Solace topic structures用来在app之间传送消息

Outlook给邮件分类

Office 365里面用word

熟悉linux, ip3的操作界面，把需要的东西都准备好

Vim的基本操作？这个不需要了把

Linux基本的system call

First week: 熟悉系统，空闲时间可以看操作系统的理论

花两个小时看一下git，书本和wiki上的信息

<https://bitbucket/projects/BNK/repos/bunker_apps/browse/bunker_apps/load_bbg_closing_prices.py>

<https://bitbucket/projects/BNK/repos/bunker_utils/browse/bunker_utils/bbg_utils.py>

i’m going to have your first project be a system that gets bloomberg data and puts it into bunker

[4:43](https://akunacapital.slack.com/archives/D011FKKNFGF/p1587591831062000)

right now, we do it where we get the data and load it into a database, but we’d like to change it in a number of ways

[4:44](https://akunacapital.slack.com/archives/D011FKKNFGF/p1587591859062400)

i’d like to be able to download the data and then put it onto solace (which is our firm-wide message bus)

[4:44](https://akunacapital.slack.com/archives/D011FKKNFGF/p1587591882063000)

this allows for hubble, our data infrastructure platform, to see the data and store it and people can replay the data and do what they want with it

[4:45](https://akunacapital.slack.com/archives/D011FKKNFGF/p1587591902063500)

then i’d like you to write a listener that listens to the solace data and writes it to the bunker database

[4:45](https://akunacapital.slack.com/archives/D011FKKNFGF/p1587591933064200)

and also, i’d like the system to be scalable so that if people want to get data they can start doing so, with approval because data pulls cost money

you should also note that we have a firewall which might prevent you from running this on your machine so that might be something we have to ask for NOC to poke a hole in the firewall for

原先是直接load到bunker\_db的，现在希望我弄到solace上面，让别人也可以接收到（这个部分我似乎不用管）。然后也还是写一个listener，接收信息，储存到bunker database上面。同时还要scalable。

bunker\_db = akdb.DBInfo(\*\*config["db"][environment]["bunker"])

from send\_bbg\_prices.py

Question to ask:

right now, the existing program gets the data and load it into the bunker database, which is used by the bunker team. From my understanding, now you also want to put the data onto solace. So why do I need to first put the data onto solace and then load it into the bunker database? The existing program already did the loading job.

Main task: (1) put the data onto solace (2) scalable

If so, the main tech difficulty for me is to get familiar with csolace package api.

You said hubble is a data infrastructure platform, but what’s hubble in specific? I didn’t find a wiki page introducing it. It seems I don’t need to worry about it for this project.

What do you mean by scalable?

Order: I should do some tests to make sure I get the right idea before start writing LWP

LWP question (how long I should write it)

Can I test and run the existing code on my machine? Or there is a database for test?

Quite a few tools that I need time to get familiar with, it’s better to have some buddy to ask around. Can I also ask ding jin for tech questions?

Scalable: 相当于分成两步走。其中一步是：

1. Download data from bloomberg

2. Convert bbg data to protocol message

3. push protocol message to solace

4. write data to database（这个是另外一步）

Tech difficulty:

Conda package issues: almost solved

Csolace api

Asyncio

aiob

Convert data to protocol message: almost solved

DB setup: not quite sure how to do it on i3wm. solved

Situate: not sure what this is

Invoke

Styling: there are four

Python memory view

Python logging

Solace里面，request和reply分别指什么？

Should be able to finish some project content part today or tomorrow

Question for Ding:

How to set up datagrip on i3wm? Need to buy a license now?

How many database do I need to use? (like akdb?)

Can I run and play around the code on my server? Does it cost money? What is a test DB?

Where is the test data?

Do I need to make a new git branch? Pr: pull request

What’s situate: automation

Ans: better use my own user account

Run code用-dev flag

Question

How to listen from Solace? How to send to Solace…

(what is a topic)

Listen, get, send

Get, subscribe, send

下面这个是用neosolace做的send，可以参考一下。数据库源不用管

https://bitbucket/projects/BNK/repos/bunker\_apps/browse/bunker\_apps/create\_otc\_grains\_options.py

Is the process like: load the database schema, download the .encode file from bbg, decode it

Meaning of ` Transferring symbol data from devbunkerdb.bunkerdevdb to /mnt/bunker/dev/bloomberg/20191120/200429103609bnkr\_eod.req…`

` Bloomberg was not able to interpret: CSZ9 Comdty`

Ans: .encode是自己写的

Invoke check报错是不是不用管

Akuna asyncio Database Library（aiob）我需要用么，和akdb区别：

Aiob: based on different library asyncio

一般用datagrip就是为了看看数据？

Estimation of time (just for reference)

What is photon

Instrument query database is another one?

Question

How to make one keep running

What’s Kunlun usage?

Project:

(0) I currently have finished the project plan and now I am doing high level designs for the library methods I will use, the database I will use and the protocol message (AIDL) part. I am running examples from the csolace/aiodb library and existing closing price script. In the previous week I had two meeting sessions with Ding to know about the system and I also get familiar with python asynchronous I/O part. So, the next steps I am going to do is what I will do for the high-level design part. As long as I test the example scripts and understand what apis I should use, the project seems straightforward if I am correct. And the running server deployment will be in the later part of the project so I am not going to worry about that now and I will ask Ding for help if needed.

(a) data fetch needs: only closing data or other data. I think it’s better to finish the closing price one first and then extend to other demands. The system we design is scalable so should be easier to extend to other utilities.

(b) run and get the data: -e dev -d 20191120 -r US. This works.

How long do you think is appropriate for the project?

UML写哪儿？应该是在system boundary diagram这边写

然后system boundary diagram不用写的很准确，但是要搞清楚和已经有的东西（比如Solace, db）的互动

Design里面参考M2的Sequence diagram and message

M2: 看Sequence diagram and message

Overall: BBGService在solace的一端，bunker在另一个端。就是两个service，不一样的名字。Duplicated data row?

Main task/difficulty: to write protocol message

如果说user是通过bunker来send request，那么我该怎么写bunker？怎么take input？take什么input？实际上明确了take什么input也就基本明确了我的message protocol要写什么？

应该使用async publisher和async request reply (subscribe and reply)？async request reply就是async RPC，在distributed system里面有

剩下的是异步问题吧：收到request后要处理ack和packet，这两步应该是异步的？

我做的实际上就是类似remote procedure call，中间多加了一层

返回的data一行可以先保存为string，之后再parse成row

我要做的应该是two async RPC (p135)吧。如果不需要最后ack的，那就一个send

What’s your expectation on me?

Where can I do better?

Question:

Does this project work on all BBG data besides closing data?

How do users input their data request to Bunker? By which means do they input their request? Like directly running bunker with some arguments or through solace?

Which Topic should I subscribe to for getting request from users?

Ans:

Receive data request in Akuna format. Hide the bbg terminology (use translater in bbg\_utilities) to transfer. Request takes about 6 minutes from bbg, so do it in batch. (like using map? A list of fields and instruments) 哦是说要查不止一个东西

Not sequence number, maybe some hash value.

Key point I think: simple and fast service crash recovery. Way to do it: stateless

Future price not option price

Get\_request

Get future price. 先做这个

问题：.req和.out有什么区别？

Ding的建议，先从bbg这边，对应trader的要求，看看哪些数据是可以得到的。以及是否可以convert成bbg request文件。

Q:

Does Solace provide request/send mode or subscribe/publish mode or both? Ideally should be both and I need to use both to work because I need to serve one request from user but broadcast to others who are interested in it as well.

Who should subscribe? Only the request user? Then we could use request\_id like a port number. If this port number is in use, reply to the user that choose a different one.

1. Is there a user who first sends to the request to bunker via Solace? Does the bunker need to answer to the user with the fetched data or Hubble will do this part and the bunker only needs to write the data to database? We have closing price database, but traders are asking for data from bbg besides closing price, do we already have database for these data?
2. If the request\_id is given by users, will there be duplicate request\_ids? How do I handle this?
3. There are publish/subscribe example, and also request/reply example. It seems I should use both. I don’t really understand request/reply one (there is also subscribe in it).

Bbg service只是一个bridge，它就是负责把数据中转到bbg

Situate launches bunker

My task: try log into Situate and see what’s happening

Server means bbg service

PJ:  
builder (director) 是什么design pattern: Builder Pattern

用subscribe还是get？difference是啥

Send to Hubble team the last two row\_num, info. And maybe enumerate the error.

看一看builder pattern

跑一下shift strikes

试一下加asyncio.event或者future来等待并且取值：event是ok的

问：为什么要用coroutine…?

Ans：在我的程序里面，我可能会await自己的函数，但最终我await的其实是csolace的方法，而这个东西已经是用non-block的方法写了，就保证了这些message传递是异步非阻塞的。对我使用的人来说，应该可以就当它是没有await的普通函数来看待。参见flags\_asyncio.py以及之后的说明。

似乎await或者说yield from要有事件循环才有意义。

问：那如果我的await是需要返回值的呢？

Ans：

现在在处理的问题是：怎么让我的service在持续subscribe。Solved: use event

什么是S法则，我是不是应该为create task和subRequest另外写一个class？

我是否要确保在我发结果过去之前，对方要先subscribe？（这个暂时可以不用考虑）

Do I need to worry about the case where subscriber is not connected (can solved by using a queue)?

是不是把topic作为一个变量分开来写，这样以后要改的话方便些。Ans:暂时看没需要

一个是写event，还有测试一下await的话会不会消息消失（不会）

要把call\_back里面的东西拿出来

you’re going to want to wait to listen to the result\_info, look at the number of rows and then wait for all of the rows to come in. probably want to use an Event synchronization primitive.

下一步：Typing, git hooks

Conda Package conflicts: solved?

第一件事：怎么改pycharm的conda环境和git branch: solved

第二件事：remote和local的branch不统一，怎么改: solved

第三件事：回到typing

关键是我还没写的怎么给参数占坑

以及是否调整一下考虑整体调用问题 solved

先还是看一下git? solved

顺便可以看下一个新的pj folder怎么创建的 Ans: cookiecutter

Git log要看下，才能知道指针怎么指的 solved

在load\_bbg那边考虑怎么给参数的问题：应该通过类似旧code里面从db获取的方式，只不过我要用aiodb而已

Dependency injection:

<https://stackoverflow.com/questions/130794/what-is-dependency-injection?page=1&tab=votes#tab-top>

找个design pattern的书看一下

正好看一下photon是什么，这是其一。其二，学习一下它的structure，然后我就可以改自己的了。

Partly solved. \_\_main\_\_.py参考shift\_strikes.py。

[用@click.group()](mailto:用@click.group())来装饰即可https://click.palletsprojects.com/en/7.x/quickstart/

RequestMsgListener应该是一个abstract base class，在service和db里面可能都有它具体的实现

第一步：看看handler或者listener是是什么design pattern，还是纯粹就是一个interface

Server继承handler

看起来我这边对应的话是让LoadBBGClosingPrices和SvcUtilApp继承？

先把db.py理清楚再动server

我猜测这个design的好处是让server的使用者自己去定义需要的callback\_func，如果再写一个不同的server（和不一样的callback\_func），就不需要改client code？比如我有server和loadbbg两个。对subscriber来说，它可以变得更generic，至少在它的code里面是和一个interface（msgHandler）打交道而不是和具体的server打交道。不过本来不也就是传一个general的handler么…

Topic builder是不是也要放进去？

但怎么解决有些subscription需要request\_id，有些不需要？注意，对于subscribe而言，它应该只关心两个东西，一个是handler，一个是command，而有没有request\_id应该是对于command/topic而言的，所以这应该是user给我的，所以应该是作为command/topic的constructor传进来的

Topicbuilder必须要拿出来单独写，不然的话typing不清楚，而且topic可能有很多种，我必须要让它单独有一个class，使得之后添加新topic可以变得容易，同时又不需要改变client code（即subscribe那边的参数）

如果我创建的topicbuilder instance太多的话，会不会占用太多内存？

但为什么要set。。。因为我wait的只是subscribe，不是收到消息

要不把topic builder拆成topicbuilder和topic fields两个cls？Bridge design pattern。不过实际上一样是让client code在使用的时候再决定给abstraction配什么implementation作初始化（现在我是直接把可能的几种implementation写进去了）。如果我的topic不会变得太多的话，没有必要弄得这么麻烦。

那我这边db里面实际得到的就是一行数据，所以我要做的事情一个是变换成closing price table的格式，参见transformer。同时还要做一个计算，也在transformer里面，这个计算就要求把前一天的data拿出来。最后做完以后要自己改写一下loader，用aiodb写进去即可。

Bbg\_symbol之类的东西看来都得靠我自己从db拿出来

如果说我要做yesterday的data的计算的话，那也得让db.py来做

这个ClosingPricesTransformer我要拆成两部分。。一部分是给service.py，一部分是给db.py

稍微看一下taxi scheduling

稍微看一下db

关于trader要的那些东西是什么，我是否能得到，我觉得暂时问题不大，我先把closing price的能写出来，然后之后再试一试。

我觉得自己不该投入过多的精力在内部doc/architecture上面，因为太vague，除非我碰到了，用上了。而是掌握general技术。这个的意思是说，我需要别人的方向性指导，但也应该凭借这个提示找到更多可以refer的资源。

Path\_factory是用了factory design pattern？

坑死了。。。akdb的get\_rows是一个generator，aiodb的返回的就是一个DBRows class

Service\_request能接受的是一个bbg\_instrument msg类型的东西，并不接受DBrows类型的。

Aidl里面给如expiry\_dt之类的设置默认值：本来aidl就有默认值，不需要。不过fixedpoint类型是不是应该改成float

还有一个就是run\_in\_executor的问题

Queryextractor改adaptor，还是改dbreader？

Sql query里面product type为什么只有s or f，不能做option么？是不是因为option没有closing price？或者akuna只做stock和future?

注：S是stock，F是future，C是call option，P是put option. WC和WP是对应HK的

能否试一下除了s和f之外的？

UnderliersQuery也要改，要把product和product type传进去

Aiodb里面PostgresClient和connection有啥区别？是否抽象类要initiate？似乎queryextractor目前就够用了。Ans:一个是异步的，一个不是

没理解：str.format(self = self)。就是位置引用。solved

To see: str.format方法，sql join, group by。solved

…我昨天看了半天sql的问题，原来原因只是因为它只显示500行搜索的结果！

是不是画个class diagram？

Error处理似乎没有做（完整），要见class diagram

为什么要create\_task？有些coroutine似乎没有被包在task里面？

两个问题：run\_in\_executor是不是一定要包在task里面？第二个，它是不是也要像管道一样一层层来？

它还不是thread safe的？？？

注意所有的static method，如果确认的话，要把self都删了

1 on 1:

What’s your evaluation on my performance? Where do you think I can improve?

是不是自己试一下ssh add，至少让我不用每次输密码？

看一下实际的key pair是保存在哪里

看起来目前位置我连到公司电脑的ssh并没有用key？Ans:有可能现在用的是纯密码模式。

一定要参考key based的话，大概可以看一下github

要不要试一下和uchicago的机器建立public key auth?同时让我不要输入passphrase. Solved

问他：一个是key放在哪儿，然后怎么能都用它

还有migrate是直接创建新的在Bunker\_apps，旧的要动么

怎么release

我要改的是ABN文件夹下面的ABNJounal里面的task包里面的common文件里面的ABNArchiveProcess，把password删了，把pkey\_path加上去即可。这个是针对load\_abn\_journal\_task

还是先要解决验证的问题

<https://stackoverflow.com/questions/49587932/i-dont-have-ssh-config-file-but-multiple-github-accounts-works-how>

这是答案

把配置文件设置一下，（看一下sftp操作，optional），（git验证方式，optional）. Solved

问题有二：一个是究竟要下载到哪里去，第二个是看一下path factory和process

卧槽，突然意识到这是两个不一样的，注意username是不同的

我就是不知道上传上去的那个locatesin.dat去哪儿了…找不到。可能是abn那边定时删除了把。

原来的staticmethod都要改掉。。。

怎么保证bunker\_apps运行的时候用的是最新的版本？Ans: conda自动安装的是最新版本。那么要用conda update

我猜测的是你上传一个文件到那个位置，它就会返回给你今天的情况

写email。。。

那个发送的邮件，实际上就是把内容在邮件body里面展示一下，然后再把内容用附件发送过去

等会儿可以看下multi…是神马

我对python的module该怎么引用并不十分清楚….

如果pull request被approve了，就可以去merge了。然后就可以invoke release version\_number，这样其它地方就可以用这个package了

先让我测试一下load吧。我要做的事情如下：第一步是自己放一个待下载的文件上去，第二步是创建一个abn的运行环境（conda），然后把代码放过去，然后尝试run一下，记得改一下邮箱名字

再测试一下upload

要解决的几个问题：

问题：假设我做了pull request之后又有更新了，那该怎么弄？是不是说如果别人更新的是非我更新的那块就没事？有点迷茫…之后再看下git

Docker暂时和我应该无关

Linux: export命令

Prod和dev有啥区别？Ans: dev是测试用的，所以跑出来的output也是放在remote里面对应的dev文件夹下（即把prod换成dev）

补充：

第一点：所有重要的数据都是放在共享的网盘上面的，即例如'/mnt/bunker/dev/abn\_files/20200610/'，所以和用哪台机器测试应该关系不大？

'/mnt/bunker/dev/bloomberg/20200611/20200611\_bnkr\_eod.out.gz.encoded'这个是load bbg里面的例子，一样的道理

所以我写的那个load\_abn\_locates其实应该是把dev(test)和prd分开来。。。不过它本来也没分开，都堆在/mnt/ndrive下面

第二点：至少bunker每个entry\_point都有dev和prd两个选项，这就导致了数据放的位置不同？

接着看abn\_locates。

Load\_abn\_locates.py干的事情很清楚，就是从ftp拿下来东西然后邮件发出去。

原来是还有一个upload\_abn\_locates.py，这个做的事情是submit request。

还有一个问题：我是否可以打tag测试。。不要每次都PR然后release了才能测试

<https://setuptools.readthedocs.io/en/latest/> python打包问题，可以看看

下午把组里的代码稍微浏览一下。也看不出太多名堂…

portfolio\_bridge里面cli.py担任了main.py的任务，就是entry point，里面可以让user传config file。还有一个kill是用linux命令来终止进程。这个service是运行在docker上面的

position\_bridge\_3也差不多

我想把entry也写进去。。不然entry那边太乱了。但其实还好，现在的diagram就是从两个真正开始的地方画起的

\_\_init\_\_ cannot be async…

等下：我写的这个东西，比如self.result里面储藏结果，会不会造成多个process同时在访问这个变量？哦这边应该没事，因为我本来就是想让它们同时访问，因为它是要subscribe多个topic的，而其实这些topic都是同一次请求的结果的分发topic

问题：是该先写test，还是先跑？我觉得一样

Collection[int]和tuple[int]难道不是包含关系？不是应该…

有两件事：第一个是看一下abc，第二个是试一下click

Click简言之就是一样从命令行call函数，只是参数是在用户输入时给定的。

解决问题：为什么from . import 不对

注意：from pktestttt import bb就可以…

记录：config在db里面只有一个地方用，就是得到solaceinfo；在service里面有两个，一个也是得到solaceinfo，另一个是得到bbg file header

Aidl的问题：把expiry\_dt换成string，fixedpoint换成float

我要把logger都加上去…这个还是很有用的

Run\_in\_executor会block？还是我用的不对？

要学习怎么写test

Logging模块就不甩attribute error了？

最后在transform里面把这些数据再还原

现在的版本并不general，应该要fetched\_bbg\_result: List[Tuple[Tuple[str, str, dt.date, float, str, str], Dict[str, str]]]

List问题，extend之后bytes就不对了

核心之一是说，当我的service开始subscribe一个topic并且提供一个callback func后，只要我一直subscribe着，当有request出现在那个topic的时候，我的callback func就会帮我抓取下来。这个我不用关心的api里面用的是机制是while loop？

使用fixture的好处之一，是在写测试时候只需要考虑核心逻辑，而不用考虑准备工作

即Given, when, then. Fixture就是given（前提）的部分

今天想做的事情：把test写一些，然后再改code？还是等到Mike明天一起看完再改？我觉得不能等，还是先不改review，把新功能实现？

附录D看Python包应该长什么样？

我是先改一些code再写test还是怎么样？先写一些test把。比如test\_main肯定可以写

能不能对datagram写一个更详细的type类型？把None或者””之类的东西梳理清楚

我觉得先把test写了肯定没关系，到时候大不了改一改名字之类的，大的test框架不会变

目前核心还是想多一些参数和例子，让test更general一些

Patch的核心就是我封装一些接口，测试的是当下我想测试的逻辑

如果我不mock一个对象，但它的初始化用了mock，那么我call它正常的方法，它会返回什么？

mock\_extract: MagicMock = mocker.patch.object(

closing\_prices\_transformer.extractor, "extract", return\_value=None

)

这玩意太坑爹了，一定是要改那个class的member的method，你改那个member本来的class (Query\_Extractor)的method是没用的

先把load\_bbg\_closing\_price.py测一下，反正是复制过来的，如果有现成testcase. Done

然后可以学photon那边对db里面的aenter和aexit也写个测试

然后把端口都对general cases进行扩展，db这边不用动，就是本来的样子

然后再给service写测试

做一个design pattern把这个转不转成None的东西剥离出来

我觉得test的核心是，对于中间逻辑，测试的就是每个函数是不是被正常的按照计划call了，最后测到底层业务函数时，再去测那个数据是不是准确处理了？

Topic builder有一个违反dependency injection的问题（？）

我其实要把on\_result\_info替换一下，做一个mock

我知道怎么回事了。。因为根本on\_result\_info就没被call，它是在等有东西发送了才会被call，而那边并没有东西发送

所以可以mock之后手动call

花一点时间再读一下第三章，看看是否有收获

在pytest里面，mock似乎兼备了stub和mock的两个功能

我现在是写test还是改代码？

我希望快速把测试框架写出来，然后refactor代码——包括添加功能，support general request

今日目标：既然要改了，是不是看一下pycharm的自动化重构？以及正好看一下那两个代码是干啥的（这个不太重要）

所以目标就主要是重构。当然老板那一手linux查询也很酷炫

Backlog是啥

Changelog

<https://keepachangelog.com/en/1.0.0/>

今天那个无非就是确认一下上传和下载的分别路径、名字即可

Mount下来的就是local的了？

解决方法，给idl\_defs.IdlVariableLengthObject里面的pack method改一改，让它support float,double之类。第二种，你直接用python自带的pack method不可以么？

等下。。我要改的是parser?不是parser，是template

注意的是，我改的其实还是protocol，所以到时候一样build一个测试一下就可以了

目标：我有一个文件夹，但我希望在不同的branch上面做事情

Ans: 很简单，你先创建一个remote branch，然后切换到那个remote branch去即可（同时也创建对应的local branch—git checkout -b <new-branch-name> <remote-branch-name>）

我觉得这个package改完之后，我完全可以本地conda build然后测试？

Ans: protocol应该不可以，还有一个.sh之类的东西运行把，不是直接变成package就行

当我已经conda build了之后，应该就是conda promote，但是它得找得到？

<https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/create-custom-channels.html>

真的要好好看一下conda整个过程

部署总是很麻烦，因为我要改一个地方，然后要别人approve，然后才能做。

所以我希望能够绕开approve也可以测试？（目前network drive我可以测试，那就没啥特别的）然后我可以build自己的tag的package，也可以用pycharm里面local引入来测试

PR到一个branch？PR不是给branch和master合并的么。。。这没问题

Serializable是什么…

如果type是确定的话，特别是长度是确定的话

下一步要造一个msg，然后试一试初始化

比较两个不同type的东西，除了有index索引的东西

变基

压缩commits已经知道了

Conda build

General conda可以看一下

Git可以再看一下

现在看一下orc\_position\_latest是从哪儿来的，

Orc\_position就是一点点加出来的

注意，orc\_position至少在load\_orc\_positions里面是最不重要的一个了。。然后每次都会被重新加载（因为没有先drop，虽然好像也没人care）。为什么要rerun，主要原因是orc\_trades和orc\_positions\_latest可能会被人为改动。

Product table里面有个namespace，和IQS里面的us\_commodity=2是对应的。

注意，这里有个问题。。。必须要import bunker\_apps，然后才能搞到config

我很想画一个datagram让整个数据流程清晰一些

是ayncio压制了错误

为什么一定要有那个跑了才行？把data diagram画出来

Git里面在不同branch的tag有啥区别？

Db不能insert into view?

