

ООО «Новые программные системы»

**Функциональные характеристики программы для ЭВМ
«Программная система NGSWizard 1.3»
(версия 1.3)**

Листов 11

2022 год

АННОТАЦИЯ

Настоящий документ является инструкцией системного программиста (администратора) программы анализа молекулярно-генетических характеристик по данным высокопроизводительного секвенирования ДНК (далее – Программа), предназначенной для мониторинга за молекулярно-генетическими показателями и поддержки принятия решений медицинских работников по лечению.

В разделе «Общие сведения о Программе» указаны назначение и функции Программы и сведения о технических и программных средствах, необходимых для работы Программы.

В разделе «Структура Программы» приведены сведения о структуре Программы, порядке взаимодействия составных частей, связях с другими программами.

СОДЕРЖАНИЕ

1. ОБЩИЕ СВЕДЕНИЯ О ПРОГРАММЕ.....	4
1.1 Наименование и назначение Программы.....	4
1.2 Функции Программы.....	4
1.3 Минимальный состав технических средств.....	5
1.4 Минимальный состав программных средств.....	6
2. СТРУКТУРА ПРОГРАММЫ	8
2.1 Структура и архитектура Программы	8
2.2 Порядок взаимодействия составных частей Программы	10
2.3 Сведения о связях с другими программами.....	10
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ	11

1. ОБЩИЕ СВЕДЕНИЯ О ПРОГРАММЕ

1.1 Наименование и назначение Программы

Полное наименование: Программа анализа молекулярно-генетических характеристик по данным высокопроизводительного секвенирования ДНК (далее - Программа).

Краткое наименование: NGSWizard 1.3.

Программа предназначена для мониторинга состояния молекулярно-генетических показателей и поддержки принятия решений по лечению.

1.2 Функции Программы

Программа выполняет следующие функции:

— формирование отчёта о состоянии спектра известных клинически значимых генетических маркеров опухоли мозга для конкретного пациента с сопутствующей актуальной референсной информацией;

— формирование отчёта о состоянии спектра гипотетических новых маркеров опухоли мозга для конкретного пациента с сопутствующей актуальной референсной информацией;

— оценка прогнозируемых характеристик развития опухоли, включая временные оценки, основываясь на собственных данных секвенирования, информации от врача (медицинского работника) и данных других анализов и наблюдений;

— импорт данных секвенирования и сопутствующей информации на сервер через пользовательский веб-интерфейс, удобный для пользователя;

— расчет метрик статистики качества данных отсеквенированного образца, составляющих его Паспорт;

— накопление и хранение данных секвенирования и результатов их обработки, а также данных об истории болезни и проведённых анализах пациентами (до 500 пациентов в год) с целью дальнейшей их выгрузки и проведения необходимых исследований, а также для использования в качестве референса.

1.3 Минимальный состав технических средств

Серверная часть Программы может быть размещена как на проприетарном сервере, так и на серверах провайдеров облачных хранилищ, таких как Yandex.Cloud, VK Cloud Solutions.

Серверная часть Программы включает в себя:

- сервер приложений;
- кластер вычислительных ресурсов в составе:
- вычислительный узел для полноэкзомного секвенирования;
- вычислительный узел для полногеномного секвенирования;
- файловое хранилище (общая файловая система POSIX compliant).

Для обработки одного образца, требуется развертывание одного кластера вычислительных ресурсов. Для одновременной обработки нескольких образцов требуется развертывание соответствующего количества узлов кластера вычислительных ресурсов.

Сервер приложений Программы должен иметь следующие характеристики:

- не менее 8 Гб оперативной памяти;
- не менее 100 Гб доступного пространства на жестком диске;
- сетевая карта, работающая на скорости не ниже 1024 Мбит/сек;
- процессор, имеющий не менее 2-х вычислительных ядер с тактовой частотой не ниже 2 ГГц.

Вычислительный узел для полноэкзомного секвенирования должен иметь следующие характеристики:

- не менее 32 Гб оперативной памяти;
- не менее 50 Гб доступного пространства на жестком диске;
- сетевая карта, работающая на скорости не ниже 1024 Мбит/сек;
- процессор, имеющий не менее 4-х вычислительных ядер с тактовой частотой не ниже 2 ГГц.

Вычислительный узел для полногеномного секвенирования должен иметь следующие характеристики:

- не менее 64 Гб оперативной памяти;
- не менее 400 Гб доступного пространства на жестком диске;
- сетевая карта, работающая на скорости не ниже 1024 Мбит/сек;
- процессор, имеющий не менее 16-х вычислительных ядер с тактовой частотой не ниже 2 ГГц.

