



ОПИСАНИЕ ФУНКЦИОНАЛЬНЫХ ХАРАКТЕРИСТИК

компонента Архивирование (код: ARCH)

продукта Platform V Archiving (код: ARC)

Содержание

Термины и сокращения	3
Назначение	7
Цель создания.....	7
Основные функции	7
Сценарии использования	9

Термины и сокращения

Термин/сокращение	Расшифровка	Определение
Archiving/ARC	-	Продукт Platform V Archiving (ARC)
DataSpace Core/DSPC	-	Серверная часть компонента « DataSpace Core » (DSPC) в составе Продукта Platform V DataSpace (APT)
БД	База данных	Массив данных, хранящихся в соответствии со схемой данных, манипулирование которыми осуществляют в соответствии с правилами средств моделирования данных
ЦС	Целевая система	Система, в которую система-источник отправляет изменения с помощью Platform V Archiving
ММТР	Межмодульный транспорт	Компонент Межмодульный транспорт (ММТР) в составе Продукта Platform V Synapse Enterprise Integration (SEI)
ТКД	Технический контроль данных	Механизм, используемый для проверки качества загруженных в целевую систему данных
ИФТ	Интеграционное и функциональное тестирование	Одна из фаз тестирования программного обеспечения, при которой отдельные программные модули объединяются и тестируются в группе
DRP	Disaster Recovery Plan	Восстановление данных
B - Pipeline	-	Фазы Build Pipeline - стадии сборки дистрибутива системы-источника
R - Pipeline	-	Фазы Release Pipeline - стадии релизного DPM конвейера системы-источника

Init	-	Инициализирующая (первоначальная) выгрузка данных
DTO	Data Transfer Object	Один из шаблонов проектирования, используется для передачи данных между подсистемами приложения
HBase	-	NoSQL СУБД с открытым исходным кодом
Hadoop	-	Свободно распространяемый набор утилит, библиотек и фреймворков для разработки и выполнения распределённых программ
Avro		Ориентированная на строки платформа удаленного вызова процедур и сериализации данных, разработанная в рамках проекта Apache Hadoop
Kafka	-	Распределённый программный брокер сообщений с открытым исходным кодом
Patroni	-	Приложение для создания высокодоступных PostgreSQL кластеров на основе потоковой репликации
Smoke - тесты	-	Минимальный набор тестов на явные ошибки
Regress - тесты	-	Тестирование уже протестированной программы, проводящееся после ее модификации
Слушатель	-	Обработчик, который привязан к определенному событию или событиям, и вызывается автоматически при его наступлении

УЗ	Учетная Запись	Совокупность данных о пользователе, необходимая для аутентификации и контроля предоставления доступов
ТУЗ	Технологическая Учетная Запись	Разновидность учетной записи, применяемая только в случаях, когда пользователем выступает автоматизированная система
ЦОД	Центр обработки данных	Специализированное помещение, в котором размещается серверное и сетевое оборудование с последующим подключением клиентов к сети Интернет
Raw Data	-	Необработанные данные, собранные из источника
СП	Сервер приложений	Программная платформа для эффективного исполнения процедур, на которых построены приложения
Batch channel	-	Канал получения запрошенных пакетов данных в процессе восстановления каких-либо повреждений в информации - часть процесса контроля качества данных
Batch request channel	-	Канал запросов пакетов данных в процессе восстановления каких-либо повреждений в информации - часть процесса контроля качества данных
Response data	-	Ответное сообщение с данными на запрос ТКД
Pipeline	-	Поток, конвейер
TLS	Transport Layer Security	Криптографические протоколы, обеспечивающие защищённую передачу данных между узлами

Дедупликация	-	Метод сжатия массива данных, использующий в качестве алгоритма исключение повторяющихся данных
UUID	Universally Unique Identifier	Стандарт для уникальной идентификации информации без центра координации
URL	Uniform Resource Locator	Стандарт унифицированных адресов и записи ссылок на объекты
Пачка	-	Пакет данных
Мапирование	-	Определение соответствия данных между различными обозначениями объектов
cURL	-	Программа командной строки для взаимодействия с серверами через различные протоколы с синтаксисом URL
Система-источник	-	Система, которая отправляет в целевую систему данные с помощью векторов изменений
ОТТ	-	Сервис авторизации One-Time Password (ОТР)/ОТТ (ОТТС) в составе продукта ОТТ Platform V One-Time-Token; входит в состав bundle Platform V Backend

Назначение

Цель создания

Для обеспечения долговременного хранения данных Platform V (как архивных, так и оперативных) и доступности этих данных в аналитических и отчетных системах необходимы механизмы, обеспечивающие передачу данных из системы-источника в ЦС.