打算做的事情：

一个是把转换类型的接口分离出来

要把error handling msg加上去

Topic是不是重写一下

注：比如是不是所有的转换都应该放在msgbuilder里面？

从sequence diagram里面再看看哪几个位置是发生了类型格式转换。大不了最简单的就每种转换都写一个toXxx的方法呗

还有一个，设置一些缩写

两个问题可待继续研究：第一个，jenkins上面build和local build有啥区别（一个当然是可以全公司都用了，那么我local build的怎么用呢）？第二个，那就是conda build出来的玩意儿有啥用…

这个神奇的硬盘位置"/net/binlib/build/docs"究竟是啥？从来没找到过。Ans：找到了，就在自己的盘上…

/net和/mnt比较像，也是一个nfs shared drive

Pytest可以指定pdb从哪儿开始么。差不多可以，用—trace。纯pdb可能没有，毕竟还有low level api

我知道了，参数化，如果有两个参数的话，它做的是m\*n次test，其中m和n分别是各自参数的长度。现在让我看看有没有不是cross的？答案就是把输入和期待输出放在一个fixture parameter里面就行了，只需要用一个参数化的fixture即可

注意-v挺好用的

不知道bbg\_header是不是要改

挺难测试的，不知道怎么测，即测试除了closing price之外的东西

究竟怎么Build一个package我可以自己测试？关键似乎是deploy那一步，还是Build那一步？

Ans:首先肯定要build，然后就是要Install，install可以让conda search local-channel，应该可以。其实channal都不用写，直接安装就是了

例子：conda install conda.dist/linux-64/bbg\_service-0.0.0-py36\_0.tar.bz2

令人头疼的python package问题

\_\_main\_\_.db.callback(\*args)可以直接call？这个callback是啥

Collections和tuple哪个是base class

With open和直接open是可能有区别的，因为前者才会call \_\_enter\_\_方法

# <https://bitbucket/projects/BNK/repos/aiodb/browse/aiodb/utils.py#69>

之后可以看下这个文档和asyncio里面的contextlib

<https://docs.python.org/3.7/library/contextlib.html#contextlib.AsyncExitStack>

目前对于哪个underlying sql的做法是先不动它。。。默认它会干正确的事情。如果真的是想查

苍天啊，Mock()弄出来的东西是没有\_\_iter\_\_方法的，很麻烦，要用MagicMock才对

<https://stackoverflow.com/questions/8166633/mocking-file-objects-or-iterables-in-python>

并行的话就可能有上下文切换。所以重点在于sql里面不能写yield？

Aiodb到底aio的是什么？我看它在dbapi.py里面最终调用的是psycopg2库的代码。我猜aio的是和数据库连接，而不是查询，因为查询都在lock里面：

<https://bitbucket/projects/BNK/repos/aiodb/browse/aiodb/postgres/dbapi.py#182>

一个是希望顺便看看怎么提高部署效率，一个是希望看一下finance，一个是希望加测试

Dockerize先？

Path那个，用过往包裹的应该没啥问题吧？Ans:如果不re-build就没问题，re-build的话就有问题

Pycharm里面设置Conda env也有点奇怪

Ans: 在existing env里面设置interpreter即可

<https://www.jetbrains.com/help/pycharm/conda-support-creating-conda-virtual-environment.html>

正解是用existing env，interpretor选python

新建包裹我知道可以用inv build来做，但觉得可以怎么样更快一点。果然不行，并没有覆盖原来的package啊。为什么conda install path/to/package并没有安装最新的？然后我发现conda install dist下面的那个就行，所以dist和conda.dist有啥区别？

给ide加plugin. Solved，好用又便捷

Enter/exit太复杂了，要不要理一理？Solved，就是包装版的\_\_enter\_\_和\_\_exit\_\_

等build了之后把bbg\_service里面依赖的版本号改了，然后make a push。不打算加test了

其实问题也不大，可以先自己tag一个，然后依赖于那个

Bitbucket里面怎么看一段的diff

以及conda install默认安装的是哪个版本？（即我怎么tag才不至于影响到别人）

Sql里面加了个extract出来的缩写就成了？

you can build whatever you want on whatever branch you want but you shouldn’t release unless tag is on master

问题：jenkins这些项目是run在什么机器上的？

还有我assume /net/binlib/build/python\_tools/repo/dev和http那个是同一个东西，只是一个显示在了网上链接里

看起来docker用来测试的核心之一是“共享”了磁盘（volumn？）

我已经build然后push到我们公司的registry了，然后捏？

对于portainer上面的app，它设定的运行时间在哪儿？它的参数是怎么给的？

话说既然有一个UI界面，说实话我觉得自己也可以直接上传？

我算是测过了还是没测过？

Add-on是啥玩意儿

看一下orc service我就知道它是怎么docker部署的，没问题。那其他没有dockerize的service是怎么弄的还得再问问

Docker基本看清楚了，不太清楚的也可以再看看doc，比如env之类的。接下来的问题是，没有dockerize的service是怎么搞的，以及如果我想要dockerize我的service然后部署该怎么搞：后者似乎可以参见orc service。

如果我可以直接docker跑的话，为什么我要那么麻烦再加那些.yml或.env文件？

一个是回答为什么要portainer，我觉得是因为可以在上面跑service，然后trader或者其他人就可以用同样的kunlun env测试？

生产用的是什么kunlun环境？

是不是复制到ch1prd就算release了？

Fast fail会不会让两个process同时访问到？像taxi-sim里面是最简单的情况，Event()没有任何side effect，只是一个dataclass。但如果Event有side effect的呢，event loop实际做的是不是context swtich的事情？显然，如果有一个global variable，然后有两个函数都访问(shared data)的话，哪怕不是asyncio的，肯定也都会更改。比如我们的service的SvcUtilApp实际上就只创建了一个而已，所以它的self变量都是会被重复访问。那关键只要on\_request每次都是像一个“新的”，那就没有问题了。这样的话，就算切换执行，其实也都是切换的不同的东西。关键还是落在callback function到底是怎么让它等候的

以及我的package怎么能像protocol一样放到dev去？

是否要给fast fail写一个strategy design pattern？貌似不合适，因为等于全部重写了

把generator视为task（因为yield可以挂起），用scheduler来做task switching。这就是一个user level的threading，其实没有什么底层玄学在里面。generator就是tasks，而yield就是“操作系统”给它们发送停止/继续运行的信号。

我觉得其实asyncio的核心就是在yield的基础上做了一些调度安排罢了，只是这个调度安排或event loop我们也看不到。（同时generator变成了coroutine）。其余从本质上说和正常运行没啥两样。

是不是说pdle上bunker的项目的Dockerfile和bunker文件夹里面的是一样的？

其实还是挺clear，整个pdle文件夹要做的其实就是两件事，build image以及create container

我就用自己的docker run在situate上跑？

问题：为什么放到pdle上面就可以弄到kunlunch了？是因为build.sh的魔力么，所以我不merge应该也可以搞上去

Image里面不用指定env，可以先搞到kunlunch上面，到创建container的时候再去设置env

为什么一定要docker build接着一个docker push才行？关键是tag貌似

先修改ssh登录方式: solved，实际上就是把key copy过去就行

在我想要jenkins跑之前本地测试package，除了inv build还能怎么做？

Git push origin tag是把tag push上去了，但没有commit？

还想加一个local docker build for test，因为快。现在看上去有两个选择，可以用if condition，或者用multi-stage build

Conda index可以删除么

How to put env in cmd in docker

这个不可行。不过我终于找到了！Env是可以在ch.yml之类的里面设置的

<https://docs.docker.com/compose/environment-variables/>

我是不是应该把entry point的flag什么都放在yml里面？

<https://docs.docker.com/compose/compose-file/#command>

对，应该放在那个里面cmd，这样旧的dockerfile就恢复回来

Db那边也想弄一个exception handling

Sending graceful shutdown heartbeat

关键今天就是弄这个

还有一个dataclass

Asyncio里面raise exception的问题: 即正常退出。这个似乎过于难了对我暂时而言

是不是考虑可以改成try/except/finally？Ans: no need

Conda activte怎么缩写? Ans: <https://askubuntu.com/questions/328162/how-to-create-shortcut-for-a-command-in-terminal>

还要加的话添加到.bash\_aliases即可

Linux cmd line里面也有写

测试通过solace直接发过来的情况，做一些logging

还有python里面能不能写ifdebug，这样我就不用每次删改了

Ans: <https://stackoverflow.com/questions/13352677/python-equivalent-for-ifdef-debug>

那问题在于打包了之后有entry point了，我还怎么写？暂时没有好办法

Ans:打包之后之所以是这样直接call，是因为setup.py里面的setuptools.setup里面的entry point已经写好了。

https://stackoverflow.com/questions/45930920/how-to-start-debugger-with-an-entry-point-script

以及很有趣地发现/sandbox/conda/envs/bbg\_service/bin里面都是些执行python的script，和ch1prd上面是一样的

去看一下上次ding说的conda环境，准备部署

\_\_init\_\_里面要写？目前应该是package引用有很大问题，有cyclic ref

看一下package

<https://stackoverflow.com/questions/16981921/relative-imports-in-python-3>

mmb，最后发现，其实都没问题，只是import的顺序问题？似乎确实是matter的，当我在\_\_init\_\_里面import的时候（cyclic import）

实际上问的问题是，如果我pip install一个editable的package，那么我import的时候import的是它还是directory里面的

这就是为什么需要relative import，看看书上string.py的那个例子

<https://www.pythoncentral.io/how-to-create-a-python-package/>这个告诉我们什么叫在package里面跑，而不是import一个module

万年未解之谜总算知道了！为什么会有这么多奇奇怪怪的问题，是因为把package和module彻底混为一谈了。

Ans:我是在build package，所以实际上内部应该都可以用from . import xxx。但更为关键的问题是，我不应该在内部，比如去运行\_\_main\_\_.py，因为从内部去运行的话，实际上它就不知道.指的是什么东西，或者parent package root，而我之前是通过from bbg\_service import，实际上这个bbg\_service是我Pip install的editable的package。所以正确的测试方式应该是用写的package的entry point去跑。

还有一个goal是熟悉bunker，确实需要

Exception. Solved

Pdb break其它module可以用<https://stackoverflow.com/questions/13589736/how-to-set-breakpoint-in-another-module-dont-set-it-on-function-definition-lin>

输出全名是ok的，import 我又失败了

<https://stackoverflow.com/questions/55695873/python3-7-module-not-found-even-though-it-is-listed-in-sys-path>

是在下学艺不精，结合上下两个看就知道。From bbg\_service import solace，同时solace的路径也要在里面加进去。

<https://docs.conda.io/projects/conda-build/en/latest/resources/variants.html>

exception的问题，似乎是因为我是在callback func里面

是否可以在我定义的exception里面退出？

一个看db view，一个看docker mount（-v mnt/bunker:mnt/bunker），还有一个Log的事情，还有一个delete query（这个不太好弄，delete要删除white\_list\_products，除非再传一个msg，不过问题不大，因为新进来的总是priority高）。证明了写的没有问题。

我猜是load到closing\_prices时没有overwrite的关系

<https://stackoverflow.com/questions/43938648/redirect-log-files-created-in-a-docker-container-to-stdout-stderr>

logging为什么不写到file的答案

<https://stackoverflow.com/questions/20240464/python-logging-file-is-not-working-when-using-logging-basicconfig>

<https://stackoverflow.com/questions/15727420/using-logging-in-multiple-modules>

还是得看一下library doc

<https://docs.python.org/3/howto/logging.html#configuring-logging-for-a-library>

最后我还是强行sudo chmod -R 777 ./加了权限，应该有更好的办法

<https://unix.stackexchange.com/questions/479890/permission-to-write-to-log>

<https://stackoverflow.com/questions/38676437/chmod-changing-permissions-of-my-script-sh-operation-not-permitted>

如果网上不容易找到答案，还是要多看看组里其它的code

Mount drive没有up to date?还是因为是我local的mount drive所以没有update？

Ans:我用mount命令看了下，应该是说在ch1devbuild05上面，实际上并没有mount，只是创建了这么个directory而已。在ch1prd和我自己的box上面，两者就是同步的

Git merge conflict只是照着书来了一遍，自己还是不太明白说实话

下面这个里面，先拉到feature branch，再通过PR的形式应该符合我需求。当然前提是你的local master branch必须是和remote一样最新的。不过这会不会有个问题，导致我的PR里面莫名其妙多了很多其他东西？

<https://community.atlassian.com/t5/Bitbucket-questions/How-can-I-solve-a-conflict-in-a-pull-request/qaq-p/577205>

下面这个是reset merge（取消merge）

https://stackoverflow.com/questions/6006737/git-merge-errors/6007083#6007083

packaging。似乎是inv release的时候，会自动把version update？

<https://packaging.python.org/tutorials/packaging-projects/>

类似的version可以参看Rec\_service，基本就是从version.\_\_version\_\_里面获取当下的

一般情况下直接attach就可以了，但Protocol那个是要build了才行，所以要local build一下再看

我是否可以看看怎么自己build一个package？应该可以conda build

TODO:设置一下akuna登陆的home directory

为什么我在repo里面搜不到自己的bbg\_service？

我想看Pyservice怎么运行，应该用pdb去弄比较好

怎么快速地熟悉一个service

Lockfile in linux: Each copy of the daemon will try to create a file and write its process ID in the file

下一步：service怎么异步或并行的原理？

看一下linux系统权限 ls -l即可。比如我可以log in ch1prd，怎么看我是否在这个名单里？

Windows权限怎么看 property--security

还有sudo到底干了啥，sudo不代表在所有group中？

Violet是干啥

Log out <https://docs.vmware.com/en/VMware-Horizon-Client-for-Windows/5.4/horizon-client-windows-user/GUID-5378232D-008C-47DC-8B61-4F44947C0419.html>

一个是能不能更好地进行搜索，找到我需要的代码

一个是需不需要看看其他service？还是说之后有ticket了再看？

Python Package怎么debug，以及它保存哪儿去了

怎么把现在改变的内容转移到另一个branch上面？

Enum类型用.\_\_dict\_\_来获得key和value?

Egg文件类似包裹

<https://stackoverflow.com/questions/2051192/what-is-a-python-egg>

如果想打开看的话unzip即可

<https://stackoverflow.com/questions/3655352/how-to-access-files-inside-a-python-egg-file>

先找到log文件存在哪儿

/var/log/akunadev/bunker/orc\_service

只要有context manager，创建postgre client时候就会自动连接

db\_client = await exit\_stack.enter\_async\_context(aiodb.create\_postgres\_client(db\_info, APP\_NAME))

<https://bitbucket/projects/BNK/repos/aiodb/browse/aiodb/utils.py#70>

wait\_for\_interruption的关键就是wait\_first，即创建了一个signal的future，当我退出的时候就会触发

看是不是handle exception，通过在我的callback func里面加一个exception即可

所以要让我的db能够handle exception并退出，只要确保有build\_asyncio\_error\_future()？

https://bitbucket/projects/BNK/repos/svcutil/browse/svcutil/direct.py#24

不过这个是对service用的，db还不太能用。要直接low level肉搏了？asyncio\_error\_future

<https://bitbucket/projects/BNK/repos/svcutil/browse/svcutil/aio.py#22>

发现核心问题是，callback里面的error并不直接体现到使用callback的stack上面。所以要么在结束的地方收集所有未处理的async error（但是我又用了wait block住了整个thread）；要么在内部handle

export PATH=/opt/akuna/trade\_bridge/current/bin:$PATH

做的事情是把新的一条path加到之前的path里面去

Source与anaconda/bin/activate

<https://stackoverflow.com/questions/27220145/anaconda-activate>

这个是不是就是conda activate？

<https://askubuntu.com/questions/849470/how-do-i-activate-a-conda-environment-in-my-bashrc>

我觉得看起来是volumn mount出现了问题。Ans：no

核心问题是run daemon，如果是在docker的话，它会直接退出。所以问题就是在于能不能不退出.

<https://stackoverflow.com/questions/25775266/how-to-keep-docker-container-running-after-starting-services>

现在确认两个东西：一个是mounted volumn，一个是scheduled time，还有改了config

可能的问题：先跑一遍全部的，然后再跑单个的

<https://medium.com/@blomsmail/linux-what-does-1-dev-null-2-1-actually-mean-14ebcc07bcdc>

conda真不是很好用…都不能跟着meta.yaml随时更新

看dependency

<https://stackoverflow.com/questions/55912363/how-to-analyze-dependency-tree-for-conda>

conda-tree whoneeds numba

conda-tree whoneeds msgpack-python

conda-tree depends msgpack-python

因为procotol 3.1.79时候没有peg

Spline bridge似乎logging error是写到前台，全部的log是写到log file

之所以spline\_bridge里面splinebridge.py会出现在/sandbox/conda/envs/orcsplinebridge/bin里面，是因为<https://bitbucket/projects/BNK/repos/orcsplinebridge/browse/setup.py#22>，和package install的setup有关

QUES: how is bbg closing price计算出来的？

很奇怪，为什么transform就不能step进去，load就可以？

Solved。因为一个是generator

SPDR就是黄金ETF

Public的ssh可以测试<https://www.sftp.net/public-online-sftp-servers>

即使是password based的sftp，确实我在local也是没办法连上去，在公司网上是可以的。

接下来要看的：sftp，trading

how to find the overflow db row elegantly?

What does it do with new\_expiry?

Merge master into feature branch

https://stackoverflow.com/questions/16955980/git-merge-master-into-feature-branch

ctypes可以看看

是以idlc为entry program，以idlparser作为parser，idl作为generator(compiler)

自动化测试

Bbg\_service之后想update的地方：

把msg的logging用decorator包起来，这样以后要不要logging某个protocol msg直接在solace的msg函数那边改就可以，不用到处找究竟哪里call了logging

Bbg header可能也要另外单独拿出来，作为config

之后要把run-db分出来，正好趁这个机会把代码重构一下（是否要动setup.py?）

要做的核心有三部分：通过input file里面的信息找到对应的在我们instrumental db里面的东西，然后拉取数据（connect with instrumental部分，似乎有点麻烦，它还是用neosolace和tornado；但不应该麻烦啊，如果只是一个topic的话；还有一个问题是我们的机器上不能connect到prd的solace？）；找到desk mapping；写出到一个csv；转界面（这个容易）。

对我来说，mapping可能最难的，我先看看excel里面的东西。

要直接引用某个库，需要在pycharm里面右键直接run

utf-8-sig的问题

https://stackoverflow.com/questions/57152985/what-is-the-difference-between-utf-8-and-utf-8-sig

yield from以及reader

<https://stackoverflow.com/questions/9708902/in-practice-what-are-the-main-uses-for-the-new-yield-from-syntax-in-python-3>

因为CSVReader其实也是一个generator，一行一行地yield，所以我们直接yield from CSVReader() = for xxx in CSVReader(): yield xxx

有趣，这个回答了yield的底层，一个是弄成了闭包，另一个是保存一个有\_\_iter\_\_ method的对象。

<https://stackoverflow.com/questions/8389812/how-are-generators-and-coroutines-implemented-in-cpython>

yield from例子，如果要看完整版的，还是要看await。

<https://stackoverflow.com/questions/8389812/how-are-generators-and-coroutines-implemented-in-cpython>

现在依然有一个matching的问题。Offset day是啥玩意？

我可以通过exchange+PF code来在abn\_products找underlying,，Underlying可能没有expiry\_dt；那么怎么通过underlying来找那个instrument，进而也就获得expiry date呢？csv里面有contract month

目前流程once i know which underlying in IQS the row refers to, i will use contract month to find the expirations and then get the future. for option i will use strike price and type as well to find the option.

45行找不到；NYM怎么处理(XNYM)。找不到是因为abn\_products里面没有，还是真的不trade；各家的symbol都不太一样？bbg和abn的，有些又都一样，有没有什么mapping除了abn。怎么看能不能booked?还有说reconcil里面看mapping

Gmi code

Neosolace

res = yield ptron\_service.get(pp.named\_desk\_spec)

# p res[1].underlying\_ids

# https://bitbucket/projects/LEG/repos/packages/browse/neosolace/neosolace/idlproxy.py#64

# 不知道哪儿设的getter方法, 很有趣的type，<class 'neosolace.idlproxy.named\_desk\_spec'>，似乎是动态生成的

# https://bitbucket/projects/LEG/repos/packages/browse/neosolace/neosolace/idlproxy.py#139 看这行里面的origclass

# ah 是这个创建了metaclass

<https://bitbucket/projects/LEG/repos/packages/browse/neosolace/neosolace/idlproxy.py#17>

is\_booked还要问一问，orc\_trades里面怎么看？

是不是建一个table

有没有return yield这个写法？

https://stackoverflow.com/questions/44051647/get-params-sent-to-a-subcommand-of-a-click-group

还是packaging的问题，python怎么分包exe包？以及为什么python的exec文件里面都是placeholder，执行不了。Pyinstaller可以么https://pyinstaller.readthedocs.io/en/stable/operating-mode.html

#!/bin/sh

'''exec' /opt/akuna/envs/bunker\_apps/versions/24.14.5/bin/python "$0" "$@"

' '''

# -\*- coding: utf-8 -\*-

import re

import sys

from bunker\_apps.load\_orc\_closing\_prices import entry\_point

if \_\_name\_\_ == '\_\_main\_\_':

sys.argv[0] = re.sub(r'(-script\.pyw|\.exe)?$', '', sys.argv[0])

sys.exit(entry\_point())

#!/bin/sh

'''exec' /sandbox/conda/envs/bunker\_apps/conda-bld/bunker\_apps\_1601053950182/\_h\_env\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_placehold\_/bin/python "$0" "$@"

' '''

# -\*- coding: utf-8 -\*-

import re

import sys

from bunker\_apps.load\_bbg\_closing\_prices import entry\_point

if \_\_name\_\_ == '\_\_main\_\_':

sys.argv[0] = re.sub(r'(-script\.pyw|\.exe)?$', '', sys.argv[0])

sys.exit(entry\_point())

一个是prd里面的，一个是自己生成的package里面的，看到区别了把

pythonw

https://github.com/chriskiehl/Gooey/issues/259

跨平台打包还是有问题

https://stackoverflow.com/questions/2950971/packaging-a-python-script-on-linux-into-a-windows-executable

pyinstaller打包的一些细节技巧

<https://blog.csdn.net/zengxiantao1994/article/details/76578421>

<https://medium.com/@peaceful0907/%E5%B0%87%E4%BD%A0%E7%9A%84python-code-%E6%89%93%E5%8C%85-pyinstaller-6777d0e06f58>

关于python打包的讨论

https://juejin.im/post/6844903936722206728

https://python-packaging.readthedocs.io/en/latest/minimal.html

这是官方教程了吧

https://packaging.python.org/overview/

这也是书的一个章节，还参考了pep，比较抽象宏观，是打包库的作者写的

https://www.aosabook.org/en/packaging.html

这是介绍实践中怎么把一个package上传到pypel

https://realpython.com/pypi-publish-python-package/

这也是一本书

<https://diveintopython3.net/packaging.html>

是不是说pyinstaller是把当前环境下的安装的package都打包进去？

<https://docs.python.org/3/library/decimal.html>

这问题难道实际变成了，如何add site-packages到pyinstaller?

