Project Phase 1

Bikash Roy
br8376@rit.edu
Rochester Institute of Technology
Rochester, NY

Ankit Jain
aj9761@g.rit.edu
Rochester Institute of Technology
Rochester, NY

Rishab Katta
rk4056@g.rit.edu
Rochester Institute of Technology
Rochester, NY

Milind Kamath
mk6715@rit.edu
Rochester Institute of Technology
Rochester, NY

ABSTRACT

A project submission for Big Data CSCI 620, Phase 1 A detailed analysis of the dataset of Daily Historical Stock Prices (1970 - 2018). Which evidently has about 20M records, for which we are going to give a relational model and a program to load the dataset onto tables

CCS CONCEPTS

• **Introduction to Big Data**; • **Daily Historical Stock Prices (1970 - 2018)**; • **Relational Model**; • **With SQL Schema creation**;

KEYWORDS

Relational Model, SQL Schema, Daily Historical Stock Prices (1970 - 2018)

ACM Reference Format:
Bikash Roy, Ankit Jain, Rishab Katta, and Milind Kamath. 2018. Project Phase 1. In *Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Of the data set of Daily Historical Stock Prices (1970 - 2018), we are to create a relational schema and provide the Relational Model and also a python code to load the data set into the table.
Basic Columns involved in our dataset are:
ticker - The symbol for the stock
open - The open price
close - The close price
adj_close - The adjusted close price
low - The low price
high - The high price
volume - The volume
date - The date

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
Woodstock '18, June 03–05, 2018, Woodstock, NY
© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

2 DESCRIPTION OF THE CONTENTS OF THE FILES

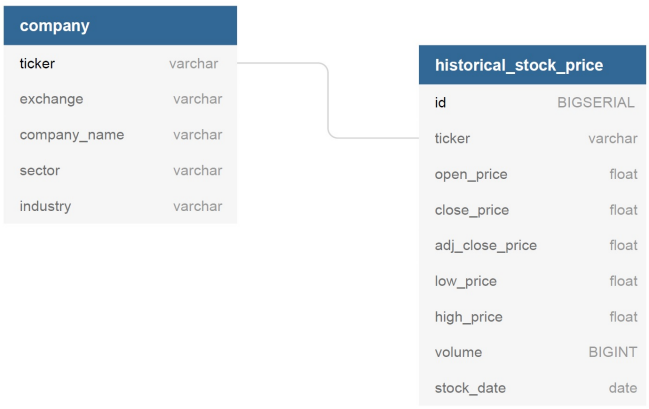
historical_stock_prices.csv

This file has information of every stock being sold, its high and low, its closing and opening price also the closing adjustments, recorded to date.

historical_stocks.csv

This file has information about the stocks of each company, and their dealers(Exchange), along with the sector to which the stock belong and also the industry it represents.

3 RELATIONAL MODEL



Design Decision

As by the current dataset, it is clear that ticker be stock identity, which will the primary key to historical_stock_price table, and a foreign key to company table.

4 SQL SCRIPTS

```
CREATE TABLE "company" (
"ticker" varchar PRIMARY KEY,
"exchange" varchar,
"company_name" varchar,
"sector" varchar,
"industry" varchar
);


CREATE TABLE "historical_stock_price" (
"id" BIGSERIAL PRIMARY KEY,
"ticker" varchar REFERENCES "company" ("ticker"),
"open_price" float,
"close_price" float,
"adj_close_price" float,
"low_price" float,
"high_price" float,
"volume" BIGINT,
"stock_date" date
);
```

5 PROGRAM TO LOAD THE DATA FROM THE TEXT FILES INTO THE DATABASE

Program file enclosed