



Publishing Dataset & Versions

Hasir Mushtaq - 2018102049

Mohsin Mamoon - 2018101029

Zia Bashir - 2018101006



Topics to be Covered

Project Objective

Upload Dataset

Issues worked on

Comment Section

Home Page

Individual Contributions

Review Dataset



Project objective

The project involves building a web application that will provide an interface for publishing datasets and updating their versions. Additional functionality includes the update and deletion of uploaded datasets. Further, the web application also allows consumers to raise comments or flags which the owners can review and edit as desired.

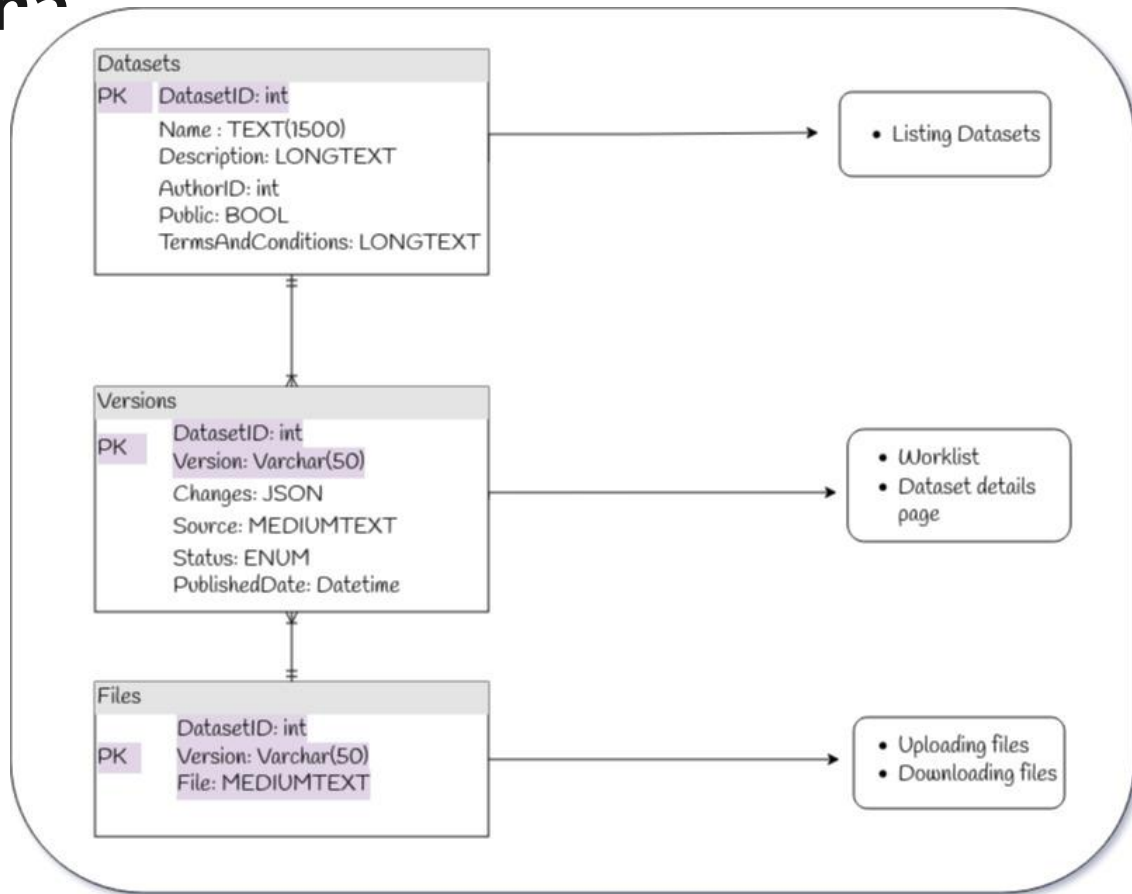


Issues worked on

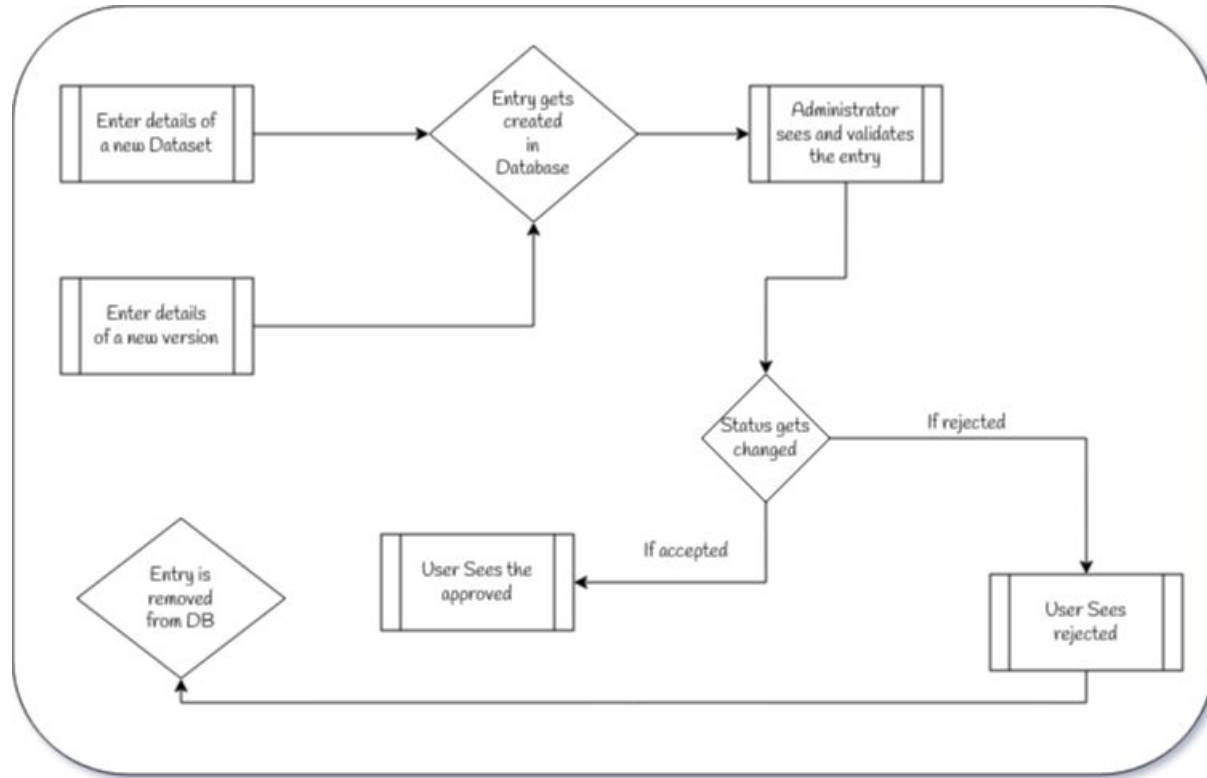
1. Comments Section UI
2. Upload Dataset UI
3. Review Dataset UI
4. Home Page UI
5. Individual Components involved in all the above.
6. Backend to integrate everything together
7. Api development for integration with other teams






Database Schema



Upload Workflow



Home Page

[Upload](#)[Review](#)

Ready to contribute?

Upload a new dataset or Review an existing dataset.

[Upload a Dataset.](#)[Review a Dataset.](#)

DATASET ID	DATASET VERSION	DATASET NAME	DATASET DESCRIPTION	STATUS	PUBLISHED DATE
39	1.0.0	COVID-19 Dataset (G)	A new coronavirus designated 2019-nCoV was first identified in Wuhan, the capital of Chinas Hubei province People developed pneumonia without a clear cause and for which existing vaccines or treatments were not effective. The virus