<https://stackoverflow.com/questions/59743234/how-can-i-add-a-python-site-package-folder-thats-not-being-included-to-a-pyin>

还有一个就是cross platform的问题

Exchange名字mapping

<https://support.quotemedia.com/support/solutions/articles/13000020527-exchange-code-group-list>

如果要静态修改.req file的话，一个是要把submit\_req里面write\_reqest\_file注释掉，然后要把path factory里面的request\_id注释掉（不然每次生成的路径都不一样），然后path factory可能得把ts也改成静态的。

qQ19941224!

# 30183565

dash

<https://www.dashregtech.com/>

<https://wiki/display/DEV/Bunker+DASH+process>

明日计划：dash，protocol。如果bbg有回的话那就继续问那个

<https://stackoverflow.com/questions/52494425/running-a-pytest-test-from-cmake-where-the-test-and-sources-are-in-different-fol>

conda.yaml似乎要加pytest在build里面。Ans:老板说不需要

git怎么在merge之前就解决conflict？以及clean PR。看起来核心是squash commits

<https://hackernoon.com/make-clean-pull-requests-version-control-and-collaboration-420ecff2e7b5>

关于git，我的问题是：第一，我知道merge会有conflict，我能不能在PR时候就solve conflict？第二，我从local push到remote的时候，能不能就squash commits（而不是在PR merge的时候）？

正确的做法是不是在local先把最新的master merge到我自己的branch上面？Ans: yes（前提是master上面并没有把你改的地方也改了把）

>>> a=[(1,2,3),(4,5,6)]

>>> b={x:(y,z) for x,y,z in a}

>>> b

{1: (2, 3), 4: (5, 6)}

从来没见过这语法

要做的事情是，替换PeriodicCallback。当然首先得知道这玩意儿究竟是干嘛的

Asyncio教程<https://github.com/jmoldow/jmoldow-python3-async-tutorial>

很糟糕

说generator和coroutine主要区别在于一个是可以决定下一步执行什么，一个不可。为啥说generator是iterator的简化版？kind of，毕竟generator是iterator的subclass。一个是写yield，一个是要explicitly写下一个循环是什么

<https://en.wikipedia.org/wiki/Coroutine#Comparison_with_generators>

希望完成一个功能：能够将一些test版本的设置保留在一个branch上面。所以git checkout -b是把当前branch上最新commits复制过去，还是没有staged也复制过去？Ans:从操作结果来看，没有staged的也复制过去了。

似乎情况是这样，checkout到new branch之后，如果new branch把修改的地方stage和commit了，旧branch上面就没有unstaged的东西了

Bbg service:目前的情况是，对于closing prices or not，我们有截然不同的做法。

不管是closing prices还是其他，我们都可以有comment line（这没问题），并且在code里面标出comment line的格式。对于没有提供expiry date之类的field，我们就不去管它就是了。

这UnderliersQuery是不是应该放到db那边去？只要我确定pass过来的名字是直接可以用的。Ans: yes。bunker\_utils.BBGSymbolConverter（在bbg\_utils.py里面）是之前用的mapping。

Hmm我还是必须要给service留一个mapping/symbol\_convertor的位置，而且这个应该是transparent to user的（意思是不影响protocol msg），所以它放的位置应该是BBGRequestTransformer这边，加一个flag即可，如果user加flag，那就必须在msg里面pass给service？但是我还是觉得应该transparent to service比较好？

现在我知道原本为什么.req里面要写comment了，目的就是为了这个mapping的事情，不需要再逆映射一回。

存一个local的mapping在那边。把这个class还是放在adaptor里面，同时让adapt函数能接受一个cache变量，缓存mapping。

我可以写一个symbol convertor的adaptor，只不过多加一个参数？

我让它先转成protocol msg然后再做converter可以么？而且db这边得把expiry什么都放进去。应该都放在adaptor里面做吧？

Bbg和db应该不用打任何交道。

怎么把比如昨天的所有数据都copy过来，但只是改一下日期

<https://unix.stackexchange.com/questions/153241/how-do-i-delete-a-folder-but-not-the-contents>

消除中间层，但记得用cp xxx/. 而不是mv

Bbg serivce别忘了改yaml里面加bbg\_convert和protocol version

Bunker\_apps也要改

一步步改会比较容易。先把db和service分开来

对于一个一路传进去的变量，有没有比较好的方法(whitelisted\_product，config也有这个问题).Ans:在上层创建

我能不能把config单独拿出来，比如放到一个file里面，然后再单独创建一个class是获取config（类似bunker\_utils里面的get\_config），然后把sftp，header，solace三个config分开来（它们本来就不该在一起）

把两个transformer拿出来另外放一个file

另外第二个tranformer出来不应该用零散的数据，要用type

我觉得bbg\_converter实际上就是一个adaptor

如果我把所有内容，包括msgbuilder都用我自己的type封装起来，这样是不是好？Done

因为你只需要两个converter，所以并没有必要用decorator

在bbg\_converter里面要加一个从db获取数据并存储的script，但是它不见得要打包到Package里面。Ans：看起来不用，因为就是从db导出来一个csv file而已。别忘了dump的话设置一下csv格式，让第一行的header保留着。

Db之前timeout的问题是因为password不对，按照<https://www.jetbrains.com/help/datagrip/reference-ide-settings-password-safe.html>重新设置成每次询问即可

如果跳出来问密码，用的就是aiodb的config里面的db密码，不是我自己的密码。

Db很卡的问题应该是cache的问题，可以看terminal里面有提示：UI latency。做法一个是关了重开，还有一个是把之前几千行的sql保存到非remote的本地，因为我猜datagrip一直在做grammer check。

怎么设置路径是个问题.Ans:solved

Mypy检查typing有问题，就用Any搪塞过去就行

如果一个msg里面有一个field要变动，是否应该拆散整个msg type？

要做的事情：mock file，以及改一下meta.yaml(solved，删一个空格即可)？或者应该把data file package进去。原来是因为我没有修改git ignore…data file从来就没上传上去过

What is a good test? Refactor service里面的这些error handling和if statement.

还有一个是在run time initialize的问题

可以用两个mapping，可能想看一下2sum…失败

Header config是只有一个instance，所以我把基本信息放在self里面，每次制作具体的header那就需要做一个hardcopy. Ans: copy.deepcopy

Docker的接口也要重新写一下

Python里面，比如一个类型原本是List[Optional[…]]，在我确认过它非空之后，把它传给一个接受List[…]的函数，怎么在type check里面体现？Ans: typing有个cast函数

excel用计算结果替换公式：比较naïve的方式是复制黏贴…黏贴的时候选择黏贴值

mmb，把fireblocks里面的bunker\_apps comment掉

pdb妙用，在package中断点（可能要先pdb里面import etl）

b etl/task.py:77

为什么会多一行，估计就是每次bank transfer产生的hash不同

原来所谓807条，说的就是全部fireblocks的记录

Crypto service的代码现在我们还看得到么？Ans:已经获得

opus也是一种加密货币。Gemini是一个比特币交易所

python or和and，即python的或

https://docs.python.org/3/reference/expressions.html#and

Crypto journal:

先从match\_crypto\_transactions.py开始，这个可以选择match一个月（会从月初跑到run script的前一天），或者一天天来。然后默认应该选clear-all flag，至于为啥我还没看懂。

注：crypto\_transfer table里面的date是我们跑script时候的每一天，而trans\_dt是withdrawl那侧的date，这个date又是直接从ts里面获得的。

然后这个script里面会先跑一下load\_crypto\_transactions.py，从各种地方下载（可能在db里面，有的是存在mount drive里面的csv，比如fireblocks，但它最后也是load到db。注意fireblocks下载的是全年的。。。虽然代码里会帮忙过滤到其它时间的。这个load\_crypto\_transactions.py本质上就是跑存在reconciliation/exch\_tasks下面的各种task，而各种task实质上就是对db先delete（如果force\_download设为true的话，但从match\_crypto\_transactions.py里面的lct看，并没有delete crypto\_transactions的东西？注意fireblocks的task没有delete，所以要人工删除），然后query对应的exchange的table，然后transform一下，然后再load到crypto\_transactions这个db table里面（load是存在reconciliation/load/transactions.py）。这些都跑完了，然后我们又回到了match\_crypto\_transactions.py，真正开始执行它里面的CryptoMatchTask。所以说实际上第一步是load，第二步是match，第0步是谁把数据放到db里面，那里面就是交易transfer数据。Match script的clear-all flag删除的是crypto\_transfer table，并不动crypto\_transactions。而且match script目前是跑一天match一天。

注：现在新加一个bybit的load task。显然bybit\_transfer里面有address，trx\_hash，还有exch\_id。

注：不知道谁写的code，之所以load\_crypto\_transaction.py里面你哪怕不set force\_download，db table也不会因为多次load而重复，是因为exchange task里面的load用的是InsertNoConflictQuery，如果key一样的话就不会复写（在crypto\_transaction table里面key是trx\_id）。不过我觉得最好的还是把force\_download设为True，同时注意，这个deletequery删除的是这个exchange除了comment带有manual insert之外的所有。

注2：想看table schema的话，找到那个table，点击下拉展开箭头就可以了…

然后这个script的核心是CryptoMatchTask。这个task一开始先populate了三个service:alias\_service, account\_service, rate\_service。account\_service在journal/account.py里面，首先把account做了一个分类，比如loan或者bank或者exchange或者pnl什么的，查询了crypto\_accounts table，然后又做了一个mapping，似乎主要这个service功能就是做个账户分类，mapping是说用(exchange—这个exchange是指crypto\_accounts table的column name, currency, acct type)获得(acct number, name, entity)，其中name像description，entity是akda akca之类的东西。注意对于每个account type它有不同的get\_xx\_account的方法。alias\_service查询的是crypto\_party\_aliases table，好像也是做了一个mapping。有四种alias type: 'otc\_wallet', 'bank\_name', 'otc\_counterparty', 'bank\_counterparty'。然后实际上就是alias的一个mapping，从一个名字可以得到另一个名字。第三个是rate\_service，查的是fx\_rates\_view table，这个是每天更新的，从bbg来，就是获得汇率（似乎是以美元为基准货币）。

然后CryptoMatchTask开始做一些pre-process，如果set了clear\_all，db里面会删除dt之后的crypto\_transfer table里面的记录。如果set了clear\_matches，除了是手工修改的（写了manual或disabled），会把dt之后crypto\_transactions table里面所有row的match\_type column都设置为null（但为什么只是设置为null而不全部删除？）。Comment里面说clear\_al会override其它flag，见create\_preprocess\_task()。

然后CryptoMatchTask开始extract，有三个extractor（即最后会返回一个三元tuple），位置都在journal/queries.py下面。要注意的是，虽然我们是按每天跑的，但第一个extractor是会把table里面的所有数据都拿出来的（尽管理论上在前面load的时候也是一天天load的，但如果你每次跑之前不把db crypto\_transactions里面非manual或disabled的东西都删干净的话，就会出错；而且第一个extractor只要求match\_type是null，这不就意味着陈年旧unmatch也显示进去了？）。第一个extractor从crypto\_transactions table里面取出所有match\_type是null的row (这个是主要的路径，我们主要要处理的transactions的位置);第二个用的是reconciliation/parse/bank.py里面的BankTransExtractor，它读取的文件路径是/mnt/bunker/prd/crypto\_transactions/bank\_transfers，但是哪一步把文件放过去的呢（答案是人为load过去的，每个月问finance的人要bank\_tranactions的文件，如果需要的话还可以按日parse一下，代码见parse.py）？这个BankTransExtractor做的事情还是比较清楚的，如果满足bank==counterparty条件的话，表示就是一个银行内部转账，我们不用管，另外如果这个csv file里面的bank或counterparty 不在我们的crypto\_party\_aliases里面(就会raise一个keyError，同时会logging报错，提醒你要更新crypto\_party\_aliases，但一般应该不会)，就返回一条bank\_transactions给后面的transformer，至于从bank\_transfer文件创建一条可以放入crypto\_transaction table的记录(是在哪儿把bank transfer放到db的？)，其中trx\_id什么的都是自己随机生成的（见create\_bank\_trans）；第三个是OTCTransExtractor，和第二个差不多，从/mnt/bunker/prd/crypto\_transactions/trader\_otc那边得到（现在一般是空的，只是每天创建一个空文件）。要注意load\_otc这边有时会load一些奇怪的数据进来，此外otc里面的数据，symbol是真正的货币，而currency是法定货币，是我们用来计算value的。看起来OTC那一条确实symbol是ETH，但值不对，应该是30000美金对应的ETH的值。

注：bank transfer是没有process去删的，这似乎是个问题（目前我在match\_transactions.py的pre\_process里面加了一步）。不过它是在哪儿load到db table的？我猜是最后的TransferLoader干的，因为只有这个和db\_conn有关系。

然后是transform过程，CryptoMatcher。先初始化三个matcher，在reconciliation/match/\*里面，其中如BankMatcher只处理银行间交易，即货币都是法定货币。exch\_matcher是用blockchain hash来match，如果不行再OTC，再不行再bank?

首先我们来看ExchangeMatcher，它扫描所有没有match的数据（就是crypto\_transactions里面的1行），建立一个transaction\_hash（trx\_id冒号后面的部分）的mapping，对每个transaction\_hash，应该有且只有两条交易与之关联，如果确实如此，那就match上了。这部分只需要用exch\_matcher.match\_with\_hash方法，还是比较直观。exch\_matcher.transform方法，先把所有已经match上的交易排除出去，然后把剩下的交易按照qty正负分为withdrawals, deposits两类，这边实际上就是在用qty来做match了，如果qty比较近的，我们就认为match上了。那么crypto\_transactions table里面的数据是从第一步load里面，从各个分散的exchange table里面来的。

而bank\_matcher也是把bank transfer先按qty是正负分成withdrawal和deposit；对从exchange那边获得的trx也做同样的事情，同时会在self.bucket里面把之前已经match上的trx去掉。然后开始match\_bank，对withdrawal和deposit各自做一次，一次是tight match，一次是loose match。核心是match\_bank()函数。

最后是load到两个table, crypto\_transactions和crypto\_transfer，在reconciliation/load/transfers.py里面。Transform返回的new legs会被load到crypto\_transactions table，不过new legs是毛线？

TODO:

把OTC，bank，fireblocks的路径都改了，不要用/mnt，老是把我自己锁住。然后再把文件都搬过去。

一个是把fireblocks load的方式还是改回有重复的就报错（而不是直接忽略掉）。还有update comments里面提到的exchange。Ans:做完了。以后第一遍跑的时候用的是旧loader，后面用新loader。

Emmm我可以把os上要操作的东西也用bash写一下，自动化了。比如改名字，parse bank\_transfer，复制来复制去的scp之类。或者不用写bash script，用python的os.system()命令？但也是和平台相关的。搞定了。

问题：为什么会suppress assertion error？是因为如果那一天的lct跑得有错，就会直接跳过？这个exception没有冒泡传上去？

TODO：把重复的tx\_id怎么可以自动改了？

Cypto journal做完之后，一个是load\_crypto\_journal.py，另一个是loan\_and\_pnl\_email.py。

先看看load\_crypto\_journal.py。

这个核心是CryptoMatchTask。它一开始先初始化四个service，分别是account, rates和两个balance（分别是月初和月末的）。

然后是两个extractor，分别是抓去transfer（从crypto\_transfer table里面抓取）和in transit（这个是从crypto\_transactions里面找match\_type是in transit的东东）。

然后是CryptoJournalTransformer。对transfer，一上来我们先用journal/transaction.py里面的identify\_unmatched\_bank\_transactions跑，目的是parse一下。Parse完之后，就把create\_trx\_entries，create\_pnl\_entries，create\_reval\_entries（pnl和reval似乎是实际上是loan\_and\_pnl\_email.py里面跑？）这些itertools.chain()起来返回，就是把iterable串起来。

我们先来看一下create\_trx\_entries。它只有在特定情况下才会跑fee。它返回的是journal里面的一行JournalEntry。我们要关注一下liquidity和description这些都是怎么产生的。其实liquidity就是amount，只是换成以美元为单位而已。同时liquidity先会看crypto\_transfer table里面src\_otc\_notional或dest\_otc\_notional，只有这两个都是0的时候，它才会尝试用rate\_service得到的rate乘以coin来得到liquidity。Description就是把src和destination以及用的currency和交易日期放了进去。Name指的是account name，即accounts里面对应的名字。

然后是create\_pnl\_entries（它是在pnl.py里面）。首先也是把balance sheet拿过来。然后是对cypto\_transfer里面的transfer进行循环，然后算的那些liquidity什么的和create\_trx\_entries很像。然后算一个pnl，这个是一些liquidity做加减法得到。最后返回的也是JournalEntry。pnl.py里面对transit是有处理逻辑的，前提是在comment里面要加上Transit这样的字眼。注意finance不要coin\_in\_transit的pnl，所以33行就把src或dest是in\_transit的都去掉了。然后这个script里面还真有一liquidity\_sum\_check，但可惜没有能曝出说是哪个不match。

最后是create\_reval\_entries，见下面loan\_and\_pnl\_email.py。

最后是SplitLoader。这个的目的是把journal entries split到它们各自的entities里面。即上面所有这些（不管是tx还是pnl之类的）都会被load到crypto\_journal\_ALL\_ENTITIES里面？

Dec经验：一个是coin\_in\_transit不要算pnl（即create\_pnl\_entries函数里面要排除掉，这个应该以后不用改了，只要是我人工加的in\_transit都没啥问题）。另一个是如果一个transfer只有一方有fee，那create\_trx\_entries里面就要改了，要变成两倍的fee。果然，因为grayscale是手工加的，手工加的tx Ding没写fee。

Transactions.py里面还有一个generate\_loan\_data()，它是在load\_and\_pnl\_email.py里面会被call。它也是直接从db读取transfer，所以如果上面create\_trx\_entries里面更新了什么，这边也要更新。

注：所以不同的account type实际上也是对应于创建journal中不同的步骤，比如pnl就是对应创建pnl sheet。当然也是对应于不同的交易类型。

注：OTC是不可能有fee，一种做法是create account anyway，另一种做法是把它的fee加到exchange那一侧，这样最后src和dest的liquidity和fee\_liquidify才会balance

注：之前没有credit from exchange的key报错，是因为之前没有碰到这个exchange的这个情况，也就没有在accounts里面建立账户。

注：notional value和market value的区别

https://www.investopedia.com/ask/answers/050615/what-difference-between-notional-value-and-market-value.asp

itertools.chain就等于这个：

def chain(\*iterables):

for it in iterables:

for each in it:

yield (each)

注：我们在不同的exchange以及银行有各种类型（account type）的账户，用来应对不同的需求。

说到loan\_and\_pnl\_email.py，跑的话传的日期应该是每个月月末，看一下crypto\_transactions/balances里面的balance文件是啥时候的就可以。核心是CryptoBreaksTask。一上来先初始化三个service: account\_service, rate\_service和balance\_service（这个根据start和end date创建了两个）。然后还弄了一个start\_dt和prev\_dt，start\_dt是把你传给它的日期里面的日变成1号，prev\_date是算start\_dt的前一天，把周末去掉，end\_dt就是你传进去的那个日子。balance\_service的任务是读取balance file，然后实际上是读两个balance file，即crypto\_transactions/balances里面的。Rate\_service就是为了获得fx\_rate，传给它两个日期，它就会把那段时间里面的所有fx\_rate都从db里面取出来。

然后又是etl，extractor做的是TransferQuery（在journal/queries里面），查的是crypto\_transfers table里面的东西，在起始结束日期之间。

Transformer是CryptoBreaksTransformer，它把前面的account\_service, rate\_service和balance\_service都作为初始化变量，还把prev\_date和end\_dt也传进去，实际上就是说从上上个月月末，跑到上个月月末。里面的transform核心是call了journal/reval里面的create\_reval\_entries。

create\_reval\_entries把crypto\_transfer db table里面取出来的transfer data，account\_service, rate\_service和start\_balance\_service以及开始结束日期作为函数变量。这个function里面report\_date就是前面的end\_dt。首先初始化一个defualtdict loan\_revals，如果key不存在就返回Decimal('0')。这里面主要是是两步。第一步是用rate\_service做一下fx转换，而我们记得rate\_service里面主要就是一个journal/queries.py里面的FXRateQuery，查的是给定时间范围里面的fx\_rates\_view table。然而因为周末的问题（即report\_dt不是在工作日，trans\_dt在工作日），所以我们要人工把report\_dt\_fx更新一下（下次是否可以考虑让finance在folder那边放一个文件，然后我们把coin\_price.py嵌入到现在每次人工复制黏贴的reval.py里面的位置）。然后又创建了一个defualtdict reval\_total。然后又把所有的loan account都拿了出来，即all\_loan\_accounts set。

最后这个create\_reval\_entries返回的也是JournalEntry，用的是reval\_total里面的值，它又是通过new\_loans函数转换loan\_revals里面的值。这个create\_reval\_entries最后是yield两条，一条loan，一条reval。

除了create\_reval\_entries，在这个transform函数里面还有\_get\_loan\_ending\_balance。所谓loan的意思是说，在不同的entity之间的transfer（记为loan）。它具体是做啥我没仔细看。

注：为啥fx\_rate\_view table里面不是所有的price都在每天的Fx email里面发出去？注意这个script是在send\_fina\_fx\_rates.py里面。这个view是由bbg\_fx\_rates，cmc\_coin\_prices，yahoo\_coin\_prices这几个合并而成。为啥QQQ也在fx\_rates\_view里面？而且这个table里面似乎只有工作日的数据，而实际上加密货币每时每刻都在交易。注意在rate.py里面，如果我们没有这个date或currency，我们的做法是找就近日期（向前找）的汇率，即get\_second\_latest\_price这个函数。Script是load\_bbg\_fx\_rates.py。

Crypto journal 常见unmatched及处理方法：

1 对bitstamp来说，它似乎不提供USD的transfer数据，所以如果bitstamp\_akda是在target的话（一般发起方其实就是银行了），我们自己创建反向的transaction。这个新加，所以没啥问题。

