

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Южно-Уральский государственный университет (национальный исследовательский университет)»
Высшая школа электроники и компьютерных наук
Кафедра системного программирования

Разработка приложения для анализа географически-распределенных данных на платформе PySyft

Научный руководитель:
доцент кафедры СП, к.ф.-м.н.,
Г. И. Радченко

Автор:
студент группы КЭ-403
В.А. Дегтярев

Челябинск, 2023 г.

АКТУАЛЬНОСТЬ

1. Огромное количество информации в мире
2. Машинный анализ данных
3. Проблемы использования информации

ЦЕЛЬ И ЗАДАЧИ

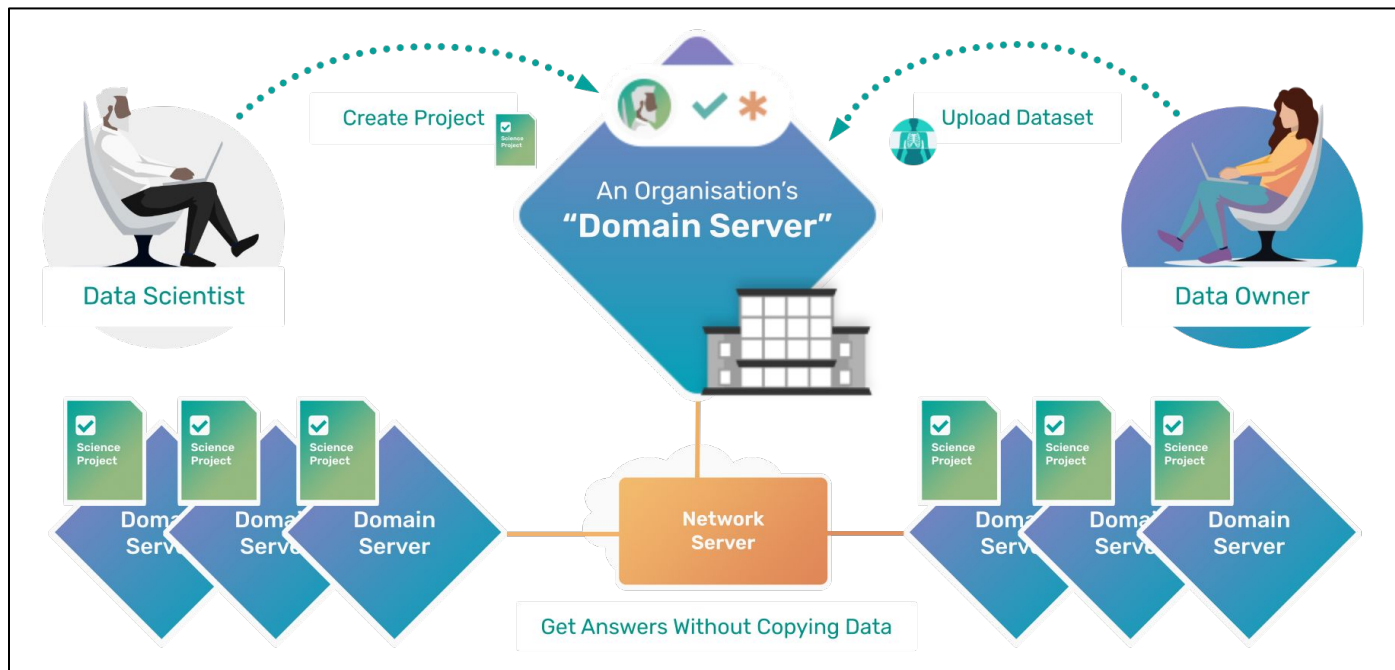
Цель:

Разработка приложения для анализа географически-распределенных данных на платформе PySyft

