>

ELK – o que não é

Não é banco de dados – armazenamento temporário de dados por n-dias

Não há garantia de consistência (ACID)

Otimizado para consulta, não alterações

Outras opções

Logz.io

Splunk

MongoDB + aplicação de visualização

Bancos de dados em memória (ex.: SAP Hana)

Trabalho prático – Dados de eventos/Sensor

Dados : <http://www.datapoa.com.br/dataset/poatransporte>

Dados sobre transporte público via API REST + enriquecimento

Dados de sensores (simulação) – REAL TIME

Elastic Conceitos

https://cdn-images-1.medium.com/max/800/1\*0gUMUCd81Oxu-npn-NZcTw.png

1. Instalação do Elasticsearch e Kibana
   1. <https://www.elastic.co/downloads/elasticsearch>
   2. <https://www.elastic.co/downloads/kibana>
   3. Pré requisitos
      1. Java instalado (java –version)
      2. JAVA\_HOME configurado
2. Configuração das ferramentas
   1. Elasticsearch.yml
   2. Kibana.yml
3. Com o Elastic Executando
   1. Criação de índices
   2. Mappings
   3. Ingestão de dados via python