2 对itbit来说，首先目前我们并没有获取它的数据的api，所以肯定是match不上的。然后它交易的是BUSD，也比较特殊一些，一般的使用情况是binance -> itbit -> silvergate，所以我们得手工补上这中间的几条记录，或者把api写了。

https://www.paxos.com/itbit/

3 对lmac来说，虽然crypto\_service写了这个exchange，但并没有lmac\_transfer table，所以我们也只能手工加上对应的交易。

4 对otc以及otc interest来说，一般都是没有trx记录的，比如blockfi,Celsius,genesis。所以手工加

5 fireblocks文件里面有notes，有的会写target，如果写了的话我们可以人工修改fireblocks load的trx的target。注意这是修改的。

6 exchange之间不太可能直接转美元，所以如果看到exchange之间转的话，看看是不是逻辑链条弄错了。应该是exchange到bank之间的。做法就是人为修改target，然后在comment里面加Manual。因为我的load的task里是insert\_no\_conflict。

Sql script见聊天记录及db table

注：match完一遍之后再post insert这是合理的。流程：把post insert放回去，跑一遍match\_crypto\_tx，再把post insert注释掉，再跑一遍match\_crypto\_tx。

1月新见：kraken，paxos busd，akuna deep vault，coinflex，nydig, sen\_usd(akuna digital sen network)

Finance说：akuna deep vault和akuna digital sen network都可以忽略

TODO：一个是自动update，一个是自动disable

再观：gemini，gdax，okex，bybit等，看db记录，fireblocks标出了target但是没match上。我先做的事情还是match fireblocks里面有明确target的。有很多是sen usd，明天问问怎么处理。说先直接忽略就行。

再更新：gdax的问题是，不少是重复的tx，我要写个script把它们挑出来，在后面disable掉。Bybit和deribit有一些是和fireblocks对应，fireblocks csv里面没有。

具体步骤：

对gdax，我只要在unmatched的里面，用exchange\_tx\_id (tx\_hash)找出那些在gemini\_transfers里面已经出现了不止一次、而且时间相近的交易，然后在crypto\_transactions table里面直接update为disabled即可。

对fireblocks里面小于0.0001的，就是dust尝试转账，依然直接disable掉即可。

对剩下没match的bybit和deribit，直接都先update它们的target为fireblocks（comment要写manual insertion），如果再没match上，那就在insertion里面自动做一下就行了，反正target都有了。

Insert的逻辑：如果之前已经跑过一次Insert，那么人工insert的rows已经在db里面了，我再一次时候，它就会先match上，然后再在insert里面实际上把这个人工insert的row给删了，同时再加上了一些其它的人工insert row，而我再跑一次的时候，其它人工insert row match上了，被删的又match不上了。这时候恐怕没办法，只能人工把manul insertion的删了再重新跑。

有趣，因为我显然每次至少要uncomment跑一遍，然后再comment掉再跑一遍script，可不可以让再一个script里面让这个script再跑一遍？应该不可，无限递归了。

Paxos是啥玩意？

这是binance usd，在paxos网站上看到的。https://www.paxos.com/busd/

注：fee rule到底是怎么样的，应不应该收两遍？

注：bakkt是一家管理公司，好像和ICE有关系，所以ice future delivery我们的target是它

注：coinflex的数据db也是missing的

注：似乎原本fireblocks不用insert\_no\_conflict的话会miss掉trx。原本因为没有说do nothing，它就会报错，但又不会退出整个process，所以其实导致那个日期后面的trx没load进去。是因为logging么？

TODO: fireblocks里面，增加一个步骤，在读取trx时，如果发现有重复id的直接就改了（比如加\_1），这样后面还是insert\_no\_conflict。

对于qty比较小的（比如小于1的），manual insert里面增加一个update row task，直接把它们批量都disable掉。

加一个select query是给所有match type =null的加reverse trx。

我们是不是要把manual insert和disabled分开来？对的。

事后人工添加：如果是external确定了，就把target在原始的csv里面更新？或者做法是写real Manual Insertion。在real Manual Insertion之后要先把之前Manul Insertion的东西删了（这个命令请手工跑一下），然后run一遍match，然后再manual insertion，然后再run一遍match（这样我们不会多加manual insertion的东西，因为manual insertion都是在match type = null里面找的）。

现在给load fireblocks的过程加一个采纳note的做法。Fireblocks的文档在

https://docs.fireblocks.com/api/#introduction

看到vault。<https://www.geeksforgeeks.org/introduction-of-vault/>

USDT\_REC（\_OMNI）之类的，都是USDT家族的？花样真多

Gemini custody，看到了注释说有这个，所以瞅一眼。<https://www.gemini.com/custody>。看起来这个是一个一般性金融监管相关的东西，不知道和书里写的比特币安全性有没有关系？

Fireblocks网站。<https://www.fireblocks.com>

姚期智提出的百万富翁问题是一个经典的multi party computation问题<https://www.fireblocks.com/what-is-mpc/>

Fireblocks的doc <https://docs.fireblocks.com/api/#introduction>

问题：什么是vault？有金库的意思，所以就是金库/银行账号的意思吧。看起来fireblocks上所有的account都叫vault？

问题：asset是什么？我猜就类似一种货币类型。

问题：address of an asset是啥？这个是什么address，url address？fireblocks文件里也没写。

问题：External Wallet是啥？是不是就是那些其他exchange里的钱包，SEN也是其中之一？

问题：exchange accounts就是指连到fireblocks workspace的交易所账号把。

补充：所以fireblocks csv里面的src type(dest type)大概也就两类，一个是vault，表示是fireblocks这边的账号，另一个是exchange，表示是与fireblocks相连的交易所的账号。而SEN\_USD这个比较特殊，实际上并不走fireblocks，所以type是空的或者写unmanaged。

补充：GET /v1/transactions，可以看看这个rest api，里面应该有很多信息，只是api doc很久没更新了吧？

问题：ERC20是什么？<https://www.investopedia.com/news/what-erc20-and-what-does-it-mean-ethereum/> 只是这不是ETH的么，怎么和USDT扯上关系了？见下<https://community.trustwallet.com/t/tether-usdt-trc20-vs-erc20/61950>

问题：有一个src名称是Akuna Digital SEN Network。所谓SEN就是这个<https://www.silvergatebank.com/solutions/digital-currency/sen> （它的用处就是可以24/7的实时转移USD，传统银行不行）

<https://docs.fireblocks.com/api/#transactiondetails>

问题：从这个doc我们可以看到，api可以获得更多的信息，是否可以真正做到更好的精准的matching？

<https://docs.pro.coinbase.com/#upcoming-changes>

coinbase pro(gdax)的api doc。

<https://github.com/danpaquin/coinbasepro-python>

还有非官方的python api library。

https://zh.wikipedia.org/zh-cn/通用唯一识别码，<https://docs.pro.coinbase.com/#ids>

UUID几乎不会重复（用作txhash？）

<https://www.investopedia.com/terms/f/fill.asp>

在doc里面看到的Fill指的是这个，意思是说被执行的（成交）order。

<https://docs.pro.coinbase.com/?python#list-deposits>

这个说明的就是deposits里面的fields。关键就是那个hash。

<https://docs.pro.coinbase.com/?python#canceled-withdrawal>

有些withdrawl是取消的，可以看到。后面还有cancel code，可以用这个来获取么？

<https://docs.pro.coinbase.com/?python#create-conversion>

比如usdc是stable coin。还可以把usd变成usdc，有这种操作？

<https://www.investopedia.com/terms/m/margin.asp>

margin trading保证金交易，意思就是借钱来交易（购买instrument）。

注：mmb，gdax基本没有说明各个field意思的doc，真是维护得很糟糕。

<https://support.cryptofacilities.com/hc/en-us/categories/115000132213-API>

Crypto facilities的api

<https://support.cryptofacilities.com/hc/en-us/articles/360000668433-Conventions-And-Definitions>

一些基本定义，比如所用的时间，有UUID，以及symbol的format。

<https://www.kraken.com/features/api>

这个是kraken的api page，所以它和crypto facilities其实是两个东西。这里面倒是对每个field都有说明，挺好的，不过也让我发现了各个exchange实在不太一样。

注：看起来各个exchange或者OTC的api都差不多，核心就是在于哪些信息（ts和hash or id）是我能用来normalize数据的。

TODO for gdax:我必须要确定各种id里面，到底哪个是有用的。

TODO:（1）能不能拿个read only api key来，看一下信息，就说留下doc来告诉后来dev fireblocks提供信息的意思。（2）为什么asset=busd那个，source type是vault？是从fireblocks走的么？（3）fireblocks还有Exchange's Internal Transfer，我要identify这些吧（目前似乎没碰到过）

<https://www.quora.com/What-is-the-difference-between-withdraw-and-deposit>

对银行来说，进入银行的钱叫deposit；对这笔交易的另一方（我们就假设也是一个银行），那就是withdrawal了。只不过我们这边另一方是exchange而不是银行。

Silvergate网站。<https://www.silvergatebank.com>

项目管理之道

https://book.douban.com/subject/30230063/

Dash:

显示run一个script，这个script会抓取各种数据，并写出到一个文件，符合DASH的格式要求。

似乎是有两个部分，一部分是自动化跑的，一部分是人工跑的？这个在bunker\_apps下面，在situate里面到middle office搜就有了

Dash script部分

核心就是这个DashRiskBasedHaircutTask。

里面首先有pre\_process：CopyOccTheoParamProcess和OccIntradayTheosArchiveProcess。

先看CopyOccTheoParamProcess。就是copy两个xml files，从/mnt/ndrive/ABN Files把RBHtheos和RBHparm两个文件拷贝到/mnt/bunker/dev/dash/import/下面，名字分别是theo.xml和parm.xml。

OccIntradayTheosArchiveProcess是从sftp那边得到两个文件['LPIDPL\_BD.txt', 'LPIDPL\_MM.txt']，放到/mnt/bunker/dev/dash/20201104/13/下面，最后两个数分别是date和hour。所以这个似乎不应该我自己跑，因为现在已经有了（situate已经每天跑了）。然后会把这两个已经存到local的file，再copy到/mnt/bunker/dev/dash/import下面。

然后是extractor：两个db query，分别是StocksAndFuturesQuery和ABNPositionsQuery。

先看StocksAndFuturesQuery。它很简单，就是从products table里面把future和stock类型的product和product拿出来。

ABNPositionsQuery比较长。不过还算直白，就看最后rbh\_pos那个，select出来的就是它里面的全部东西，entity, product, product\_type,expiry\_dt, strike\_price, price, sum\_qty之类的东西。

注：之前人们都是在dev里面跑的，如果我想跑上个月的，等几天dev db也就更新同步了

然后是一个transformer：PositionTransformer。

其中一部分是调整生成的file的格式。前面StocksAndFuturesQuery得到的东西变成{product:product\_type}的mapping即可，主要是辅助作用。PositionTransformer取出来的东西叫position\_rows，它要进行一系列的类型转换，然后yield出去，叫fields，看上去就是把每条trx都记录了。

然后是一个file loader。把前面的数据放到/mnt/bunker/dev/dash/20201104/13/ourpos.txt的位置。第一个post\_process CopyPositionsToDashProcess把这个file copy到/mnt/bunker/dev/dash/import下面

最后还有一个post\_process：CopyPositionsToDashProcess和TriggerDashRBHProcess。

CopyPositionsToDashProcess上面已经讲了。

TriggerDashRBHProcess先在/mnt/bunker/dev/dash/import下面创建一个空的RBHBatch.txt，然后等Dash program自动去运行，如果运行成功的话，RBHBatch.txt应该会被删了。不过事实上我看log，似乎从来就没成功过。如果dash program成功的话，我们就会把/mnt/bunker/dev/dash/print/下面的DASH\_OUTPUT\_FILES复制到/mnt/bunker/dev/dash/20201104/13下面

结论：因为situate每天都跑，所以对我来说，我自己其实都不需要跑，直接就把它跑的在/mnt/bunker/prd/dash/20201102/16里面月初的三个txt放到dash software里面就行了（更确切地说，是放到/mnt/bunker/dev/dash/IMPORT下面）。还有从/mnt/ndrive/ABN Files把RBHtheos和RBHparm两个文件拷贝到/mnt/bunker/dev/dash/import/下面，名字分别是theo.xml和parm.xml。记得用正确的日期的，方法就是看txt和xml的datemodified都是一样的即可。非theo和parm的文件，选的是每个月第一个工作日（即提供的当月第一个folder），而parm和theo，就按照文件后缀来match，是前个月最后工作日（或按date modified看，就是每个月第一个工作日），同时可以通过看产出的summary.txt里面写的日期。发邮件的时候说是月末。然后跑的是devRBH，import时间有点长是正常的。

https://wiki/display/~cheng.hu/WFH+Set+Up

hmmm怎么umount？似乎只是把/etc/fstab里面的注释掉然后再跑一遍mount -a并没有用，umount我没有权限

dell旋转外接屏幕的问题，通过在mac的setting的display里面设置即可。

关于屏幕全屏的问题，解决办法就是在monitor上直接打开windows remote machine，而不是mac打开之后再拖过去

此外还希望两个屏幕能做到，一个屏幕锁定只显示一个程序，比如windows remote desktop，其它工作在另一个屏幕上完成。

Ssh那个挺直观，一个是ssh-copy-id，还有一个就是在.ssh下面改一下config file就可以创建shortcut了

Csv pivot table还是很好用的

看起来logging的config可以写成”DEBUG”这种

前面logging不显示看来是没有把console和local\_file加到config里面，现在都对了

我应该预留一些时间，如果需要去看reconcile的整个过程（应该不会很复杂）。

我是不是需要把config放到trade support的acc里面？

比如trade bridge，在ch1prdtdb01的/opt/akuna/envs/trade\_bridge/cfg里面。那么它是怎么release的？

docker create --name=${AppName}\_${INSTANCE} ${CommOpt} -v /opt/akuna:/opt/akuna -v /var/log:/var/log -v /tmp:/tmp ${RegServer}/${REPO}/${AppName}:${VERSION} ${AppName} --config-file /opt/akuna/envs/trade\_bridge/cfg/trade\_${INSTANCE}\_config\_docker.xml

从release.sh可以看出，实际上它做的事情，就是在BASE\_DIR/envs/$PKG/versions下面安装已经在网上index里面的package，然后把这个版本放在current version里面。核心问题应该已经封装在conda create里面了。至于用的是什么conda，在里面也有说明。

似乎这个是symlink过去的，ls -l一下就看出来了

如果man一个东西没有的话，可以再试一下help，比如help set

Pycharm有个Source code navigation

<https://www.jetbrains.com/help/pycharm/navigating-through-the-source-code.html>

但看起来不是很好用。。control+click可以跳转

Set -x里面一坨，你按了没啥反应，但是之后再跑其它的就都是反应了。。。幸好这个只是临时设置的，新开一个terminal就没了。

又想到一个问题，你这个config放到外面之后，portainer那边怎么弄？看一看trade bridge是怎么处理config的

Tag dev version on feature branch is good?

Rec\_service里面的akuna\_trade和非akuna\_trade可以看看

然后几个file都用的常量是不是可以放到一个单独的constants文件里？Ans：目前已经这么放了

conda verify为什么会出错？为什么PR不报错，但我inv release就报错？等会儿，因为inv build的时候，它build的确实是只有0.0.0的东西…并没有你那个0.0.0.dev0，所以verify说找不到也是很正常的？但是仔细看

[2020-11-09T20:47:02.993Z] anaconda upload conda.dist/linux-64/bbg\_convert-0.0.0\_dev-py36\_0.tar.bz2

[2020-11-09T20:47:07.143Z] cp conda.dist/linux-64/\*.tar.bz2 dist/linux-64

[2020-11-09T20:47:08.061Z] tar (child): dist/linux-64/bbg\_convert-0.0.0.dev0-\*.tar.bz2: Cannot open: No such file or directory

说明已经弄出来了而且复制过去了啊？

有两个问题。一个是说，既然是在例如cofdevbuild04上build的，那么package最开始是怎么弄过去的？复制过去的么？另一个是说，这个path是在哪儿…

Setup.py里面的version是不是会动态改变的？

注意，ctx.run会把version tag里面的下换线变成点（.）

rc0是成功的。。。

logging也还是有问题

注：BRN是brent oil，https://zh.wikipedia.org/zh-cn/布蘭特原油。是在ICE上交易的https://www.theice.com/products/219/Brent-Crude-Futures

注：利用python进行数据分析的中文版就看这个网页吧：

https://github.com/iamseancheney/python\_for\_data\_analysis\_2nd\_chinese\_version

check breaks:

<https://wiki/display/DEV/Intraday+Transaction+Reconciliation+v1>这个是intraday transaction recon，暂时雨我无瓜

最后send email是在send\_breaks\_email.py:793。这个workflow在situate里面是同名的，每天跑三次。因为file\_name = f'breaks\_{abn\_region}\_{date.strftime("%Y%m%d")}.csv'。似乎应该是看这个。里面一大坨check\_eod\_accounts.py → check\_eod\_deleted\_trades.py → check\_eod\_expirations.py → check\_eod\_positions.py → check\_eod\_transactions.py → send\_breaks\_email.py。

然后这一次我们看的应该是check\_eod\_transactions.py。而这个script实际上运用的内容就是check\_expirations文件夹下expirations.py或queries.py或warrant.py里面的东西。

check\_eod\_transactions.py。核心是EODTransactionCheckTask，其中依然是etl。

Extraction部分有三个，分别是TRXORCTradesQuery，TRXABNTradesQuery以及TRXABNPreviousDayTransactions。

TRXORCTradesQuery的sql是查询orc\_trades table，不知道为啥这个query写的特别长，里面有一个nfx alias的应该不用管了，因为nfx都关闭了。不过这边根据price的不同，qty也是不会合并的，但email里面的file或db table abn\_transaction\_breaks就没有price这个column了？至少在后面transform的第一步clean\_data里面，price column还是保留的。Mismatch\_df里面也是有Price的。得到match\_df和mismatch\_df之后，有一个get\_mismatch\_summary，按照一些column group by了，这里面price莫非就不管了（即不同的price也算同一个group）？Ans:不是，如果price不一样，在concat的时候就不会对应同一个key而是在两行里面。所以20201127里面显示wheat不match，虽然qty是一样的，但其实price不同。

TRXABNTradesQuery是查询abn\_transactions table。

TRXABNPreviousDayTransactions也是查询abn\_transactions table，只不过对trans\_dt这个column要求的是查询日期以前的，而前面一个query是查询日期当日的。

Transformer第一步是clean\_data，第二步是将abn\_df和orc\_df做match。第一步中，把数据导入pandas的dataframe，然后做一些基本清理（比如改变列名，把qty=0的去掉，因为这个不计入sum），但要注意，这边更关键的是设置了两个column qty和trade\_value，同时把其它的原先column都设置为这个df的index（相当于二维数组里面的行号）。现在看第二步get\_match\_and\_mismatch\_df()，进去之后先把abn\_df和orc\_df做pd.concat(axis=1)，这种concat就是把两个df的列名都变成新的列名了（即原本各两列，现在新的就变成4列了），然后把concat之后里面NA的部分用0替换掉，然后加一个match column，注意到这一步结束为止，依然是和sql返回出来的没有什么大差别。第二步后，得到match\_df和mismatch\_df之后，有一个get\_mismatch\_summary()，按照一些column group by了。然后按照get\_mismatch\_summary里面group by的column，加上sum之后的column，加上diy添加的两个column，就得到了最后返回的东西。这一步里面最重要的就是dataframe的groupby函数，这边其实就是和sql里面的做法是一样的，所以核心在于prev\_abn（nooo）？

Transform并不return东西，而是里面一大堆yield。注意yield返回的总是generator object，所以实际上关键点在于load(data)这个函数怎么处理，而它就是把data=list(data)，generator变成list罢了。Transformer yield一堆大杂烩给load，甚至包括header，空行之类的（用来提示是next section）。还有就是如果你在load()函数设置断点，会很明显看到它是在等data yield过来。

注：关键还是在concat那一步，当把abn和orc的数据分别从sql拿出来后，concat就是根据给的key column进行match，我们把match上的就类似扔了？没match上的就是一个里面有而另一个里面没有的数据，放在了unmatched里面，如果恰好只是一方多出来来了，那么会看到另一方在unmatched summary里面是0，否则两方都不是0。

Loader就还是把breaks加载到db table abn\_transaction\_breaks以及保存到文件'/mnt/bunker/dev/reconciliation/US/transaction\_check/202011/TransactionsCheck\_20201127.csv'

里面。

check\_eod\_ expirations.py。首先注意date是bunker\_utils.previous\_business\_day(country\_code="US")。然后得到'/mnt/bunker/prd/reconciliation/US/expiration\_check/202011/ExpirationCheck\_20201109.csv'为file\_path，db table是"abn\_expiration\_breaks"。

check\_eod\_expirations.py整个的流程套路依然是preprocess,etl,postprocess。

Preprocess就是把abn\_expiration\_breaks表里面所有这个date(实际上是前一个工作日)和region的数据先删了。

Extractor是check\_expirations.ExpirationCompositeExtractor，在query.py里面。这个复合db extractor里面由四个sql query组成。

看一下ExpORCTransactionsQuery。它是从orc\_trades和products（这不太重要）和abn\_products和orc\_portfolios四个一起查的，当然orc\_trades其实是最重要的，其它也可以不看。

然后是SingleDayExpirationCheckTransformer。这个实际上就是ExpirationCheckTransformer，其实在expiration.py里面。这个transformer真Tm长。这个transform最开始读取data，这个data有四个部分，里面最重要的其实是abn\_data和orc\_data，也就是前面四个extractor里面后两个抓取出来的东西。还有process\_orc\_row和process\_abn\_row，这两个实际上是作为lambda函数传给self.container，因为它以lambda函数作为value类型。这个transform最后返回的就是report，是一个collections.defaultDict，从str->list。其中key有match和mismatch两种。

最后是ExpirationLoader。实质上就是BreakLoader，在expiration.py里面。拿过来的data说实话就是已经像db里面的row差不多了。Load的file\_path就是前面说的那个。除了file loader还有一个db loader，会把break\_data load进去abn\_expiration\_breaks。所以说email里面的File和db里面的这个table内容是一样的，至少就expirations break来说

A picture containing graphical user interface

Description automatically generated

这次是position breaks。是从check\_eod\_positions.py开始。核心是这个EODPositionCheckTask。最后发现的breaks是存放到PositionsCheck\_{dt}.csv和abn\_position\_breaks table里面去的。

最开始显示extract，一个是查orc\_position query (for orc positions)，是查product\_aliased\_orc\_positions\_latest这个玩意儿，这个玩意儿又是实际上从orc\_positions\_latest这个view里面查出来的（这个view和orc\_positions有点点区别，似乎是去掉了重复）。另一个是abn\_position query，是从abn\_positions里面查出来的(for abn positions)。

注：注意这里有个问题就是，abn\_product的名字都是怪怪的，可能要从abn\_products table里面找到那个名字。比如select \* from abn\_products where product ~ ‘OZC’。而在abn\_position\_breaks里面，product\_group又变成正常的Corn了。应该还有一个表存了mapping。

注2：abn\_positions表里面的数据是从abn/abn\_sod\_load/load\_abn\_postions.py里面导入的。原始文件是在prd/abn\_positions里面