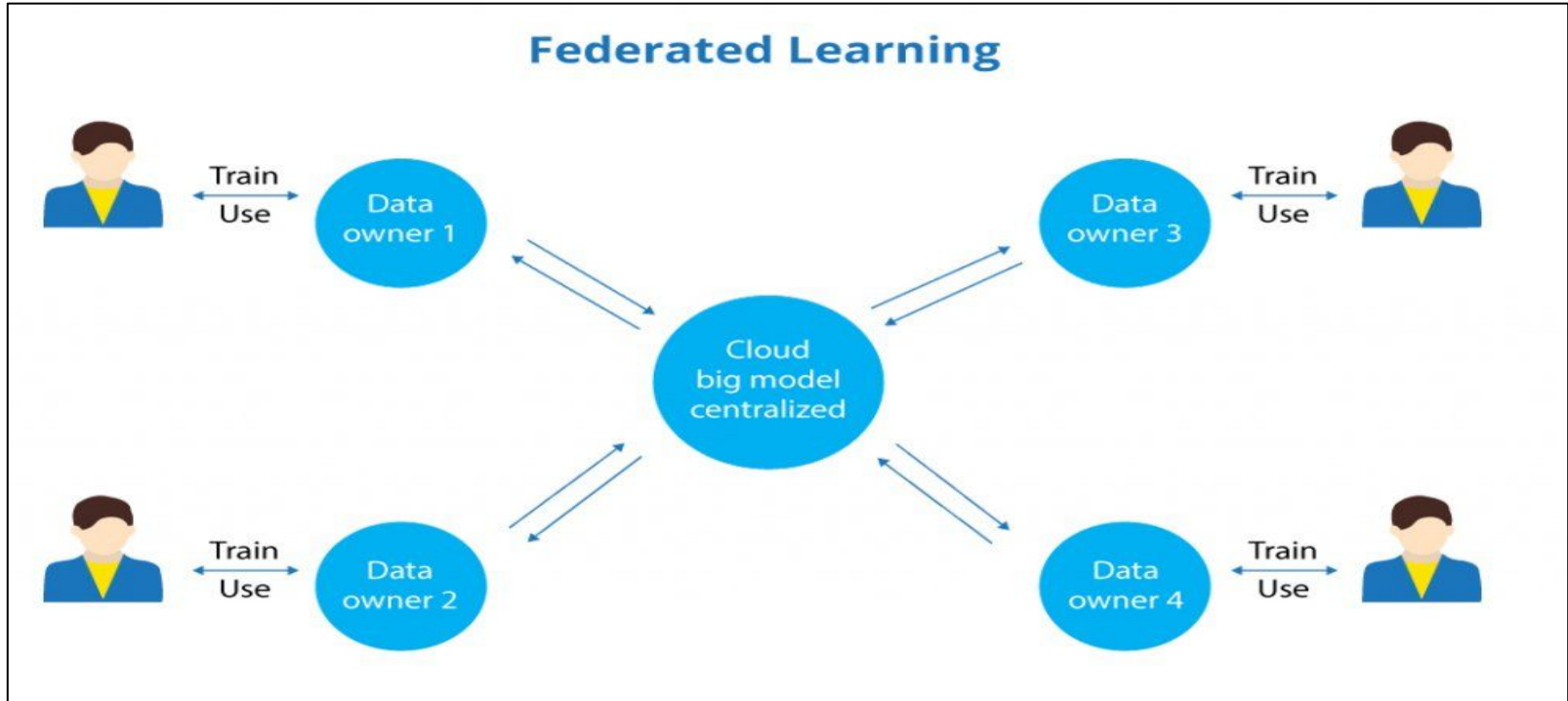
Задачи:

1. Выполнить обзор литературы и существующих аналогов
2. Спроектировать систему
3. Реализовать функционал для управления географически-распределенными узлами системы
4. Выполнить запуск не менее двух географически-распределенных узлов системы
5. Провести тестирование системы

