

Типовой протокол подключения поставщика данных к бирже данных 1DMS

Оглавление

Определения.....	1
Кратко.....	2
Технические условия типовой интеграции.....	2
1. Формат данных.....	2
Приведение данных к формату Таксономии Биржи.....	2
Формат файла.....	3
2. Передача данных.....	3
3. Получение доступа.....	4
4. Отладка интеграции.....	4
5. Проверка поставки данных.....	5

Определения

- **Биржа данных 1DMS, Биржа** - независимая технологическая платформа, объединяющая поставщиков и потребителей данных для обмена деперсонализированными знаниями о своей аудитории.
- **Поставщик данных** - участник Биржи, который предоставляет данные для продажи Потребителям данных.
- **Таксономия интегрального профиля Биржи, Таксономия** - двухуровневое дерево, на первом уровне которого находятся атрибуты, кодируемые числовым кодом атрибута rcode (например Пол, Возраст, Гражданство и так далее). Каждый атрибут имеет справочник допустимых значений. На втором уровне кодируется значение атрибута (числовой код scode из соответствующего справочника допустимых значений). Таксономия представлена в виде электронной таблицы MS Excel (для человека) и JSON-файла (для компьютера). Пример: "мужской пол" кодируется как rcode=10045, scode=10001, где 10045 определяет атрибут "Пол", а 10001 - запись "Мужской пол" из справочника значений этого атрибута.
- **CSV** - (англ. Comma-Separated Values) текстовый формат файлов для представления табличных данных, в котором одной записи таблицы соответствует одна строка файла, а отдельные значения записи отделены друг от друга запятой или другим символом-разделителем.
- **SFTP** - (англ. Secure File Transfer Protocol) сетевой протокол для передачи файлов поверх надёжного и безопасного соединения (например, SSH-тоннеля).

Работает по клиент-серверной модели: клиенты могут загружать файлы на сервер или скачивать их с него, а SFTP-сервер физически хранит файлы и управляет доступом к ним и т. д.

Кратко

Поставщик может быстро подключиться к Бирже данных по типовому способу интеграции: загрузить и регулярно обновлять свои данные в виде CSV-файлов на SFTP-сервер Биржи данных.

Для этого необходимо:

1. Определить формат CSV-файлов - раздел 1,
2. Настроить их загрузку на SFTP-сервер Биржи данных - раздел 2,
3. Получить доступ к SFTP-серверу Биржи данных - раздел 3,
4. Отладить поставку данных на тестовом окружении - раздел 4,
5. Проверить поставку данных на производственном окружении - раздел 5.

Более подробно условия типовой интеграции описаны в следующих разделах.

Если типовая интеграция не подходит, то поставщик и биржа данных могут по договоренности организовать поставку данные другим способом.

Технические условия типовой интеграции

1. Формат данных

Приведение данных к формату Таксономии Биржи

Данные поставляются в табличном виде. Таблица определяется перечнем столбцов, значения которых указываются в каждой записи таблицы. Поставщику доступно создание произвольного количества таблиц с произвольным набором столбцов.

Каждый столбец отвечает за один из атрибутов Таксономии Биржи (первый уровень дерева). В строках передаются значения атрибутов (второй уровень таксономии, *scope*).

При этом в одной строке может быть передано только одно значение по каждому столбцу. Если необходимо по одному профилю пользователя передать несколько значений одного атрибута Таксономии, то возможны две реализации:

- Значения передаются отдельными записями, а идентификаторы конкретного профиля в всех этих строках дублируются. При обработке на Бирже данные из всех строк

объединятся в один профиль, так как строки склеиваются по значениям в столбцах идентификаторов.

- Значения передаются отдельными столбцами. Если заранее известно количество значений атрибута в профиле, то в структуру таблицы добавляются по столбцу на каждое значения атрибута. Тогда в одной строке передаются все значения для профиля.

После определения таблицы Поставщик передает описание её структуры Бирже: какой по порядку столбец относится к какому атрибуту таксономии (rcode). На своей стороне, Биржа по описанию таблицы настраивает правила её обработки.

Таким образом, в файле Поставщик передаёт только атрибуты второго уровня Таксономии - числовые коды scode из справочников. Исключение составляют пользовательские идентификаторы (номер телефона, email, cookie и т. д.) - для них в файле передаются значения идентификаторов.

Формат файла

Данные в файле должны быть представлены в CSV-формате. Требования к CSV файлам:

1. Разделитель значений колонок - , (запятая)
2. Значение поля может быть заключено в двойные кавычки ("")
3. Символ экранирования - \ (escape)
4. разделитель строк - \n
5. кодировка файла — **UTF-8**,
6. Строка с заголовком колонок не указывается.