Orc positions的query里面重点是做了两个left (outer) join。关于join参见book，里面讲的很清楚了。

Left and right outer join差别就在于（左右）哪一张是主表。

Full join: 在Postgre中full join = full outer join。

Outer和inner的区别在于，outer它会把两种表中，只要存在于某一张表中的数据，都取出来。而inner只是取两种表中同时有的。（那如果说是用outer join，但其中一张表里有重复的数据，最后会出现什么？）

那么在没有给出left或right的时候，是否默认left？答：inner join不用给left, right；outer join必须给left,right,full三个中的一个。Full outer join见<https://www.postgresql.org/docs/8.2/queries-table-expressions.html>。

注：我有一个困惑，为什么orc\_positions里面根本查不到BRN-CME这个product，后面它又是怎么冒出来的？答案就在coalesce以及case when？||竟然是字符串连接符，怎么设计的…coalesce是SQL 特有的函数。该函数会返回可变参数A 中左侧开始第1个不是NULL的值。终于知道BRN-CME从哪儿来的了，是从products里面查product\_gorup得到的，它的product是BZO

查完就开始transform（其实最重要的是开始的那两个表的数据是怎么进去的）。Match的过程和之前上面看过的check\_transactions很像。

在abn\_positions table里面，除了date，还有一个trans\_dt，那是啥？不太清楚。

Canola future的break证明是account break，意思是放错账号了？见3.22邮件作为一个example。

这次是account breaks。这是啥玩意，应该是看check\_eod\_accounts.py。最后的break是放在abn\_account\_breaks table里面。

它有两个task，PositionAccountCheckTask和TransactionAccountCheckTask。

先看PositionAccountCheckTask。它先查询db，AccABNPositionQuery，这个应该是获得abn\_positions，这个table里面是有account这个colume的。然后是一个PositionAccountCheckTransformer。

所有break process的最后一步都是run send\_breaks\_email.py。比如对account break而言，它extract的是abn\_account\_breaks\_view这个db table。果然是这个view不对，和abn\_account\_breaks不相符合。原因：在create view的sql代码里，最后加了一个group by，这就是剔除了重复的（但实际上有意义的）row。

注：multiple group by的意思。<https://www.postgresql.org/docs/8.2/queries-table-expressions.html>

<https://stackoverflow.com/questions/2421388/using-group-by-on-multiple-columns>

multiple group by的意思很简单，说的就是把group by后面的columns的每一个都相同的row放在同一组返回。

问题：postgresql的view是自动更新的么？即depends on的db table更新了它就会更新？Ans:是的。

<https://www.postgresql.org/docs/current/sql-refreshmaterializedview.html>

看起来这个命令是把所有的旧数据都删掉然后重新算。

<https://hashrocket.com/blog/posts/materialized-view-strategies-using-postgresql>

eager or lazy materialized view

Orc\_service

此次考虑的问题是，orc\_trades table的数据是从orc\_service来的，但不是时时更新？首先TradeTableInfo这个class是在orm.py里面，而用到这个class是在handles/ trade.py里面的TradeHandler class，这里面有一个uploader，这个uploader又是定义在base.py里面的base.TableUploader。而从base.TableUploader里面，又可以看出这个东西实际上是继承自async\_db.TableUploader，这个async\_db=akdb.aio。

Orc\_trades里面的ts是utc时间。

Db里面go to ddl可以看这个table的各种type信息

什么叫materiallized view？就是既有view的那些metadata(query)，又有实在的数据（而不是每次查询的时候再生成）。它的更新应该是通过refresh把？

occ\_market\_share\_view是从occ\_market\_volume等一些创建而来。

Occ\_market\_volumn里面有volumn，不过没有Open interest。两者应该还是有区别的吧

https://www.investopedia.com/ask/answers/050615/what-difference-between-open-interest-and-volume.asp

check\_orc\_base\_instruments\_traders\_auto.py：

核心task是CheckClosingPricesTask。

Extractor又是有四个。positions\_data, instruments\_data, desk\_email\_groups, base\_offsets。以instruments\_data为例，它查询的是orc\_instruments table，而这个table又是orc\_service load进来的。

然后是transformer。把四个data拿过来之后，全都转成dataframe。

然后是一个EmailLoader。

Bunker\_apps的meta.yaml里面要加bbg\_convert

新的mac上面保护键要设置一下，不然每次cmd q都是退出整个remote desktop。

之前的test写得不好…不应该test过程吧

为什么git一部分sync了，一部分没有？似乎不是这个问题。而是说，一个在我这个branch根本不存在的文件，却出现在inv test测试的某个文件的全路径搜索里面。

能否选取两个db中的两个table？似乎不需要。让我之后去看一下怎么sync两个table。

Bbg\_symbol和product不是一对一的关系。所以会有一些closing prices没有得到，因为cache吞掉了。原来是之前的convert逻辑出错了

服了。。因为grep没有用\*号。不是，是因为对.gz的file没有用zgrep。。。我服了

Bunker\_apps.egg-info这个咋回事

一直报test\_apps出错，后来我把souces.txt和entry\_points.txt里面的eod\_mark的东西都删了就好了。对git也没影响，因为egg-info里面的东西应该是不上传的。那么egg-info里面的东西最开始从哪儿来的？

原来在async def里面可以定义async def，有趣

把future transfer account原来是ABN发个邮件

绝了，看来exchange的网页是个学习的好地方。<https://www.cmegroup.com/trading/agricultural/livestock/cme-pork-cutout-futures-and-options.html?gclsrc=aw.ds&&gclid=CjwKCAiA17P9BRB2EiwAMvwNyI3V4JnKd9zwtORIcGSRAoa7LGZfPfNFhUjEqF_4ln018c3_xj8_-RoCafsQAvD_BwE&gclsrc=aw.ds>

在这边<https://www.cmegroup.com/markets/agriculture.html>

对期货产品进行学习，在cme这种地方很合适

https://www.cmegroup.com/education/courses/introduction-to-agriculture/learn-about-agricultural-markets.html

有些地方，要做code refactor的时候，因为不知道究竟有没有问题，所以宁愿往上加东西（哪怕是不合理），也不敢轻易改，这就是问题。Refactor那本书里记得有说过怎么加test先测试自己改了不会有影响，然后再改。

果然改了之后test就过不了，所以应该写test，看看mike怎么写的。

Typing里面，要去掉optional，可以用assert，<https://github.com/python/typing/issues/645>

有的test是测试过程，有的是测试实际behavior。

还有一个singleton可以看一下

一件件来。先把这个快捷键的问题搞定。现在看起来mod键都不管用，连2.2换排列布置也不行了。前面3节主要是介绍，讲配置从第4节开始。Mike之前应该帮我配置过文件，在~/.config/i3/config位置。4.3键盘绑定是我最关心的。为什么明明keyboard shortcuts里面也设置了退出键，依然是cmd+q就退出？<https://discussions.apple.com/thread/7634486>这个不行啊，虽然cmd+q是不退出了，但在i3里面也没用啊。对键盘来说还有key symbol和key code两个东西，如果经常换不同地区的键盘的话，应该用key code比较好。还可以man xmodmap。实际上还可以设置鼠标（4.4）。又比如4.6的这个设置，按道理说我可以通过按mod+鼠标来拖动窗口（比如pycharm里面的弹窗），但实际上不行。。。4.15是设置变量。第5部分是设置i3bar，虽然config里面也有，但我觉得应该没怎么用。第6部分是说命令，即我们是把命令bind到keypress的。分号是用来chain cmds，即从左至右一个个执行。

Ans：太傻了，最后发现的结果是，我没有把mac里面的keyboard shortcuts的名字的大小写设置对。。

基本键盘配置应该都在/etc/i3/config这边。<https://i3wm.org/docs/userguide.html>

前面很奇怪的反应似乎是类似大写键被按住了。但实际上又没有，我再按了一次shift就好了。

Scp还可以scp到别人的电脑上？

Pycharm可以checkout到branch，也可以checkout到tag。https://www.jetbrains.com/help/pycharm/use-tags-to-mark-specific-commits.html#jump-to-tag

Logging还是什么level都显示。Ans：下面这两种方式是等价的

# logging\_config = copy.deepcopy(bunker\_utils.DEFAULT\_CONFIG)  
 # logging\_config["root"]["level"] = logging.INFO  
 # bunker\_utils.configure\_logging(environment, APP\_NAME, logging\_config)  
 bunker\_utils.configure\_logging(environment, APP\_NAME)  
 logging.getLogger().setLevel("INFO")

核心是root的level被设置为了什么，即logging.getLogger()返回的是root logger。在bunker\_utils.DEFAULT\_CONFIG里面，root的level是被设置为DEBUG的，所以会有之前的种种问题。

再看一遍logging的package

<https://docs.python.org/3/library/logging.html#logging.basicConfig>

最后我还是用之前版本的办法，把default config复制过来，修改level，再传给config\_logging函数https://bitbucket/projects/BNK/repos/bbg\_service/browse/bbg\_service/\_\_main\_\_.py?at=refs%2Ftags%2Fbbg\_service-0.2.9#38

https://www.cnblogs.com/hf8051/p/11727520.html

logging说可以配置来track where msg come from

说logging主要有四个玩意儿

Text

Description automatically generated

Logger是树状结构的？

要设置logging到什么位置，以及消息格式，是在basicConfig()里面设置。

从logging flow里面可以看出，它是不停地call自己的parent logger来处理消息的。

Logger有三个任务。

Logging.getLogger(name)返回一个logger instance，如果没有specify name的话，返回的就是root logger；如果specify name了，只要name一样，返回的总是同一个logger。

某个logger如果没有设置level的话，就会按照parent的来执行，直到递归到root，root的level是warning（所以这么看的话，我的level应该是某个import的module决定了，不过import module里面的logger和我的logger是在同一个树里面么？）。正因为有这样的传递机制，所以没必要每层logger都设置一下，只需要设置top-level logger（当然你也可以认为把传导机制关了）。

Handlers负责把logging msg分发到合适的地方，每个logger可以有多个handlers。Handlers也有setlevel，不过它的setlevel是决定它处理哪个level的消息。

用户一般有三种（其中两种）config的用法。一种是直接人工用api config logger, handler之类的；另一种是用config file or config dict。

如果没有做configuration，那么就要按照python版本来看啥情况，3.2之后是默认一个输出到sys.stderr的warning level的handler，所以warning以下的信息是看不到的（不输出的）。

最后还说了以下library应该怎么设置logging。认为应该把handlers的配置留给library user来设置。

问题：logging.info()这种用法，用的就是root logger？Ans:对的，lib里面\_\_init\_\_的2038行注释讲了

问题：module里面logger的hierarchy是怎么确定的？Ans:The logger name hierarchy is analogous to the Python package hierarchy, and identical to it if you organize your loggers on a per-module basis using the recommended construction logging.getLogger(\_\_name\_\_). That’s because in a module, \_\_name\_\_ is the module’s name in the Python package namespace。试了一下确实是这样，如果是在module里面，print(\_\_name\_\_)显示的就是由点号分隔的、显示hierarchy的名字。

原来java也有logging，看起来接口api还差不多。

Openfigi可以查这些符号。

Pytest的问题。首先的问题是，我们需要什么样的测试？其次的问题是，我们怎么去达成（做到）这些测试？

Pytest官方文档。<https://docs.pytest.org/en/stable/contents.html#toc>。

它先讲了assert，基本上沿袭了python的assert，常用的还有怎么assert一个exception是否发生。

然后讲test fixture，即测试的基本构件。但是为什么我一定要用@pytest.fixture?直接用正常的python function不好么？Ans:原因是dependency injection，把test\_function（被测试函数）需要的东西给它，而不是让它自己创建。多个地方要用同一个fixture的话，创建一个conftest.py，pytest内部写好了回去搜索这个file来找需要的fixture，有一个搜寻顺序在那边。Fixture里面还可以用fixture，也有一个顺序。Fixture还有参数化。

然后是monkeypatching和mocking，猴子补丁。教程里面只介绍了monkeypatching而没讲mocking，虽然两者干的事情应该是一样的。哦因为Mock（至少我们用的）是unittest.mock.Mock，pytest并没有Mock class。Unittset是python built in lib。https://docs.python.org/3/library/unittest.html。还有sentinel也是。

Pytest还提供tmp\_directory这样的自带临时文件夹。

然后讲了stdout/stderr是怎么handle

Unittest <https://docs.python.org/3/library/unittest.html>。

Fixture的功能应该是setup/cleanup for tests。

TestCase是一个individual unit of testing

Mock。After performing an action, you can make assertions about which methods / attributes were used and arguments they were called with。Mock还提供了sentinel, MagicMock, patch等东西。

写test的问题是，路都不会走就得开始跑，本来这种又有异步又和各种外部东西打交道的test就不好写。

之前那个问题什么加了return value也没用，部分是因为说，我该的是新的mock\_db\_reader，而不见得是之前传给load\_bbg\_closing\_prices初始化用的那个mock\_db\_reader

似乎终极原因是fixture function的问题？我把本来function返回的东西复制黏贴直接传进去，就对了？见db\_reader

比如直接import bunker\_apps.bbg.handler.transform as transform不行（因为没有在\_\_init\_\_里面写），那就可以试试from bunker\_apps.bbg.load\_closing\_prices import LoadBBGClosingPrices

看起来一定要先在conda dev repo里面有了之后才能docker build。。。那还是有点慢（当然你更新dockerfile是不用管的，因为它不是在dev repo中的）。

为什么本地跑的时候没有问题，在docker上就会有no module found: from bbg.infra import tyeps?我确实在infra那个文件夹下面没加\_\_init\_\_，难道本地动态build的package和conda build那个不一样？

python -m pip install –editable .是实时更新的么？还是我更新时候需要重新跑一下这个命令？好像是的，跑完之后conda/envs/bunker\_apps/bin里面的东西就更新了。

但为什么每次我改完code之后，import bbg\_service，它里面的东西却是最新的？

分析思路：我每次import 的package是在conda环境下的，即/sandbox/conda/envs/bunker\_apps里面（例如）。最开始创建conda环境的命令是conda create。可以指定储存环境的路径，默认的path是你安装conda的位置下面（而prd里面，你可以看到它就不是在conda下面，那是因为deploy.sh里面create conda env时故意指定了别的路径）。但deploy.sh里面没有activate环境，这是什么时候做的？因为不activate的话，situate workflow还是不能跑的，而且一个box上面有好几个package，不可能每次都人工activate env吧。补充：在dev situate里面有deploy bunker，.sh和prd差不多。

Conda create之后是inv develop和inv bootstrap，这两个都是安装package。Bootstrap安装的是meta.yaml里面的东西，没有这个包本身。第二个是用pip install包本尊，conda里面也有讲。<https://docs.conda.io/projects/conda/en/latest/user-guide/tasks/manage-environments.html#using-pip-in-an-environment>

为什么我手工装的conda env里面，在/sandbox/conda/envs/bunker\_apps/bin下面就没有conda？Ans:因为没有安装conda-build包。https://docs.conda.io/projects/conda-build/en/latest/install-conda-build.html

注：可以看一下别人解释conda activate是怎么做的。https://stackoverflow.com/questions/48585040/conda-what-happens-when-you-activate-an-environment

还是merge的问题，正确的做法是不是先merge到feature branch然后再push？我觉得应该是的（如果不能自动merge的话）。

Bbg symbol的话题：

https://www.cmegroup.com/month-codes.html

小救星https://www.cmegroup.com/education/courses/introduction-to-futures/understanding-contract-trading-codes.html

Git为什么checkout到master后，会把另一个branch的修改都带过去？

Dataclass怎么设置defualt value？为什么原本跑就没任何问题…看来原本用的不知道是哪家的DBInstrument

当我大部分数据都只有一个，小部分数据有多个的时候怎么处理？用list如果我需要遍历的话

Bbg\_service别忘了之后acc里面每次也要pr

有空找一下project应该怎么structure。

Prometheus:

<https://prometheus.io/docs/introduction/first_steps/>

metric name就是一个time series的名字。还有label。

有几种类型的Metrics。只能单调递增的counter，可增可减的guage，histogram，summary。

一个endpoint就是一个instance？一群endpoint就叫job？

根据监测需求，可以把services分为以下三类<https://prometheus.io/docs/practices/instrumentation/>，对于每一类service应该用什么样的检测metrics，里面都有介绍。

这个是python client library的介绍：<https://github.com/prometheus/client_python>

Prometheus是怎么获取它要检测的数据的？这句话是核心：Prometheus collects metrics from targets by scraping metrics HTTP endpoints。所以是用http获取target发来（pull or push）的数据？

Prometheus跑的时候有一个config file。等会儿，这个是prometheus server的config file把？不是client side的。

我写的应该是对的，打开的那个网页里面的数据并不是增加新条目，而是直接修改原本数据，所以可以看bbgservice\_solace\_recieve\_message有了变化。

如果result info传过来的数目和self.cache里面的数目不对那就有问题了。但现在我们是把空的数据也会传回来的。

还有就是为什么await delete的时候，似乎它又会回到await on\_result\_row的状态？为啥我没有在bbg这边unsubscribe user request的topic？你不subscribe它当然会持续监听，而send过去的信息并不是说有人接收了它就结束了，而是依旧会在那儿。之前因为说之后哪怕改了self.results，但后面也不再用了，因为update\_results是把result作为变量传过去的。所以应该一get完result就unsubscribe。

Node exporter是监测啥来着？

<https://prometheus.io/docs/guides/node-exporter/>

kafka是用来replace solace的？不是，是内部一些通信在用的

这个是hubble的home page

<https://wiki/display/HUB/Solace+Capture+Utility>

<https://github.com/confluentinc/confluent-kafka-python>

opendirectoryd代替/etc/passwd？

<https://superuser.com/questions/191330/users-dont-appear-in-etc-passwd-on-mac-os-x>

最后是用getent passed | grep danny.hu在linux上看到了启动的shell

https://www.geeksforgeeks.org/getent-command-in-linux-with-examples/

在terminal里面输一个exit，就会得到

(base) hushengchaodeiMac:consul hushengchao$ exit

logout

Saving session...

...copying shared history...

...saving history...truncating history files...

...completed.

注意到，当我没有exit的时候，terminal上方栏目显示的是-bash，其实表示它在运行一个叫bash的程序，比如输入python3的话，上方栏目显示的是python3。

很有趣的一点，对于external cmd，shell实际上都是create一个sub shell来执行它。比如ps，比如mkdir，任何在例如/bin目录里的都是。

持久化设置PATH变量，是要在这两个start up file里面的任意一个改一下。

https://stackoverflow.com/a/14638025

Python sort的key原来是take one argument，但可以返回多个值（用来进行多pass的比较）

Crypto\_service是用来做crypto交易的，现在是否还用？如果这样的话，基本可以从它窥探一下trading engine的模样。Orc service也是，不过是古老破旧版本。

Crypto\_service里面还读取一个api key文件，大概是为了安全考虑。

然后就是Context class里面，对各个exchange创建对应的service（不过应该是每个instance跑一个吧？对，见create\_service() method）。

我们以bitmex为例。它是一个交易平台<https://www.bitmex.com/?lang=zh-CN>。

BitmexService这个class先是初始化，用的是有账号密码的key\_config，连接到它的官网，连接用的class是BitmexRESTConnector，这是neosocket包里面的东西。它的parent class就叫RESTConnector。

注：因为用的是交易平台的rest api <https://restfulapi.net>?

一个简单的rest介绍<https://www.ruanyifeng.com/blog/2011/09/restful.html>

有一本书叫 rest实战

不过后面还连了一个websockethandler和bitmexwsconnector，为啥要两个connector？

再看看deribit。Deribit用的是neosocket里面的WSConnector。On\_msg是callback函数，处理从web socket stream里面获得的msg。然后service里面还有各种handler，比如positionhandler。然后这些handler里面都会被pass一个db\_client，这个handler里面会有insert rows到db table的步骤，insert各种东西，position也有，account也有，等等。而我们选择load进去的只是transfer，而没有真实的trade（这一点我们可以通过TradeHandler看到，trades是在deribit\_trades table里面的，而且好像不是实时更新）。

Deribit里面似乎也是采用某种类似solace的topic的获取消息模式，websocket是open source的包裹。

<https://docs.deribit.com/?python#public-get_instruments>

可以看出instrument handler做的事情是Retrieves available trading instruments. This method can be used to see which instruments are available for trading, or which instruments have recently expired.是market data的一部分，而不是下单。

<https://docs.deribit.com/?python#private-get_position>

而position handler做的事情也是获得一个private market data，即用户在某个instrument上面的position。这个数据会被放到deribit\_positions里面。

<https://docs.deribit.com/?python#private-get_account_summary>

account handler里面，get\_accounts也是从deribit那边获得account summary data，然后是每隔一段时间就load一条到deribit\_accounts table里面。至于db table里面的数据分别对应的是deribit传过来的哪一条，见accounts.py里面create\_account\_row函数，它会做一个转换。

<https://docs.deribit.com/?python#private-subscribe>

Trade handler里面有subscribe，这看起来和solace很像。

<https://docs.deribit.com/?python#private-buy>

这个link就是buy的topic。在neosocket里面它也确实implement了，但我没看到具体哪里用了。

需要注意的是，accounts.py里面的account handler里面是load deribit\_transfer table的地方。而具体的是在get\_transfers\_per\_currency()里面，更具体是self.ws\_client.desposites()或self.ws\_client.withdrawl()里面。

<https://docs.deribit.com/?python#private-get_deposits>

然后我们再看一个fireblocks把。

它比较简单，只有一个instrument handler。里面的request是通过rest\_connector来发出的，是直接向fireblocks发送http request。

再看看gdax，现在已经是coinbase pro了。

像gdax没有提供api的话，就是直接用rest的GET之类的东西获得数据，见neosocket/gdax/rest.py里面如account\_history，从这里面获得gdax\_transactions。至于这个http获得的数据怎么转成db table数据，见account.py里面的create\_account\_row()，这是update到gdax\_accounts table的。

再看一下gemini。

Transfer是在transfer\_handler里面处理的，db table是gemini\_transfers。从gemini获取的方式是直接网站上获取(rest api)。至于加载到db gemini\_transfers table里面的row长什么样，见transfer.py里面的create\_transfer\_row()函数。

看一下sync\_db.py。

既然crypto journal关心的是transfer（当然实质上应该也是transaction），那么也得明白一下在crypto coin的意义下，transfer是什么。

Orc\_service:

进入点是在cli.py。orc\_client就是和orc相连的玩意儿。整个orc\_service和db，solace和orc相联系。然后设置为基本配置之后就是service.start了。Start也不是重点，主要是OrcService里面初始化的各种handler开始处于callback状态了。有trade，instrument和两个portfolio的handler，我不知道它们在干啥…每一个handler创建完后，都是让orc\_client来register callback，大概是传递过去？ioloop是tornado的东西，然后把flush这些handler拿到的数据到db里面是靠加入periodicCallback来做，每隔一秒做一次。Upload到db实质上call的是handler的flush\_db方法，而这个方法是base里面的抽象方法，实质上在每个handler里面个性化的是uploader（包含了不同handler如instrument, portfolio, trade）要upload到的目的地db table信息。

然后每个handler都老长老长的，但实质上传给orc\_client的callback只是里面的一个method，比如对instrument来说是on\_instrument\_feed。

