

# Защита проекта

Приложение для аналитики рынка ценных  
бумаг (котировки и рейтинги)

Орловский А.Е.  
Ведущий разработчик  
ООО “Большая Тройка”

# Цели проекта

1. Загрузка в реальном времени данных об основных показателях акций (цена открытия торгов, цена закрытия торгов и другие базовые показатели), а так же о новостях, касающихся фондового рынка и торгующихся компаний.
2. Сохранение этих данных в PostgreSQL для дальнейшей подготовки этих данных для расширенной аналитики и машинного обучения.
3. Сохранение этих данных в Elasticsearch для дальнейшей визуализации в Kibana. Графики основных показателей акций дают понимание того, как ведут себя те или иные акции на временной кривой. В табличном виде такие данные очень трудны для восприятия.

# Используемые технологии

1. Jupyter Notebook + Python



2. Kafka 

3. Spark + Spark Streaming



4. ElasticSearch + Kibana



5. PostgreSQL



6. Telegram (Логгирование)



# Исполняемый код

1. Весь код выполняется и хранится в виде ноутбуков для Jupyter Notebook

# Источники данных

1. Alpha Vantage API (данные о показателях торгуемых акций)
2. FMP API (новости из мира фондовых рынков)

# Загрузка данных в реальном времени

1. Для загрузки данных в реальном времени используется связка Apache Kafka + Apache Spark (Spark Streaming).
2. Данные из разных источников пишутся в разные топики Kafka. Spark Streaming Job читает данные из топиков и определяет, что нужно сделать с этими данными и куда их необходимо записать (разные индексы ElasticSearch, разные таблицы в PostgreSQL)

# Визуализация данных

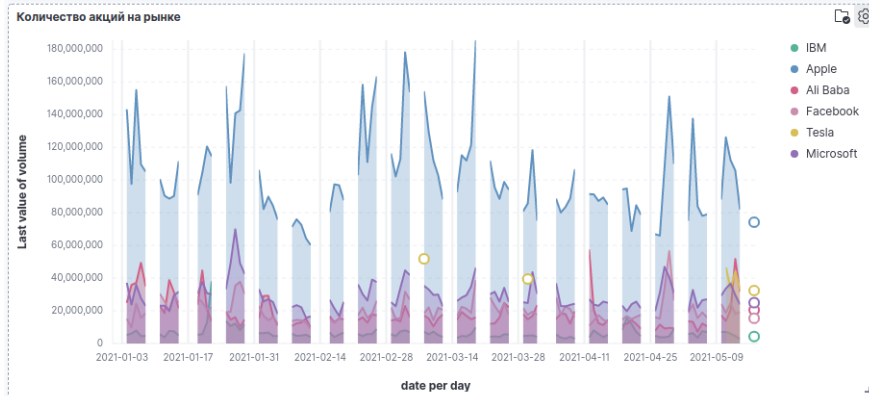
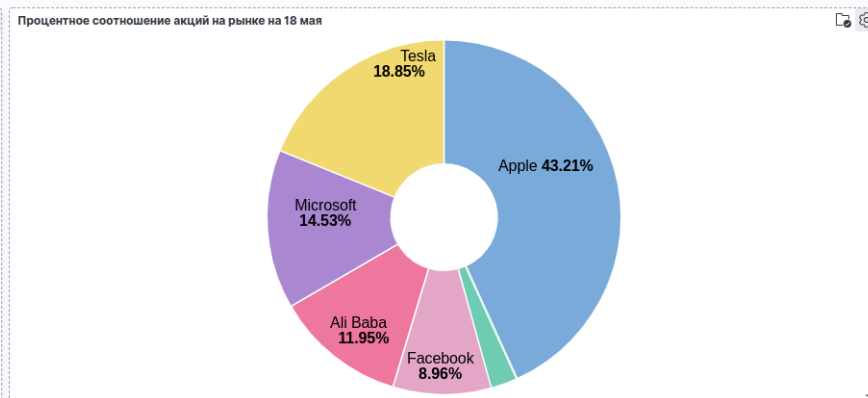
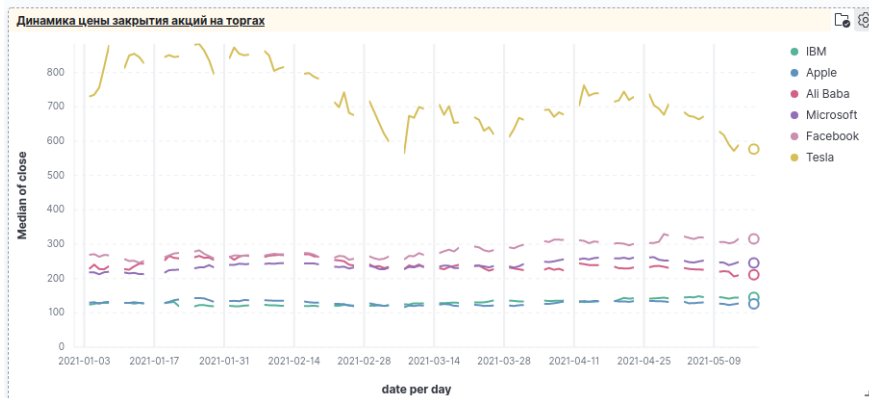
1. Для визуализации данных используется ELK стек.
2. Нет необходимости для использования LogStash.
3. Данные для визуализации хранятся в ElasticSearch.
4. Визуализация данных сделана в Kibana. Создан дашборд с графиками.