has shown evidence of human-to-human transmission Transmission rate (rate of infection) appeared to escalate in mid-January 2020 As of 30 January 2020, approximately 8,243 cases have been confirmed	APPROVED	2022-05-01T00:11:48.000Z
39	2.0.0	COVID-19 Dataset (G)	A new coronavirus designated 2019-nCoV was first identified in Wuhan, the capital of Chinas Hubei province People developed pneumonia without a clear cause and for which existing vaccines or treatments were not effective. The virus has shown evidence of human-to-human transmission Transmission rate (rate of infection) appeared to escalate in mid-January 2020 As of 30 January 2020, approximately 8,243 cases have been confirmed	PENDING	2022-05-01T00:25:25.000Z
40	1.0.0	IDD Segmentation-part 1	10,000 images and fine semantic segmentation annotation; released in 2018	PENDING	2022-05-01T00:54:28.000Z
41	1.0.0	IDD Segmentation-part 2	20,000 images and fine semantic segmentation annotation (14K Train, 2K Val, 4K Test) from 350 drive sequences, in 2 part downloads - this data and the data released in 2018 as IDD - Segmentation (IDD 20k Part I)	REJECTED	2022-05-01T00:55:46.000Z
42	1.0.0	IDD Temporal Train	The IDD temporal dataset consists of the temporally nearby frames (+- 15 frames) from the IDD Segmentation and Detection Datasets. The 30 nearby frames for many of the frames in the IDD Segmentation dataset, is provided in the correspondingly named subfolders in the IDD temporal dataset.	APPROVED	2022-05-01T00:59:07.000Z

