

Executing Analytical Queries Through Aggregations



Janani Ravi

CO-FOUNDER, LOONYCORN

www.loonycorn.com

Overview

Learn how Elasticsearch can be used beyond search as an analytical engine

Know the different kinds of aggregations that can be performed

Implement queries for metrics and bucketing aggregations

Work with multi-level nesting of aggregations

Aggregations

Four Types of Aggregations

Metric

Bucketing

Matrix

Pipeline

Metric

Metric

Aggregations over a set of documents

- All documents in a search result
- Documents within a logical group

Bucketing

Bucketing

Logically group documents based on search query

A document falls into a bucket if the **criteria matches**

Each bucket associated with a **key**

Matrix

Matrix

Operates on multiple fields and produces a matrix result

Experimental and may change in future releases

Not covered in this course

Pipeline

Pipeline

Aggregations that work on the output of other aggregations

Experimental and may change in future releases

Not covered in this course

Four Types of Aggregations

Metric

Bucketing

Matrix

Pipeline

Four Types of Aggregations

Metric

Bucketing

Matrix

Pipeline

Demo

Metric aggregations such as avg, and the multi-value stats aggregation

Demo

Cardinality, the number of unique values in a field

Enable `fielddata` for text fields via mappings

Search vs. Aggregations

Search

**Inverted index of the terms
present in documents**

**The terms themselves can be
hashed and stored in the
index**

**“Which documents contain
this term?”**

Aggregation

**Actual value of fields present in
documents**

**Actual values of the terms are
needed, hash values do not
suffice**

**“What is the value of this field
for this document?”**

Getting the Value of a Text Field



Text field values are stored in an **in-memory** data structure called **fielddata**

Getting the Value of a Text Field



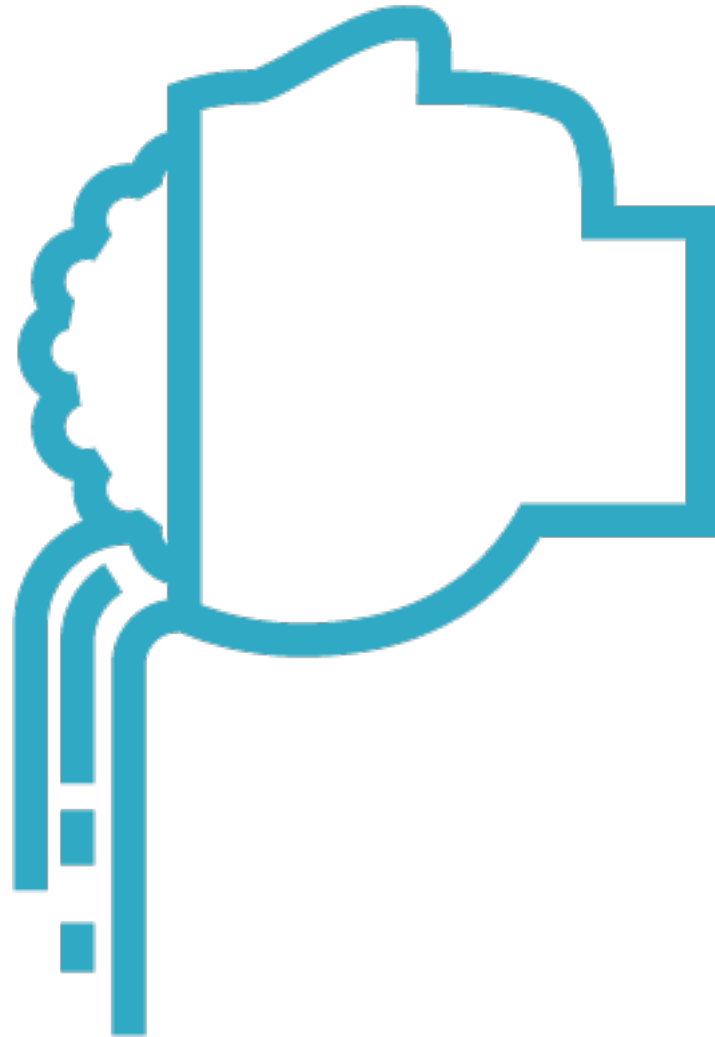
`fielddata` is built **on demand** when a field is used for aggregations, sorting etc

