# INFO 611 Fall 2020

## Week 10, Oct 27, 2020

Elasticsearch - Fun Tuesday

Elasticsearch for SQL Experts

chrisfauerbach.github.io/info610_fall_2020/

# Elasticsearch

https://www.elastic.co/what-is/elasticsearch

Elasticsearch is a distributed, open source search and analytics engine for all types of data, including textual, numerical, geospatial, structured, and unstructured.

Elasticsearch is built on Apache Lucene and was first released in 2010 by Elasticsearch N.V. (now known as Elastic).

Known for its simple REST APIs, distributed nature, speed, and scalability, Elasticsearch is the central component of the Elastic Stack, a set of open source tools for data ingestion, enrichment, storage, analysis, and visualization.

Commonly referred to as the ELK Stack (after Elasticsearch, Logstash, and Kibana), the Elastic Stack now includes a rich collection of lightweight shipping agents known as Beats for sending data to Elasticsearch.

# Elasticsearch - When do you use it?

The speed and scalability of Elasticsearch and its ability to index many types of content mean that it can be used for a number of use cases:

- Application search
- Website search
- Enterprise search
- Logging and log analytics
- Infrastructure metrics and container monitoring
- Application performance monitoring
- Geospatial data analysis and visualization
- Security analytics
- Business analytics

# How does it work

Elasticsearch is built on Lucene, the de facto standard Java Full Text Search Library.

The basic flow of data into Elasticsearch is as follows:

1. Raw data flows to a collector (beats, logstash)
2. Data is transformed and modified to a standard format
3. Data is ingested into Elasticsearch
4. Data is mapped to a specific data type (Document Type)
5. Data is then persisted in ES
6. Data is organized (Indexes)
7. Data is then made redundant for safety/scaling (Shards)

# Index

An Index in ES is a collection of documents. Typically of the same type, but not required.

Indexes can be named and organized by convention to represent things like time ranges (date, month, year) etc.

Elasticsearch uses a data structure called an inverted index, which is designed to allow very fast full-text searches. An inverted index lists every unique word that appears in any document and identifies all of the documents each word occurs in.

# Kibana

Kibana is a user interface that can show information about Elasticsearch.

Kibana is a data visualization and management tool for Elasticsearch that provides real-time histograms, line graphs, pie charts, and maps. Kibana also includes advanced applications such as Canvas, which allows users to create custom dynamic infographics based on their data, and Elastic Maps for visualizing geospatial data.

# Interaction

Custom application development can occur in almost any programming language and interact with ES. All major programming languages have 'libraries' that allow easier integration.

In the end, primary integration occurs with REST APIs over HTTP.

# DEMO

# See you Thursday!