



PICODATA

*Распределенный сервер приложений со встроенной  
распределенной базой данных*

**Техническое задание на доработку**

## **Аннотация**

Документ «Техническое задание на доработку Picodata» содержит сведения, позволяющие определить основные цели и задачи мероприятий, направленных на доработку, совершенствование и достижение новой функциональности программного комплекса Picodata (далее — *Программа*). Документ также определяет требования к этапам и составу доработок, срокам их исполнения, а также описывает методы оценки проведенных работ.

Структура документа соответствует основным требованиям к оформлению технических заданий на основании требований ГОСТ 19.201-78. Единая система программной документации (ЕСПД).

## Содержание

1. Наименование программы.....	4
2. Наименование разработчика программы.....	4
3. Область применения программы и ее назначение.....	4
4. Основания для проведения доработки.....	5
5. Назначение доработки.....	5
6. Требования к программе.....	6
6.1. Требования к среде исполнения.....	6
6.2. Требования к среде сборки.....	6
6.3. Требования к поддержке отдельных ОС.....	7
6.4. Требование к аппаратной совместимости.....	7
6.5. Требования к функциональным характеристикам.....	7
6.6. Требования к производительности.....	8
6.7. Требования к программной документации.....	8
7. Техничко-экономические показатели.....	8
7.1. Описание востребованности и экономических преимуществ доработки.....	8
7.2. Сравнение Программы с другими решениями.....	9
8. Стадии и этапы доработки.....	11
9. Порядок контроля и приемки.....	12

## **1. Наименование программы**

Полное наименование дорабатываемого ПО:

PICODATA. Распределенный сервер приложений со встроенной распределенной базой данных.

## **2. Наименование разработчика программы**

Полное наименование и основные сведения о компании-разработчике дорабатываемого программного обеспечения:

ООО «Пикодата»

Компания-разработчик находится в реестре аккредитованных организаций, осуществляющих деятельность в области информационных технологий.

Номер решения о государственной аккредитации: АО-20210827-266606-3

Номер в реестре аккредитованных организаций: 15177

Продукт Picodata включен в единый реестр российских программ для электронных вычислительных машин и баз данных.

Номер регистрации: 2021680169

Дата регистрации: 07.12.2021

## **3. Область применения программы и ее назначение**

Основным назначением продукта Picodata является горизонтально масштабируемое хранение структурированных и неструктурированных данных, управление ими, предоставление среды вычислений внутри кластера, состоящего из реплицированных отдельных узлов (инстансов). Данная комбинация возможностей позволяет эффективно реализовать сценарии управления наиболее востребованными, часто изменяющимися, горячими данными. В традиционных корпоративных архитектурах для ускорения и повышения надёжности доступа к данным классических, универсальных СУБД используются кэши и шины данных. Использование ПО Picodata позволяет заменить три компонента корпоративной архитектуры — кэш, шина и витрина доступа к данным — единым, высокопроизводительным и строго консистентным решением.

Кластер Picodata обеспечивает быстрый доступ к данным внутри распределенного хранилища. Это позволяет использовать его в следующих областях:

- управление телекоммуникационным оборудованием;
- банковские и в целом финансовые услуги, биржевые торги, аукционы;
- формирование персональных маркетинговых предложений с привязкой ко времени и месту;

- обработка больших объемов данных в реальном времени для систем класса "интернет вещей" (IoT);
- игровые рейтинговые таблицы;
- прочие смежные области применения.

## **4. Основания для проведения доработки**

Основанием для доработки Программы является решение генерального директора ООО «Пикодата» о целесообразности развития функциональности Picodata с учетом особенностей отечественного ИТ-рынка и актуальности резидентных СУБД в целом и Picodata в частности.

Актуальность решения продиктована растущей необходимостью в отечественной распределенной СУБД, способной эффективно заменить Oracle Exadata в качестве коммерчески перспективного решения как для коммерческих, так и государственных структур. Программный продукт Picodata — это современная кластерная СУБД, реализующая распределенные транзакции и бескомпромиссную отказоустойчивость. Это надежное решение по хранению и обработке данных при предоставлении цифровых услуг.

Picodata к 1 июня 2024 года станет полностью отечественным конкурентоспособным программным решением для задач распределенного хранения данных и выполнения приложений в адресном пространстве СУБД на языке программирования Rust.

Доработка программного продукта повысит его конкурентоспособность за счет лучшего соответствия стандарту ISO/IEC 9075 “Database Language SQL” (Язык баз данных SQL). На данный момент не существует СУБД, полностью отвечающей данному стандарту, однако чем полнее соответствие, тем больше ценностью обладает СУБД с точки зрения надежности хранения данных и совместимости с другими СУБД. Именно надежности сохранения данных и возможности работы с ними в кластере не хватает клиентам в настоящий момент.

Доработка программного комплекса “Picodata” позволит реализовать более полную поддержку части 2 стандарта ISO/IEC 9075 (ISO/IEC 9075-2 Основа (SQL/Foundation)) в ядре базы данных, используемой в “Picodata” не только на одном узле, но и в рамках всего распределенного кластера.