Review Dataset

Dataset Information

Details about the dataset.

Dataset Name	COVID-19 Dataset (G)		
Dataset Identifier	39		
Source URL	www.kaggle.com/datasets/mdevskp/corona-virus-report		
Dataset Description	A new coronavirus designated 2019-nCoV was first identified in Wuhan, the capital of Chinas Hubei province. People developed pneumonia without a clear cause and for which existing vaccines or treatments were not effective. The virus has shown evidence of human-to-human transmission. Transmission rate (rate of infection) appeared to escalate in mid-January 2020. As of 30 January 2020, approximately 8,243 cases have been confirmed.		
Terms and Conditions	The Service provided by the Owner allows Users to run a series of technological tools in a self-service manner, allowing them to generate legal documents based on Users' input for their online activities (eg. website, mobile app, etc.). As part of its Service, the Owner allows Users to generate, host and keep one or more document templates up to date online. Under no circumstance will ibenda's staff or any counsel assist Users in making the correct choice or in drafting the correct custom clauses. It is therefore the Users' sole responsibility to pick the correct choices for their respective scenario or activity, to verify compatibility of the generated documents with applicable law and to ensure that their activity complies with all applicable legal provisions. In light of the above, the Service offered by ibenda cannot be regarded as, nor does it substitute any legal advice given by a professional or expert.		
Dataset Availability	Public		
DATASET VERSION	CHANGES	FILES	PUBLISHED DATE
1.0.0	1. Initial version	<div> Gandhi_Data-20210427T035504Z-001493f2f90.zip </div>	2022-05-01T00:11:48.000Z Edit
2.0.0	1. Sepearted Oct and Dec Versions	<div> 42.zip </div>	2022-05-01T00:25:25.000Z Edit
	2. Added more data	<div> covid19_Dec.zip </div>	
		<div> covid19_Oct.zip </div>	

Upload Dataset



Upload

Review



Upload a New Dataset

Please input the details about the dataset.

Name for the Dataset

Source URL

Version

Major

Minor

Patch

Privacy

☒ Public

☐ Private

Description of the Dataset

Enter a brief description of the dataset you wish to upload.

Terms and Conditions

Enter the terms and conditions for the usage of the dataset.

Dataset files


Upload a file or drag and drop
ZIP, TAR, upto 100GBs



Selected Files

No Files selected!

Upload

Comments Section

UploadReview



Add a comment

Comment

User #1729

Lorem ipsum is simply dummy text of the printing and typesetting industry. Lorem ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem ipsum.

ReplyDelete

Admin

Lorem ipsum is simply dummy text of the printing and typesetting industry. Lorem ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem ipsum.

User #1729

Lorem ipsum is simply dummy text of the printing and typesetting industry. Lorem ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem ipsum.

ReplyDelete

Admin

Lorem ipsum is simply dummy text of the printing and typesetting industry. Lorem ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem ipsum.

User #1729

Lorem ipsum is simply dummy text of the printing and typesetting industry. Lorem ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem ipsum.