Пример:

```
"d54hgfdt576bdftrw83",10000,10001,10001,10010,10008
```

```
"ghf65h4gf532i2bfdhs","ghfjty674h5vdhfu3",10002,10003,10005,10002
```

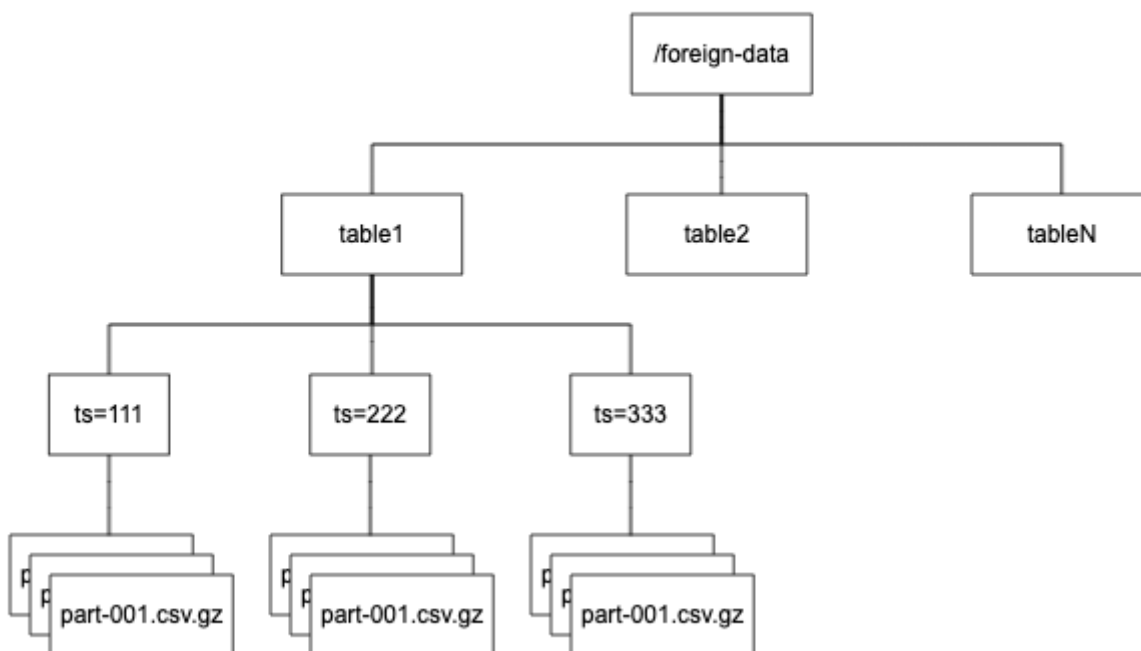
2. Передача данных

На SFTP-сервере Биржи для каждого Поставщика создан свой корневой каталог. Поставляемые данные должны сохраняться в каталог /foreign-data. Данные каждой таблицы Поставщика загружаются в отдельный подкаталог с именем таблицы в названии.

Внутри каталога таблицы Поставщик должен создавать отдельный каталог для каждой передачи данных таблицы - с временем загрузки в названии каталога (вида /ts=1234567890, где 1234567890 - время в Unix-формате в секундах). То есть, для следующей передачи данных таблицы (за очередной период времени и т. п.) внутри каталога таблицы Поставщик создает новый каталог с временем передачи.

Для удобства данные одной передачи таблицы могут быть разбиты на несколько файлов.

Пример итогового дерева каталогов и файлов:



Рекомендуется использовать компрессию загружаемых файлов. Поддерживаются кодеки: bzip, gzip, lz4, snappy, deflate.

Биржа обрабатывает только те передачи данных таблицы, что отмечены как готовые. Для этого в подкаталоге с временем загрузки данных Поставщик создает пустой файл с именем _SUCCESS.

Готовые данные Поставщика обрабатываются Биржей автоматически, раз в час. Сведения об аудиторрии в Личном кабинете Поставщика обновляются с периодичностью раз в сутки.

Файлы и каталоги после обработки не удаляются и не перемещаются. Результаты обработки данных на SFTP-сервер не сохраняются и автоматическим способом Поставщику не сообщаются. Поставщик может узнать результат обработки в Личном кабинете Биржи.

3. Получение доступа

После предоставления Бирже описания своих таблиц Поставщику предоставляются данные доступа к SFTP-серверу (адрес сервера в Интернет, логин и пароль) на двух окружениях: тестовом и производном.

Производное окружение используется для поставки данных на Биржу, тестовое - для безопасной отладки интеграции.

4. Отладка интеграции

Поставщик передает данные на Биржу в тестовом окружении по согласованному формату таблиц, после чего проверяет корректность обработки данных на Бирже.

После отладки интеграции Поставщик и Биржа по договоренности принимают решение о начале поставки данных на производственном окружении.

Рекомендуется отлаживать интеграцию на тестовом окружении при внесении любых изменений в состав поставляемых данных: добавление таблиц, изменение их структуры, значимые изменения в источниках данных на стороне Поставщика.

5. Проверка поставки данных

Результат поставки данных на Биржу Поставщик может контролировать в своем Личном кабинете, в разделе Моя аудитория (сведения обновляются раз в сутки).

Контроль возможен по объему аудитории по разным видам идентификаторов и атрибутов за разные периоды времени (день, неделя, месяц и все время).

Также доступны отчеты со статистикой заполнения атрибутов по дням, в виде CSV-файлов.