## **5. Назначение доработки**

Назначением доработки Программы является своевременное и верифицируемое достижение основных целей. В них входит:

- соответствие отдельным требованиям стандарта SQL:2016;
- добавление кластерного распределения данных, что означает повышение функциональности и надежности горизонтально масштабируемого хранения структурированных и неструктурированных данных и управления ими;

- предоставления среды вычислений внутри реплицированного кластера;
- повышение отказоустойчивости кластера, работающего под управлением Программы.

В результате доработки, программный комплекс будет иметь следующие функциональные возможности:

- Работа распределенной СУБД на удаленных вычислительных узлах;
- Работа сервера приложений совместно с распределенной СУБД;
- Улучшение отказоустойчивости работы программного комплекса за счет репликации вычислительных узлов и узлов хранения данных: комплекс сможет продолжить работу при отказе до половины узлов кластера;
- Реализация удобного упрощенного управления распределенным программным комплексом с поддержкой распределенного SQL (в рамках всего кластера).

Доработка позволит вывести на рынок отказоустойчивую СУБД с поддержкой распределенного SQL и встроенным сервером приложений — уникальное предложение, не имеющее прямых равноценных аналогов.

## **6. Требования к программе**

### **6.1. Требования к среде исполнения**

Основным требованием к среде исполнения Программы является программная среда на основе GNU/Linux, способная выполнять приложения, написанные на языках Rust, Python, C, C++. Программа также требует работающей сетевой подсистемы, включая как минимум локальное пространство адресов, предоставляемое интерфейсом loopback.

### **6.2. Требования к среде сборки**

Для сборки/компиляции Программы из исходного кода требуются:

Rust и Cargo 1.59 или новее

Cmake 3.16 или новее

gcc, g++

libstc++-static

glibc 2.17 или новее

### 6.3. Требования к поддержке отдельных ОС

Программа должна поддерживать полноценное функционирование в следующих операционных системах на основе GNU/Linux:

- Debian 10 и новее
- Ubuntu 20.04 и новее
- CentOS 7 и новее
- Astra Linux 1.6 и новее
- Alt Linux p9/p10

Данное требование относится также и к прочим разновидностям ОС на базе GNU/Linux, основанным и совместимым с указанными выше образцами.

### 6.4. Требование к аппаратной совместимости

Программа должна работать на аппаратных комплексах, использующих архитектуру x86\_64 (64-битная операционная система на Intel-совместимом оборудовании).

Автоматизированное рабочее место с установленной Программой должно соответствовать следующим рекомендуемым характеристикам:

- ЦП Intel/AMD с частотой не менее 2 ГГц и минимум двумя ядрами;
- 8 ГБ ОЗУ или более;
- основной накопитель емкостью не менее 30 ГБ;
- ОС на основе Linux x86\_64 с glibc 2.17 или новее

### 6.5. Требования к функциональным характеристикам

Программа должна поддерживать следующие функции:

- успешный запуск и предоставление Lua-интерпретатора СУБД Tarantool
- работающий сервер экземпляра приложения (успешная загрузка модуля box.cfg)
- возможность записи и чтения данных из СУБД Tarantool
- объединение нескольких (2 и более) экземпляров приложения в единый кластер
- возможность исполнения основных SQL-запросов (SELECT, INSERT) в рамках распределенного кластера
- отказоустойчивая работа кластера (сохранение работоспособности при наличии 50% или более узлов распределенного кластера)

## **6.6. Требования к производительности**

Целевым показателем производительности Программы является число транзакций за единицу времени внутри распределенного кластера. Программа обеспечивает 1000000 (1 миллион транзакций) в секунду при соответствующей нагрузке.

## **6.7. Требования к программной документации**

Программная и эксплуатационная документация на Программу представляют собой один комплект документов, который должен включать в себя следующие обязательные документы:

- Руководство по эксплуатации
- Руководство по установке

## **7. Техничко-экономические показатели**

### **7.1. Описание востребованности и экономических преимуществ доработки**

Компания-разработчик Программы создает, совершенствует и продвигает в ИТ-отрасли рыночное решение по внедрению, поддержке и эксплуатации специализированной СУБД Tarantool, которая доступна на условиях open source. Компания предлагает собственное решение на основе Tarantool для применения в качестве in-memory data grid (IMDG), т.е. СУБД, выполняющейся резидентно (в ОЗУ) в виде т. н. «решетки данных», распределенных по нескольким физическим узлам. Решение имеет двойную функциональность: с одной стороны, это база данных, с другой — среда сильно распараллеленных вычислений.

