

Data Analysis of Historical Stock Data



TEAM#4

Fadi Tabet

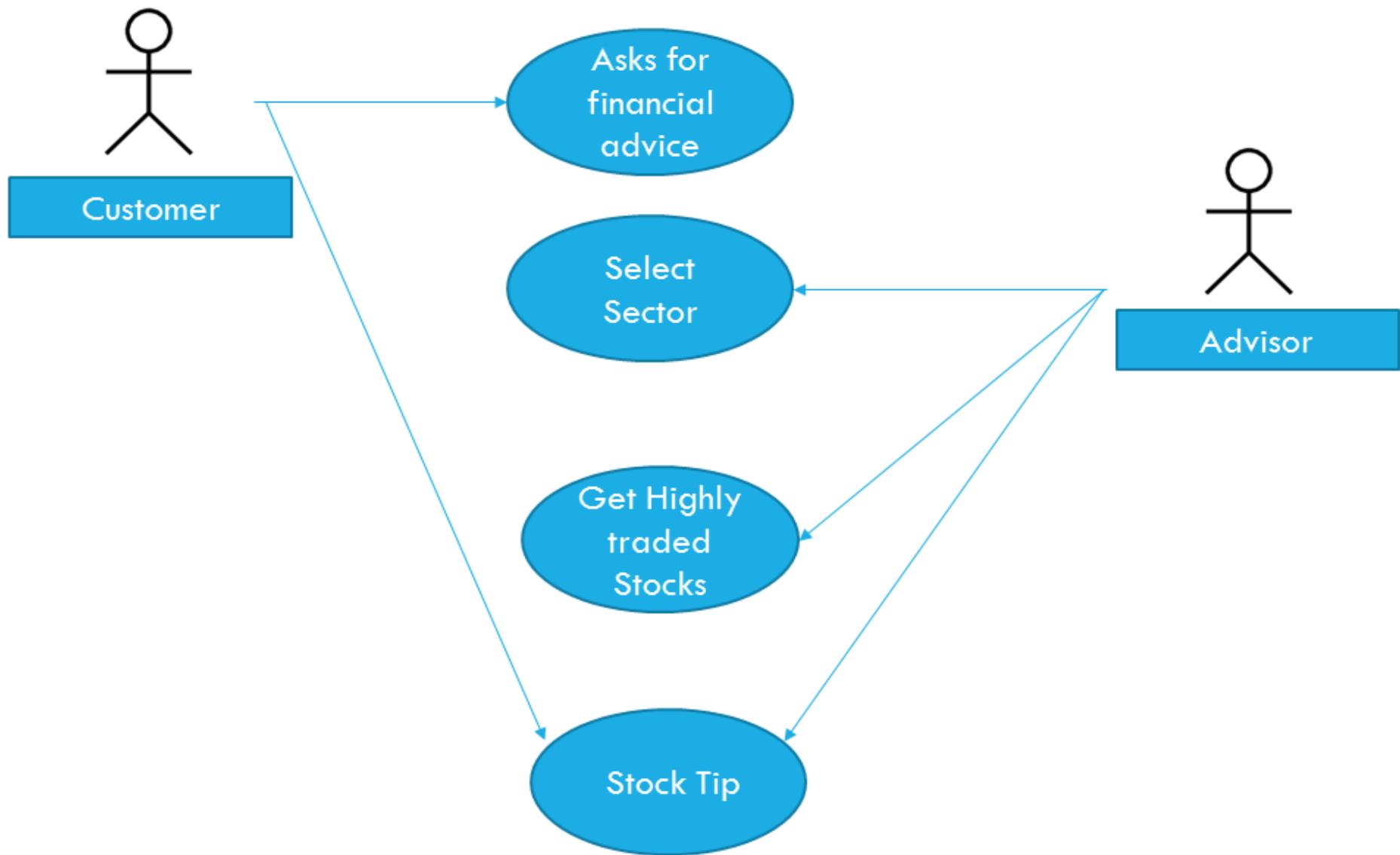
Vandana Iyer

Omkar Daphal

PROJECT GOAL

Project goal is to build a resilient and high efficient Scala/Spark application that is able to parse & analyze big records of financial stock data to provide meaningful and accurate analysis result and prediction.

GENERAL USE CASE



METHODOLOGY: PHASE 1

1. Collecting Stock Data from 'Yahoo Finance' on the fly using their API.

Data Collected Specification:

- EOD Stock Data for 10 Years.
- 5 Major Sectors (Computer & Technology, Finance, Retail, Industrial & Medical).
- Each sector contains data related to 5 different companies.

METHODOLOGY: PHASE 2

2. Filtering and Grinding of Data Collected.

This Phase is Done Using SCALA:

- Sorting each company by sector.
- Mapping stock data to a data model.
- Applying unit test to maintain stable behavior of each code block.
- Feed data into spark.

METHODOLOGY: PHASE 3

3. Performing Insightful Financial Analysis on Collected Data.

This Phase is Done using SPARK:

- Earning for each company per day.
- High & Low Companies by Volume per day.
- Company performance over time.
- Company VS Company performance.
- Company VS Sector performance.
- Sector VS Sector performance.
- Grading & Ratio per company & sector.

METHODOLOGY: PHASE 4

4. Displaying Result & Prediction.

Based on Data Collected & Analyzed:

- Display result of analysis in a Web UI.
- Generating a short term prediction for each company & sector based on the highs and lows of a stock(Ex: Using Linear Regression we can predict the next day high/low based on previous day).