Файловое хранилище должно иметь следующие характеристики:

- не менее 1 Тб доступного пространства на жестком диске;
- сетевая карта, работающая на скорости не ниже 1024 Мбит/сек.

Между сервером приложений, вычислительными узлами и хранилищем файлов должно быть организовано сетевое соединение со скоростью не ниже 1 Гбит/с.

Для нормальной работы клиентской части Программы рабочие места пользователей должны иметь следующие характеристики:

- не менее 4 Гб оперативной памяти;
- интернет-соединение со скоростью не ниже 100 Мбит/сек.
- процессор, имеющий не менее 2-х вычислительных ядер с тактовой частотой не ниже 2 ГГц.

1.4 Минимальный состав программных средств

Для обеспечения нормального функционирования серверной части Программы на технических средствах должно быть установлено следующее СПО:

- Операционная система GNU/Linux;
- СПО Docker 19;
- СУБД PostgreSQL 12.5.

Тестирование системы проводилось на следующих операционных системах: Debian 10 "Buster", Debian 11 "Bullseye", Ubuntu 18 "Bionic Beaver", Ubuntu 20 "Focal Fossa".

Для обеспечения нормального функционирования клиентской части Программы на рабочих местах пользователей должно быть установлено следующее СПО:

— WEB-браузер (Google Chrome версии не ниже 50.0.2661, Mozilla Firefox версии не ниже 52, Safari версии не ниже 10.0, Microsoft Edge версии не ниже 41.16299.15, Microsoft Internet Explorer версии не ниже 11).

2. СТРУКТУРА ПРОГРАММЫ

2.1 Структура и архитектура Программы

Программа реализована в классической архитектуре клиент-сервер, за исключением кластера вычислительных узлов, предназначенных для выполнения ресурсоемких операций. К таким операциям относятся процедуры очистки исходных данных, картирования, поиска мутаций и вариаций (данные задачи могут выполняться от нескольких минут до нескольких часов в зависимости от размеров исходных данных и используемых аппаратных ресурсов). Архитектура Программы представлена на Рисунке 1.