# Логгирование

1. Для удобства отладки приложения также реализовано логгирование системных сообщений и ошибок в Telegram



# Что получилось и как это выглядит (визуализация в Kibana)

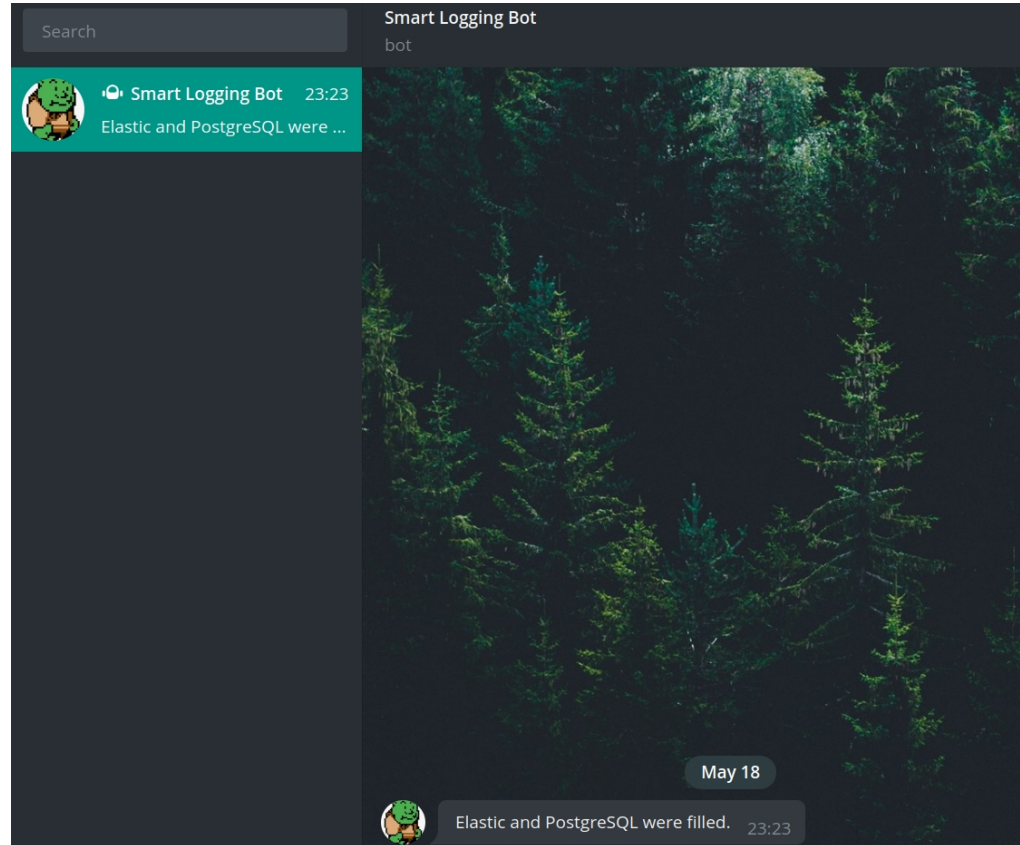


# Что получилось и как это выглядит (PostgreSQL)

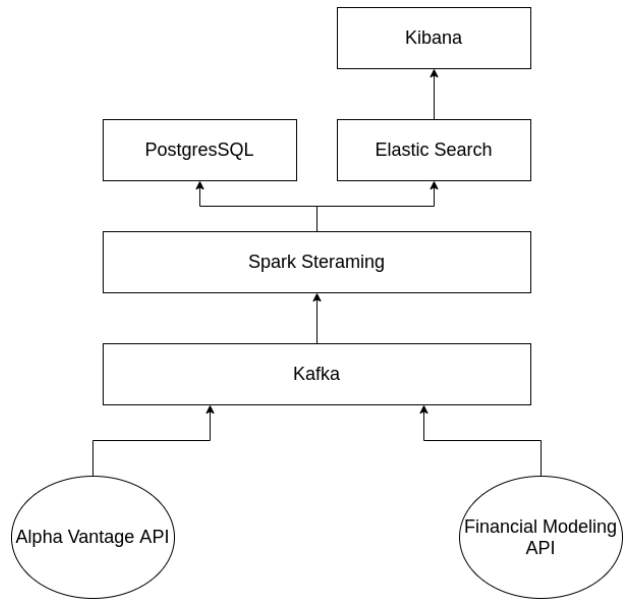
id	image	title	site	news_text	url	published_date	symbol
1	https://c...	Amazon Goes ...	Schaeffer...	Reports came o...	https://www.schaeffersrese...	2021-05-18	AMZN
2	https://c...	Apple report...	CNBC	Apple is plann...	https://www.cnbc.com/2021/...	2021-05-18	AAPL
3	https://c...	The Latest S...	The Motle...	A lot more peo...	https://www.fool.com/inves...	2021-05-18	AMZN
4	https://c...	How Google's...	CNBC	Google is the ...	https://www.cnbc.com/2021/...	2021-05-18	GOOG
5	https://c...	Amazon in Ta...	Barrons	The industry's...	https://www.barrons.com/ar...	2021-05-18	AMZN
6	https://c...	Amazon Goes ...	Schaeffer...	Reports came o...	https://www.schaeffersrese...	2021-05-18	AMZN
7	https://c...	Apple report...	CNBC	Apple is plann...	https://www.cnbc.com/2021/...	2021-05-18	AAPL
8	https://c...	The Latest S...	The Motle...	A lot more peo...	https://www.fool.com/inves...	2021-05-18	AMZN
9	https://c...	How Google's...	CNBC	Google is the ...	https://www.cnbc.com/2021/...	2021-05-18	GOOG
10	https://c...	Amazon in Ta...	Barrons	The industry's...	https://www.barrons.com/ar...	2021-05-18	AMZN
11	https://c...	Amazon Goes ...	Schaeffer...	Reports came o...	https://www.schaeffersrese...	2021-05-18	AMZN
12	https://c...	Apple report...	CNBC	Apple is plann...	https://www.cnbc.com/2021/...	2021-05-18	AAPL
