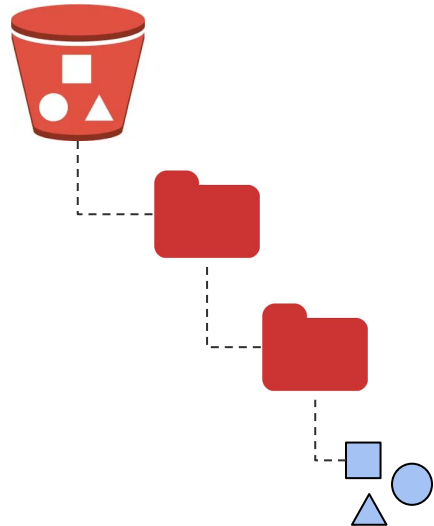


Technology Crash Course



S3 = Simple Storage Service



Bucket. *Name:* **data.mycompany.com** *Prefix:* **s3://data.mycompany.com**

Key. *Name:* **users** *Prefix:* **s3://data.mycompany.com/users**

Key. *Name:* **users/2018** *Prefix:* **s3://data.mycompany.com/users/2018**

Object. *Name:* **data.json** *Prefix:* **s3://data.mycompany.com/users/2018/data.json**

S3 = Simple Storage Service

Amazon S3 > [data.public.bdatainstitute.com](#) / [dam18](#) / [query_logs](#) / [2018](#) / [09](#)

2018_09_18452212_log.json [Latest version](#) ▼

Overview **Properties** **Permissions** **Select from**

[Open](#) [Download](#) [Download as](#) [Make public](#) [Copy path](#)

Owner
szota.rafal

Last modified
Nov 5, 2018 11:22:08 PM GMT+0100

Etag
309bffc483379fdbd6376c69ef8950d

Storage class
Standard

Server-side encryption
None

Size
232922

Link
https://s3-eu-west-1.amazonaws.com/data.public.bdatainstitute.com/dam18/query_logs/2018/09/2018_09_18452212_log.json

https://s3-eu-west-1.amazonaws.com/data.public.bdatainstitute.com/dam18/query_logs/2018/09/2018_09_18452212_log.json

AWS Command Line Interface

<https://docs.aws.amazon.com/cli/latest/userguide/cli-chap-welcome.html>

```
$ aws configure
```

```
AWS Access Key ID [None]: AKIAIOSFODNN7EXAMPLE
```

```
AWS Secret Access Key [None]: wJalrXUtnFEMI/K7MDENG/bPxRfiCYEXAMPLEKEY
```

```
Default region name [None]: eu-west-1
```

```
Default output format [None]: ENTER
```

```
$ aws s3 ls s3://data.public.bdatainstitute.com
```



AWS Service



command

<https://docs.aws.amazon.com/cli/latest/reference/s3/>

AWS CLI Documentation

Assignment 1: Get familiar with S3.

Task:

- **Online Travel Company** uses S3 to store statistics on successful and failed searches data for the hotel reservation. As a data analyst you will help the Head of Product.
- Head of Product wants to know: **What is the estimated monthly cost, the company needs to pay for storing this data?**
- You should answer the question and prepare a brief summary that justify your answer.

Useful context:

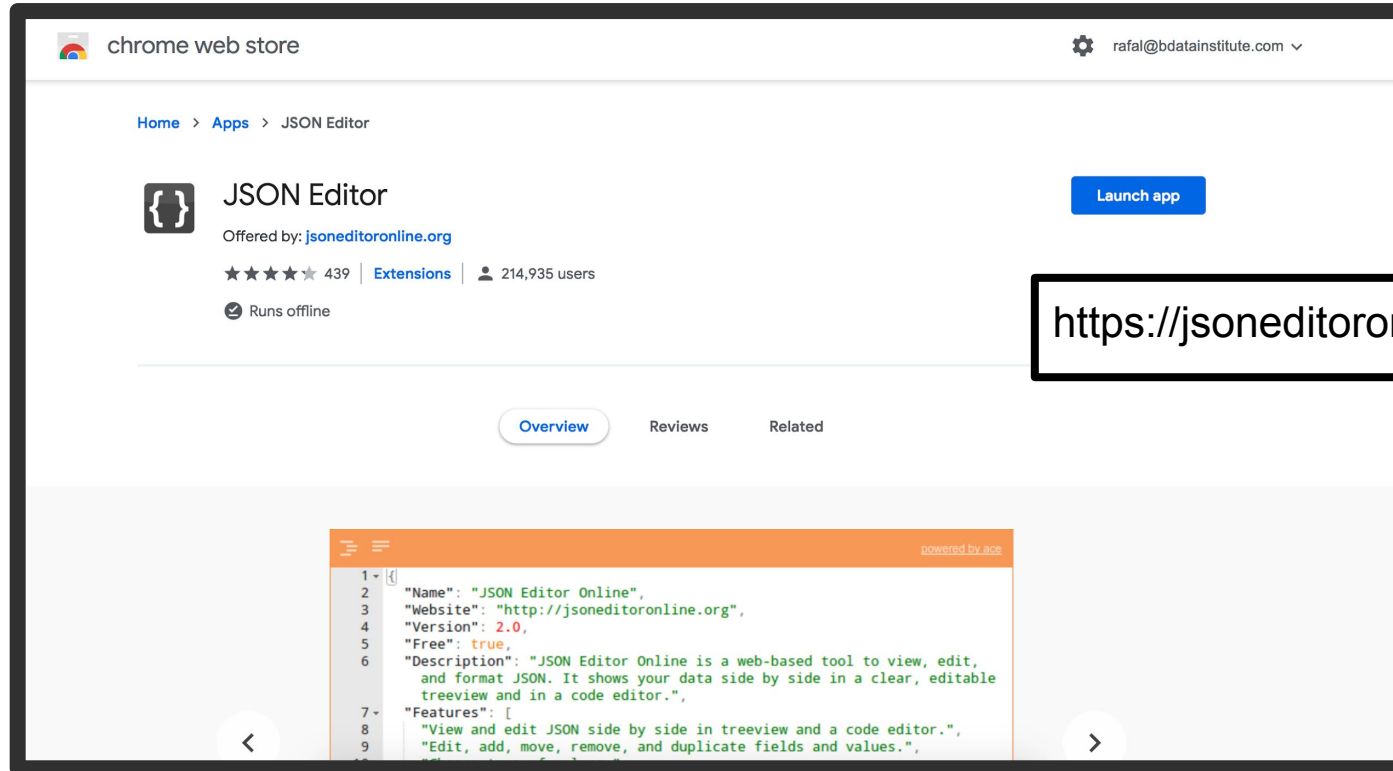
- The company operates approximately with 5000 different destinations and 5 years (60 months) of data.
- Searches are stored by location and time. Every location and month is represented by a file similar to :

`s3://data.public.bdatainstitute.com/dam18/search_logs/2018/09/2018_09_18452212_log.json`

(filename contains: Year, Month and ServiceID)

{ JSON }

JSON Editor Online / Chrome Plugin



The screenshot shows the Chrome Web Store page for the 'JSON Editor' extension. The page header includes the 'chrome web store' logo and a user profile 'rafal@bdatainstitute.com'. The breadcrumb trail is 'Home > Apps > JSON Editor'. The extension's icon is a black square with a white curly brace. The title 'JSON Editor' is prominently displayed, followed by 'Offered by: jsoneditoronline.org'. It has a rating of 439 stars and 214,935 users, and a 'Runs offline' status. A blue 'Launch app' button is located to the right. Below the extension details are tabs for 'Overview', 'Reviews', and 'Related'. The 'Overview' tab is active, showing a preview of the extension's interface. The preview displays a code editor with a JSON object:


```
{  "Name": "JSON Editor Online",  "Website": "http://jsoneditoronline.org",  "Version": 2.0,  "Free": true,  "Description": "JSON Editor Online is a web-based tool to view, edit, and format JSON. It shows your data side by side in a clear, editable treeview and in a code editor.",  "Features": [    "View and edit JSON side by side in treeview and a code editor.",    "Edit, add, move, remove, and duplicate fields and values."  ]}
```

 The interface is powered by ace, as indicated in the top right corner of the preview window.

chrome web store

rafal@bdatainstitute.com

Home > Apps > JSON Editor

 **JSON Editor**

Offered by: jsoneditoronline.org

★★★★★ 439 | [Extensions](#) | 214,935 users

🔒 Runs offline

[Launch app](#)

[Overview](#) | [Reviews](#) | [Related](#)

```
1 {  2   "Name": "JSON Editor Online",  3   "Website": "http://jsoneditoronline.org",  4   "Version": 2.0,  5   "Free": true,  6   "Description": "JSON Editor Online is a web-based tool to view, edit, and format JSON. It shows your data side by side in a clear, editable treeview and in a code editor.",  7   "Features": [  8     "View and edit JSON side by side in treeview and a code editor.",  9     "Edit, add, move, remove, and duplicate fields and values."  ]}
```

powered by ace

<https://jsoneditoronline.org/>

JSON Path Tester

<https://codebeautify.org/jsonpath-tester>

JSON Path Example

`$.health_checks[*].timestamp`

**What is the world's most common
data analytics tool?**

JSON → CSV ?

