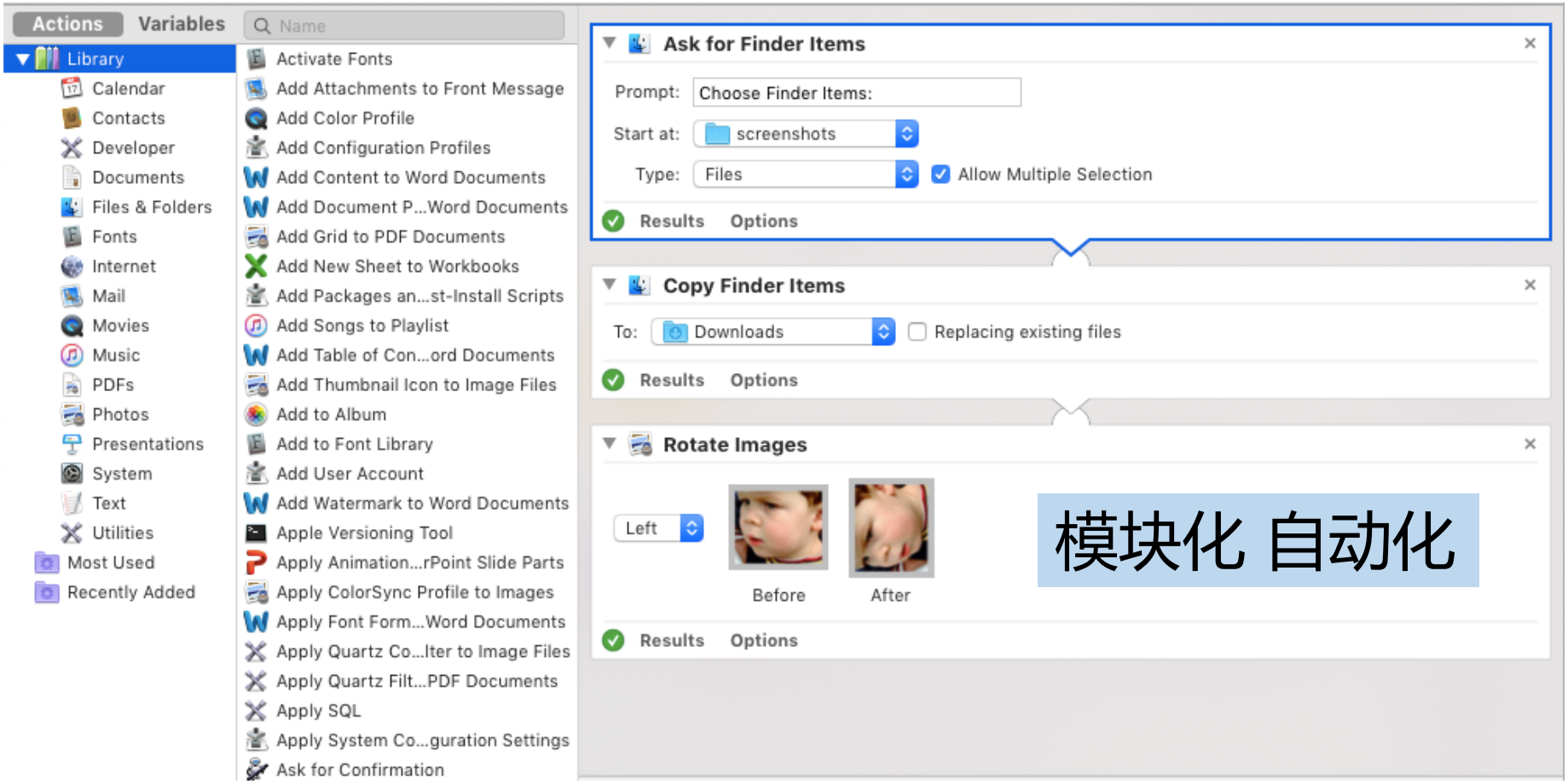
工作流教程

熊景放

什么是工作流

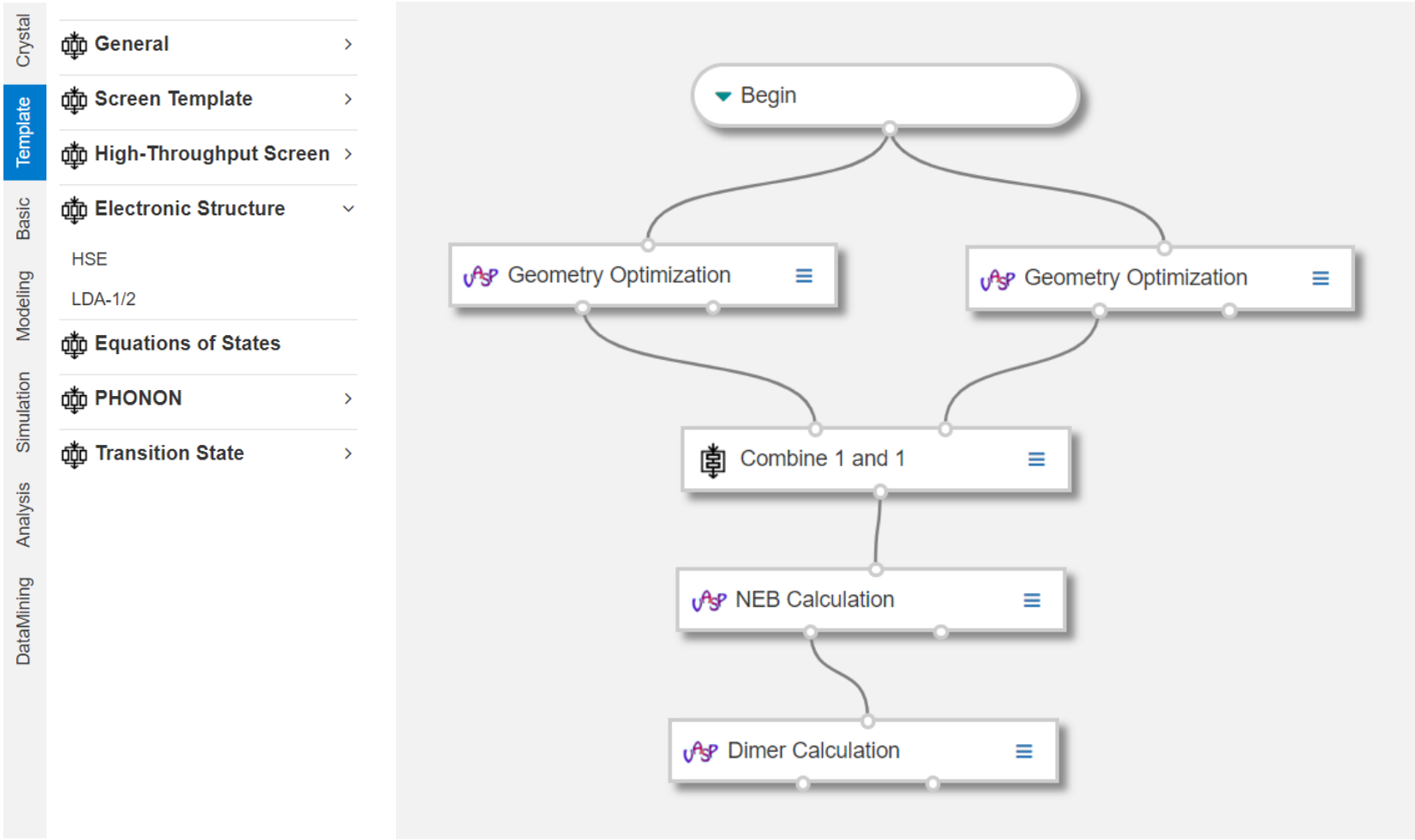

Actions
Variables
Name
Library
Calendar
Contacts
Developer
Documents
Files & Folders
Fonts
Internet
Mail
Movies
Music
PDFs
Photos
Presentations
System
Text
Utilities
Most Used
Recently Added
Activate Fonts
Add Attachments to Front Message
Add Color Profile
Add Configuration Profiles
Add Content to Word Documents
Add Document P...Word Documents
Add Grid to PDF Documents
Add New Sheet to Workbooks
Add Packages an...st-Install Scripts
Add Songs to Playlist
Add Table of Con...ord Documents
Add Thumbnail Icon to Image Files
Add to Album
Add to Font Library
Add User Account
Add Watermark to Word Documents
Apple Versioning Tool
Apply Animation...rPoint Slide Parts
Apply ColorSync Profile to Images
Apply Font Form...Word Documents
Apply Quartz Co...lter to Image Files
Apply Quartz Filt...PDF Documents
Apply SQL
Apply System Co...guration Settings
Ask for Confirmation
Ask for Finder Items
Prompt:
Choose Finder Items:
Start at:
screenshots
Type:
Files
Allow Multiple Selection
Results
Options
Copy Finder Items
To:
Downloads
Replacing existing files
Results
Options
Rotate Images
Left
Before
After
Results
Options
模块化 自动化