Назначение Archiving — выгружать данные из системы-источника в ЦС и далее сохранять и поддерживать в актуальном состоянии данные, устранять на стороне ЦС технические инциденты качества данных, которые могут возникать в процессе передачи и обработки данных, восстанавливать данные в случае сбоев на стороне ЦС или в процессах актуализации данных.

Основные функции

Init. Это тип взаимодействия Archiving и системы-источника, во время которого обеспечивается первоначальная инициализация (заливка) данных системы-источника в ЦС. В процессе первоначальной загрузки Archiving обеспечивает запрос по всем типам, по которым выполняется взаимодействие — запрос всех актуальных объектов у системы-источника и передача этих объектов в ЦС. В процессе первоначальной загрузки в ЦС создаются первичные структуры для хранения (таблицы) в соответствии со сгенерированной Archiving схемой данных. Первоначальная загрузка со стороны Archiving представлена отдельным механизмом и отдельным API, реализуемым на стороне системы-источника. Первоначальная загрузка выполняется, как правило, один раз для каждого типа данных, применяемого системой—источником, в определенных случаях допускается проведение повторного Init.

Поток (обработка векторов изменений). Этот тип взаимодействия всегда происходит со стороны системы-источника. Поток обеспечивает непрерывную актуализацию данных на стороне ЦС по мере их изменения на стороне системы-источника. В процессе выполнения бизнес-активности система-источник генерирует поток векторов изменений. В набор изменений, соответствующий транзакции, выполненной на стороне системы-источника, входят векторы изменения, выполненные в рамках этой транзакции. Обработка потока в Archiving представляет собой получение

векторов изменений, генерируемых системой-источником через отдельный механизм на стороне Archiving, не требующий какого-либо API на стороне системы-источника при условии, что система-источник интегрирована с сервисом приема векторов изменений.

ТКД. Этот тип взаимодействия инициируется ЦС в одном из двух случаев:

- В процессе обработки потока выявлены ситуации неконсистентности данных на стороне ЦС («инцидент качества»). ЦС инициирует запрос данных по проблемному объекту у Archiving, которое в свою очередь выполняет запрос полной текущей версии объекта у системы-источника — по типу и идентификатору объекта, вызвавшего инцидент. Получив ответ, Archiving направляет полную версию этого объекта в ЦС (с учетом применения правил сокрытия, черного и белого списков).

- Выполняются плановые сверки (запросы списков объектов) у Archiving со стороны ЦС. Плановая сверка выполняется ЦС, исходя из регламентов и процедур, определенных на их стороне, при этом ЦС инициирует запрос по типу (типам) и списку идентификаторов, по которым требуется получить полные актуальные версии объектов. Archiving в свою очередь запрашивает у системы-источника выборку идентификаторов на его стороне. Таким образом проверяется и актуальное состояние объектов на стороне ЦС (идентификаторы от ЦС), и наличие нереплицированных объектов, которые есть у системы-источника, но отсутствуют у ЦС (идентификаторы от системы-источника). Далее список идентификаторов от системы-источника и от ЦС объединяются на стороне Archiving, и выполняются запросы соответствующих объектов у системы-источника по объединенному списку ключей. Получив эти объекты, Archiving обрабатывает их и направляет полные версии в ЦС (с учетом применения правил сокрытия, черного и белого списков).

ТКД представлен отдельным механизмом и отдельным API, реализуемым на стороне системы-источника. Работа ТКД выполняется постоянно по мере необходимости (инициирования ЦС).