注：orc总不会用的也是solace了吧？从asyncOrcClient里面来看，用的是tcp连接，连了之后获得一个self.stream，用这个stream来write。信息传输的数据类型是用class表示的。

然后handler里面还有down\_instrument方法，这个是在service的prepare step被call的，这是干啥用的？也是从orc获取某个东西。

此次考虑的问题是，orc\_trades table的数据是从orc\_service来的，但不是时时更新？首先TradeTableInfo这个class是在orm.py里面，而用到这个class是在handles/ trade.py里面的TradeHandler class，这里面有一个uploader，这个uploader又是定义在base.py里面的base.TableUploader。而从base.TableUploader里面，又可以看出这个东西实际上是继承自async\_db.TableUploader，这个async\_db=akdb.aio。

Orc\_trades里面的ts是utc时间。

<https://www.nasdaq.com/solutions/nasdaq-futures-nfx>。看起来nasdaq future exchange已经没了？

TODO:安装homebrew。然后安装sublime。

Conda env删了之后却还在，可能是因为当初安装的位置就不对。我安装在了~/.conda/envs下面。

https://stackoverflow.com/questions/40180652/condavalueerror-value-error-prefix-already-exists/40435573

wedbush:

situate: middleoffice/wedbush

有趣，遇到这个问题

<https://stackoverflow.com/questions/25609153/paramiko-error-reading-ssh-protocol-banner>

现在wedbush reconn了position，我要写一个reconn trades的。

问题：我reconn trades和reconn position有啥区别？

答：我依然没看出来，不过可以从orc\_trades里面，按照WEDBUSH\_PRODUCTS = ("BAKTBTM", "BTC", "ETH"), WEDBUSH\_PORTFOLIOS = ("ABN\_X", "ice\_btc")等条件去搜，会看到这个orc\_trades里面的数据应该是和我从prltrades3\_\* file里面读到的是一致的，这样就match上trades了。

pip不存在就试试pip3，没有权限就试试pip3 –user。

Ice and akuna volumes是在send\_ice\_volumes.py里面。Situate名称是load\_and\_send\_ice\_volumns。另一个相关script是load\_ice\_volumns.py。

先看一下load\_ice\_volumns.py。核心是ICEVolumnTask。相关的db table是ice\_volumns。

pdb跳出loop可以用until

https://stackoverflow.com/questions/6731119/python-pdb-skip-loop

找ip地址

<https://opensource.com/article/18/5/how-find-ip-address-linux>

似乎第一个是正确的

Position Bridge:

和portfolio bridge似乎是兄弟。

i think you might need to -u 1000:1000 when you docker create to get the logs to be owned by the akuna userhttps://zh.wikipedia.org/wiki/日志轮替

umount的话要用sudo umount dir。

我们的机器用的文件系统的协议似乎是CIFS？<https://cifs.com>。因为在mount里面我们的cmd有cifs

<https://linuxize.com/post/how-to-mount-cifs-windows-share-on-linux/>

<https://linux.die.net/man/8/mount> 搜索cifs就看到了，在-t type里面

https://www.dell.com/community/入门级和中端/分享-CIFS和NFS的区别/td-p/6934849

因为是distributed file system，所以我可以理解为什么account locked了，因为读取文件过程中到处都是authentication。

https://linux.die.net/man/8/mount.cifs

看起来global和nonlocal的区别在于，一个是用在function套function，一个是用在module套function。都是为了告诉一个function去哪儿refer这个变量（module或外层function，因为会有x+=1的用法）

不过我看起来觉得不应该用nonlocal和global，因为它实际上破坏了代码的可读性，人为强行改变了scope的自然状态（就像用goto）。

设置port forwarding

https://www.hellotech.com/guide/for/how-to-port-forward

<https://www.whatismyip.com>

hdal client是什么下次可以仔细看看，我猜是包装了一个solace client在里面？是在hdal package里面

load\_ice\_volumes.py

数据似乎是放在/data下面。

核心是ICEVolumeTask。它先是从ice\_products表里面取出查询日的product。

在extractor里面似乎是从多个地方（比如db，h5）拿数据，比如/data/derived/chrysalis/\*/ice\_md\*/{:%Y%m%d}.h5。

Transform挺简单，就是把数据变成db table row的形式，load到ice\_volumes表。

Brew可以安装的好用工具：

<https://orianna-zzo.github.io/AllinOne-html/blog/2018-01/mac%E5%B0%8F%E7%99%BD%E6%8E%A2%E7%B4%A2%E8%AE%B01-%E5%9F%BA%E6%9C%AC%E8%AE%BE%E7%BD%AE%E4%B8%8E%E5%B8%B8%E7%94%A8%E8%BD%AF%E4%BB%B6/>

搜狗好像还是不能安装。

brew install –cask

https://brew.sh/

Svcutil包裹解析。

以build\_signal\_future为例，它用的是python的AsyncExitStack和signal library，以及asyncio.AbstractEventLoop。关键是看看asyncio.AbstractEventLoop.create\_future，以及add\_signal\_handler等一系列方法。这里的signal就是常规的signal.SIGINT, signal.SIGTERM之类的东西。

处理build\_asyncio\_error\_future也是类似的，用的是loop.set\_exception\_handler，这是asyncio封装好的。

不过没看懂，为什么在wait\_for\_interruption里面又重新call了一次build\_signal\_future和build\_asyncio\_error\_future

Metrics需要

Grafana

Consul

Docker

https://github.com/prometheus/prometheus/wiki/Default-port-allocations

为什么需要consul的核心原因是因为prometheus是要export http port的，于是就要管理这个寻址的事情。

是怎么在windows机器上面跑Linux代码？

Cookiecutter是一个package。

<https://github.com/cookiecutter/cookiecutter>

所以用起来是要先自己安装一下conda install cookiecutter，然后它会自动替换生成package框架。

有个困惑：我是不是能从log里面看出前后commits的继承关系？好像并不能？

如何解决doc的git冲突问题，对我的code也有借鉴意义，如果我确定应该是用哪儿的话，不需要手动再删除conflicts了，直接可以用ours/theirs来做。

<https://stackoverflow.com/questions/278081/resolving-a-git-conflict-with-binary-files>

boto3是aws sdk的python package

<https://boto3.amazonaws.com/v1/documentation/api/latest/index.html>

linux怎么scp里复制hidden file(即.xxx文件)。最简单的就是在folder后面加/

https://serverfault.com/questions/21580/how-to-make-scp-copy-hidden-files

pre-hook里面set -x的意思

http://linuxcommand.org/lc3\_man\_pages/seth.html

pre-hook里面-d的意思是测试一下是否存在这个file/folder，如果存在的话才run test或doc

<https://unix.stackexchange.com/questions/90112/usage-of-d-in-a-if-condition>

<https://tldp.org/LDP/abs/html/fto.html>

原本docker image是在docker\_image = f"{DOCKER\_REGISTRY}/{repo}/{PACKAGE}:{version}"，现在"{DOCKER\_REGISTRY}/{repo}”都不需要了？

看起来似乎是我们现在没有对prd repo的access了，可以看到tasks里面的各种都是.dev而没有prd

build\_docker就是在本地build一个image而已，用docker image ls就可以看到了，也可以用tasks里面的命令清理。

很好用，可以把commit bring back add后的状态

<https://stackoverflow.com/questions/7214039/how-do-you-move-a-commit-to-the-staging-area-in-git>

可以直接忽略掉一群type check警告

https://stackoverflow.com/questions/53607737/flake8-to-ignore-warning-f405-on-multiple-lines

rebase和squash

https://thoughtbot.com/blog/git-interactive-rebase-squash-amend-rewriting-history

不如参考megan说的

git rebase -i HEAD~N

git push --force

为什么package已经在某个环境的conda list里了，而我用python3打开一个session之后依然不能import？

似乎和用的python确实有关系。

https://github.com/conda/conda/issues/8716

这也是太奇怪了，在pycharm里面，which -a python返回的就是只有两个环境。然后我把原先terminal关掉再重启也对了？

只修改部分文件的git情况

<https://stackoverflow.com/questions/6119036/how-to-revert-a-folder-to-a-particular-commit-by-creating-a-patch>

先git reset …..再git checkout…

似乎除了你给出一个.hooks/pre-push文件外，在.git/.hooks里面也会自动生成一个，但似乎并没有总是同步？比如我修改了了inv check –no-typing，但.git里面的却没改。

这看起来bunker\_web里面的inv format有毒啊？它是有remove-all-unused-imports flag，不过好像是乱弄的。

明明autoflake里面也有--remove-unused-variables，为啥没起作用？是不是因为check用的package和autoflake并没有就什么是unused variable达成共识？

Pycharm倒是设置了是否要标出unused variable，但是却没有写怎么删除？是故意没有还是什么？可能是因为删除比较难。

<https://stackoverflow.com/questions/36724041/using-pycharm-ide-for-python-unused-variables>

有一个针对java的ide的答案，不过我似乎也可以用用？

https://www.jetbrains.com/help/pycharm/running-inspections.html

<https://stackoverflow.com/questions/30669093/remove-unused-variable-declarations-automatically-android-studio-inteiij>

虽然我没发现ide可以自动删除，但至少它会帮我锁定在哪儿了，不用我自己翻页。

还有Code/Reformat File这个这么强的

Windows task manager任务管理器可以让人安排计算机自动执行任务，不知道linux有没有？Ans:有的，linux命令行那本书的16.6。

https://www.windowscentral.com/how-create-automated-task-using-task-scheduler-windows-10

突发idea：像crypto journal这种，一个好的code design应该是可以让每一步可以独立分开而不互相影响。即我修改crypto transfer table应该不会被删把。

对crypto journal来说，涉及银行的账目我是不能作假，因为finance直接看得到。其它的话，如果数额不大，没match上的直接删掉也可？毕竟会计允许一个误差在那边。

那些说的deltaone2之类的是portfolio的名字，是akuna自己的内部分类

今天试一下wget

wget --tries=1 --http-user=dmorley4 --http-password=Akuna@33\!3?$ --auth-no-challenge -O clearing\_edit\_deltas.csv "https://www.batstrading.com/cone/account/clearing\_edit\_deltas/download/?executingFirmId=ZOML&dt=2021-01-14"

有个问题是password parsing不对。var1=Akuna@33\!3?$，所以得用一个反斜杠

Akuna@33!3?$

是否可以直接download到他的machine上？

Linux bash太坑爹了，对数值和字符串还要用不同的比较符号（-lt是给数值用的，<是给字符串用的），你要是用错了还不报错。以及date还不能直接比较，要变成string比较。

<https://unix.stackexchange.com/questions/84381/how-to-compare-two-dates-in-a-shell>

为什么ls -l并不显示文件夹内所有文件的大小？Ans:因为对directory要用du命令。

https://stackoverflow.com/questions/679567/unable-to-get-the-size-of-a-directory-by-ls

ssh包括scp的原理是什么？我觉得依然是网络传输。像wget，我觉得背后其实也是用c写的socket编程。Ans:对的，这些只是高层的protocol。

怎么把package里面的config file用作文件参数？可以这样做么？

https://wiki/display/DEVOPS/Migrate+Python+Repo+to+Artifactory

Conda搜索哪些channel是自己在~/.condarc里面设置的

<http://docs.pyinvoke.org/en/stable/concepts/invoking-tasks.html#per-task-help-printing-available-flags>

inv --help deploy-conda就可以，inv deploy-conda –help就不行

inv release里面git push就会trigger Jenkins task，然后在jenkins里面它会handle version的问题吧？所以在新的CI里面，只要我inv release一个version，它就会自动帮我跑所有的流程，并根据是不是dev version来做相应的变化。YES：在Jenkindsfile里面有一个reg expression来判断是不是prd的version number来决定是不是要运行prd的步骤。

新docker在Included Repositories里面，有点难找。

<https://artifactory.us.akuna.dev/ui/repos/tree/General/docker>

接下来要做的事情是，看看是不是要update Docker image的位置？ANS: artifactory.us.akuna.dev/docker/akuna/python/bbg\_service:1.0.4，找位置的方法是直接在那个箭头下面点出来，会看到repo的位置（而不是url的链接）。有意思，所以这个不是url，那代表的是什么呢？网络是如何找到/路由到这个registry呢？

git每次merge到master之后再checkout到之前的branch，需要再git merge master一次？

原来还有这个market identifier的ISO标准，这也算统一度量衡的一部分了？

https://www.iso20022.org/market-identifier-codes

Photon

我还是没看懂它究竟是在哪里真正运行那个code或者发送请求的。是在photon的solace.py里面？（搜索bloomberg会看到）

bbg convert

模式识别的问题还是reg的问题？

现在这个算法问题是说，我知道一些parse的规律，要从string里面把需要的元素parse出来。这不就是compiler里面第一部分parser做的事情嘛？

TODO: 在dev里面研究一下bbg的docker的log到底是怎么搞的，老是missing.

Decorator。

<https://en.wikipedia.org/wiki/Exponential_backoff>

就是重试的等待时间按指数增长，和申请cache按指数要是一样的。

问题，怎么把self放到decorator的参数里面？答案是不可能，因为一个是run time发送的事情<https://stackoverflow.com/questions/1231950/how-can-i-use-a-class-instance-variable-as-an-argument-for-a-method-decorator-in>

答案似乎应该是这个<https://stackoverflow.com/questions/7590682/access-self-from-decorator>，因为有一个functools.wraps，所以在那里面的是可以access self？成功了。

为啥子我用键盘发送了term之后它还是继续运行，好可怕…

还有我这么修改，那个sleep就变成blocking sleep了，会有问题么？为什么它现在就download不下来了？答：不是的，是因为你没有return bool….

<https://www.saltycrane.com/blog/2009/11/trying-out-retry-decorator-python/> 这个也可以用映射来写，最后还是拥有和原来的f一样的参数列表，只是已经做了一些binding和变化。

<https://www.geeksforgeeks.org/creating-decorator-inside-a-class-in-python/> Cong xu建议这么写。而对这个来说，hmmm，我还是没太搞懂，func(self)是啥玩意儿。这不会是class和func decorator两者的结合吧？答：不奇怪啊，method(self,)本来就是这样的，只不过call的时候，我们一般用instance.method()来写，就不用把self放进去了，一种design选择而已。

Python的decoration，实际上就只是在function和class定义时运行一些额外的语法而已。

A picture containing text

Description automatically generated

即把被装饰函数传给decorator，然后把decorator跑完的结果再assign给原来的被装饰函数名。如果是在class里面，它就是这样，这也说明了“装饰”这一步是在何时运行的：看起来是instance创建时候。

只要是return一个callable的函数都可以当函数装饰器？准确地说，装饰器是一个callable，然后它返回一个callable。

比如函数装饰器的目的就是add a layer of logic

@staticmethod and @classmethod。

当我们想要处理一些与class而非instance有关的数据时，我们就该考虑使用特殊的方法了。

一般来说，正常的、普通的instance method里面第一个argument是instance object（self）;@staticmethod则不传入self;@classmethod则传入cls（这个用的很少）。而有时我们需要处理的data是和class而非instance相关，当然我们也可以选择把这个处理函数直接放在class外面，但放在class里面是更为合理的设计想法。而放在class里面，用的就是staticmethod，它**旨在操作类属性而非instance（self）属性**。

注意，private method(在函数名称前加\_\_或\_)和static method完全是两码事。<https://stackoverflow.com/a/1301409>. 这个讲的清楚了，单下划线，双下划线，以及前后双下划线的区别（表示这是python system用的naming）。

如果你只是准备用somecls.staticmethod()来call，可以不用加@staticmethod，同时也不用在函数签名中有self；但如果准备用inst.staticmethod()来call，就必须要加@staticmethod（此时函数签名中依然没有self参数，因为如果有self的话就是普通的method了）。这就是为什么你可以有@staticmethod，但也可以没有@staticmethod这个装饰器。

@staicmethod其实就相当于在类里面，用somemethod=staticmethod(somemethod)重新绑定这个somemethod

现在来解释classmethod。<https://stackoverflow.com/questions/33128325/how-to-set-class-attribute-with-await-in-init>. 这是一个骚操作，来设置class中有coroutine的初始化变量。

看起来classmethod做的事情就是在于明确传入cls变量，这样你就不需要hardcode class名字。同时这个cls指的是最上层（最近，最specific）的那个class。下面这个就是说class method和staticmethod的最主要区别：

Graphical user interface, text, application

Description automatically generated

descriptor binding是啥玩意？（在书里32和38章有讨论descriptor）

所以现在的核心问题是，decorator的rebind是发生在什么时候。

Decorator也可以take argument，以function decorator为例，一般做法就是用nest function，在最外层那个就负责take argument（来保存state info），里面才是真正的decorator，它又返回一个callable。这里面每一层都可能是function或者class。

Timeline

Description automatically generated with low confidence

Text

Description automatically generated

这个例子不知道为啥，就是TypeError: 'staticmethod' object is not callable。

改成@ha.\_\_func\_\_(3)就work了。因为ha只返回一个不能call的函数名字而已，见下。

<https://stackoverflow.com/questions/6412146/python-decorator-as-a-staticmethod>

说python的decorator，实质上就只是在函数和类定义的时候做一些额外的处理步骤



只是在外面再包一层逻辑而已。

<https://stackoverflow.com/questions/37835179/how-can-i-specify-the-function-type-in-my-type-hints>

typing里面如果一个argument的类型是函数的话，怎么写

像IndexError这种，本来就是type名（是一个class），所以要用Type[Exception]来说明类型才行，直接用Exception是不行的。

<https://docs.python.org/3/library/typing.html#typing.Type>

Class A and class B

<https://www.investopedia.com/ask/answers/062215/what-difference-between-class-shares-and-other-common-shares-companys-stock.asp>

一般和投票权有关。

Mike的项目update/refactor的方式，就是从头开始一步步参与，对每一步，可以思考它按照已有的项目经验应该怎么做，明确这个工作应该归于哪个组是当下最为妥帖的。我觉得重点是在于模仿，即模仿已有的项目处理方式，比如fireblocks就应该归于用一个situate自动化来download过程，不管是rest api也好，其它也好。至于之后的每一步，可以通过老大来压阵讨论，来明确说责任落实在哪个组，或者是否采用其它方式。

我的灵机一动的想法：reconciliation这个东西在finance里面是有，但在其他software development的地方也应该更有，就像分布式系统里面要做多来源数据核对、检查的过程一样，只是面对的数据不见得是金融数据而已。

<https://exactpro.com/ideas/research-papers/reconciliation-testing-aspects-trading-systems-software-failures>

显然trading里面是最多的，见这篇论文。

<https://www.guru99.com/what-is-data-reconciliation.html>

这是general的软工的data reconciliation介绍。注：所以其实我们做的例如crypto journal或者wedbush recon都是data recon的一部分，因此对data recon这个问题，工业界肯定是有一些主流（“标准”）的做法，但是没有教科书（可能有论文），这就和做marketing的有自己的套路、左投行ipo的有自己的流程是一样的，或者工匠师傅雕花工艺流程，或者怎么拍电视剧电影纪录片。这就是一个很实践的问题，根据前人的经验得到了一些”best practice”，比如组与组之间的责任分配，用什么方式获得原始文件比较好，以及怎么recon，没有什么太玄乎的，因为它是一个已经被界定清楚的问题，怎么解决也都基本是套路，Mike做过一遍也就都知道了。

<http://www.uml.org.cn/bigdata/201811291.asp>

基于 Apache Ignite 日均十亿数据对账实践应用。没啥用，主要是解决分布式的问题。

<https://patents.google.com/patent/CN101458705A/zh>

比如阿里为了支付问题和银行对账，就开发了类似系统，还有专利。注：这简直了，支付宝的pain不就是我（们）现在组的pain么，支付宝还和非常多银行有海量对账信息，简直了。

在会计上面，reconciliation就是对账。https://zh.wikipedia.org/zh-cn/銀行對賬

<https://www.solvexia.com/blog/finance-reconciliation-how-to-step-by-step-process>

上面这个是金融reconn，没啥大用，不是给dev看的

<http://blogxin.cn/2018/09/30/Distributed-Consistency/>

<https://xie.infoq.cn/article/c9bd3866ed5f76c555314aefa>

这个说的是在分布式系统，data reconn叫数据一致性问题。提到了一些做法，比如保证可以回滚（即我们为什么要用disable而不是直接删掉）。这个又和network里面的TCP之类的连接协议好像有点像？

突然想到，整个bunker的check breaks系统，不就是data reconn么？它能够做到几乎实现了自动化，这就非常不容易了，只需要偶尔的人工support和维护。核心并不是用了什么高深的cs代码技巧，而是一个工程选择吧？但也有可能那个数据本来就比较clean？

公司发的mac上面用的是zsh而不是ssh，所以wild char \*不work，要加引号才行。我自己的电脑上是ssh。

https://unix.stackexchange.com/questions/130958/scp-wildcard-not-working-in-zsh

sql两个表join，怎么让多个列按照同一个列join进去。

https://stackoverflow.com/questions/12912537/sql-how-to-make-multiple-joins-to-the-same-column-of-a-table-without-overriding

现在每次查crypto\_journal的错误都是用sql，但它不是代码，以至于每次都要重新改一下和复制黏贴，很糟糕，我能不能考虑把它变成script然后参数化？或者把常用的flow写成代码，做sanity check

我想查一下所有的transfer类型，是不是必然是有fireblocks和bank参与。Itbit是个例外。

Ssh真是万能，反正可以connect到remote，然后想做什么都可以。<https://unix.stackexchange.com/questions/17466/how-to-delete-a-file-on-remote-machine-via-ssh-by-using-a-shell-script>

Generator的核心是，当被调用时，它们并不直接返回result，而是返回一个result generator（可以循环调用或者直接call next()）。Generator里面也可以有return，它的作用是相当于结束了整个函数（所以return之后的任何代码，包括yield，都是不会执行的）。Generator的使用主要是出于performance和memory的考虑。

而yield from，从简单意义上说，就是yield from g=for v in g: yield v。

Generator function就体现为里面有yield，函数被call的时候，返回的是一个generator（可以循环调用或者直接call next(generator)来读取yield右侧的值）。但函数的运行是在yield那里suspend，向**调用者**返回yield右边的值，而且在下一个iter的时候再resume。

问题：在yield左侧的值有什么用，可以得到yield右侧表达式返回的值么？

Ans：yield左侧的值，一般是generator.send(x)传进去的x。

注：我觉得tornato或者ayncio里面用的await或yield左边可以返回一个值，已经和上面说的不是一个意思了。

问题：yield一个generator会怎么样？

答：会再返回一个generator的generator

Text

Description automatically generated

Generator似乎是Iterator的一个subclass，在typing doc里面可以看到<https://docs.python.org/3/library/typing.html#typing.Generator>

至于说为什么可以在表达式右侧是result generator，左侧是多个值，是因为在这种情况下，python自动对右侧做了一个循环来unpack，相当于左侧多个值等于右侧一个list。

Any加generator一起的时候碰到的奇怪问题，有个performance差异。

<https://stackoverflow.com/questions/44802682/python-any-unexpected-performance>

为什么yield就不能用typing了，比如variable: str = yield something()

sql中要特别注意，不能对NULL使用比较运算符，见mick-sql基础教程65页。因为sql认为，null field的column根本就无法进行比较，所以总是不会被选中非返回。