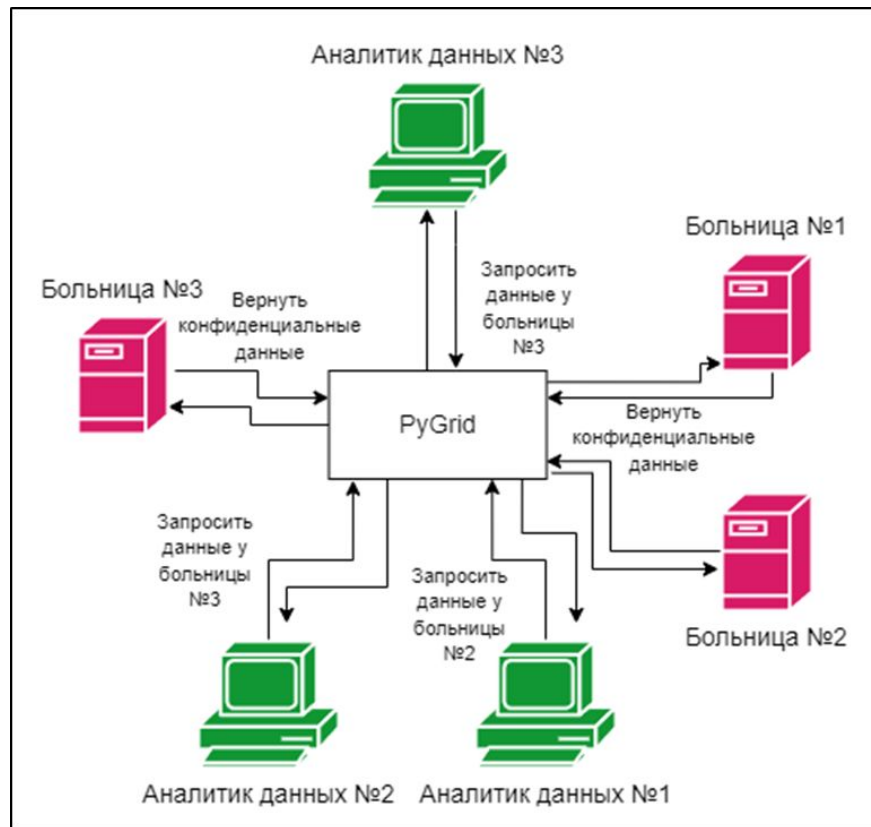
PySyft



TensorFlow Federated



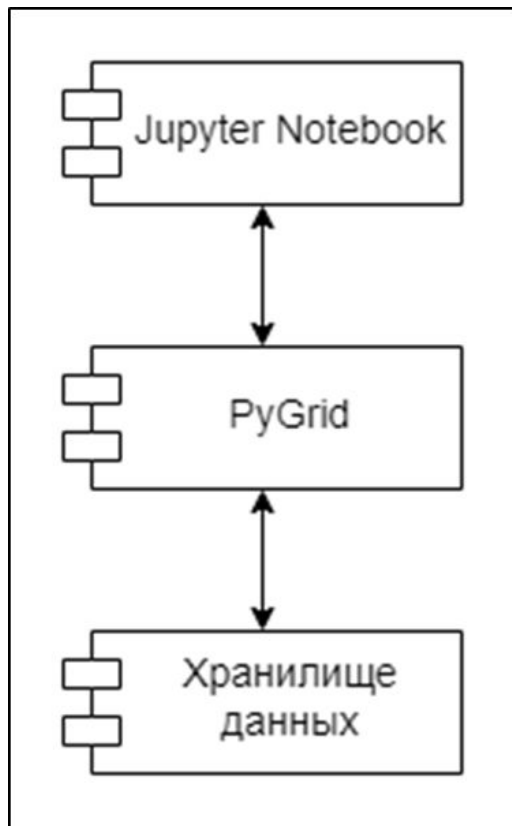
ЛОГИКА РАБОТЫ СИСТЕМЫ



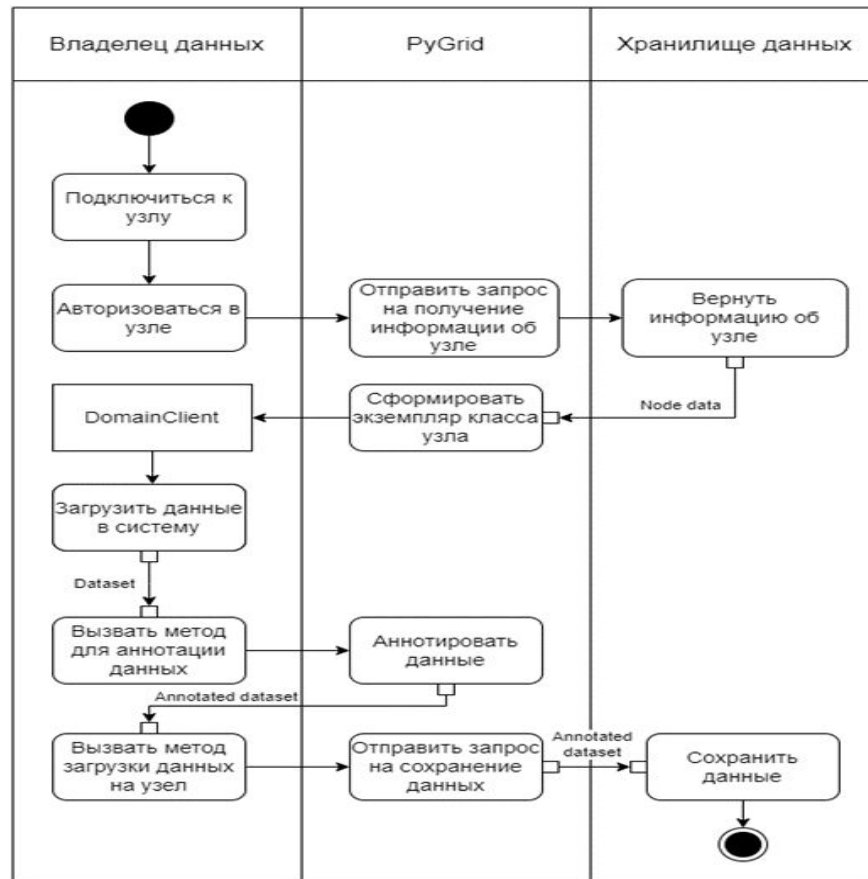
ВАРИАНТЫ ИСПОЛЬЗОВАНИЯ СИСТЕМЫ



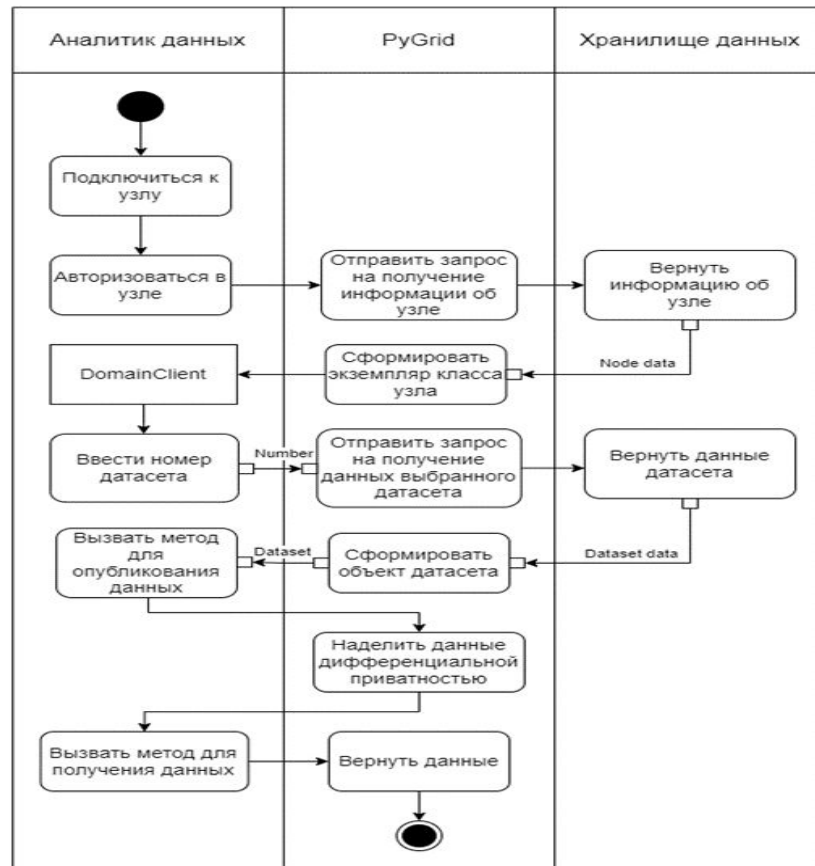
КОМПОНЕНТЫ СИСТЕМЫ



ЗАГРУЗКА ДАННЫХ



ПОЛУЧЕНИЕ ДАННЫХ



СРЕДСТВА РАЗРАБОТКИ

- Язык программирования Python
- Среда разработки Jupyter Notebook
- Платформа PySyft
- Инструмент для развертывания NaGrid

ПРИКЛАДНАЯ ЗАДАЧА

Цель прикладной задачи заключается в том, чтобы на основе данных о ядрах клеток опухоли молочной железы выявить, злокачественной или доброкачественной является клетка опухоли.

