|  |
| --- |
| **Проектное решение «HDP Озера данных БЛПС»** |

|  |  |
| --- | --- |
| **Код проекта** | **Наименование проекта** |
|  | Мониторинг технологических отклонений, включая параметры тех. режима и показатели качества. Разработка бизнес-системы «HDP Озеро данных БЛПС» |

|  |  |
| --- | --- |
| **Наименование Бизнес-решения** | **Наименование Бизнес-системы** |
| МО | HDP Озеро данных БЛПС |

|  |  |
| --- | --- |
| **Название документа:** | **Проектное решение «HDP Озера данных БЛПС»** |
| **Версия документа:** | 1.0 |
| **Дата документа:** | 05.06.2018 |
| **Назначение документа:** |  |
| **Аудитория:** |  |

**История изменений**

|  |  |  |  |
| --- | --- | --- | --- |
| **Версия** | **Дата** | **Комментарий** | **Автор** |
| 1.0. | 05.06.18 | Первоначальная версия подрядчика ЗАО «КРОК Инкорпорейтед» | Егоров П.В. |
| 1.1 | 16.05.19 | Актуализированная версия документа (с учетом работ по ЗНИ) | Егоров П.В. |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Связанные документы** (этот документ должен читаться вместе с)

|  |  |  |
| --- | --- | --- |
| **Название документа** | **Номер версии / Имя файла** | **Дата** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Согласовано:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Начальник [Управления технологической эффективности](http://portal.gazprom-neft.local/HR/Employees/Pages/search.aspx#c=1&fl=MPhotoUrl,Name,PublicUrl,NameLink,PublicUrlLink,JobTitle,Company,WorkPhone,MobilePhone,Email,City,Office,Location,DepartmentChain,ExtensionPhone&p=5&flv=,,,,,,,&d=76028) ЦУЭ БЛПС | Рыков Р.В. | Единое ответственное лицо |  |  |
| Главный специалист дирекции корпоративной защиты ДКЗ ББ | Горбунов А.А. | Эксперт |  |  |
| Начальник управления систем автоматизации бизнеса ДСУ БЛПС | Макеенко Д.В. | Куратор от ДСУ БЛПС |  |  |
| Начальник управления промышленной автоматизации БЛПС | Стариков В.А. | Куратор от ДСУ БЛПС |  |  |
| Руководитель направления трансформации бизнес – процессов ДСУ БЛПС | Боровикова Л.В.. | Руководитель проекта от заказчика |  |  |
| Руководитель направления  Инженерный центр, Управление систем операционной деятельности  ООО «Автоматика-сервис» | Еременко А.В. | Руководитель проекта от исполнителя |  |  |
| Руководитель направления ИТ архитектуры ДСУ БЛПС | Кобец А.В. | Архитектор БЛПС |  |  |
| Главный специалист ООО «ИТСК» Дирекция переработки и сбыта | Буткин Р.Ю. | Архитектор ИТ-решения 1С |  |  |
| Начальник управления аналитических исследований и метрологического обеспечения ООО «Автоматика-сервис» | Васбиев Д.Р. | Архитектор ИТ-решения HADOOP |  |  |

Оглавление

[1. Общие положения 6](#_Toc10119993)

[2. Термины и сокращения 7](#_Toc10119994)

[3. Архитектура решения 8](#_Toc10119995)

[3.1. Функционально-техническая архитектура решения 8](#_Toc10119996)

[3.2. Компоненты решения 8](#_Toc10119997)

[3.2.1. HBase 8](#_Toc10119998)

[3.2.2. Phoenix 9](#_Toc10119999)

[3.2.3. NiFi 9](#_Toc10120000)

[3.2.4. Kafka 9](#_Toc10120001)

[3.2.5. Spark 9](#_Toc10120002)

[3.3. Управление ресурсами 9](#_Toc10120003)

[3.4. Структура данных 11](#_Toc10120004)

[Реестр таблиц с описанием назначения 11](#_Toc10120005)

[Структура таблицы «Pi\_value\_tbl» 12](#_Toc10120006)

[Структура таблицы «lims\_value\_tbl» 12](#_Toc10120007)

[Структура таблицы «lims\_requirements\_tbl» 13](#_Toc10120008)

[Структура таблицы «pi\_requirements\_tbl» 14](#_Toc10120009)

[Структура таблиц «lims\_file\_tbl» 16](#_Toc10120010)

[3.5. Источники данных 16](#_Toc10120011)

[3.6. Расчетные алгоритмы 19](#_Toc10120012)

[3.6.1. Схема информационных потоков 20](#_Toc10120013)

[3.6.2. Получение данных по статистике значений 21](#_Toc10120014)

[Структура «Process1CValueExport» 22](#_Toc10120015)

[Структура «Process1CStaisticExport» 22](#_Toc10120016)

[Структура «Process1CDetailedStaisticExport» 22](#_Toc10120017)

[Структура «Process1CExport» 22](#_Toc10120018)

[4. Отказоустойчивость 23](#_Toc10120019)

[4.1 Zookeper 23](#_Toc10120024)

[4.2 HDFS 23](#_Toc10120025)

[4.3 YARN 24](#_Toc10120026)

[4.4 HBase 24](#_Toc10120027)

[4.5 NiFi 24](#_Toc10120028)

[4.6 Kafka 24](#_Toc10120029)

[Приложение 1 «Описание форматов сообщений» 25](#_Toc10120030)

[1.а Структура входящих данных из системы PI 25](#_Toc10120031)

[1.б Структура входящих данных из системы LIMS 25](#_Toc10120032)

[1.в Структура входящих данных по тех. режиму из системы 1с 26](#_Toc10120033)

[1.г Структура входящих данных по показателям качества из системы 1с 27](#_Toc10120034)

[1.д Структура исходящих данных по статическим показателям тех. режима 27](#_Toc10120035)

[1.е Структура исходящих данных по статическим показателям качества 28](#_Toc10120036)

[1.ж Структура исходящих данных по событиям (технологические показатели) 28](#_Toc10120037)

[1.з Структура исходящих данных по событиям (показатели качества) 29](#_Toc10120038)

[1.и Структура исходящих данных по детализированной статистике событий (показатели качества) 29](#_Toc10120039)

[1.к Структура исходящих данных по детализированной статистике событий (показатели тех. режима) 29](#_Toc10120040)

[1.л Структура исходящих данных по значениям показателей тех. режима 29](#_Toc10120041)

[1.м Структура исходящих данных по значениям показателей качества 31](#_Toc10120042)

# Общие положения

Настоящий документ является описанием архитектурной модели «HDP Озера данных БЛПС»