<https://askubuntu.com/questions/45521/how-to-navigate-long-commands-faster>

在命令中快速移动

<https://stackoverflow.com/questions/19327556/get-specific-line-from-text-file-using-just-shell-script>

从文本中读取特定行

https://www.investopedia.com/terms/e/ex-dividend.asp

https://stackoverflow.com/questions/1367373/python-subprocess-popen-oserror-errno-12-cannot-allocate-memory

太傻了，free -g显示的是内存，df -h显示的是空间，两码事

woc situate居然用了我15GB的内存（不对，是27！！）

<https://unix.stackexchange.com/questions/166558/how-to-check-which-process-is-using-most-memory>

检查哪个process用了最多的memory

free -g

pip install -e .似乎并没有更新git pull下来的变化？还是说，它真的是会追踪变化？（我觉得似乎不追踪）

<https://stackoverflow.com/questions/35064426/when-would-the-e-editable-option-be-useful-with-pip-install>

<https://setuptools.readthedocs.io/en/latest/userguide/development_mode.html>

反正我觉得这里面肯定有什么bug。至少比如setup.py里面的更新，似乎pip install并没有及时追踪（和git pull没关系，你就是在本地直接改，但别run inv develop也会有这问题）

解决方案：inv develop。

有趣，csv格式的数据老有问题。。比如年份在格子里显示的和在栏目里显示的不一样。用sublime打开就可以看得挺清楚的。

refactor的动力是什么？这问题挺有意思，因为有些写得实在太broken了，逻辑不清楚，变量名字也不清楚。

TODO:摆脱global variable，改为用函数参数的形式传入、返回

注：OOD里面讲的是class抽象，category说的是函数抽象？都很重要，比如对library writer。

db ts field has type ts with timezone. while you pass me is a utc ts string without timezone (i am not to blame you at all). by default db think it's a CDT ts and auto converts it to utc ts

postgre ts真是太绝了，默认转换ts的timezone，错的神不知鬼不觉。

https://stackoverflow.com/questions/31715078/postgressqlalchemy-converting-time-to-utc-when-using-default-func-now

<https://stackoverflow.com/questions/4770297/convert-utc-datetime-string-to-local-datetime>

crypto journal里面，之所以要把bank的date转换成CDT的ts，就是因为读入db table的时候，postgresql会自动对ts的tz做一个转换，转换成utc时间。而之所以那个里面不用先做一个改变default tz的操作，是因为反正只要落在同一天里就可以了，差几个小时（即cdt和utc差五个小时也无所谓），太粗糙了。

卧槽太可怕了，为什么三月份数据那么奇怪，是因为还有daylight saving time，之前是-600，现在是-500，这么看的话就make sense了。

crypto\_journal里面还有reconciliation/utils.py:get\_date()为什么是4pm？

crypto\_transaction table里面的dt是芝加哥时间得到的date，而且还加了一天。所以按天跑的话，dt应该显示的都是同一天（因为前一天的会被加一天，这一天的还是这一天）。

我要删除一个变量，得看它首先是什么意思，然后它在哪里被用到了，删除了是否会有影响。

太绝了，在python3.6里面居然还有timezone不能strptime的问题。

<https://stackoverflow.com/questions/53291250/python-3-6-datetime-strptime-returns-error-while-python-3-7-works-well>

为什么在现有环境下不能装python 3.8？很简单，因为conda创建那个环境的时候就指定了python的version，所以会有这个问题。

先升级python。这还有个先来后到的顺序，bunker\_utils没升级所以我还不能升级。

Help field was set for param(s) that don't exist: ['docs']看起来是deploy\_conda这边的问题。Ans:确实，注释掉那一行"docs": "Deploy docs"就没事了，虽然还不知道为啥。

然后还有一个原因就是没有在正确的目录下运行inv…

Ts太可怕了，ts > '{date:%Y-%m-%d} 15:00:00 America/Chicago'"就是因为搜索的timezone不对，所以在load之后又误删了一些数据。

因为update\_accounts里面也有一些yield，我猜在一些情况下，可能后面的journal部分先运行了，而db account table还没load，所以可能会出现问题。

nb，居然有新currency叫1inch。

https://www.coinbase.com/price/1inch

things to do:

1. 厘清crypto\_transactions里面的删除query和date，保证它等价是最好的了，不知道有没有办法能让前一步（添加）完全等价可逆。Ans: FB and exchange checked. TODO: bank. partly checked
2. 把bank从matching process里面load挪出来，这不符合single responsibility。而且之后要每天跑，现在是每次extract所有unmatched的，可能matching效率会好一些，但太难维护，还是要做成只做每一天的数据，只对每一天的数据产生影响。现在继续check：bank side也只有同一天的数据（yes finance side check），我可以加sanity check了。checked。所以现在对bank side我可以很简单地按照dt和comment来删除bank以及bank对面加的数据。checked。我觉得account service还是很好用的，应该说这个接口设计得很好，就是用来和外面交流，也是最经常改动的地方。
3. 然后对crypto\_transfers也做同样的事情。处理match好的东西的幂等性。checked。现在dt就等于chicago\_cutoff\_date，一定是同一天。
4. 现在我只match同一天里面的。checked。
5. 可以容易地核对：从文件里读的数字=我从db里拿出来要做matching的数字，这就是保证pipes的接口完全是严丝合缝的，没有意外地漏水。checked
6. 修改db table name。
7. bank side的那个uuid一点帮助也没有，要改掉，并不能帮助我们辨认tx。

问题：这个bank里面的rate service和alias service怎么处理。这个couple实在太严重了。我觉得应该区分这两个的“层次”，这两个应该说是更容易改变的，所以应该放在更显著/high level的地方？

现在trader看到的是一套数据，bunker看到的是一套数据，finance在journal里面看到的又是另一套数据。应该至少要提供更多的信息来让后两者看到的数据类似。然后naming也是一个问题。

写代码怎么还要做到易于debug，所以得记录必要的中间步骤，这个是工程设计的一般原则。这就是why cpp这么难debug。

为什么这个项目写起来难，还是接口的问题。

注：幂等性确实很重要，比如在支付领域，如果没有幂等性这简直太可怕了。

注2：怪不得有些支付页面说不要刷新，因为它自己side的幂等性做的不好…原来如此

注3：看起来幂等性在分布式系统中用的比较多？<https://zhuanlan.zhihu.com/p/30493445> 看起来在http中也很重要<https://www.i3geek.com/archives/841> 但应该这两个讲的是一回事，因为分布式系统里面的RPC用的原理也就是http的消息问答机制。

靠，原来是assertions的dependency版本问题，可以看Mike的secure的PR来改。也不对，就是把assertions升级到1.12.0即可，原来这个是In house的library

现在要把branch merge到master，那么哪些merge哪些不被merge，怎么merge呢？

我想要在linux上怎么快速删除光标前的所有内容

<https://stackoverflow.com/questions/44846908/how-to-delete-from-current-position-to-the-beginning-of-the-line>

ctrl u确实是work的。但其实我是在gnu里面。

查看我目前用的是什么shell。

<https://unix.stackexchange.com/questions/9501/how-to-test-what-shell-i-am-using-in-a-terminal>

<https://askubuntu.com/questions/590899/how-do-i-check-which-shell-i-am-using>

linux在单词之间跳来跳去

<https://stackoverflow.com/questions/81272/how-to-move-the-cursor-word-by-word-in-the-os-x-terminal>

强行把feature branch里面的东西merge到master。

<https://stackoverflow.com/questions/30105210/git-overwrite-master-with-branch>

先建PR，然后git checkout，然后用下面这个git pull，然后再确认一些local changes。

git pull origin new\_matching -s recursive -X theirs

<https://stackoverflow.com/questions/4785107/resolve-conflicts-using-remote-changes-when-pulling-from-git-remote>

这个确实好用，免去了我自己在那边手动删除，如果我已经确认了merge里面哪一侧是我需要的。不过它似乎在没有conflict的地方还是会自动merge，这是个问题。

python的black不负责remove comment里面的空格W291

<https://github.com/psf/black/issues/144>

还有一个package引用的问题，又来了。要么在init里面import下面file里面的class并且加入all=[“xxx”]，要么就不import也不用all，就像bbg\_service里面做的那样。

Ans:发现其实是因为之前做的是引用到directory这一级，而不是file这一级。所以要import file里面某个class，其实是需要在init里面引入。

<https://stackoverflow.com/questions/35727134/module-imports-and-init-py>

而根据这个可以看到，直接让某个directory下面所有东西都引入并不见得是一个很好的选择。所以init还是空着比较好。

我目前做的事情是让旧的branch merge到新pj上面。第一步这个git pull已经完成了。现在看的应该是pj的结构。

pj的structure或者叫layout也是专门有讨论的

<https://docs.python-guide.org/writing/structure/>

这个包含internal packages的结构和我比较像

<https://realpython.com/python-application-layouts/>

还有书专门讨论开源软件的结构

<https://book.douban.com/subject/6430747/>

 结论：看起来代码位置怎么设计，还是一个design patter里面的问题。

mypy的type checking还是很有用的，尤其是针对Optional的情况，确实人有时候会忽视这些Null情况的edge cases。

TODO：可怜，可能还是要看一下git。

git push为什么我已经改了pre-push，依然没有用

Ans: 因为有两个pre-push，一个在.hooks/pre-push，一个在.git/hooks/pre-push。但为什么有两个我还没搞清楚。

next step：我要把已经改好（解决完conflicts）的新弄到一个branch上，然后所有都采用我的标准即可。

阻止push到master

<https://stackoverflow.com/questions/46146491/prevent-pushing-to-master-on-github>

看起来好像是在hooks里面设置的

<https://stackoverflow.com/questions/19021978/how-to-block-push-to-master-branch-on-remote>

我最后是在Bitbucket的branch permission里面设置了之后，直接push到master的。

为什么gitignore不work

<https://stackoverflow.com/questions/25436312/gitignore-not-working>

因为有些文件一直就在cache里面，或者说一直被持续track了

第一个问题，用了ssh验证而非密码验证后，是不是我就不能登录了？

我现在的问题其实很简单，如果我想要登录abn sftp，怎么做，能不能做？

Ans: sftp -I ~/.ssh/id\_rsa\_akchicago [akunasftp@sftp.us.abnamroclearing.com](mailto:akunasftp@sftp.us.abnamroclearing.com)，明确key的位置

<https://serverfault.com/questions/273967/how-to-use-openssh-sftp-command-with-a-rsa-dsa-key-specified-from-the-command-li/273984>

我猜是因为，默认用的是id\_rsa秘钥，但这个我是用来自己登录其它prd/dev box的。Ans:对的，默认情况下看的是id\_rsa秘钥，这就是为什么在config里面有一个pkey\_path的field，python secure连接的时候用的是这个路径里面的pkey，它会用paramiko.RSAKey.from\_private\_key\_file(pkey\_path)来读取。奇怪，我改了pkey文件的名字就不能连了？好像改别的名字都行，除了id\_rsa这个默认名字。

woc我一不下心删除了自己的id\_rsa秘钥。

<https://serverfault.com/questions/183664/recover-an-ssh-private-key>。没得救了，只能重新生成一个，然后用ssh-copy-id -i key\_file [user@]server\_name重新复制到我连接的prd box上面。

在sftp上面如果要搜索某些文件的话，可以用ls \*2021-04-01.txt

ssh原理就是说，有个人生成一个pkey加public key，这个public key发给另一方作为验证，自己保存这个pkey作为验证的钥匙。至于这个key是danny还是其他人生成，一点也不重要，反正只要akuna这一侧需要连接ftp的都用这个pkey就行了，其实就是一把钥匙，public key就是一把锁。

我目前思考了认为想做的几件事：

1. 每一个pipe都产生并保存一个T字路口型的中间数据结果，并且每一步都能接受对中间结果的修改。即journal那一步，会先读取transfers csv的结果。
2. 把MCV里面的view部分，即常常需要修改的account和alias拿出来，要做到可以单独维护。
3. 把matching策略做到黑板模式里面的能单独分离，可以拆卸修改。
4. 统一数据接口，比如像这种有transaction又有target\_transaction的就可以删掉了。
5. 剩下的可以再按照小的design pattern进行优化设计。
6. typing有空处理下。

Question:那么在T字路口输出中间数据，应该顺理成章地跟在T字路口的下高速口，而读取可能被修改过的中间数据，应该在T字路口的上高速口侧？

这个要check一下design的书。

注：这样设计不仅是因为我现在知道说这个地方会要改，哪怕我不知道，一个好的design依然是让每个pipes之间可以有“喘息”的空间，以防万一要做点啥事，就像高速有休息站和应急停车点一样。如果先期不设计好，或者没有按照pipes设计，后期要断开就很麻烦。如果前期断开好了，哪怕没建休息区，后期要建也是很容易的。

 注：我的pipes里面很有用的logging信息，是否有办法弄到email里面去？这个以后再说了，先把架构弄好。

把python dictionary转换成namedtuple，只要用namedtuple(\*\*dict)就可以了，非常简单。原理在书的18章，意思说在func(\*\*dict)会pass all key/value pairs in dict as individual keyword arguments

<https://stackoverflow.com/questions/43921240/pythonic-way-to-convert-a-dictionary-into-namedtuple-or-another-hashable-dict-li>

问题：那如果是反过来想把namedtuple pass给另一个namedtuple呢？或者func(\*\*namedtuple)会得到什么？这个问题的来源在于，我查清楚了，akdb.connections.DBRow就是一个collections.namedtuple。

Ans：想把namedtuple pass给另一个namedtuple要another\_named\_tuple(\*\*somenamedtuple.\_asdict())这样处理。传给函数作为参数的话，有这两种骚用法<https://stackoverflow.com/questions/19707267/how-can-i-apply-a-namedtuple-onto-a-function>

注：在我不知道Low level的情况下，说实话我已经忘了\*和\*\*了。所以说不看low level没啥意义，如果也不是天天用的话。

问题：如果我只有最后一个或者其中某一个field需要特别处理，怎么做？

问题：dict里面为什么读进来都是string？这样typing是不对的（好像也不是不对，db table读进去的时候会自动convertion）。

注：这个其实涉及到一个更广的问题，即dump和load的interface处理。最理想的情况是，我希望直接\*\*即可，因为dump的type也是我决定的。

Ans：可以啊，在akdb insert里面有一个self.vendor.formatter，它是用来处理type的？看了看，好像也不是，只不过是加了个引号而已。看来还是有postgre的底层做了类型转换。要小心了。这就像之前ts传进入，它自动转换timezone一样。

Ans：从原理上说，python和db进行交互，就像和csv进行交互一样，只是csv是通过一个file pointer，而db是通过一个cursor。据我观察，除非是Bool这种类型比较特别，其它postgre里面的text,date,numeric都可以用string直接insert进去，它会自动转换。

<http://www.postgres.cn/v2/news/viewone/1/481>

在官方文档里面有这一行：

<https://www.postgresql.org/docs/9.5/sql-insert.html>

If the expression for any column is not of the correct data type, automatic type conversion will be attempted.

还有一章是专门讲conversion

<https://www.postgresql.org/docs/current/typeconv.html>

还有这个讨论讲到了postgresql auto conversion的机制

<https://dba.stackexchange.com/questions/194975/how-to-avoid-implicit-type-casts-in-postgresql>

所以namedtuple就有点像tuple和dict和class的混血产物。它既可以用position来访问field，也可以用.attribute访问。namedtuple是一个iterable。

tuple和list的区别之一在于，tuple是不能原地修改的，所以namedtuple也是immutable的。

<https://www.geeksforgeeks.org/python-difference-between-list-and-tuple/>

python书上还真讨论过（Why You Will Care: Dictionaries Versus Lists），是用List还是dict来表示db record。这确实很有意义，因为和数据库经常打交道的程序员，难道不应该思考一下data structure么？好的数据结构能提升工作效率和代码效率。

Namedtuple和Decimal（但我看第五章又说这个是core numeric type？？）一样，都是python的extension types而不是core types。

Decimal是有固定精度的浮点数。用Decimal的主要原因是，python float计算并不是精确的，因为储存空间问题。例如：

>>> 0.1+0.1+0.1-0.3

5.551115123125783e-17

有时间的话，还想看一下Iterable和Iterator的区别。

csv里面读进来的是str，怎么能简单转换类型？<https://stackoverflow.com/a/54863733>

dataclass可以，namedtuple还要再看看。 <https://stackoverflow.com/questions/49673688/elegant-way-for-getting-field-defaults-on-a-typing-namedtuple>

不同文件之间的读取的类型接口是个大问题。这大概就是为什么用Json file了，可以很好地保存type。

为什么python不能封装好说date(str)就能得到date。

check db table自动conversion。Ans:基本check了，不属于我能控制的，要看底层数据库实现了。

给matching加for loop的date，看看是不是要重新写，把matching和loading的entry point script也完全分开来。Ans: checked

加note column。Ans: checked

<https://www.postgresqltutorial.com/postgresql-add-column/>

<https://www.postgresql.org/docs/current/datatype.html>

postgresql官网有专门讲data types的一节，书里面只是讲了一些比较不常见的类型。postgresql里面的boolean有三值，true,false,unknown(null)，默认是Null。

然后再是用新的design pattern重写matching那一块。

由于删除只是process里面的一步，如果其他步失败了就没法保证原子性。最好能保证原子性，不行的话问题也不大，只是可能会让人出乎意料而已。

和Ben的这个测试过程可能是大公司里面的上下游测试（集成测试）？

是否follow了正确的软件开发/重构流程？

以及软件开发中，似乎每一次改动都能被测试，这个是很重要的，因为越到后面bug积累越多。这可能是python的优势？调试容易。

<http://staff.ustc.edu.cn/~sycheng/sst/lectures/ch09_Introduction_of_Software_Testing.pdf>

crypto journal TODO:

1. 把until\_ts之后就用start\_date的end\_ts来获得就行，不用传入until\_dt
2. 如果有空的话check一下email format，但这个不是很重要。ignored
3. 能不能给表格加一列序号。ignored
4. 既然都有对business date的check，以及要做一个date loop能不能放一块做。这是一个design问题。或者封装一个date range loop的函数。以管道取代循环？好像用不上。实际上etl也是这样的过程，只要是每一步都被封装了而且看起来很清楚即可。
5. 把root路径单独拿出来，这样每一次只要单改一个root就可以修改跑的文件的位置了。

python class里面的protected variable就是那些变量名前面加了\_的变量。

<https://www.geeksforgeeks.org/protected-variable-in-python/>

在ssh通信里面，cipher就是encryption algorithm。用法就是，对dash sftp，要用sftp -oCiphers=aes256-cbc user@ipaddress来登录

见这个<https://superuser.com/questions/918118/how-do-i-fix-a-cipher-mismatch-issue-with-sftp>

问题：pysftp.Connection用的默认cipher是什么？我怎么觉得是none。<https://pysftp.readthedocs.io/en/release_0.2.8/pysftp.html>确实是None，但为什么它是None却还可以连接，我直接用sftp就不能连要提示cipher错误？

outlook email要高亮显示的话，似乎就是用editor里面的东西即可：高亮，加粗，斜体，什么都有。

<https://slack.com/help/articles/202288908-Format-your-messages>

<https://slack.com/help/articles/202528808-Search-in-Slack#browser-1>

我猜spot说的是这个

<https://www.investopedia.com/terms/s/spot.asp>

design问题：query放在哪儿？Path放哪儿？

<https://realpython.com/python-application-layouts/>

所以这个其实是一个python module design的选择问题？即思考在module层面上，把什么class放到什么module下面比较好，其实每个module就相当于一个子package，如果有跨package import，那肯定就不太对了。

<https://pythonguidecn.readthedocs.io/zh/latest/writing/structure.html>

这里面还讲到把纯函数和有side effect的函数分开。

<https://learnku.com/docs/python-guide/2018/structure/3260>

我刚才改的一个地方，就只是把逻辑理顺了而已，这是真的叫Refactor，因为实际运行结果不该有任何改变。

又改了一个地方，把原来的函数改成pure function。yield应该会返回return的东西把？

把”FUT”这样的magic string封装成constant

还是想把那个逻辑条件改的清楚一点

TODO：

1. 看一下python class method rebinding。<https://stackoverflow.com/questions/1301346/what-is-the-meaning-of-single-and-double-underscore-before-an-object-name>. 更具体地说，我想看的是name mangling。name mangling实际上就是通过将\_\_method自动转化为\_classname\_\_method来避免naming conflicts，没有什么特别神奇的地方，也并不是真正的完全private了，因为有心人依然可以通过\_classname\_\_method这个名字来直接访问该变量。
2. 网卡有没有办法解决
3. bbg\_service是不是能再重构一下，依然看起来不是很清晰？至少private method那什么的可以写的更清楚一点。

可以按照ts来删除某个时间范围内的文件

<https://stackoverflow.com/questions/33091013/how-do-you-delete-files-older-than-specific-date-in-linux>

bound method就是说把一个class method和self绑定在一起，然后就可以当一个正常的、普通的function来用了。

python的scope主要是为了解决name clash问题。

骚操作：比如众多内置函数，python会自动查找我们用的某个变量是不是来自builtins这个package，如果它在其它范围内没找到的话。再换句话说，如果你希望的话，可以半路截下这个call。不过这是挺可怕的practice。

Graphical user interface, text

Description automatically generated

函数之间如何通信（如果不用global variable的话）？简言之，依赖参数和返回值，即做成一个pure function。

闭包：一个函数记住了enclose它的函数的状态（变量之类的东西）。闭包是函数式编程的技巧，因为也有状态记忆能力，所以可以作为class的替代。即使外部函数已经退出了，内部函数还是能记住一些状态，这就是最别致的地方

A picture containing text

Description automatically generated

比如说这个N就会在action内部被记住。

下面这个写法和上面是等价的，因为反正是nested函数，也不需要名字



提一个问题：假如说我想要通过class method初始化一个class variable，而且仅仅需要做一次。那么我可以在init函数里完成，但如果这个method是coroutine怎么办？

managed attributes的意思是说，用getter, setter methods来管理attribute access。

怎么把string形式的attribute用来get，用getattr(x, 'foobar')=x.foobar。正好就是我上面看过的这个managed attributes一章里面讲的东西。

<https://stackoverflow.com/questions/3253966/python-string-to-attribute>

defaultdict就是一个对不存在的key提供默认值的dict，初始化的时候给它传入的是一个callable，用来在key不存在时用callable()生成这个key的value。

简便起见这个callable有时可以是lambda。

drop copy是什么。看上去像一个trading数据copy的东西？

[https://www.cmegroup.com/globex/trade-on-cme-globex/dropcopy-faq.html#](https://www.cmegroup.com/globex/trade-on-cme-globex/dropcopy-faq.html)

<https://secure.fia.org/downloads/FIA-Drop_Copy(FINAL).pdf>

这个crypto\_journal每次exchange side matching的逻辑（不管是用transaction id或者不用），都是先比对一番，如果能match上的话，就建transfer并加到list，把id加到matched\_id里面，返回transfer list和matched\_id。

把一系列判断条件放到一个返回True or False的函数里面，就像\_trans\_matches那样。

有不同的match策略，但又有一些基本的相同的东西（方法），我不想为每个策略设计一个子类。我希望未来在策略上，以及方法上是容易扩展的。是否考虑两个类？还是用策略模式把。