ReplyDelete

Admin

localhost:3000/?comment=*



Project objective

Live Demo

APIs

```
const sql = require('mysql');

var con = sql.createConnection({
  host: "localhost",
  user: "mohsin",
  password: "password",
  database: "dfs"
});

var usersTable = `
CREATE TABLE users(
  UserID int NOT NULL AUTO_INCREMENT,
  Name MEDIUMTEXT NOT NULL,
  PRIMARY KEY (UserID)
);

con.query(usersTable, (err, res) => {
  if (err) throw err;
  console.log("Users Table created!");
});

var datasetsTable = `
CREATE TABLE datasets(
  DatasetID int NOT NULL AUTO_INCREMENT,
  Name TEXT(1500) NOT NULL,
  Description LONGTEXT NOT NULL,
  AuthorID int NOT NULL,
  Public BOOL NOT NULL,
  TnC LONGTEXT,
  Source MEDIUMTEXT,
  PRIMARY KEY (DatasetID),
  FOREIGN KEY (AuthorID) REFERENCES users(UserID)
);

con.query(datasetsTable, (err, res) => {
  if (err) throw err;
  console.log("Datasets Table created!");
});

var versionsTable = `
CREATE TABLE versions(
  DatasetID int NOT NULL,
  Version varchar(50) NOT NULL,
  Changes JSON,
  Status ENUM("PENDING","APPROVED","REJECTED") DEFAULT "PENDING",
  Published Datetime,
  PRIMARY KEY (DatasetID, Version),
  FOREIGN KEY (DatasetID) REFERENCES datasets(DatasetID)
);
```

```
con.query(versionsTable, (err, res) => {
  if (err) throw err;
  console.log("Versions Table created!");
});

var filesTable = `
CREATE TABLE files(
  DatasetID int NOT NULL,
  Version varchar(50) NOT NULL,
  Filename varchar(256) NOT NULL,
  PRIMARY KEY (DatasetID, Version, Filename),
  FOREIGN KEY (DatasetID, Version) REFERENCES versions(DatasetID, Version)
);

con.query(filesTable, (err, res) => {
  if (err) throw err;
  console.log("Files Table created!");
});

var commentsTable = `
CREATE TABLE comments(
  DatasetID int NOT NULL,
  CommentID int NOT NULL AUTO_INCREMENT,
  Comment LONGTEXT NOT NULL,
  AuthorID int NOT NULL,
  ReplyID int,
  PRIMARY KEY (CommentID),
  FOREIGN KEY (DatasetID) REFERENCES datasets(DatasetID),
  FOREIGN KEY (ReplyID) REFERENCES comments(CommentID),
  FOREIGN KEY (AuthorID) REFERENCES users(UserID)
);

con.query(commentsTable, (err, res) => {
  if (err) throw err;
  console.log("Comments Table Created!");
});
```

APIs

```
router.get('/', (req, res) => {

  console.log(req.query);
  let {DatasetID, AuthorID, Status} = req.query;

  if(DatasetID){
    var resp = {}
    console.log("Get dataset details")
    execSql('SELECT * FROM datasets WHERE DatasetID = ' + DatasetID + ''').then(rslt => {
      resp['dataset'] = rslt[0];
      execSql('SELECT * FROM versions WHERE DatasetID = ' + DatasetID + ''')
        .then(async rslt2 => {
          let versions = rslt2
          for (v of versions) {
            v['files'] = []
            await execSql('SELECT filename as file FROM files WHERE DatasetID = ' + DatasetID + '' AND Version = ' + v["Version"] + ''').then(rslt3 => {
              for (file of rslt3) {
                // console.log(file)
                v['files'].push(file['file'])
              }
            });
          }
          resp['versions'] = versions;
          res.json(resp);
        })
      })
  }

  else if (AuthorID && Status){
    console.log("List all datasets of User AuthorID and Status");
    execSql('SELECT * FROM datasets JOIN versions ON datasets.DatasetID = versions.DatasetID WHERE versions.Status = ' + Status + '' AND datasets.AuthorID = ' + AuthorID + ''').then(rslt => res.json(rslt));
  }

  else if (AuthorID){
    console.log("List all datasets of User userID");
    execSql('SELECT * FROM datasets JOIN versions ON datasets.DatasetID = versions.DatasetID WHERE datasets.AuthorID = ' + AuthorID + ''').then(rslt => res.json(rslt));
  }

  else if (Status){
    console.log("List all datasets with Status");
    execSql('SELECT * FROM datasets JOIN versions ON datasets.DatasetID = versions.DatasetID WHERE versions.Status = ' + Status + ''').then(rslt => res.json(rslt));
  }

  else {
    console.log('List all datasets');
    execSql('SELECT * FROM datasets').then(rslt => res.json(rslt))
  }

});
```



Thank you.