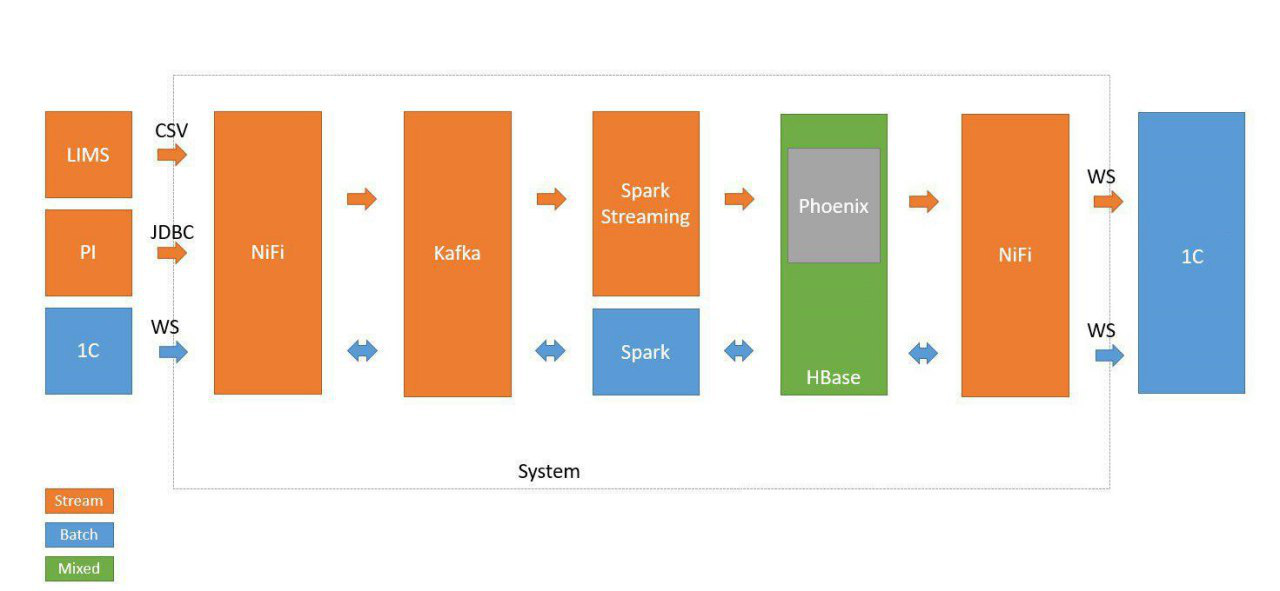
Добавления и изменения в проектное решение вносятся по согласованию Заказчика и Исполнителя и регистрируются в Листе регистрации изменений. По результатам проведения опытно-промышленной эксплуатации системы при необходимости выполняется актуализация проектного решения.

# Термины и сокращения

|  |  |  |
| --- | --- | --- |
| **№** | **Термин/**  **сокращение** | **Определение** |
|  | Заказчик | ООО «Автоматика-Сервис» |
|  | Исполнитель | ЗАО «КРОК инкорпорейтед» |
|  | БЛПС | Блок логистики, переработки и сбыта |
|  | Система МО | Система мониторинга технологических отклонений, включая тех. режимы и качества |
|  | Внешние системы, системы-источники данных | Смежные информационные системы к Системе МО, по которым будут разработаны интеграционные требования (PI, LIMS, лента событий ЦУЭ и др.) и которые будут подключены к Системе МО |
|  | ДСУ | Департамент систем управления Компании |
|  | МНПЗ | Московский нефтеперерабатывающий завод |
|  | ОНПЗ | Омский нефтеперерабатывающий завод |
|  | НТР | Нормы технологического режима |
|  | ТБС | Таблица блокировок и сигнализаций (раздел технологических регламентов установок содержащий перечень блокировок и сигнализаций) |
|  | ЦУЭ | Центр управления эффективностью БЛПС |
|  | HDP Озеро данных БЛПС | Бизнес-система, входящая в бизнес-решение МО (Мониторинг отклонений). |
|  | НМД | Нормативно-методическая документация, в т.ч. технологические регламенты, организационные стандарты |
|  | Отклонение | Отклонения фактических параметров от норм, зафиксированных в НМД |
|  | Параметры технологического режима, Параметры | Величина, значение которой, характеризуют технологический процесс: давления, температуры, расходы, и т.п. |
|  | PI-System | Система хранения фактических параметров технологических установок на ОНПЗ и на МНПЗ |
|  | LIMS | Система хранения фактических данных лабораторного контроля качества на ОНПЗ и на МНПЗ |
|  | ЕСУ НСИ | Единая система управления НСИ |
|  | ОПЭ | Опытно-промышленная эксплуатация |

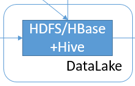
# Архитектура решения

## Функционально-техническая архитектура решения



## Компоненты решения

## HBase

HBase – нереляционная распределённая колоночно-ориентированная, мультиверсионная база данных типа «ключ-значение», с открытым исходным кодом. Используется для задач, в которых необходим доступ на чтение/запись данных в режиме near real-time.

Данные HBase организованы в таблицы, которые проиндексированы RowKey (первичный ключ). Для каждого RowKey ключа может храниться неограниченный набор атрибутов (или колонок).

Column Family – набор колонок таблицы HBase. Как правило в одну Column Family объединяют колонки, для которых одинаковы паттерн использования и хранения.

Для каждого атрибута может храниться несколько различных версий, каждая из которых имеет свой timestamp.

Физически записи в таблице HBase хранятся в отсортированном по RowKey порядке. При этом данные, соответствующие разным Column Family, хранятся отдельно, что позволяет, при необходимости, читать данные только из нужного семейства колонок.

Атрибуты, принадлежащие одной Column Family и соответствующие одному ключу, физически хранятся как отсортированный список. Любой атрибут может отсутствовать или присутствовать для каждого ключа, при этом если атрибут отсутствует — это не вызывает накладных расходов на хранение пустых значений.

Схематичное представление Table Hbase (частный случай) приведено на рисунке (Рисунок 1).

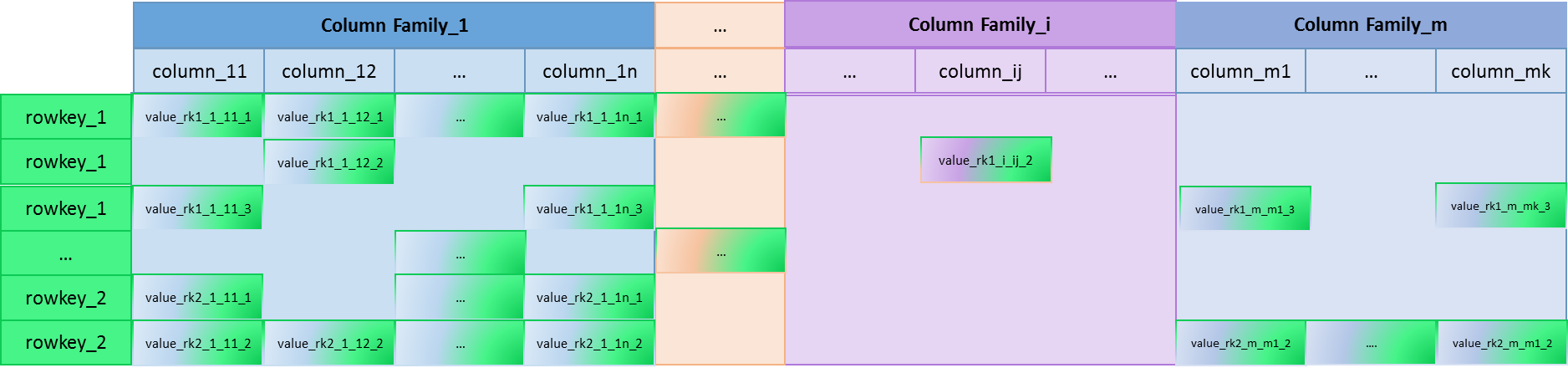


Рисунок 1 - Схематичное представление Table Hbase (частный случай)

## Phoenix

Используется как дополнительный слой над HBase. Реализует JDBC интерфейс над HBase с целью исполнения SQL-запросов над данными. Кроме этого добавляет две весьма важных для проекта функциональных возможности.