Oct

Nov



R语言中文社区

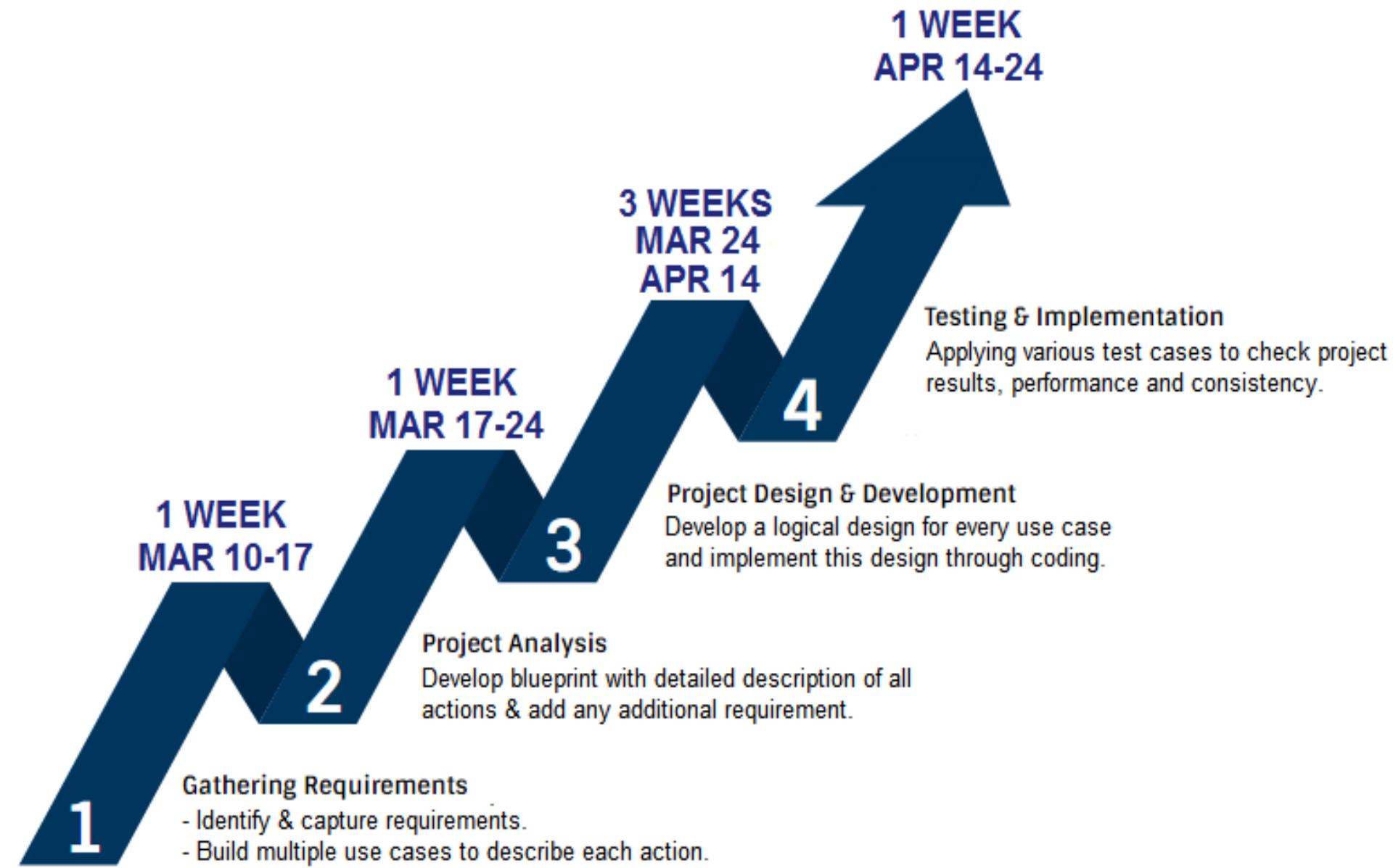
THE DATA SOURCE (~70,000 rows)

10 years of stock data (open, high, low, close and volume) for 5 companies in 5 different sectors received from Yahoo Financial API.

Date	Open	High	Low	Close	Adj Close*	Volume
Mar 17, 2016	736.45	743.07	736.00	737.78	737.78	1,860,800
Mar 16, 2016	726.37	737.47	724.51	736.09	736.09	1,624,400
Mar 15, 2016	726.92	732.29	724.77	728.33	728.33	1,721,000
Mar 14, 2016	726.81	735.50	725.15	730.49	730.49	1,718,300
Mar 11, 2016	720.00	726.92	717.13	726.82	726.82	1,970,800
Mar 10, 2016	708.12	716.44	703.36	712.82	712.82	2,833,500
Mar 09, 2016	698.47	705.68	694.00	705.24	705.24	1,421,500
Mar 08, 2016	688.59	703.79	685.34	693.97	693.97	2,076,300
Mar 07, 2016	706.90	708.09	686.90	695.16	695.16	2,985,100
Mar 04, 2016	714.99	716.49	706.02	710.89	710.89	1,972,100

<http://download.finance.yahoo.com/d/quotes.csv?s=%40%5EDJI,GOOG&f=ns1op&e=.csv>

MILESTONES



GITHUB REPOSITORY

- Contributions done by any team member will be posted/committed immediately to the project repo.
- Daily repository backups will be performed.
- <https://github.com/tabet-f/Analysis-of-Historical-Stock-Data> [\[LINK\]](#)

ACCEPTANCE CRITERIA FAIL OR PASS

1. Project is delivered on time.
2. Project deliverable is ~ error free.
3. Project meets all requirements specified in “Requirements gathering & analysis steps”.
4. Project is documented.
5. Project KPI is close to popular benchmarks (Esignal, Ninjatrader...).
6. Project Scalability and Extensibility over different data sets (performance with data for 20 years x 20 companies x 20 sectors => ~2M+ Rows).



Thank you