MatCloud+

中国科学院计算机网络信息中心：杨小渝

http://www.matcloudplus.com/

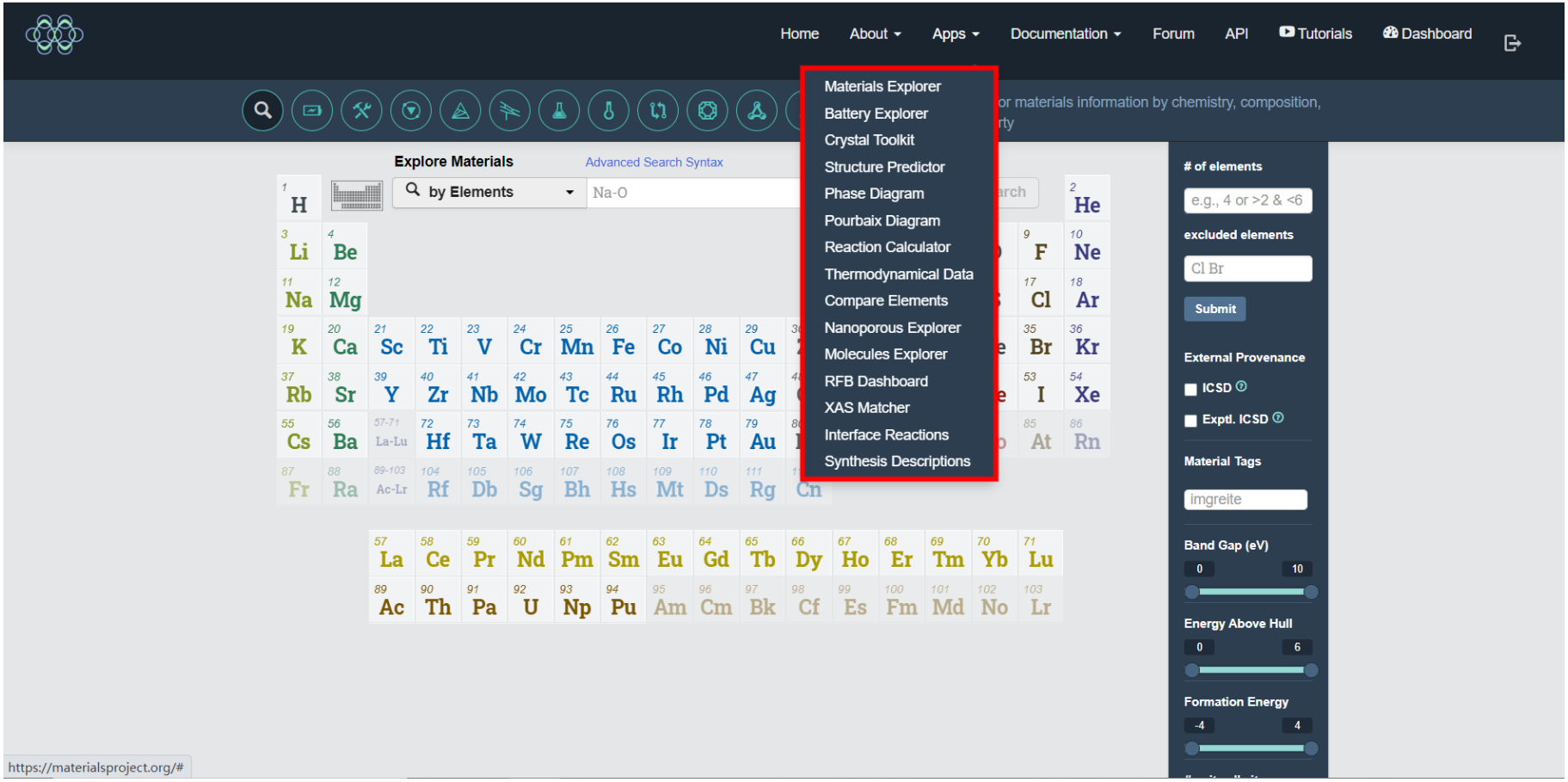
迈高科技
www.matcloud.com.cn



Materials Project