1. Позволяет хранить типизированные данные в системе, а не только byte array.
2. Позволяет строить вторичные индексы над данными. То есть добавляет возможность доступа к данным не по первичному ключу.

## NiFi

NiFi реализует концепцию Data Flow Programming, которая включает в себя набор процессоров – асинхронных модулей обработчиков, и связей – фактически очередей между ними. Данный продукт не исполняется в рамках YARN и рекомендуется к установке на выделенный кластер. Однако в отличии от большинства других систем обработки потоковых сообщений в экосистеме Hadoop, он масштабируется в первую очередь по количеству сообщений, а не по количеству интеграционных интерфейсов. Это позволяет реализовать большое количество интеграционных интерфейсов средней интенсивности на достаточно кластере умеренной мощности.

## Kafka

Высокопроизводительная отказоустойчивая распределенная MOM (Message Oriented Middleware) система. Предназначена для организации асинхронного взаимодействия компонентов посредство асинхронного обмена сообщениями.

## Spark

Spark – быстрый и общий механизм для крупномасштабной распределённой обработки [неструктурированных](https://ru.wikipedia.org/wiki/%D0%9D%D0%B5%D1%81%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D0%B8%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5) и слабоструктурированных данных (открытый исходный код), для которого верны следующие утверждения:

* скорость наличие усовершенствованного механизма выполнения DAG, который поддерживает ациклический поток данных и вычисления в памяти;
* простота использования – возможность использования большого количества операторов высокого уровня для упрощения сбора параллельных приложений, в том числе в интерактивном режиме из Scala, Python;
* generality – возможность использования стека библиотек, в том числе следующих:

## Управление ресурсами

В текущей архитектуре существует несколько основных компонентов, которые потребляют вычислительные ресурсы кластера – это HBase, YARN, Kafka и NiFi.

При этом выделяются несколько основных сегментов кластера – сегмент Operational Profile и сегмент DataLake. Каждый из этих сегментов выполняет свою роль в кластере, имеет свой профиль нагрузки и поэтому желательно выделить отдельные физические сегменты кластера для них.

В случае небольшой нагрузки большинство этих сегментов допустимо располагать на одном физическом кластере. Однако при возрастании нагрузки, данные сегменты рекомендуется физически разносить.

Для составления требований к оборудованию кластера, необходимо провести нагрузочное тестирование с профилями нагрузки, близкими к целевым показателям системы.

На уровне сегмента Data Lake управление ресурсами кластера происходит на уровне YARN. Именно YARN управляет основными ресурсами кластера и именно он выделяет ресурсы приложениям, которые требуют их для исполнения локальных задач. При планировании распределения ресурсов необходимо придерживаться рекомендации вендора и оставлять достаточно памяти на ОС, JVM и другие сервисы системы.

На уровне YARN также необходимо распланировать распределение и приоритет ресурсов по очередям. Каждое запускаемое приложение или процесс должно относиться к соответствующей очереди и для нее должен выделяться соответствующий пул ресурсов. Детальное планирование очередей ресурсов должно производиться на уровне детального проектирования системы.

На уровне архитектуры для управления ресурсами кластера предлагается использование Fair Scheduler и создание соответствующих очередей в нем. Особенностью fair scheduler является то, что в рамках очереди все процессы получают равное количество ресурсов (см. Рисунок 6). Однако важно учитывать, что при исполнении Spark приложений в распределении ресурсов участвует дополнительный механизм управления ресурсами самого Spark и высвобождение ресурсов внутри приложения может происходить с существенным запаздыванием или не происходить вовсе до окончания работы Spark приложения.

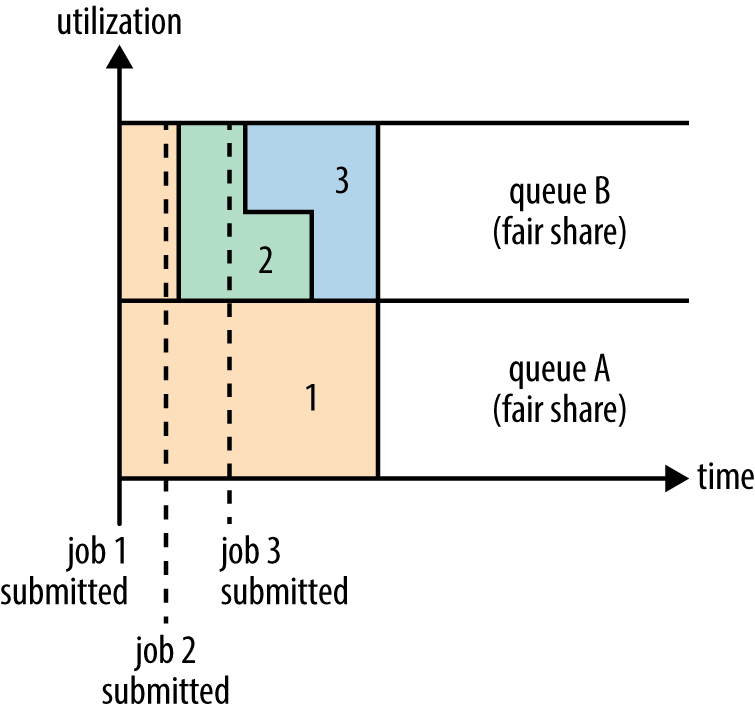


Рисунок 6 - Распределение ресурсов. Fair scheduler

## Структура данных

## Реестр таблиц с описанием назначения

|  |  |
| --- | --- |
| **Название таблицы** | **Назначение и описание** |
| lims\_Value\_tbl | Таблица предназначена для хранения данных по всем событиям пришедшим из системы Lims |
| Pi\_Value\_tbl | Таблица предназначена для хранения данных по всем событиям пришедшим из системы Pi |
| pi\_requirements\_tbl | Таблица предназначена для хранения данных Нормативам PI |
| Lims\_requirements\_tbl | Таблица предназначена для хранения данных Нормативам LIMS |
| Pi\_Range\_output\_tbl | Таблица предназначена для хранения периода выхода (событий) показателей PI |
| Lims\_Range\_output\_tbl | Таблица предназначена для хранения периода выхода (событий) показателей LIMS |
| Lims\_file\_tbl | Таблица предназначена для хранения информации об обработанных данных, поступающих из LIMS во избежание дублирования |

## Структура таблицы «Pi\_value\_tbl»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Ключевое поле | Описание |
| tag\_name | varchar | Да | Да | Наименование показателя |
| tag\_value | varchar |  |  | Значение показателя |
| unit | varchar |  |  | Единицы измерения |
| Date\_request | timestamp | Да | Да | Дата запроса |
| Date\_response | timestamp | Да | Да | Дата ответа из ГПН |
| status | Integer |  |  | Статус (Quality - good,bad) |
| Refinery\_ID | integer |  |  | НПЗ |