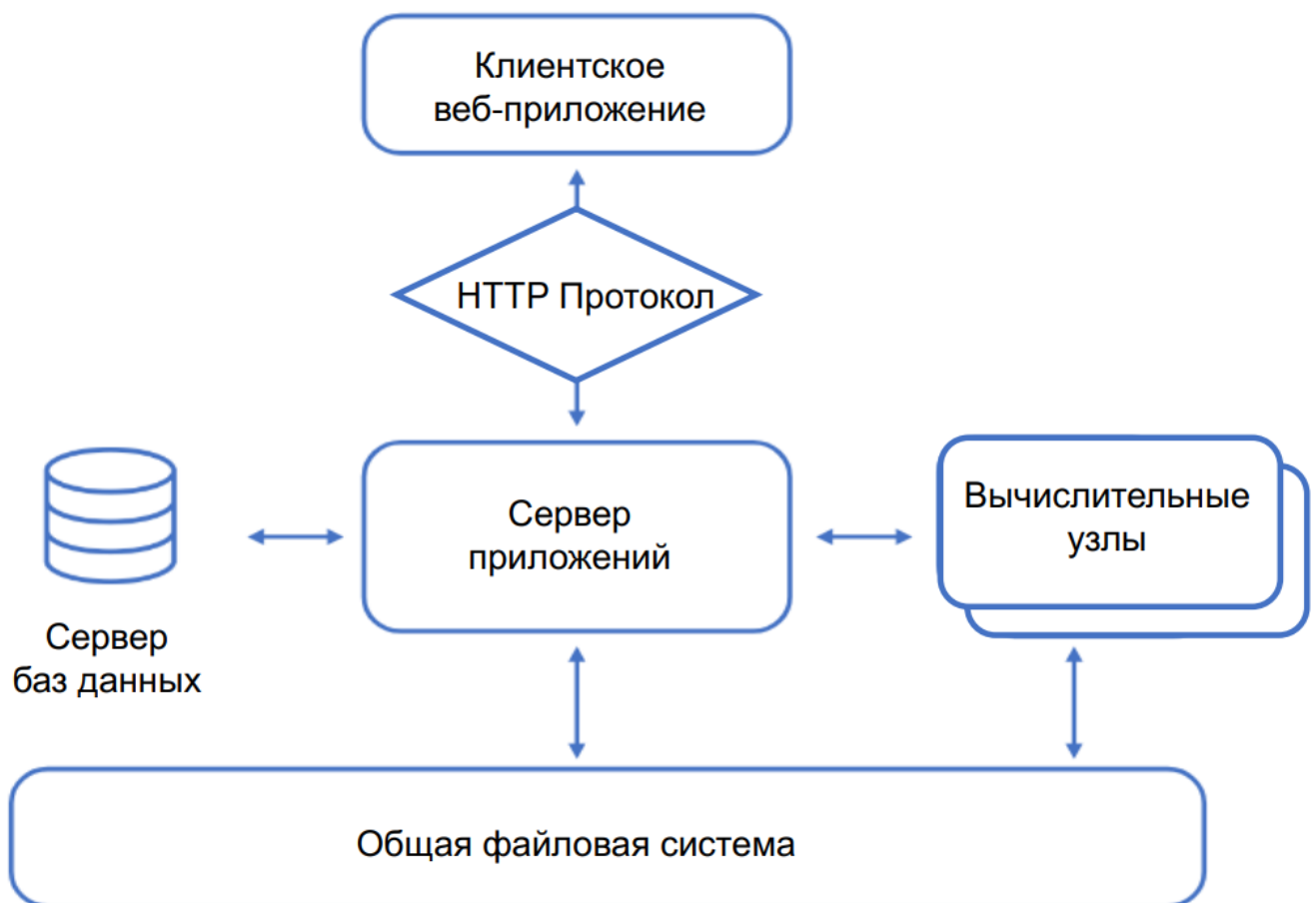


Рисунок 1 — Архитектура Программы

Использование кластера вычислительных узлов позволяет снять нагрузку с основного сервера приложений и обеспечить горизонтальное масштабирование для параллельной обработки множества файлов. Сервер приложений выступает в роли клиента кластера, вычислительные узлы – в роли серверов кластера в терминах Apache Ignite. Фреймворк решает следующие задачи:

— `service discovery` – сервер приложений и вычислительные узлы находят друг друга в сети и регистрируются в кластере;

— `task execution` – вычислительные узлы получают задачи на выполнение и возвращают на сервер приложений результаты;

— `publisher/subscriber communication` – вычислительные узлы отправляют сообщения о текущем статусе (прогносте) выполнения задач;

— `failover` – если задача не была выполнена из-за ошибки кластера (сетевой сбой, сбой на вычислительном узле), то она передается на выполнение на другой узел. Узел автоматически перезапускается, в случае сбоя.

В качестве СУБД используется PostgreSQL 12.5 (<https://www.postgresql.org>) и отвечает за хранение данных о пользователях, настроек приложения, метаданных исходных файлов (формат, тип секвенатора, имя, размер, время загрузки) и протоколов их обработки.

Для итеративного обновления структуры БД используется библиотека Flyway (<https://flywaydb.org>).

Для хранения исходных файлов и файлов, полученных в результате обработки, используется общая файловая система POSIX compliant, доступная как серверу приложений, так и вычислительным узлам.

Клиентская часть представляет собой одностраничное web-приложение (SPA, Single Page Application). Поддерживается работа на всех современных браузерах: Google Chrome, Mozilla Firefox, Safari, Microsoft Edge, Microsoft Internet Explorer.

Клиентская часть взаимодействует с серверной используя REST API серверной части по протоколу HTTP. В качестве транспорта используется JSON.

2.2 Порядок взаимодействия составных частей Программы

Взаимодействие компонентов Программы происходит в следующем порядке:

1. Пользователь в web-интерфейсе Программы выбирает для загрузки один или несколько файлов. Файлы передаются на сервер приложений по протоколу HTTP и сохраняются на общей файловой системе.

2. Метаданные файлов (уникальный идентификатор, формат, тип секвенатора, имя, размер, время загрузки) сохраняются в СУБД.

3. Асинхронно запускаются задачи идентификации, верификации, оценки качества, очистки данных, картирования, поиска соматических и герминальных вариантов на вычислительных узлах кластера. Сервер приложений при этом выступает в роли координатора процесса обработки файла.

4. При выполнении задачи вычислительные узлы передают серверу приложений текущий процент выполнения задачи и логи обработки. Эти данные протоколируются в СУБД и доступны в web-интерфейсе через REST API.

5. При обработке файлов результаты всех задач протоколируются в СУБД и доступны в web-интерфейсе через REST API.

6. После завершения обработки исходные файлы, отчеты по оценке качества, и файлы, полученные в результате обработки доступны в web-интерфейсе через REST API.

2.3 Сведения о связях с другими программами

Взаимодействие с другими программами обеспечивается в рамках поддержки файлов данных биологических последовательностей форматов *.fastq, *.bam и *.vcf.

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

Сокращение	Расшифровка
ДНК	Дизоксирибонуклеиновая кислота
СПО	Специальное программное обеспечение
СУБД	Система управления базами данных
БД	База данных
API	Application program interface
REST	Representation State Transfer
JSON	JavaScript Object Notation
SSL	Secure Sockets Layers