UCB: Kristin Persson

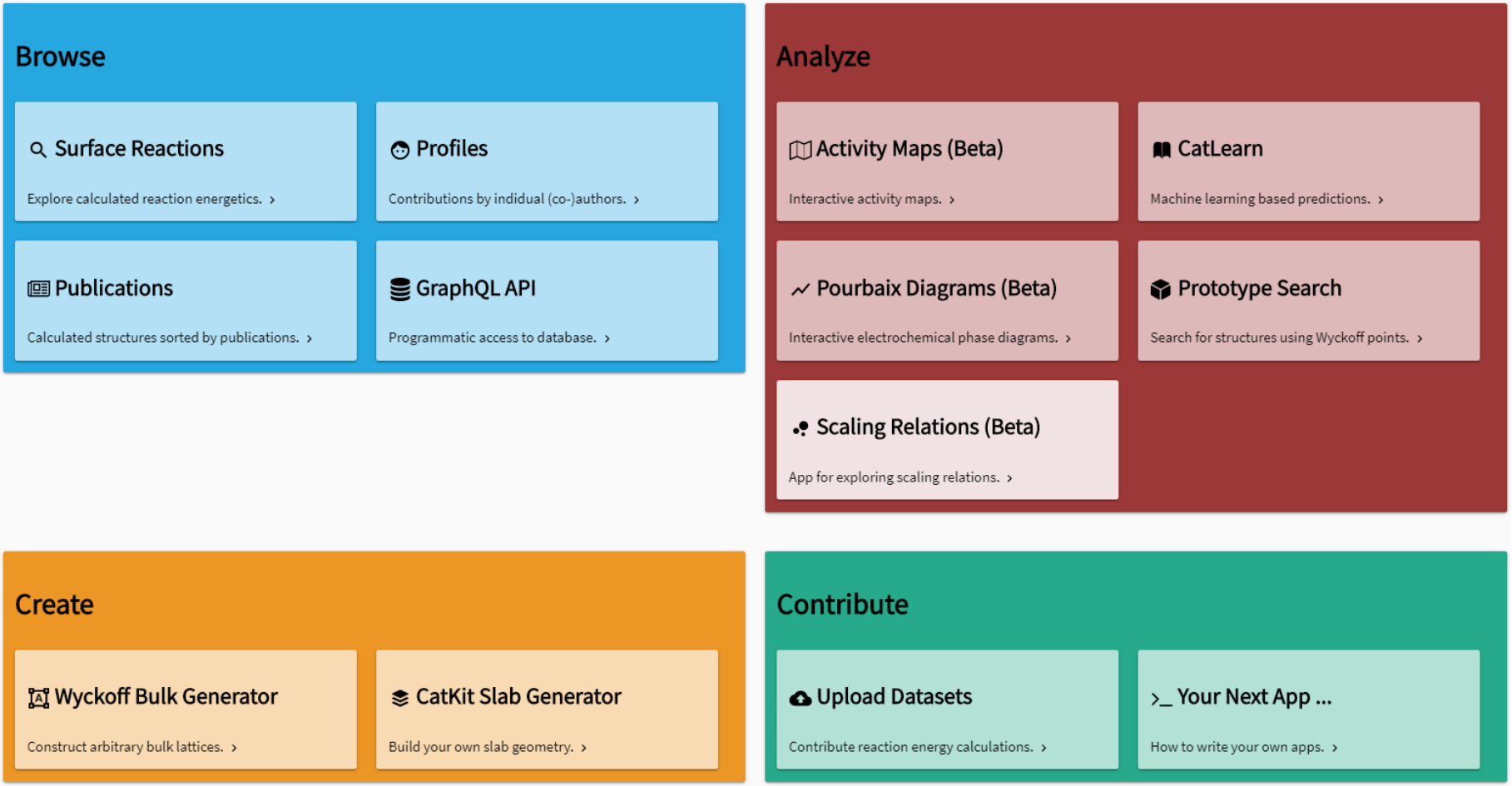
https://materialsproject.org/



Catalysis-Hub

Stanford: Thomas Bligaard

https://www.catalysis-hub.org/appsIndex