## Структура таблицы «lims\_value\_tbl»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Ключевое поле | Описание |
| Date | Timestamp | Да | Да | Дата/время |
| refinery\_ID | Integer | Да | Да | Код НПЗ |
| unit\_name | varchar |  |  | Наименование установки в ЛИМС |
| selection\_point\_ID | varchar |  |  | Код точки отбора в ЛИМС |
| selection\_point\_name | varchar | Да | Да | Наименование точки отбора в ЛИМС |
| product\_code | varchar |  |  | Код продукта |
| product\_name | varchar | Да | Да | Наименование продукта в ЛИМС |
| product\_sort | varchar |  |  | Сорт детализирует продукт. Например, Продукт  - Бензин, Сорт - фр. бензиновая НК 80-180, или Продукт - АИ-95, Сорт - Бензин автомобильный АИ-95 Кл. 5 |
| quality\_index\_code | varchar |  |  | Код показателя качества в ЛИМС |
| quality\_index\_name | varchar | Да | Да | Наименование показателя качества в ЛИМС |
| quality\_index\_type | varchar |  |  | Группа, к которой относится показатель (например, фракционный состав, плотность и т.п.) |
| quality\_index\_value | varchar |  |  | Значение показателя качества |
| quality\_index\_low\_level | Double |  |  | Нижняя граница нормы по показателю качества |
| quality\_index\_hi\_level | Double |  |  | Верхняя граница нормы по показателю качества |
| quality\_index\_units | varchar |  |  | Единицы измерения показателя качества |
| selection\_no | varchar |  |  | № образца (пробы) |
| quality\_certificate\_no | varchar |  |  | Номер паспорта товарного резервуара, либо пустая строка |
| passport\_tank\_mass | Double |  |  | Тоннаж, на который резервуар был паспортизован |
| passport\_tank\_level | Double |  |  | Уровень наполнения, на который резервуар был паспортизован |
| defect\_tank\_comments | varchar |  |  | Комментарий к некондиционному резервуару |
| Product\_brand | varchar |  |  | Марка продукта |

## Структура таблицы «lims\_requirements\_tbl»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Ключевое поле | Описание |
| Date | timestamp | Да | Да | Дата |
| id\_metrics | varchar | Да | Да | Показатель |
| id\_product | varchar |  | Да | Продукт |
| id\_point | varchar |  | Да | Точка сбора |
| norma\_min | varchar |  |  | Есть Значение Мин |
| norma\_min\_value | double |  |  | Значение Мин |
| norma\_max | varchar |  |  | ЕстьЗначение Макс |
| norma\_max\_value | double |  |  | Значение Макс |
| Unit | varchar |  | Да | ед. изм. |
| Activity | varchar |  |  | Активность записи |
| Refinery\_ID | integer |  |  | Код завода |
| Product\_sort | varchar |  | Да | Сорт продукта |
| Product\_brand | varchar |  | Да | Марка продукта |
| Quality\_index\_type | varchar |  | Да | Показатель качества |
| create\_date | timestamp |  |  | Дата создания |

## Структура таблицы «pi\_requirements\_tbl»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Ключевое поле | Описание |
| Date | timestamp | Да | Да | Дата начала действия |
| id\_metrics | varchar | Да | Да | Показатель |
| Unit | varchar |  | Да | ЕдиницаИзмерения PI |
| Ratio | double |  |  | Коэффициент пересчета ед. изм. |
| ratio\_formula | integer |  |  | Коэффициент пересчета ед. изм. |
| ratio\_formula\_pi\_id\_1 | varchar |  |  | Коэффициент пересчета ед. изм. Показатель 1 |
| ratio\_formula\_pi\_id\_2 | varchar |  |  | Коэффициент пересчета ед. изм. Показатель 2 |
| norma\_min | varchar |  |  | ЕстьЗначение Норма Мин |
| norma\_min\_value | double |  |  | Значение Норма Мин |
| norma\_max | varchar |  |  | ЕстьЗначение Норма Макс |
| norma\_max\_value | double |  |  | Значение Норма Макс |
| Formula | integer |  |  | Формула Вид |
| formula\_pi\_ id\_1 | varchar |  |  | Формула Показатель 1 |
| formula\_pi\_id\_2 | varchar |  |  | Формула Показатель 2 |
| criticallevel\_min | varchar |  |  | Есть Значение Критическое Мин |
| criticallevel\_min\_value | double |  |  | Значение Критическое Мин |
| criticallevel\_max | varchar |  |  | Есть Значение Критическое Макс |
| criticallevel\_max\_value | double |  |  | Значение Критическое Макс |
| alarmlevel\_min | varchar |  |  | Есть Значение Сигнализация Мин 1 |
| alarmlevel\_min\_value | double |  |  | Значение Сигнализация Мин 1 |
| alarmlevel\_max | varchar |  |  | Есть Значение Сигнализация Макс 1 |
| alarmlevel\_max\_value | Double |  |  | Значение Сигнализация Макс 1 |
| blockinglevel\_min | varchar |  |  | Есть Значение Блокировка Мин 1 |
| blockinglevel\_min\_value | double |  |  | Значение Блокировка Мин 1 |
| blockinglevel\_max | varchar |  |  | Есть Значение Блокировка Макс 1 |
| blockinglevel\_max\_value | double |  |  | Значение Блокировка Макс 1 |
| Activity | varchar |  |  | Актуальность записи |
| Refinery\_ID | integer | Да | Да | Код завода |
| Workmode | varchar |  | Да | Режим работы |
| Accuracy | integer |  |  | Точность (число знаков после запятой) |
| Create\_date | timestamp |  |  | Дата создания |

**Структура таблиц «Pi\_Range\_output\_tbl»**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Ключевое поле | Описание |
| id\_metrics | varchar | Да | Да | ID Показатель |
| Refinery\_id | integer | Да | Да | НПЗ |
| Date\_start | timestamp | Да | Да | Дата и время возникновения события |
| Date\_end | timestamp |  |  | Дата и время завершения события |
| Event\_code | integer | Да | Да | Событие |
| Metrics\_value | double |  |  | Значение показателя |
| workmode | varchar |  | Да | Режим работы |
| difference | double |  |  | Разница между значением и границей нормы |
| flag | Integer |  |  | Значение полученное выгрузкой из 1С (0-значение не получено, 1-значение получено) |

