



ChEMBLActivityPipeline

ChEMBLCommonPipeline ,
ChEMBLPipelineContract

bioetl/pipelines/chembl/
activity/run.py

Конкр
ChEMBL
контро
оркест
данны
трансф
сохран

Класс	Предок	Модуль	Назна
ChemblClient	ConfiguredHttpClient	bioetl/clients/chembl/ client.py	Высоки ChEMBL унифи инкап ChEMBL запро содер "chem Chembl

Класс	Предок	Модуль	Назна
RequestsBackend	HttpBackend (протокол)	bioetl/clients/chembl/ factories.py	Реализация библиотечной API Chembl, по запросу закрытия
ChemblRequestBuilder	(нет; dataclass)	bioetl/clients/chembl/ client.py	Конструктор Инкапсуляция конфигурации (Source) строит запросы различ

Класс	Предок	Модуль	Название
ActivityExtractor	(нет; <code>dataclass</code>)	<code>bioetl/pipelines/chembl/activity/stages.py</code>	Класс: <code>ActivityExtractor</code> ChEMBL извлекает данные через запросы, итеративно собирая результаты, реализован частично
ActivityTransformer	(нет; <code>dataclass</code>)	<code>bioetl/pipelines/chembl/activity/stages.py</code>	Класс: <code>ActivityTransformer</code> нормализует активные вещества, поле <code>name</code> колонки <code>name</code> названий нормализовано
ActivityWriter	(нет; <code>dataclass</code>)	<code>bioetl/pipelines/chembl/activity/stages.py</code>	Класс: <code>ActivityWriter</code> результаты на диск, контроль унификации записей, Parquet (manifest) задан
ActivityParser	(нет)	<code>bioetl/pipelines/chembl/activity/parsers.py</code>	Утилиты: <code>ChEMBLParser</code> Преобразование от ChEMBL нормализованного DataFrame колонки имен

Класс	Предок	Модуль	Назначение
ActivityNormalizer	(нет)	bioetl/pipelines/chembl/ activity/normalizers.py	Утилита для активации нормализации (BaseNormalizer) настроек колонок перед его к DataFrame. Обеспечивает заполнение умолченных ключей.
BaseChemblNormalizer	(нет явно; базовый класс для нормализации ChEMBL)	bioetl/clients/chembl (модуль нормализации данных)	Базовый класс для нормализации данных. Принимает схему колонок, полную спецификацию (ColumnNormalizationSpec) основанную на DataFrame, колонки, значения которых отсутствуют, вычисленные по ключевым словам.
ColumnNormalizationSpec	(нет; dataclass)	bioetl/clients/chembl (спецификации колонок)	Спецификация одной колонки с атрибутами: данные, умолченные значения, опции (функции), значения, исполняемые (BaseColumnNormalizationSpec) последовательно для каждой колонки.

Класс	Предок	Модуль	Назначение
ColumnMapping	(нет; <code>dataclass</code>)	<code>bioetl/clients/chembl</code> (утилиты разбора JSON)	Структура для парсинга данных, соответствующая структуре результатов API или не соответствующая в исходном формате. Например, <code>ColumnMapping</code> ("active") означает, что колонка "active" в исходном формате API соответствует записи в базе данных.
ChemblExtractionDescriptor	(нет; <code>dataclass</code>)	<code>bioetl/pipelines/chembl/common/descriptor.py</code>	Описание параметров для извлечения данных из Chembl. ID's индексов параметров пагинации, режим загрузки, содержимое запроса для упорядочивания результатов по числу соединений. И т.д. Разбивка на логики.
BatchPlan	(нет; <code>dataclass</code>)	<code>bioetl/clients/chembl/descriptor_factory.py</code>	Простая структура пакетов для загрузки данных. Размер пакета, число соединений, запросы. Например, <code>BatchPlan</code> ("chun") означает, что запись в базе данных, влияющая на загрузку данных, на загрузку данных.

Класс	Предок	Модуль	Назначение
ChEMBLActivitySchema	<code>pandera.pandas.DataFrameSchema</code>	<code>bioetl/schemas/chembl_activity_schema.py</code>	Валидация данных. Описание итогов (строки) также поля. Соответствие ожидаемым настройкам сортировки при со
ActivityArtifactPlanner	<code>ArtifactPlanner</code>	<code>bioetl/pipelines/chembl/activity/run.py</code>	Специализированный планировщик для данного компонента, как бы, размер реализации формирует именованные модули и файлы активности. Основная обработка детерминированного каталога
ActivityWriteService	<code>WriteService</code>	<code>bioetl/pipelines/chembl/activity/run.py</code>	Специализированный сервис для записи активности. Метод <code>ActivityWriteService.write</code> интерпретирует контекст выполнения, данные, учётные данные, выходные файлы активности

Класс	Предок	Модуль	Назв
UnifiedOutputWriter	<i>(нет)</i>	bioetl/core/io/output.py	Унифи вывод атомар диск в файла выпол по зар сортир детерм основн .parc файл, генера (meta выгруз (run_ размер создан 74 75
DefaultValidationService	<i>(нет явно; реализация сервиса валидации)</i>	bioetl/core/pipeline/ services	Серви пайпл вокруг исполн соотве ожид пустог необх пайпл напри пустог извлеч либо д дрейф исключ

Класс	Предок	Модуль	Назв
SchemaRegistry	(нет)	bioetl/core/io/artifacts.py	In-me ETL. Хр зарег Pander различ вместе инфор эталон настро сортир (busine регист (ChEM реестр затем валид перед

1234567891011121314151617181920212266676869707172

73run.py

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/pipelines/chembl/activity/run.py

232425263435client.py

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/clients/chembl/client.py

27282930313233factories.py

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/clients/chembl/factories.py

36373839404142434445464748495062stages.py

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/pipelines/chembl/activity/stages.py

515253parsers.py

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/pipelines/chembl/activity/parsers.py

5455565758normalizers.py

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/pipelines/chembl/activity/normalizers.py

59606163descriptor.py

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/pipelines/chembl/common/descriptor.py

6465chembl_activity_schema.py

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/schemas/chembl_activity_schema.py

74 75 76 **output.py**

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/core/io/output.py

77 78 **strategies.py**

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/pipelines/chembl/common/strategies.py

79 80 81 82 83 **artifacts.py**

https://github.com/SatoryKono/bioactivity_data_acquisition/blob/98f17f432532cddd56d0a07fd4796b37c54677ec/src/bioetl/core/io/artifacts.py