DRP. Данный тип взаимодействия инициируется ЦС в случае неудовлетворительной оценки качества данных по одному или нескольким типам. Заключается в отправке от ЦС в Archiving списка идентификаторов всех объектов из БД ЦС по каждому из выбранных типов. Далее Archiving запрашивает все эти объекты у системы-источника. Полученные объекты отправляются в ЦС и обновляются в ее БД. Объекты, которые не были получены,

помечаются как удаленные. Для взаимодействия используется то же API, что при ТКД, так как на стороне системы-источника процессы DRP и ТКД не отличаются.

Название функции	Потребитель функции	Аргументы функции	Результат
DRP	ЦС	Набор идентификаторов объектов данных, которые необходимо запросить	Получены объекты данных, идентификаторы которых были присланы от ЦС в Archiving
Init	Администратор системы-источника	Набор типов данных, которые необходимо выгрузить из системы-источника в ЦС	Данные, отправленные системой-источником, выгружены в интеграционный шлюз между Archiving и ЦС
Поток (обработка векторов изменений)	Интеграционный шлюз	Вектор изменений – информация об изменениях, которые нужно реплицировать	Данные из вектора изменений отправлены в интеграционный шлюз между Archiving и целевой системы
ТКД (техническое качество данных)	ЦС	Набор идентификаторов объектов данных, которые необходимо запросить	Получены объекты данных, идентификаторы которых были присланы от ЦС в Archiving

Сценарии использования

Сценарий «Init»

Для построения реплики системы-источника в ЦС необходимо провести Init.

- Администратор системы-источника запускает процесс Init.
- Archiving запрашивает данные у системы-источника.
- Система-источник отправляет (порциями) данные в Archiving.

- Archiving принимает данные (так же порциями), обрабатывает и отправляет их в интеграционную шину, откуда их вычитывает ЦС.

- После получения и обработки всех данных, данные выгружаются в базу данных ЦС, пользователь получает сообщение об успешном завершении Init.

Сценарий «Потоковая выгрузка данных»

Для поддержания актуальности данных в БД ЦС системы-источника отправляет информацию обо всех изменениях в данных, выгруженных в ЦС, через сервис приема векторов изменений в Archiving. Сценарий состоит из следующих шагов:

- В базе данных системы-источника происходит изменение. По их итогам формируется вектор изменений.

- Сервис приема векторов изменений обрабатывает данные и отправляет в Archiving.

- Archiving принимает данные, обрабатывает и отправляет в ЦС.

- ЦС получает данные и вносит изменения в своей базе данных, после чего отправляет подтверждение о получении данных.

После выполнения шагов 1-4 сценарий успешно выполнен.

Сценарий «Технический контроль данных»

Для контроля актуальности данных в БД ЦС используется механизм технического контроля данных (ТКД), который поддерживается Archiving.

Выполняется он по следующему сценарию:

- ЦС формирует список идентификаторов объектов данных, которые необходимо проверить на актуальность и отправляет запрос ТКД в Archiving.

- После получения запроса на ТКД Archiving отправляет запрос на получение данных к системе-источнику.

- Система-источник на основе параметров запроса формирует ответ, содержащий запрошенные данные.

- После получения данных от системы-источника Archiving обрабатывает их и отправляет в ЦС.

- По полученным данным на стороне ЦС проводится оценка по актуальности данных в его БД.

После выполнения шагов 1-5 сценарий успешно выполнен.

Сценарий «Восстановление данных»

В случае низкой актуальности (качества) данных в базе данных ЦС (по сравнению с базой данных системы-источника) выполняется сценарий восстановления данных.

Он состоит из следующих этапов:

- На стороне ЦС формируется запрос, содержащий список идентификаторов объектов данных, которые необходимо восстановить. После этого запрос отправляется в Archiving.

- На стороне Archiving принимается запрос на восстановление данных, после чего формируется и отправляется запрос данных на систему-источник.

- Система-источник принимает запрос, формирует ответ и частями отправляет его на Archiving.

- Archiving принимает данные. После принятия очередной части данных формирует ответ для ЦС и отправляет его.

- После принятия всех запрошенных данных на стороне ЦС сценарий восстановления данных успешно выполнен.