

# DataOps.onETL



Unit#6

Коннекторы к файловым системам (ФС, fs):  
какие они бывают, какие методы предоставляют



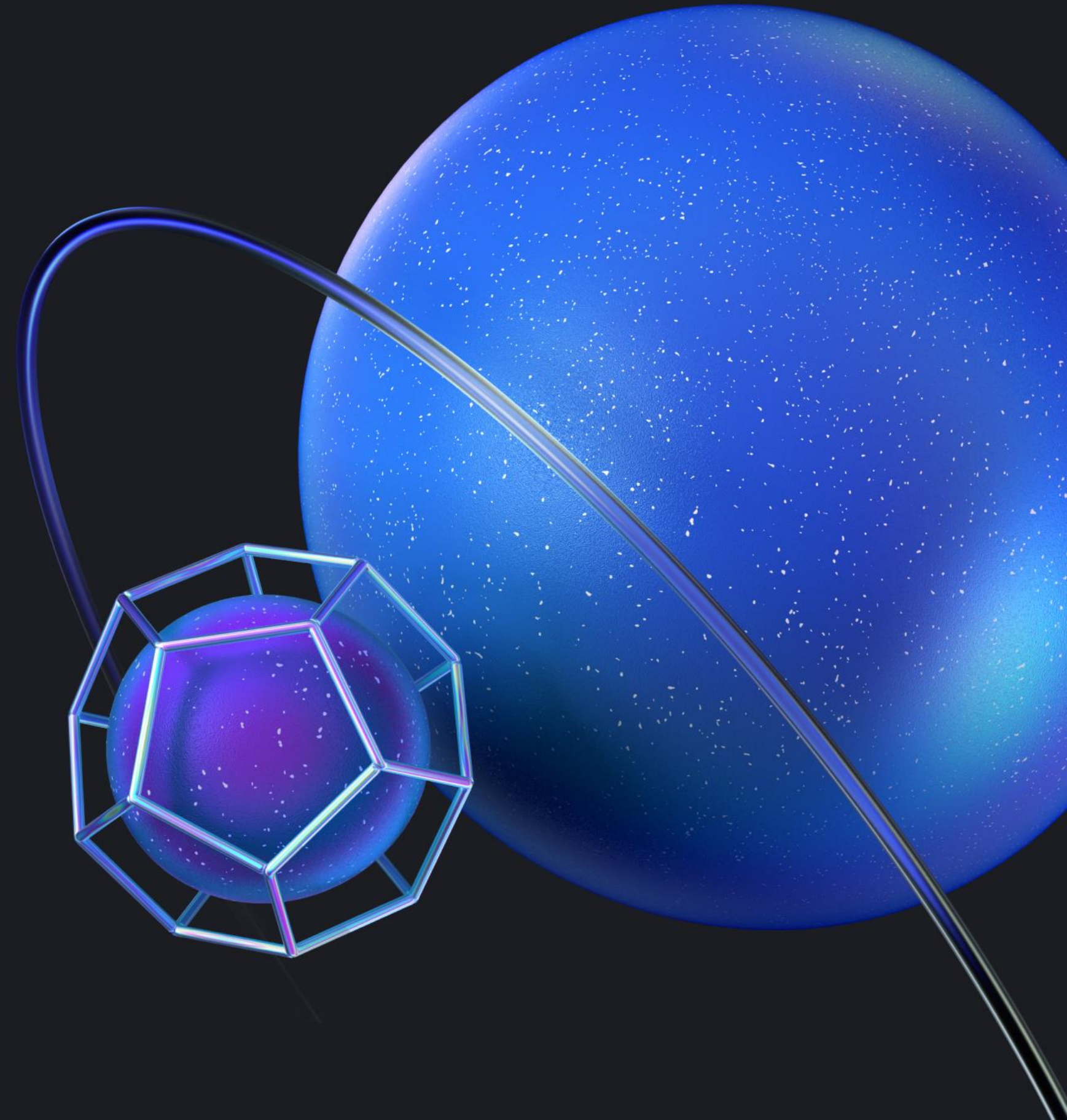
**Саттар Гюльмамедов**

РО команды ETL

МТС Тета

х

DataOps Platform



# onETL урок # 6



- Коннекторы для выполнения файловых операций
- Как их импортировать
- Методы коннекторов
- Отличия объектов манипуляции файлами
- Отличия от коннекторов для работы с DataFrame

# Коннекторы к файловым системам

→ HDFS

→ S3

→ FTP

→ Webdav

→ FTPS

→ Samba

→ SFTP

```
pip install onetl[files]
```

Shell



# Методы файловых коннекторов



Метод	Объяснение	Где поддерживается
check()	Проверить соединение	Везде
create_dir()	Создать директорию	Везде
download_file()	Скачать файл на локальную ФС	Везде
get_stat()	Получить статистическую информацию об объекте	Везде
is_dir()	Проверить является ли объект директорией	Везде
is_file()	Проверить является ли объект файлом	Везде
list_dir()	Получить список объектов в директории	Везде
path_exists()	Проверить существует ли путь	Везде
remove_dir()	Удалить директорию	Везде
remove_file()	Удалить файл	Везде
rename_file()	Переименовать файл	Везде
resolve_dir()	Получить объект директории	Везде
resolve_file()	Получить объект файла	Везде
upload_file()	Загрузить файл на файловую систему	Везде
walk()	Получить генератор, позволяющий итерироваться по переданному пути	Кроме Samba
rename_dir()	Переименовать директорию	Кроме Samba, S3 и WebDAV