Why?

<http://www.convertcsv.com/json-to-csv.htm>

Classroom Assignment: Analyze search data.

Task:

- Head of Product approaches you urgency asking:
- **What time of a day the people search for Barcelona most?**
Morning, Afternoon, Evening, Night? What is the most “busy” hour for searches?
- The answer for the question, should be in some visual form (summary table, plot, etc) so that you can use it during a meeting.

Useful context:

- ID for Barcelona destination = 18452212
- Answer the question with the most recent data available (September)





git

vs



GitHub

<https://git-scm.com/download/>

Install **git** on your computer

<https://github.com/>

Create GitHub user


Create a new repository

Create a new repository

A repository contains all the files for your project, including the revision history.

Owner

Repository name


 rafalszota ▾

 /


dam18 ✓

Great repository names are short and memorable. Need inspiration? How about **legendary-octo-rotary-phone**.

Description (optional)

☒  **Public**

Anyone can see this repository. You choose who can commit.

☐  **Private**

You choose who can see and commit to this repository.

☒ **Initialize this repository with a README**

This will let you immediately clone the repository to your computer. Skip this step if you're importing an existing repository.

Add .gitignore: **None** ▾

Add a license: **None** ▾ ⓘ

Create repository

```
$ git clone https://github.com/rafalszota/dam18.git
```

```
$ git add myfirstfile.txt
```

```
$ git commit -m "My first commit"
```

```
$ git push
```

In case of fire



1. git commit



2. git push



3. leave the building

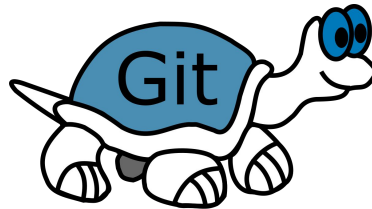
Assignment 2: Visualize a Git repository.

Context:

Working with git in command line might be very inefficient. There are multiple tools, which facilitate usage of multiple repositories.

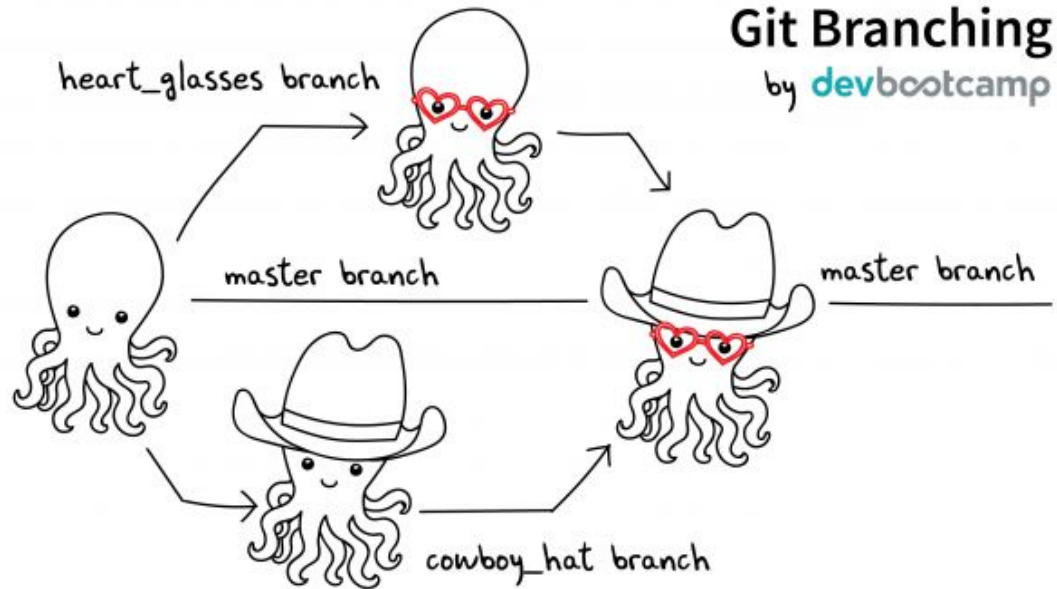
Task:

- Explore and get familiar with one of the popular git clients. Install and configure the tool of your choice. For example (<https://git-scm.com/downloads/guis>):