对等待match的tx，用状态机？再创建一个State类？

Atlassian居然只能用paragraph那个选项来弄格式，太傻了

<https://www.youtube.com/watch?v=HuiicbL1SpY>

原来table of contents是根据heading自己生成的，nice。

sql IN谓词后面跟的是一个括号（tuple），所以postgresql是有tuple类型的？<https://www.postgresql.org/docs/current/functions-comparisons.html>

答：这里面只是说了要有括号而已，没说类型。我猜测就是说在python string里面，其实你也看不出是不是tuple了，只是表示为括号而已，而这就恰好和postgresql里面要求的括号一致。

怎么样给tuple of variable length写typing

<https://stackoverflow.com/questions/54747253/how-to-annotate-function-that-takes-a-tuple-of-variable-length-variadic-tuple>

原来三个点还有这种说法

<https://stackoverflow.com/questions/772124/what-does-the-ellipsis-object-do>

所以说三个点主要是和\_\_getitem\_\_方法有关咯？

<https://stackoverflow.com/a/753260>

所以说python3定义了一个ellipsis，其实只是说在各种内置方法里面或者typing里面，如果看到这个object/class的时候，知道怎么处理它，仅此而已。

所谓运算符重载，实际上就只是说用user自己定义的方法代替python提供的默认内置方法而已。

blockchain的tx id实际上是sha-256，即16进制数，所以严格来说是只有数字加1abcdef的，当然前面还有0x表示这是十六进制。

<https://zh.wikipedia.org/zh-cn/SHA-2>

regex官方教程

<https://docs.python.org/zh-cn/3/howto/regex.html>

我想要返回regex matching的最后一个组，怎么做

TODO:

1. 看一下tx hash究竟是用什么组成的
2. 再看一下合成pipe，怎么做的更好（如果要我合成itbit的话）。现在不是很重要。
3. 有个问题，为什么不直接给table加一个column，名字叫is\_synthesized，type是 bool呢？如果这样的话，我同时也要看看，怎么样让named tuple的改动最小。check
4. 怎么把一些对input的sanity check放到其他地方，这样不会那么碍眼

还要提供一个monthly或者说compiled email。这里要做的一个清理，实际上就是对email format的格式进行清理，现在其实是static的email msg（self.msg）和格式渲染混合在一起，非常不好。

TODO:

1. 清理所有的path。答：这个是不是在问（或者应该参考）怎么design一个file management system？（？这玩意儿是啥，我查了下也没看到存在，我心里想的是linux的file system，所以估计是两个东西把，上面这个信息应该基本无效）现在的问题，是用function还是class的design比较好？基本解决了。
2. 使用account service来替代原有的旧条件判断。checked。
3. 清理email format和content

下面这两个都是设计问题啊：

有一个设计问题：我应该在哪里储存放置sanity check？因为它只会越来越多，而且我希望它可以reuse。

为什么这些service不把populate放到init里面去？尽管我看所有后面的操作都要先populate。（原本是有performance的考虑？我不知道why）Ans：congxu不愧是老法师，一眼就看出问题所在，如果要弄asyncio的话就可能会有问题，这是一个python语言特性的问题。

我在某一小点上做refactor，实际上也是遵循了重构那本书里面的一些步骤。

python还有一个tempfile module

<https://stackoverflow.com/questions/8577137/how-can-i-create-a-tmp-file-in-python>

python有typing是好处还是坏处？它是不是破坏了多态？

为什么要super().\_\_init\_\_而不只是super.\_\_init\_\_。

书里32章讨论了super

对于matching而言，把strategy class和负责matching基本工作的class分开。

那么matching class负责做什么？负责调用那个strategy class的match函数？现在这个工作是在transformer里面的，确实应该分出来。

要解决的两个问题：

1. 返回什么，目前看起来似乎还是返回合成的transfer比较好
2. 用每种策略match一下试试，又像是一个个pipe，是否还用一个global set来记录哪些已经match上了？

想要检测nested list是否为空。这是一个简单的算法了。

<https://stackoverflow.com/questions/1593564/python-how-to-check-if-a-nested-list-is-essentially-empty>

是不是改成说，如果不是iterable，就return False？

解释了TypeError是什么

<https://docs.python.org/3/library/exceptions.html#TypeError>

map的函数签名是take iterable

<https://docs.python.org/3/library/functions.html#map>

python set怎么一次update多个值

<https://stackoverflow.com/questions/24091455/how-to-add-multiple-strings-to-a-set-in-python>

问题：

1. 如果strategy design pattern里面有一次组合多个strategies怎么搞？这个责任放在哪里？
2. 每个strategy的match是不是能做到像pipes一样？答：现在考虑要让input和output一样，这实际上就是取一个input加output的并集（transfer, trx, matched\_id）。唯一缺点就是可能performance会比较差，如果数据量很大的话。如果准备用Pipe的话，要不要用map filter reduce写，比较清楚一点。比如找到unmatched，这就是个filter。
3. 以及shared state（trx\_ids）准备怎么处理？答：如果用上面那个方法，就没有函数间通信的问题了，不需要global variable/shared state。closure和decorator实际上是相关的
4. 能不能把checks都变成deco
5. 以及cross date transfer的回滚怎么做？

python里面parent class里面的某个函数如果用了self.someattribute，这个self指的应该是call它的那个class instance把？

python是用abc来标识super abstract class？

parent class的virtual method能不能不指定函数signature?

有意思，有没有composite loader？即返回的数据需要用多个不同的loader来处理。Ans:目前给的解决方法是，写一个自己的Task的class，里面能接受两个loader的逻辑。

基本想法是说：match\_date应该是由extractor来控制，而transformer, loader, strategy这边不需要考虑date的事情。

现在一种强制的做法是，要求monthly in transit matching，一定要跟在整个时间段里面的matching后面跑，这是make sense的。

目前准备是先让journal跑起来，然后再一点点调试和refactor。

先是path。给path加一个logger的decorator？

答：难点主要在于，我是否能得到返回的这个变量的名字？<https://stackoverflow.com/questions/2749796/how-to-get-the-original-variable-name-of-variable-passed-to-a-function> 看起来是不可以的，这和变量是在什么时候被evaluate有关？

naming得想个办法。

market segment gateway

<https://www.tradingtechnologies.com/xtrader-help/cme-gateway/configuring-order-management/market-segment-gateways/>

所以sorted的key要求的就是提供一个函数能返回一个变量，代码里面返回的是tuple。而tuple的比较就是按位置从左往右一个个来<https://stackoverflow.com/a/5292332>

怎么给sorted里面的key传递超过一个参数。这其实就是用了闭包或者命名空间、函数参数域的概念。也算一种奇技淫巧？我现在倒觉得，命名空间其实是程序员打交道最多的东西之一了。

<https://stackoverflow.com/questions/22483867/how-to-pass-multiple-arguments-to-the-method-key-of-sorted>

symbolic link

<https://askubuntu.com/questions/522051/how-to-list-all-symbolic-links-in-a-directory>

我账号被锁还可能是credential没更新？/sandbox/.smbcredentials

能不能试一下mound shared drive在local上？

sshfs？反正我现在失败了。

我现在怀疑，python formatted string的底层实现可能也是regex。

下面这个链接说明了，strftime是date特有的方法，而另一种方式是用了general string format。

<https://stackoverflow.com/a/22842734>

其实改完date，我可以用一个regex来check的，甚至可以直接用regex改。

是不是应该看python source code？可以考虑

我需要一个ide能让我在c代码里面跳来跳去，pycharm好像只能在python代码之间跳。

pycharm可以右键一个文件点compare，这样就可以比较一个文件里两个函数了。

能不能增加一个功能，缺啥就发email，相当于把logging发出去，但是得判断好哪一步出问题。这就像是编程语言的error msg呗，或者os里面的报错。不然总不能出一个问题，os就直接崩溃吧？后者编程语言编译错误也不发送错误提示，要开发这个编程语言的人自己去看。

目前的一个目标是说，如果在Journal里面看到一行，我能不能很快在transfer里面找到。

naming真的很难啊，比如上个月的最后一个工作日，last business date of the last month，这变量怎么命名

在自己的box上面也可以定时跑程序啊，unix上用cron就行。所以situate我猜也是基于这个做了个ui而已。

我觉得看OOD不仅看看模式，是不是还应该看看总结use case那一部分？这个我倒是没去总结。

说不定以后可以从daily break check再到real time break check？

我之所以一定要用一个class method decorator，就是为了要利用那个instance attribute来保存一些state information。

<https://stackoverflow.com/a/13852431>

把一个class method decorator放在class里面或者外面的主要区别在这里。看一下functools.wrap主要是做了啥。

<https://github.com/python/cpython/blob/049c4125f8a2b482c6129db68463f58e20c31526/Lib/functools.py#L35>

可以直接看source code，wraps实际上就是update\_wrapper函数，只不过用partial包装了一下。update\_wrapper做的实际上就是把一些attribute用getattr从原函数获取，然后用setattr赋给新函数。

为什么\n不能直接用在html里？

<https://developer.mozilla.org/zh-CN/docs/Web/CSS/white-space>

<https://www.jianshu.com/p/a580568e0055>

不知道为什么现在设置的是white-space:non-wrap，我也不改了。

我觉得前端这个css太可怕了，封装得又不好用。

有空把所有name都重新命名一下。

是不是那个fill\_missing\_data重新写一下，在其他exception出现的地方也都可以用？

ftp看一下

怎么check一个Ip地址是不是whitelist了？Ans:好像可以用telnet ip\_address。telnet要加port。。。

可以用Ping来测试一个ip是不是能用。

<https://unix.stackexchange.com/a/283185>

telnet也是一个application，built on telnet protocol？通过man telnet可以看到。

现在要看一看，有没有讲sanity checks应该怎么放的design pattern

<https://softwareengineering.stackexchange.com/a/261328>

这个有点玄学，specification pattern规格模式？好像不是。

看起来还是应该用visitor模式？看起来visitor的核心是double dispatch

还有registry，是什么？

目前我想要解决的，一个是出于对用户输入不信任的check，这个应该放在另外的visitor class里面，因为它和我的核心逻辑是毫无关系的。一个是matching strategy里面的check，这个应该用chain of responsibility这样类似的东西来handle，因为它就是matching的一部分。

我仔细看了一圈，觉得好像没有很好的唯一的放之皆准的解决方式，visitor pattern不见得是最合适的。

<https://stackoverflow.com/a/18318439>

方案：我准备先解决大部分is\_business\_date的assertion。checked

然后fireblocks和exchange的loading task可以合并

现在来解决matching strategy里面的validation。

目前我是整理出了一个check，并且假设不管是什么matching strategy，都必须通过这个check。目前看起来这个要求是合理的，但如果未来我希望不同的matching用不同的check呢？这时候是不是就可以用visitor模式了？因为不同的matching strategy是不同的class。

由于现在check是作为一个visitor class传给strategy class，所以我哪怕是用同一个strategy，也可以用不同的check标准。

会不会太复杂了？而且这也不make sense，难道每次我在double for loop里面，都要调用一个外面的check函数？

当然咯，最简单的一个方式就是直接用composite，check的逻辑反正也已经独立出来了，无非就是把validator作为一个变量在strategy init的时候传给它就行。

需不需要要factory？不是，需要的是Builder。

我再仔细想了想，这个matching实质上应该有这么几个部分：search（double for loop），criteria/validation，create matched transfer。只是现在hash matching里面把第一和第二部分混起来了。

还得有一个额外考虑：现在基于的假设都是transfer必须有一方是fb,bank或itbit。未来如果这个假设不成立了，我也得要做好扩展的准备。

第一步：extractor这边最好不要基于这个假设，就是普通的拿出所有unmatched。

第二步：search部分也不该动。create transfer也不该动。只需要动criteria/validation。所以是否第一部分和第二部分可以串成一个meta class？

注：hash相同不也是一个criteria么？很合理啊。主要问题是，如果对hash相同的也用double for loop，性能就下降太多了。要不把search也变成可以变动的？毕竟一个O(n)一个O(n^2)。

注：如此一来，之前在create transfer里面的那条assert就可以放到criteria class里面了。

第三步：肯定用strategy pattern，是否用factory或其他？

还有一个问题。我们有好几个check，是不是可以放在专门的check class里面？还要确定的是，此check非彼matching里面的check，前者更应该叫assertion。

答：先不管这个。

针对in transit matching，不论是把in transit transfer放在哪一天，都需要一个对csv进行update的封装函数。

在一个class里面用它自己作为typing

<https://stackoverflow.com/a/33533514>

为什么self不能绑定instance名字？

<https://unix.stackexchange.com/a/6985>

如果想数一数grep返回了多少结果。

<https://stackoverflow.com/a/1537226>

Text

Description automatically generated

这个评论讲的对的。那么如果是instance variable，我当然是在init里面初始化是更合理的。但问题是我不想在subclass里面要多加一步call super().\_\_init\_\_()，应该怎么做？

<https://stackoverflow.com/questions/19029634/how-to-make-child-class-call-parent-class-init-automatically>

看起来好像偷不了这个懒。

<https://stackoverflow.com/questions/22522929/return-none-when-attribute-does-not-exist>

或者设置python，让它对不存在的attribute默认返回None。

注：明天可以看看python语法。

接下来主要就是一个build的design了。

对于cross date matching，我可以要求最严格的标准。

先看看是不是大多数cross date都是用hash match的？是的。我可以先只match这一步。

以及我可以要求bank不进行cross date match。

不知道还有什么比copy.deepcopy更优雅的处理set赋值的问题

能不能check一下整个service有什么危险的ref instead of copy的地方。

注：有些地方我却实际上是要利用这个pass by ref，比如整个matching中更新trx\_id\_set

partition返回的是两个list，这是很合理的，但我希望返回两个object。

怎么优雅明晰地把一个单元素list转变为单元素。singleton

lambda能不能像decorator一样叠加nesting？

Ans：这个是不是函数式编程的写法？

我现在的partition无非这几个需求：按src来（bank, exchange），qty，是否matched。

Ans：这个待check，不是first priority。

fuzzy matching的出处

https://zh.wikipedia.org/wiki/字符串近似匹配

问题：git怎么比较两个版本之间的差别

这个关于partitioner的问题其实在于，如果一个class在两个level上的classes里面都被用到，应该怎么design

注：我实际上是否需要是一码事，这个技术问题要解决也是一码事。

我想到的一个解释是，做成module，这样不需要在class里面传来传去。

现在看起来真的需要一个Builder class，不然每次改起来实在太麻烦了。

如果它是空的时候，我依然希望能给它一个值（哨兵值），并且能和任意值比较。

注：要不我overload comparison？

hmm说起来只要overload \_\_lt\_\_一个就行，因为据说sort就只用这一个比较。虽然这其实不合理，因为这样定义的\_\_it\_\_并不满足自反性

python是否能做到说，我把一个abstract method继续传递（继承）下去？

以及我对一个class和它的component都abstract东西出来，是否合理？

能否想一个办法解决naming的问题？

情况总是在input那边有一个名字，我这边有一个名字，我的名字就是账号的名字。

还有一个0x的问题，总是要解决。

怎么解决external monitor上面屏幕切换的事情？目前看来还是传统方法。

投影。在setting到arrangement里面去选，如果不选mirror的话，那就是两个窗口拖来拖去。

<https://www.gechic.com/cn/how-to-set-up-dual-monitors-on-macbook-os/>

有没有办法让一个屏幕黑下去？

怎么设置椅子高度？solved。

ftp wildcard。用get不行，得用mget？

<https://www.smartfile.com/blog/the-ultimate-ftp-commands-list/>

目前crypto\_journal里面我是把所有的data都储存在db里面。

是不是可以考虑用msg queue？一个优势是快，另一个优势是不需要清理，在消费完之后就会自动消失。Ans：no，因为现在它还只是一个批处理的script，而不是real time service。

容错其实是crypto\_journal这样的项目很需要考虑的问题。只要是经常可能发生错误的script，都应该有容错的考虑。

语言层面考虑。一个函数返回一个Bool再加一个值，看上去很糟糕，为什么不能直接返回一个值？不知道date里面有没有类似于0的值？是不是就是None，这就是通用的False？

答：好像不是，None和False还是两回事。<https://www.w3schools.com/python/ref_keyword_none.asp>

有时还用…或者None作为哨兵值（比如coroutine）。或者在函数原址修改时让这个函数返回None，来提醒user这是原址修改（比如sorted）。

猜测：None应该是有学理上的意义的。

看起来如果有值的话返回值，没值的话返回None，这个做法还是很常用的管理，比如re.match就是这样子的。

(True or False)返回的还是Bool而不是tuple，(True or False,)返回的就是tuple

typing里面List[str]需不需要写成List[Optional[str]]

<https://stackoverflow.com/questions/56801431/how-to-annotate-empty-list-on-python>

<https://stackoverflow.com/questions/45264194/what-is-the-correct-type-hint-for-an-empty-list>

看起来pep484说应该是前者，虽然我觉得不是很好，这有点误导人。不过它给的一个理由make sense，[]并不等于[None]。所以说是不是应该有一个类似null的东西？idk，这个和上面关于None和False的讨论也是一致的。

<https://realpython.com/null-in-python/>

还真有这一章。在很多语言里面，null实际上就等于0。但在Python里面，如果check type(None)得到的是<class 'NoneType'>。所以None是一个Object，它是一个built in变量。None还是一个singleton，即世界上只会有一个None。

Graphical user interface, text, application, email

Description automatically generated

这个trick很有趣，或许我当初应该这么建立哨兵值，这样也保证了它是一个单例。

还有教你怎么创建singleton的帖子

<https://stackoverflow.com/questions/6760685/creating-a-singleton-in-python>

所以cme\_market\_share\_view实际上是来自cme\_volumes和cme\_products。所以我要在products和cme\_products里面加上这个东西，还要修改wrap\_products。是对的。看来只是说明materialized view么有更新。

我是想用形式逻辑来证明确定性，根据我的最新的design。

你要auto config是不是跳过No break那一天的email好像有点难，在命令行加一个flag是最简单的。而且flag pass进来的距离也不是很远。

还有个问题，email现在还是一个临时储存文件的方式，这不好，还是要持久化，所以还是要储存在local。

另外我还准备给持久化的数据加一个ts in the directory，这样它就不会被覆盖。显然我是把一次run的东西视为一个batch，所以我认为它们应该在一个folder下面？还是说可以按每一天自己来？后者似乎更合理，符合我的设计原则。但问题在于，我还是得让transfer和breaks放在一起看才有意义？同时我要有ts让自己好找。

但如果我加了ts或随机数，后面的读取（马上的email以及再往后的生成journal）也是个问题啊。答：这个问题已经通过单独做一个current data system solved了

问：怎么处理ts？如果是确实是希望每次有一个新的time，怎么样能够让读的人一眼就看出来？就比如我只看Loader，我完全不知道它会被call几次，是每个task里面只被call一次还是可能多次？

what about一个master folder以及一个archive folder？也就是说，我在背后多了一步工作，来把历史数据archive/move到合适的位置。

dump\_break和dump\_transfer是否应该在一起做？现在它是分开来的。现在我有两个数据节点，一个是db，一个是files。其中files是最全的。但其实很奇怪，前一个pipe写入的是db table，但archive的却是files。也就是说，在db和master file之间是主从复制关系，而archive files是备份数据关系。

之所以要有一层current，其实是为了给user做修改。

语法问题。怎么用一个list中的变量来初始化一个object？

或许把ReplicationTask作为observer？而不是下一个pipe。所以这两者区别是啥？是不是一个有顺序一个没顺序？但我觉得还是得让observer有一个顺序，除非真的没有dependency，毕竟我也不是真的并行在跑，还是顺序在跑。ideally最后我希望让每一个组件都变成一个个pipe或者task，而且可以很容易地添加或移除，而数据接口是交给持久化数据系统（由data source加（potential）replicas service）来完成的，而不是直接由一个task递给下一个。

那么这种显然在针对要添加新功能时非常extensible（而且要重跑某一段task也很容易；要scale某一个pipe也很容易），但是否也能handle数据分割问题properly？

所以是不是考虑etl+observer而不是et+observer？ETL加observer也可以递归嵌套。

而且这也的话，我可以把整个项目的pipes连起来，而不用单独建一个script来call，而且中断或者加其他任务也很容易。

现在所有的task假设都是基于最小单位一天来Build的。

这个太头疼了，我尽管是每天跑，但并不是说一天跑完了就是match完了。

还有一个db和file的区别是，我的file是不会修改的（至少historical里面是不会的），但db里面是会的。那这个实际上就涉及到数据同步的问题。一种解决方法是，只要db修改了，我就也修改file的数据，但这个不是很好，也做不太到。

问题是：有没有办法能够在只重跑一天数据的情况下，不影响其他天的cross date matching。这个似乎是不可能的，因为现在有一个cross date tx\_hash based matching，而且未来可能还有其它标准下的cross date matching。我们可以思考，如果说数据修改，不管是做了增加还是修改还是减少，都可能会影响其他日子的matching，而且这种影响还可能因为matching的逻辑做了修改而发生修改，我们不能对此做很强的假设。要不就放弃这个想法吧？再思考一下，能只重跑一天，除了performance上的可能优势，还有什么其他优势么？好像没有？也不对，也有，当希望以小于一个月为粒度来处理某些事务时就有了，现在cross date的transfer对于db清理和添加都是个麻烦（想想每次遇到的duplicate ids in transfer table）。

先解决一个简单点的问题，一个合理的要求是，我希望把cross date的matching的结果放在一个合理的地方（这个要求很make sense吧），同时如果是同一天的fuzzy matching，我希望日期就是那一天。因此，对creator，我希望自己是可以configure date的。目前creator, validator, partitioner都是pass给matcher的。

让执行顺序变成daily matching（甚至可以并行），cross date matching，data replica。这些task之间可以用observer连起来。这样的设计把data replica和matching分开来，哪怕要多一个et的过程。是否合理（我觉得合理，因为它是不一样的任务）？再看一看。这个想法的假设是，把可以分开重跑的部分，和不能分开重跑的部分区分开来，哪怕对于用户来说没什么区别，因为他要看到的是最后的结果，但对于软件维护来说还是大有帮助。对cross date，似乎有理由把transfer date定为前一天。

上述也就是说，增量修改而不是全局重跑（意味着我可以通过check input file ts来看），把可以单独跑的东西分离出来，而且可以并行化；把不能单独跑的、每次都一定要修改的东西也分离出来。

注：所以我现在要修改的其实只有一个trans dt的选择问题，以及把replica步骤单独拿出来（可能新建一个repo）。然后还有就是把整个service弄成task based，组装task（pipe）的地方可以经常变，但task是在另外的地方。

之所以有目前的cross date问题，一种可能也是data provider的ts并不准确，这也是分布式系统的问题。有趣，所以谷歌提供的全局时间就是一个时间的置信区间而不是一个单点值。

我们现在做到了按月跑，或累计跑（增量），都没问题。意思是说，现在如果一个月里面某一天数据出问题了希望重跑，我们必须要重跑从月初到上次跑到过的end date的所有数据。

注：主从复制不就是有持久化数据要解决的问题么？即“读自己的写”。这是“复制滞后问题”里面的一个。

注2：前端和后端要选择就是同步还是异步复制。

之所以一开始升级package不行，是因为要把环境删了重新装，只重新运行inv bootstrap那些还不够。

原来git也是分布式的，怪不得我看到reuse觉得这么熟悉