

# Data Recipes

Dean Hunter  
Msc Big Data



## Problem & Aim

Most code sharing sites have no focus on particular topics, making navigation for a particular type of sample difficult. Big Data technologies also lack in example code.

This project aims to deliver a code sharing website specifically for Big Data samples, focusing on delivering an intuitive design to quickly navigate and locate relevant samples and content.

The site offers a number of unique features to provide users with an intuitive way to navigate all relevant examples.

## Latest uploads

The latest section allows users to navigate all recently uploaded samples and rate each sample based on its value.

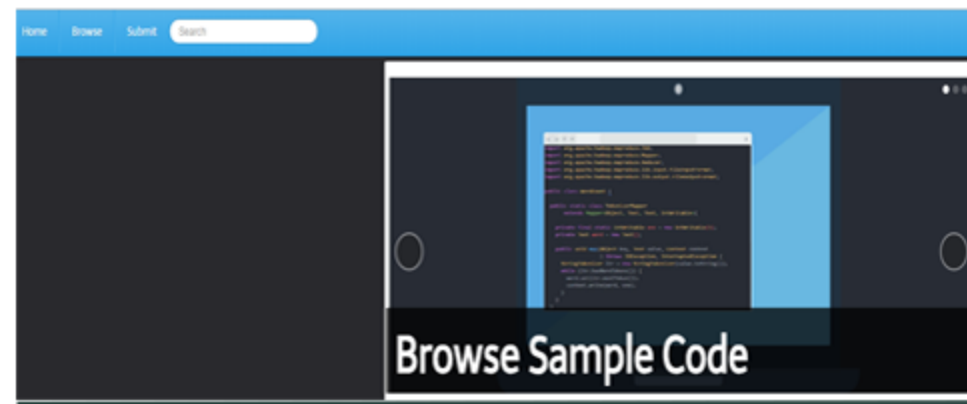
Each sample contains content dynamically generated by the back-end, allowing users to find samples that share key features. This directs users to relevant examples that may be more helpful than their current search.

## Front-End



## Simple and Intuitive

The Data recipes site uses a modified version of Twitter-Bootstrap to provide a robust, responsive and stable user experience.



## Code Editor

Upload a file or use the ACEJS code editor to paste your code, generate suggested client side tags and search for similar samples.

```

1 public class Processunits
2 {
3     //Mapper class
4     public static class E_BMapper extends MapReduceBase implements
5     Mapper<LongWritable, <input key type />
6     Text, <input value type />
7     Text, <output key type />
8     IntWritable> <output value type />
9     {
10
11     //map function
12     public void map(LongWritable key, Text value,
13     OutputCollector<Text, IntWritable> output,
14     Reporter reporter) throws IOException
15     {
16         String line = value.toString();
17         String lasttoken = null;
18         StringTokenizer s = new StringTokenizer(line, "\t");
19         String year = s.nextToken();
20
21         while(s.hasMoreTokens())

```

## Suggested Samples

Unlike other code sharing sites, Data recipes will recommend other samples based on analytics to help solve users' Big Data questions, reducing time spent manually searching.

Similar:

- Reducer
- Hadoop Reduce
- core.reducer

## Back-End

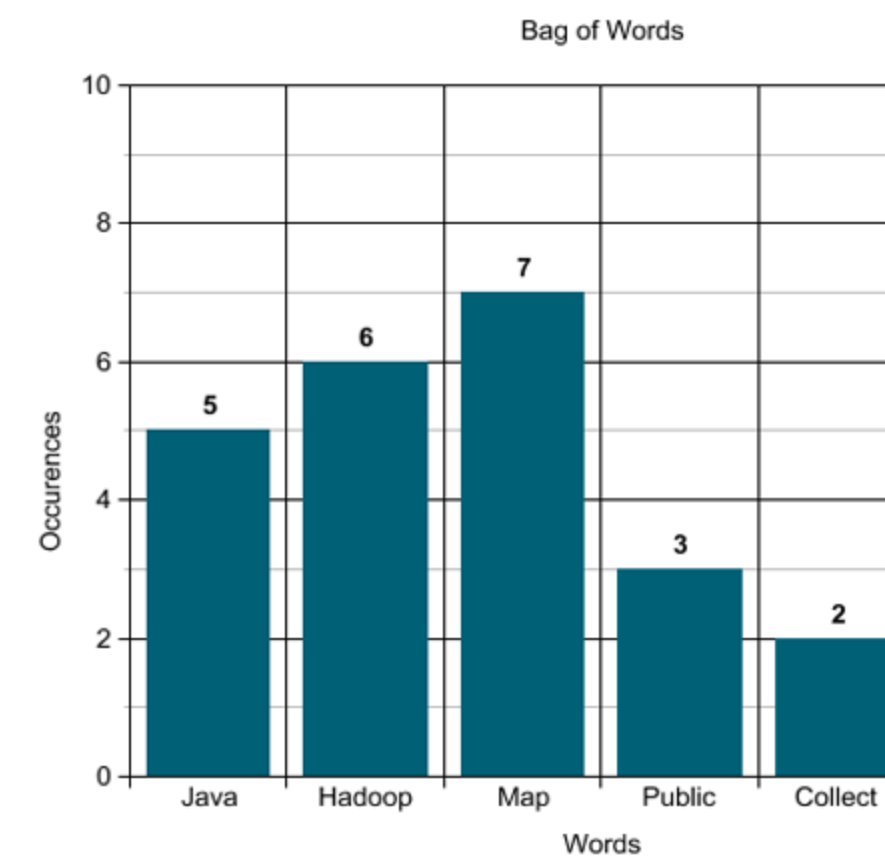


## Suggest and Vote

Each sample uploaded is processed by both the Front-end and Back-end, ensuring enough data is collected to provide a fluid and relevant navigation experience.

## Bag of Words

Both Front-end and Back-end utilize a bag of words model to provide realtime suggestions to end users both in tagging and uploaded project descriptions to ensure results are as accurate as possible.



## MongoDB

Each sample uploaded is stored in MongoDB as a document in a collection. MongoDB provides a robust schema-less document model and deep query-ability using a JSON like format. This provides Data recipes with scalability and dynamic searching capabilities.

# UNIVERSITY of STIRLING



## Tagging

Each submission is tagged either manually or automatically by the back-end after parsing the uploaded content.

## Rating

Rate other submitted samples using a score based system to promote the sharing of useful and relevant samples and aid navigation through the site.