# HDFS

```
from onetl.connection import HDFS
Hdfs = HDFS(
    host="namenode1.domain.com",
    user="someuser",
    password="*****",
)
```

Python



# Параметры конструктора HDFS



Параметр	Объяснение
cluster	Имя кластера
host	Имя хоста, на котором развернута namenode кластера
port	Порт по которому namenode принимает обращения по протоколу WebHDFS (по умолчанию 50070)
user	Имя пользователя
password	Пароль пользователя
keytab	Путь, по которому расположен keytab-файл
timeout	Количество секунд в течение которого допустимо отсутствие ответа от namenode (по умолчанию)

# FTP, FTPS, SFTP

```
from onetl.connection import FTP

ftp = FTP(
    host="ftp.domain.com",
    user="someuser",
    password="*****",
)
```

Python



```
from onetl.connection import FTPS

ftps = FTPS(
    host="ftps.domain.com",
    user="someuser",
    password="*****",
)
```

Python



```
from onetl.connection import SFTP

sftp = SFTP(
    host="192.168.1.19",
    user="someuser",
    password="*****",
)
```

Python



МТС

# Параметры конструктора [S]FTP[S]

Параметр	Объяснение
host	Имя или адрес хоста, к которому необходимо установить подключение
port	Порт (для ftp и ftps - по умолчанию 21, для sftp - 22)
user	Имя пользователя
password	Пароль пользователя
key_file	Имя файла с частным ключом для подключения (только для sftp)
timeout	Количество секунд в течение которого допустимо отсутствие ответа (только для sftp)
host_key_check	Флаг для активации поиска частных ключей при подключении (только для sftp)
compress	Флаг для активации сжатия данных (только для sftp)



# WebDAV

```
from onetl.connection import WebDAV

wd = WebDAV(
    host="webdav.domain.com",
    user="someuser",
    password="*****",
    protocol="https",
)
```

Python



# Параметры конструктора webdav



Параметр	Объяснение
host	Имя или адрес хоста, к которому необходимо установить подключение
user	Имя пользователя
password	Пароль пользователя
port	Порт по которому необходимо установить подключение
ssl_verify	Флаг активации защищенного подключения, по умолчанию `True`
protocol	Протокол, по которому будет происходить подключение: `http` или `https`; по умолчанию `https`

# S3

```
from onetl.connection import S3

s3 = S3(
    host="s3.domain.com",
    protocol="http",
    bucket="my-bucket",
    access_key="ACCESS_KEY",
    secret_key="SECRET_KEY",
)
```

Python



Параметр	Объяснение
host	Имя или адрес хоста, к которому необходимо установить подключение
port	Порт по которому необходимо установить подключение
bucket	Имя бакета для подключения
access_key	Идентификационный ключ
secret_key	Секретный ключ
protocol	Протокол, по которому будет происходить подключение: `http` или `https`; по умолчанию `https`
session_token	Сессионный токен для идентификации
region	Имя региона, в котором расположен бакет

# Samba

```
from onetl.connection import Samba

samba = Samba(
    host="mydomain.com",
    share="share_name",
    protocol="SMB",
    port=445,
    user="user",
    password="password",
)
```

Python



Параметр	Объяснение
host	Имя или адрес хоста, к которому необходимо установить подключение
share	Имя ресурса на Samba-сервере, к которому будет происходить подключение
protocol	Протокол, по которому будет происходить подключение: `SMB` или `NetBIOS`; по умолчанию `SMB`
port	Порт по которому необходимо установить подключение, по умолчанию 445
domain	Имя домена, используемое для подключения, по умолчанию имя хоста
auth_type	Тип аутентификации, который будет использоваться: `NTLMv2` или `NTLMv1`; по умолчанию `NTLMv2`
user	Имя пользователя
password	Пароль пользователя

# Методы файловых коннекторов



Метод	Объяснение	Где поддерживается
check()	Проверить соединение	Везде
create_dir()	Создать директорию	Везде
download_file()	Скачать файл на локальную ФС	Везде
get_stat()	Получить статистическую информацию об объекте	Везде
is_dir()	Проверить является ли объект директорией	Везде
is_file()	Проверить является ли объект файлом	Везде
list_dir()	Получить список объектов в директории	Везде
path_exists()	Проверить существует ли путь	Везде
remove_dir()	Удалить директорию	Везде
remove_file()	Удалить файл	Везде
rename_file()	Переименовать файл	Везде
resolve_dir()	Получить объект директории	Везде
resolve_file()	Получить объект файла	Везде
upload_file()	Загрузить файл на файловую систему	Везде
walk()	Получить генератор, позволяющий итерироваться по переданному пути	Кроме Samba
rename_dir()	Переименовать директорию	Кроме Samba, S3 и WebDAV

# Отличия от объектов манипуляции файлами

## Объекты манипуляции данными:

- FileDownloader
- FileUploader
- FileMover

## Вспомогательные объекты:

- FileFilters
- FileLimits



# Спасибо!



[onetools@mts.ru](mailto:onetools@mts.ru)  
<https://t.me/c/1511728757/5>

MTC Тета

x

DataOps Platform