13	https://c...	The Latest S...	The Motle...	A lot more peo...	https://www.fool.com/inves...	2021-05-18	AMZN
14	https://c...	How Google's...	CNBC	Google is the ...	https://www.cnbc.com/2021/...	2021-05-18	GOOG
15	https://c...	Amazon in Ta...	Barrons	The industry's...	https://www.barrons.com/ar...	2021-05-18	AMZN

id	open	high	low	close	adjust...	volume	dividen...	split...	series_date	symbol
1	61553	144.44	145.80	144.33	145.11	4275172.00	0.00	1.00	2021-05-17	IBM
2	61554	145.11	145.69	143.96	144.68	2747385.00	0.00	1.00	2021-05-14	IBM
3	61555	141.45	144.90	141.28	144.17	4598920.00	0.00	1.00	2021-05-13	IBM
4	61556	143.84	144.15	141.14	141.30	5959579.00	0.00	1.00	2021-05-12	IBM
5	61557	144.99	145.19	142.90	144.22	7126404.00	0.00	1.00	2021-05-11	IBM
6	61558	145.80	148.30	145.80	146.17	6983377.00	0.00	1.00	2021-05-10	IBM
7	61559	145.92	146.14	144.57	145.46	7003467.00	1.64	1.00	2021-05-07	IBM
8	61560	145.26	148.52	145.18	148.42	7583487.00	0.00	1.00	2021-05-06	IBM
9	61561	145.94	145.94	144.97	145.22	3622000.00	0.00	1.00	2021-05-05	IBM
10	61562	144.41	146.59	144.33	145.75	6350479.00	0.00	1.00	2021-05-04	IBM
11	61563	143.81	145.75	143.55	144.75	5733463.00	0.00	1.00	2021-05-03	IBM
12	61564	143.70	143.83	140.55	141.80	8072181.00	0.00	1.00	2021-04-30	IBM
13	61565	144.13	148.74	142.98	144.24	4353800.00	0.00	1.00	2021-04-29	IBM
14	61566	142.92	143.40	142.10	143.00	3768129.00	0.00	1.00	2021-04-28	IBM
15	61567	141.69	142.56	140.75	142.01	3084037.00	0.00	1.00	2021-04-27	IBM
16	61568	142.42	143.63	141.48	141.57	4718763.00	0.00	1.00	2021-04-26	IBM

# Что получилось и как это выглядит (логгирование)



# Схема проекта



Спасибо за внимание!