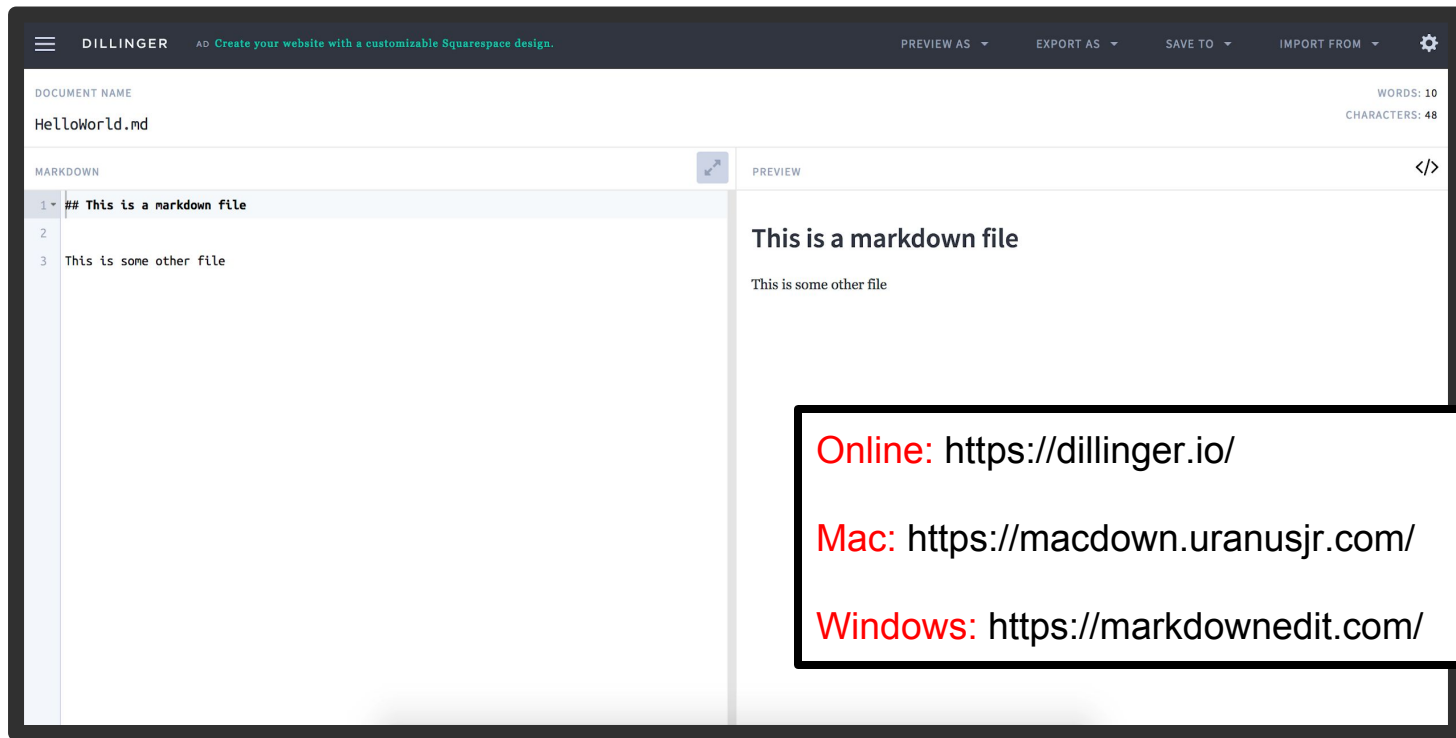
Optional Assignment:

Get familiar with **Branch** and **Merge**





Classroom Assignment: Write .md report



Assignment: Decision making

Task:

- Clients complain that the searches for a destinations sometimes fail. Head of product decided to address the issue, and ask development team to work on fix.
- The team committed to work out the solution during August. It was agreed that the team's bonus payout would depend on effectiveness of the solution.
- **The Head of Product ask you to analyze the data and help him to decide whether the team deserve the bonus?**
- Your answer should be documented and submitted to the GitHub so that any changes can be easily tracked.

Useful context:

- Data for Barcelona destination (18452212) is representative so that any conclusions can be extrapolated for other destinations.

Assignment: Collaborative work

Task:

- Since bonus payout is a very sensitive decision, the Head of Product asked similar question to several analysts. In this exercise you will collaborate with your peers and enhance their findings.
- **Fork** and **Clone** repository of any member from your group. You should enhance his/her analysis with your own findings, building on top of the report.
- Once finished - **Commit** and **Push** improved results to your own repository.