**Структура таблиц «LIMS\_Range\_output\_tbl»**

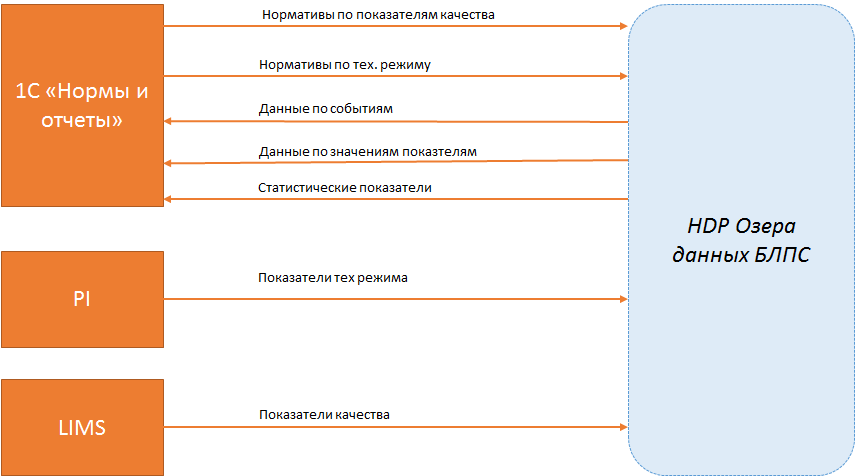
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Ключевое поле | Описание |
| id\_metrics | varchar | Да | Да | ID Показатель |
| Id\_product | varchar | Да | Да | ID Продукта |
| Id\_point | varchar | Да | Да | ID Места отбора пробы |
| Refinery\_id | integer | Да | Да | НПЗ |
| Date\_start | timestamp | Да | Да | Дата и время возникновения события |
| Date\_end | timestamp |  |  | Дата и время завершения события |
| Event\_code | integer | Да | Да | Событие |
| Metrics\_value | double |  |  | Значение показателя |
| Unit\_name | varchar |  | Да | Единицы измерения |
| difference | double |  |  | Разница между значением и границей нормы |
| flag | Integer |  |  | Значение полученное выгрузкой из 1С (0-значение не получено, 1-значение получено) |
| Product\_sort | varchar |  | Да | Сорт продукта |
| Product\_brand | varchar |  | Да | Марка продукта |
| Quality\_index\_type | varchar |  | Да | Показатель качества |

## Структура таблиц «lims\_file\_tbl»

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Ключевое поле | Описание |
| date | date |  |  | Дата файла |
| filename | varchar | Да | Да | Наименование полученного файла из системы LIMS |

## Источники данных

Схема интеграционных интерфейсов



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **№** | **Бизнес-процесс** | **Шаг процесса** | **ИС отправитель** | **Интерфейс (отправителя)** | **Сообщение (отправителя)** | **Формат (отправителя)** | **ИС получатель** |  |
| **Описание взаимодействия, примечания** |
| 1 | Получение данных значений показателей тех режима (PI) | Системная операция 1 | PI SYSTEM ОНПЗ, PI SYSTEM МНПЗ | Rest сервис (gRPC gate way) | Структура представлена в приложении 1.а | Получение данных через интерфейс системы | HDP Озера данных БЛПС | Получение данных каждые 2 минуты |
| 2 | Получение данных значений показателей качества (LIMS) | Системная операция 1 | LIMS ОНПЗ, LIMS МНПЗ | Выгруженные текстовые файлы в папке | Структура представлена в приложении 1.б | Текстовый файл | HDP Озера данных БЛПС | Получение данных каждые 2 минуты |
| 3 | Получение нормативов (1с) по тех. режиму | Системная операция 1 | 1С «Нормы и отчеты» | Rest сервис | Структура представлена в приложении 1.в | Json | HDP Озера данных БЛПС | Отправка данные по времени изменения в системе источнике |
| 4 | Получение нормативов (1с) по показателям качества | Системная операция 1 | 1С «Нормы и отчеты» | Rest сервис | Структура представлена в приложении 1.г | Json | HDP Озера данных БЛПС | Отправка данные по времени изменения в системе источнике |
| 5 | Отправка данных по статистике отклонений показателей тех. режима | Системная операция 1 | HDP Озера данных БЛПС | Rest сервис | Структура представлена в приложении 1.д | Json | 1С «Нормы и отчеты» | Отправка данных по запросу |
| 6 | Отправка данных по статистике отклонений показателей качества | Системная операция 1 | HDP Озера данных БЛПС | Rest сервис | Структура представлена в приложении 1.е | Json | 1С «Нормы и отчеты» | Отправка данных по запросу |
| 7 | Отправка данных по событиям (технологические показатели) | Системная операция 1 | HDP Озера данных БЛПС | Rest сервис | Структура представлена в приложении 1.ж | Json | 1С «Нормы и отчеты» | Отправка данных по запросу |
| 8 | Отправка данных по событиям (показатели качества) | Системная операция 1 | HDP Озера данных БЛПС | Rest сервис | Структура представлена в приложении 1.з | Json | 1С «Нормы и отчеты» | Отправка данных по запросу |
| 9 | Отправка данных по детализированной статистике отклонений показателей тех. режима | Системная операция 1 | HDP Озера данных БЛПС | Rest сервис | Структура представлена в приложении 1.и | Json | 1С «Нормы и отчеты» | Отправка данных по запросу |
| 10 | Отправка данных по детализированной статистике отклонений показателей качества | Системная операция 1 | HDP Озера данных БЛПС | Rest сервис | Структура представлена в приложении 1.к | Json | 1С «Нормы и отчеты» | Отправка данных по запросу |
| 11 | Отправка данных по значениям показателей тех. режима | Системная операция 1 | HDP Озера данных БЛПС | Rest сервис | Структура представлена в приложении 1.л | Json | 1С «Нормы и отчеты» | Отправка данных по запросу |
| 12 | Отправка данных по значениям показателей качества | Системная операция 1 | HDP Озера данных БЛПС | Rest сервис | Структура представлена в приложении 1.м | Json | 1С «Нормы и отчеты» | Отправка данных по запросу |

## Расчетные алгоритмы

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Номер п/п | Наименование алгоритма | Среда выполнения | Таблицы источники | Ключевые поля для связи | Взаимодействие |
| 1 | Расчет нормативов и отслеживание событий | Spark streaming | **По PI:**  Поток PI (1а) и pi\_requirements\_tbl  **По LIMS:**  Поток LIMS (1б) и lims\_requirements\_tbl | **По PI:**  RefeneryID, TagName  WorkMode  **По LIMS:** SelectionPointName  ProductName  QualityIndexName | Hbase – в части чтения нормативов, оконные операции Spark |
| 2 | Получение данных по статистике значений | Spark | **По параметрам тех. режима:** pi\_requirements\_tbl, pi\_warning\_tbl, pi\_range\_output\_tbl  По показателям качества:  lims\_requirements\_tbl, lims\_warning\_tbl, lims\_range\_output\_tbl | **По параметрам тех. режима:** RefeneryID, TagName  WorkMode  **По показателям качества:** SelectionPointName  ProductName  QualityIndexName | Nifi – при запросе данных со стороны 1С «Нормы и отчеты» рассчитываются статистические показатели |