Getting the Value of a Text Field



`fielddata` on text fields take up lots of
heap space

Getting the Value of a Text Field



`fielddata` is disabled by default on text fields

Four Types of Aggregations

Metric

Bucketing

Matrix

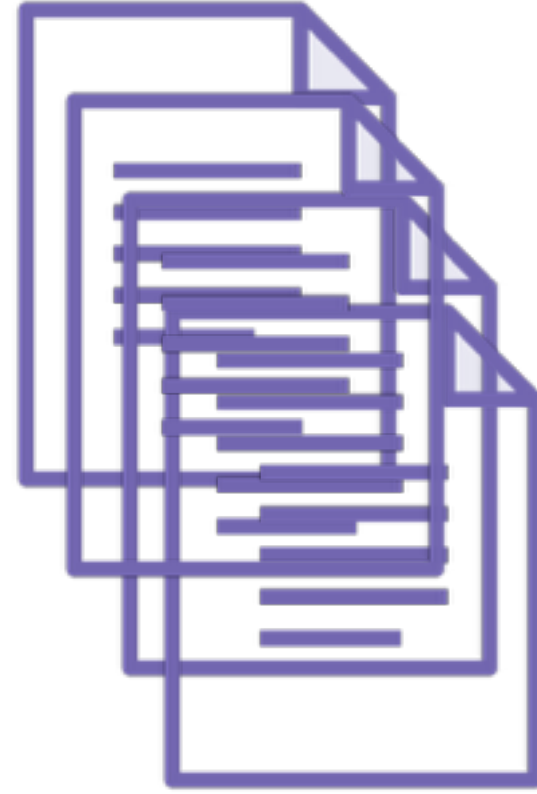
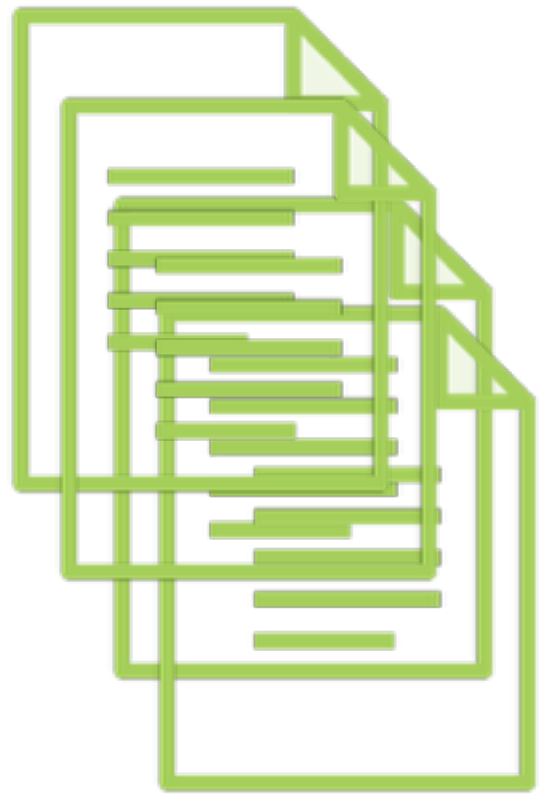
Pipeline

Indexed Documents



**Logically group these
documents into buckets**

Types



**Each bucket satisfies some
criterion**

Demo

Bucket aggregations by field values

Demo

Multi-level nested aggregations

Demo

The “filter” aggregation to filter results

The “filters” aggregation to specify multiple filter matches

Summary

Perform analytics queries on Elasticsearch

Metric, bucketing, matrix and pipeline aggregations

Implemented queries for metrics and bucketing aggregations

Worked with multi-level nesting of aggregations