ЗАПУСК УЗЛА

[illegible]

ОПИСАНИЕ ЗАГРУЖЕННЫХ ДАННЫХ НА УЗЛЕ

Idx	Name	Description	Assets	Id
[0]	testDataset_v3	Our dataset contains test data. There are 31 columns and 250 rows in our dataset.	["radius_mean"] -> Tensor ["texture_mean"] -> Tensor ["perimeter_mean"] -> Tensor ...	a4614973-9288-4a94-9019-cb06f42d7bdf
[1]	testDataset_v3_diagnosis	Dataset contains diagnosis result. There are 1 column and 250 rows in our dataset.	["diagnosis"] -> Tensor	a62e69d7-0ecb-4e49-a168-e2c94764f050

ИСХОДНЫЕ ДАННЫЕ

radius_mean	texture_mean	perimeter_mean	area_mean	smoothness_mean	compactness_mean	concavity_mean	concave points_mean
17.99	10.38	122.80	1001.0	0.11840	0.27760	0.30010	0.14710
20.57	17.77	132.90	1326.0	0.08474	0.07864	0.08690	0.07017
19.69	21.25	130.00	1203.0	0.10960	0.15990	0.19740	0.12790
11.42	20.38	77.58	386.1	0.14250	0.28390	0.24140	0.10520
20.29	14.34	135.10	1297.0	0.10030	0.13280	0.19800	0.10430
...
10.48	19.86	66.72	337.7	0.10700	0.05971	0.04831	0.03070
13.20	17.43	84.13	541.6	0.07215	0.04524	0.04336	0.01105
12.89	14.11	84.95	512.2	0.08760	0.13460	0.13740	0.03980
10.65	25.22	68.01	347.0	0.09657	0.07234	0.02379	0.01615
11.52	14.93	73.87	406.3	0.10130	0.07808	0.04328	0.02929

ПРЕОБРАЗОВАННЫЕ ДАННЫЕ

	0	1	2	3	4	5	6	7	8	9	...	20	21
0	16.132178	8.434578	134.104971	1100.135552	0.087600	0.280242	0.298157	0.153714	0.226333	0.071672	...	34.266756	15.488959
1	17.078989	17.094781	127.665273	1272.185442	0.105969	0.089305	0.032212	0.048682	0.154750	0.063071	...	36.662754	23.940783
2	15.622685	15.774698	162.606136	1219.693821	0.103326	0.171460	0.152722	0.105217	0.220147	0.063774	...	18.326853	18.217385
3	10.051632	22.897328	55.129015	359.715497	0.160802	0.288066	0.231524	0.052108	0.190165	0.093511	...	12.677428	18.982321
4	20.652202	17.420749	126.831422	1330.350017	0.100585	0.117940	0.205592	0.097690	0.126316	0.048576	...	20.672949	19.742730
...
245	10.478143	17.744799	80.199168	281.126391	0.119582	0.058763	0.050669	-0.000533	0.211496	0.069394	...	7.003120	34.228511
246	13.277229	11.052537	120.697787	579.991444	0.045019	0.052506	0.023185	0.038215	0.102785	0.059122	...	20.335097	33.831021
247	13.609391	12.406311	89.819592	332.250218	0.133811	0.123512	0.107841	0.053611	0.173542	0.065869	...	12.548284	17.728222
248	9.921877	26.998829	88.430236	475.303414	0.091708	0.073301	-0.026388	0.005190	0.159686	0.054961	...	7.946641	27.365150
249	6.742087	19.200117	80.417527	363.016130	0.099238	0.067331	0.005977	0.022594	0.138886	0.057071	...	14.359081	23.946826

АНАЛИЗ ПОЛУЧЕННЫХ ДАННЫХ

- Была создана модель дерева решений для задачи классификации клетки опухоли молочной железы.
- Модель обучалась на данных первого узла, а также повторно обучалась на данных второго узла

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

1. Выполнен обзор литературы и существующих аналогов
2. Спроектирована система
3. Реализован функционал для управления географически-распределенными узлами системы
4. Выполнен запуск двух географически-распределенных узлов системы
5. Проведено тестирование системы