## Схема информационных потоков

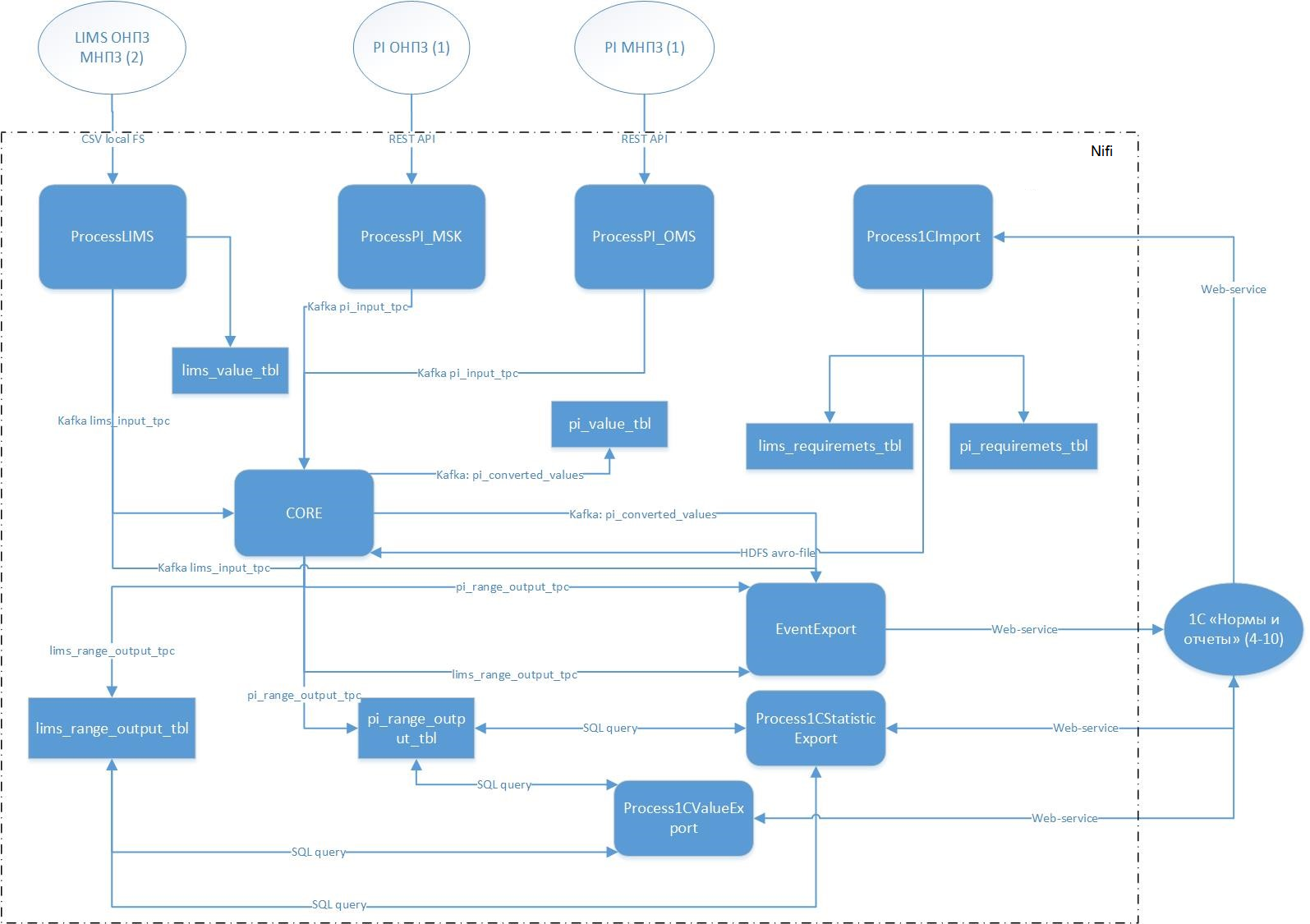


Таблица расшифровки

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Наименование процесса | Наименование компонента | Описание | Компонент |
| 1 | Получение нормативов по показателям качества | Process1CImport | Интеграция 1С «Нормы и отчеты» | Nifi, phoenix, hbase |
| 2 | Получение нормативов по показателям технологического режима | Process1CImport | Интеграция 1С «Нормы и отчеты» | Nifi, phoenix, hbase |
| 3 | Получение значений показателей качества | ProcessLIMS | Интеграция LIMS | Nifi, kafka, phoenix, hbase |
| 4 | Получение значений показателей технологического режима | ProcessPI\_OMS  ProcessPI\_MSK | Интеграция PI | Nifi, kafka, phoenix, hbase |
| 5 | Получение архивных значений показателей технологического режима | ProcessPIArchive\_OMS  ProcessPIArchive\_MSK | Интеграция PI | Nifi, kafka, phoenix, hbase |
| 6 | Расчет значений | Core | Расчет отклонений для PI и LIMS реализованный на базе Spark | Spark streaming |
| 7 | Хранение данных | Pi\_value\_tbl, Lims\_value\_tbl, Pi\_range\_output\_tbl, Lims\_range\_output\_tbl, Pi\_reqirments\_tbl, lims\_requirments\_tbl | Хранение нормативов и расчетных данных | Hbase с интерфейсом Phoenix |
| 8 | Экспорт данных по статистике отклонений в 1с | Process1CStatisticExport | Экспорт значений по запросу из 1с | Nifi, phonix, hbase |
| 9 | Экспорт данных по текущим отклонениям в 1с | EventExport | Экспорт отклонений в 1с в режиме near real time | Nifi, kafka |
| 10 | Экспорт данных по текущим значениям в 1с | Process1cValueExport | Экспорт текущих значений в 1с в режиме near real time | Nif, kafka |

## Получение данных по статистике значений

Расчет статистических значений будет производиться во время обращения из 1С «Нормы и отчеты» в HDP Озера данных БЛПС (интеграционный интерфейс №6 по таблице «Реестр интеграционных интерфейсов»).

Источник данных для расчета: Pi\_value\_tbl, Lims\_value\_tbl

Параметры отбора: Период данных для расчета, НПЗ

Показатели расчета:

* Количество выходов показателя за пределы нормативного значения
* Количество выходов показателя за пределы критического значения
* Количество выходов показателя за пределы уровней сигнализации
* Количество выходов показателя за пределы уровней блокировок
* Время нахождения показателя за пределами нормативного значения
* Время нахождения показателя за пределами критического значения
* Время нахождения показателя за пределами уровня сигнализации
* Время нахождения показателя за пределами уровня блокировок
* Минимальное/максимальное значение показателя за пределами нормативного значения
* Минимальное/максимальное значение показателя за пределами критического значения
* Минимальное/максимальное значение показателя за пределами уровня сигнализации
* Минимальное/максимальное значение показателя за пределами уровня блокировки

## Структура «Process1CValueExport»

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| Date\_start | string | Да | Дата начала периода |
| Date\_end | string | Да | Дата окончания периода |
| Refinery\_ID | Integer | Да | Код завода |
| tags | string | Нет | Имя тэга |
| method | string | Да | Функция (pi\_values, lims\_values) |

## Структура «Process1CStaisticExport»

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| Date\_start | string | Да | Дата начала периода |
| Date\_end | string | Да | Дата окончания периода |
| Refinery\_ID | Integer | Да | Код завода |
| tags | string | Нет | Имя тэга |
| method | string | Да | Функция (pi\_metrics\_statistics\_export, lims\_metrics\_statistics\_export) |

## Структура «Process1CDetailedStaisticExport»

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| Date\_start | string | Да | Дата начала периода |
| Date\_end | string | Да | Дата окончания периода |
| Refinery\_ID | Integer | Да | Код завода |
| tags | string | Нет | Имя тэга |
| method | string | Да | Функция (pi\_metrics\_statistics\_export, lims\_metrics\_statistics\_export) |