Предлагаемый компанией-разработчиком программный продукт является программой нового поколения, поскольку он интегрирует уже используемые в отрасли технологические решения. Конкретно, речь идет о следующих решениях:

- хранении данных с помощью SQL или NoSQL СУБД. Каждый экземпляр «Picodata» запускает инстанс СУБД Tarantool и поддерживает язык локальных запросов к SQL, кластерные транзакции и запросы к данным посредством распределенного SQL являются предметом доработки программного комплекса;
- поддержке сервера приложений. Принцип IMDG предполагает выполнение бизнес-логики непосредственно внутри «решетки данных», что позволяет создавать клиентские приложения и управлять их жизненным циклом. В «Picodata» поддерживаются такие приложения на языках Rust и Lua;



- шине данных, используемой для обмена информации между узлами распределенной системы. Распределенный кластер «Picodata» подразумевает высокопроизводительный обмен данными как внутри кластера, так и между внешними потребителями;
- системе кэширования, используемой для ускорения доступа к часто используемым данным. Распределенный кластер «Picodata» обладает отказоустойчивостью, благодаря чему доступ к данным осуществляется без сбоев и задержек;

Picodata — интегрированное программное решение, которое обеспечивает высокую производительность и надежность: от 10000 запросов в секунду и uptime не ниже 99,99%.

## 7.2. Сравнение Программы с другими решениями

Российские и зарубежные разработчики, работающие на российском рынке, в настоящий момент не предлагают программные продукты, которые бы эффективно автоматизировали управление распределенным кластером хранения данных. Тем не менее, среди конкурирующих решений, можно выделить:

- СУБД Tarantool — изначальный продукт, созданный в Mail.ru и позднее предоставленный публично в качестве open source-решения. Используется как в «Picodata», так и некоторыми сторонними вендорами. Предоставляет среду выполнения SQL-запросов на локальном узле, но не на распределенном кластере. В рамках доработки «Picodata» создается кластерная реализация SQL для Tarantool;
- СУБД Redis — зарубежное решение, которое не отвечает потребностям рынка и заказчиков в РФ;
- СУБД MongoDB — решение, не обеспечивающее выполнение клиентских приложений в кластере;
- СУБД PostgreSQL — решение на основе объектно-реляционной СУБД, не раскрывающее своих преимуществ в резидентном режиме.

Ниже представлена таблица сравнения Программы с наиболее близкими аналогами по характерным ключевым параметрам.

<b>Характеристики/ параметры</b>	<b>Picodata *</b>	<b>Tarantool **</b>	<b>Redis</b>	<b>MongoD B</b>	<b>PostgreSQ L</b>
Поддержка кластера и репликации	+	+	+	+	+
Поддержка сервера приложений	+	+	+	+	+
Поддержка локального SQL	+	+	+	+	+
Поддержка кластерного SQL	+	-	-	-	-
Возможность управлять топологией кластера при выходе из строя одного из узлов	+	-	-	-	-

## 8. Стадии и этапы доработки

План доработки Программы предполагает выпуск серии промежуточных версий, последовательно добавляющих новую функциональность в соответствии с перечнем, указанном в разделе «Требования к функциональным характеристикам». Каждая промежуточная версия соответствует стадии доработки. Компания-разработчик Программы отмечает следующие стадии:

### Q4-2022 (22.11.0)

Добавление модуля vshard, который обеспечит распределение сегментов данных между различными репликасетам. Поддержка распределения данных по узлам кластера в соответствии с заданными критериями. Доступ к данным с любого узла кластера.

### Q1-2023 (23.01.0)

Доведение Picodata API до функционального состояния, обеспечение возможности создавать и удалять таблицы (spaces) в СУБД посредством Picodata API.

### Q2-2023 (23.04.0)

Реализация автоматического балансировщика данных в кластере, перемещающего данные от более наполненных узлов кластера к менее наполненным для сохранения равномерного распределения данных по узлам кластера.

Реализация частичной поддержки стандарта SQL:2016 в рамках всего кластера (поддержка распределенного SQL), включая элементы из подразделов: E011. Числовые типы данных, E011-05. Числовые сравнения, E021. Типы символьных строк.

### Q3-2023 (23.07.0)

Реализация распределенного механизма управления схемой данных СУБД (таблицы, хранимые процедуры, пользователи, привилегии), гарантирующего идентичную схему данных на всех узлах кластера.

Расширение поддержки стандарта SQL:2016 в рамках всего кластера (поддержка распределенного SQL), включая элементы из подразделов: E031. Идентификаторы, E051. Спецификация базовых запросов, E061. Базовые предикаты и условия поиска, E071. Базовые выражения с запросами, E101. Базовая обработка данных.

### Q4-2023 (23.10.0)

Интеграция Tarantool-Rust-модуля в основное приложение Picodata.

Реализация механизма выполнения задач в кластере в семантике “точно один раз”, “не более одного раза”.

Реализация механизма ролей — распределения вычислительных и прикладных задач по узлам кластера на основании их предназначения (роли), предоставляющего централизованное управление программой, выполняемой кластером.

Расширение поддержки стандарта SQL:2016 в рамках всего кластера (поддержка распределенного SQL), включая элементы из подразделов: F041. Базовое объединение таблиц, F471. Скалярные значения подзапросов, T631. Предикат IN с одним элементом списка.

## **9. Порядок контроля и приемки**

Контроль и приемка доработки осуществляются в соответствии с документом «Программа и методика испытаний», который разрабатывается на стороне испытательной лаборатории в рамках мероприятий по сертификации Программы.