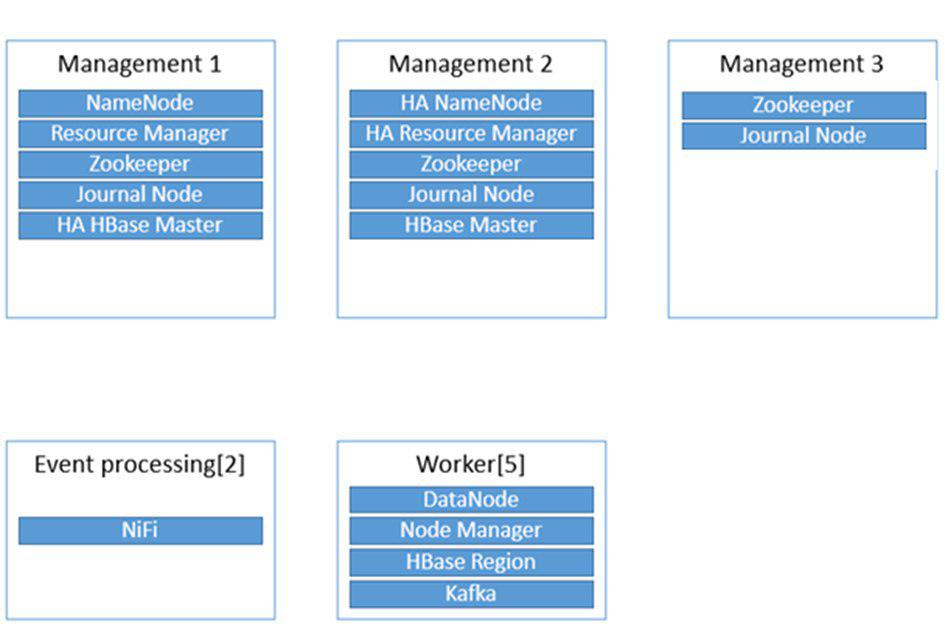
## Структура «Process1CExport»

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| Refinery\_ID | Integer | Да | Код завода |
| method | string | Да | Функция (pi, lims) |

# Отказоустойчивость

Основная идея данной схемы заключается в том, что есть три управляющих узла (Management 1‑3) и два рабочих сегмента (Event Processing, Worker), которые несут основную нагрузку. В случае отказа основного узла или какого-либо из управляющих приложений, расположенных на этом узле, происходить перераспределение обращений к failover-сервисам. Каждый из сегментов также допускает потерю как минимум одного узла, что обеспечивает высокий уровень отказоустойчивости всей системы.

Рисунок 7 - Диаграмма развертывания компонентов на кластере и роли узлов.



# Zookeper

Необходимо наличие не менее трех экземпляров Zookeeper в кластере, что даст возможность потери одного узла. Данный компонент обеспечивает гарантию отказоустойчивости других сервисов.

# HDFS

Для обеспечения отказоустойчивости HDFS необходимо наличие standby-экземпляра сервиса NameNode. Данный экземпляр обозначен на рисунке HA NameNode. Также на каждый из узлов, обеспечивающих функциональность NameNode, устанавливается по одному экземпляру JournalNode. Но для отказоустойчивости и возможности автоматического переключения активной NameNode на standby, необходимо наличие дополнительного сервиса JournalNode. Эти сервисы выполняют функцию write-ahead логов и позволяют корректно отследить изменения в случае падения NameNode в процессе операции записи.

# YARN

Для обеспечения отказоустойчивости менеджера ресурсов необходимо наличие standby-экземпляра Resource Manager. Он обозначен на рисунке как HA Resource Manager

# HBase

Для обеспечения отказоустойчивости мастера HBase необходимо наличие standby-экземпляра. Он обозначен на рисунке как HA HBase Master. Узлы обработки данных HBase обозначены на схеме как HBase Region.

# NiFi

NiFi рекомендуется разворачивать на отдельном сегменте кластера, состоящем не менее чем из двух узлов. Отказоустойчивость обеспечивается наличием второго и последующих экземпляров сервиса. Однако важно отметить, что в случае отказа одного из узлов сегмента, данные которые в данный момент находятся в обработке сервисом NiFi на данном узле не реплицируются на другие узлы, и в случае падения будут обработаны только после перезапуска упавшего узла.

# Kafka

Размещается на сегменте кластера Worker. Отказоустойчивость обеспечивается за счет использования одноранговой репликации данных между узлами. При установке крайне рекомендуется установка на отдельные диски, объединенные в RAID 1.

# Приложение 1 «Описание форматов сообщений»

### 1.а Структура входящих данных из системы PI

Функция: GetCurrentValues(CurrentValuesRequest) returns (ValuesResponse)

Пример запроса:

Запрос:

{"tags":["AT\_9:FIR1007","AT\_9:FIRC1070"]}

Структура возвращаемых значений ValuesResponse

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| tag | string | Да | Имя тега (показателя) |
| unit | string | Да | Ед. изм |
| value | string | Да | Значение тега (показателя) |
| date | date | Да | Дата/Время |
| stat | integer |  | Статус |

Пример ответа:

{"values":[{"tag":"AT\_9:FIR1007","unit":"м3/ч","values":[{"value":{"dval":377.2156982421875},"stat":0,"timedate":"2018-07-03T09:34:41Z"}]},{"tag":"AT\_9:FIRC1070","unit":"м3/ч","values":[{"value":{"dval":9.532805442810059},"stat":0,"timedate":"2018-07-03T09:31:41Z"}]}]}

### 1.б Структура входящих данных из системы LIMS

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| date | date | Да | Дата/время |
| refinery\_ID | integer | Да | Код НПЗ |
| unit\_name | string | Да | Наименование установки в ЛИМС |
| selection\_point\_ID | string | Да | Код точки отбора в ЛИМС |
| selection\_point\_name | string | Да | Наименование точки отбора в ЛИМС |
| product\_code | string | Да | Код продукта |
| product\_name | string | Да | Наименование продукта в ЛИМС |
| product\_sort | string | Нет | Сорт детализирует продукт. Например, Продукт  - Бензин, Сорт - фр. бензиновая НК 80-180, или Продукт - АИ-95, Сорт - Бензин автомобильный АИ-95 Кл. 5 |
| quality\_index\_code | string | Да | Код показателя качества в ЛИМС |
| quality\_index\_name | string | Да | Наименование показателя качества в ЛИМС |
| quality\_index\_type | string | Нет | Группа, к которой относится показатель (например, фракционный состав, плотность и т.п.) |
| quality\_index\_value | string | Да | Значение показателя качества |
| quality\_index\_low\_level | double | Да | Нижняя граница нормы по показателю качества |
| quality\_index\_hi\_level | double | Да | Верхняя граница нормы по показателю качества |
| quality\_index\_units | string | Да | Единицы измерения показателя качества |
| selection\_no | string | Нет | № образца (пробы) |
| quality\_certificate\_no | string | Нет | Номер паспорта товарного резервуара, либо пустая строка |
| passport\_tank\_mass | Double | Нет | Тоннаж, на который резервуар был паспортизован |
| passport\_tank\_level | Double | Нет | Уровень наполнения, на который резервуар был паспортизован |
| defect\_tank\_comments | String | Нет | Комментарий к некондиционному резервуару |

### 1.в Структура входящих данных по тех. режиму из системы 1с

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| date | date | Да | Дата начала действия |
| id\_metrics | string | Да | ID показателя |
| Unit | string | Нет | ЕдиницаИзмеренияPI |
| norma\_min | boolean | Да | Есть значение Норма Мин |
| norma\_min\_value | double | Нет | Значение Норма Мин |
| norma\_max | boolean | Да | Есть значение Норма Макс |
| norma\_max\_value | double | Нет | Значение Норма Макс |
| Formula | integer | Да | Формула Вид |
| formula\_pi\_id\_1 | string | Нет | Формула Показатель 1 |
| formula\_pi\_id\_2 | string | Нет | Формула Показатель 2 |
| criticallevel\_min | boolean | Да | Есть значение Критическое Мин |
| criticallevel\_min\_value | double | Нет | Значение Критическое Мин |
| criticallevel\_max | boolean | Да | Есть значение Критическое Макс |
| criticallevel\_max\_value | double | Нет | Значение Критическое Макс |
| alarmlevel\_min | boolean | Да | Есть значение Сигнализация Мин 1 |
| alarmlevel\_min\_value | double | Нет | Значение Сигнализация Мин 1 |
| alarmlevel\_max | boolean | Да | Есть значение Сигнализация Макс 1 |
| alarmlevel\_max\_value | double | Нет | Значение Сигнализация Макс 1 |
| blockinglevel\_min | boolean | Да | Есть значение Блокировка Мин 1 |
| blockinglevel\_min\_value | double | Нет | Значение Блокировка Мин 1 |
| blockinglevel\_max | boolean | Да | Есть значение Блокировка Макс 1 |
| blockinglevel\_max\_value | double | Нет | Значение Блокировка Макс 1 |
| Ratio | Double | Да | Коэффициент пересчета |
| Ratio\_Formula | Double | Да | Тип формулы пересчета ед. изм |
| Ratio\_Formula\_id1 | Double | Да | Показатель пересчета 1 |
| Ratio\_Formula\_Id2 | Double | Да | Показатель пересчета 2 |
| Refinery\_Id | Integer | Да | НПЗ |
| Active | Boolean | Да | Признак активности нормы |
| Workmode | String | Нет | Режим работы |
| Accuracy | Integer | Нет | Точность (число знаков после запятой) |

### 1.г Структура входящих данных по показателям качества из системы 1с

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| date | Date | Да | Период |
| UnitName | string | Да | Наименование установки |
| QualityIndexName | string | Да | ID показателя |
| ProductName | string | Нет | ID продукта |
| SelectionPointName | string | Нет | ID места отбора пробы |
| RefinaryId | string | Нет | НПЗ |
| Active | string | Нет | Действительность норматива |
| norma\_min | boolean | Да | Есть Значение Мин |
| norma\_min\_value | double | Нет | Значение Мин |
| norma\_max | boolean | Да | Есть Значение Макс |
| norma\_max\_value | double | Нет | Значение Макс |

### 1.д Структура исходящих данных по статическим показателям тех. режима

|  |  |  |
| --- | --- | --- |
| Название поля | Тип данных | Комментарий |
| id\_metrics | String | ID показателя |
| out\_norma\_times | Integer | Количество выходов показателя за пределы нормативного значения |
| out\_crit\_times | Integer | Количество выходов показателя за пределы критического значения |
| out\_alarm\_times | Integer | Количество выходов показателя за пределы уровней сигнализации |
| out\_block\_times | Integer | Количество выходов показателя за пределы уровней блокировок |
| out\_norma\_duration | Double | Время нахождения показателя за пределами нормативного значения |
| out\_crit\_duration | Double | Время нахождения показателя за пределами критического значения |
| out\_alarm\_duration | Double | Время нахождения показателя за пределами уровня сигнализации |
| out\_block\_duration | Double | Время нахождения показателя за пределами уровня блокировок |
| out\_norma\_max\_value | Double | Минимальное/максимальное значение показателя за пределами нормативного значения |
| out\_crit\_max\_value | Double | Минимальное/максимальное значение показателя за пределами критического значения |
| out\_alarm\_max\_value | Double | Минимальное/максимальное значение показателя за пределами уровня сигнализации |
| out\_block\_max\_value | Double | Минимальное/максимальное значение показателя за пределами уровня блокировки |
| Refinery\_ID | Integer | НПЗ |

### 1.е Структура исходящих данных по статическим показателям качества

|  |  |  |
| --- | --- | --- |
| Название поля | Тип данных | Комментарий |
| id\_metrics | String | ID показателя |
| id\_product | String | Продукт |
| id\_point | String | Точка отбора пробы |
| out\_norma\_times | Integer | Количество выходов показателя за пределы нормативного значения |
| out\_norma\_duration | Double | Время нахождения показателя за пределами нормативного значения |
| out\_norma\_max\_value | Double | Минимальное/максимальное значение показателя за пределами нормативного значения |
| Refinery\_ID | Integer | НПЗ |

### 1.ж Структура исходящих данных по событиям (технологические показатели)

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| id\_metrics | String | Да | Имя показателя |
| Date\_start | string | Да | Дата |
| Date\_end | string |  |  |
| metrics\_value | Double | Да | Значение |
| event\_code | Integer | Да | Код события |
| Refinery\_ID | Integer | Да | НПЗ |
| Difference | string |  | Разность |
| Workmode | string |  | Режим работы |

### 1.з Структура исходящих данных по событиям (показатели качества)

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| id\_metrics | string | Да | Показатель |
| id\_product | String | Да | Продукт |
| id\_point | String | Да | Точка отбора пробы |
| Date\_start | datetime | Да | Дата старта |
| Date\_end | datetime |  | Дата окончания |
| metrics\_value | Double |  | Значение |
| event\_code | Integer | Да | Код события |
| Refinery\_ID | Integer | Да | НПЗ |
| Difference | string |  | Разность |
| Unit | string |  | Ед. имзмерения |

### 1.и Структура исходящих данных по детализированной статистике событий (показатели качества)

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Комментарий |
| id\_point | string | Да | места отбора пробы |
| id\_product | string | Да | Продукт |
| id\_metrics | string | Да | Показатель |
| date\_start | Date | Да | Дата и время возникновения события |
| date\_end | Date | Да | Дата и время завершения события |
| event\_code | integer | Да | Событие (вышел за нормативное значение) |
| metrics\_value | double | Нет | Значение показателя |
| Refinery\_ID | integer | Да | НПЗ |

### 1.к Структура исходящих данных по детализированной статистике событий (показатели тех. режима)

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Ключевое поле | Описание |
| Refinery\_ID | integer | Да | НПЗ |
| id\_metrics | string | Да | ID показателя |
| date\_start | Date | Да | Дата и время возникновения события |
| date\_end | Date | Да | Дата и время завершения события |
| event\_code | integer | Да | Событие (вышел за нормативное значение) |
| metrics\_value | double | Нет | Значение показателя |

### 1.л Структура исходящих данных по значениям показателей тех. режима

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| id\_metrics | string | Да | Показатель |
| Value | Double | Да | Значение показателя |
| Date | Date | Да | Дата/Время |
| Status | String |  | Статус |
| Refinery\_ID | Integer | Да | НПЗ |

### 1.м Структура исходящих данных по значениям показателей качества

|  |  |  |  |
| --- | --- | --- | --- |
| Название поля | Тип данных | Обязательность | Описание |
| id\_metrics | String | Да | Показатель |
| Value | Double | Да | Значение показателя |
| Date | Date | Да | Дата/Время |
| Status | String |  | Статус |
| Refinery\_ID | Integer | Да | НПЗ |
| id\_product | String | Да | Продукт |
| id\_point | String | Да | Точка сбора показателя |