工作流平台对比

迈高科技
www.matcloud.com.cn

专注于模块化

GUI 注重创建工作流

降低开发门槛

闭源

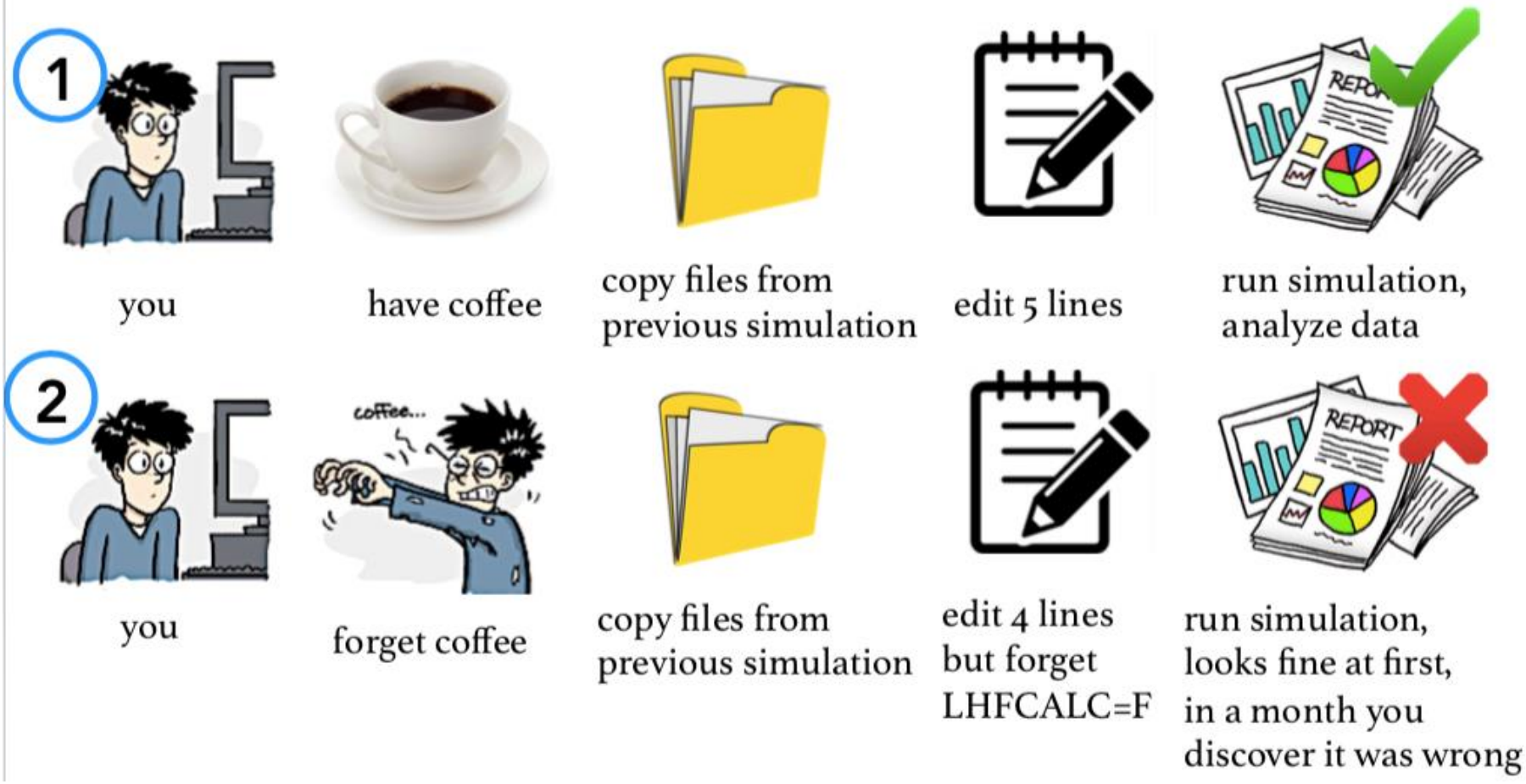
专注于自动化

GUI 注重引导用户使用工作流，创建工作流时一般通过代码创建

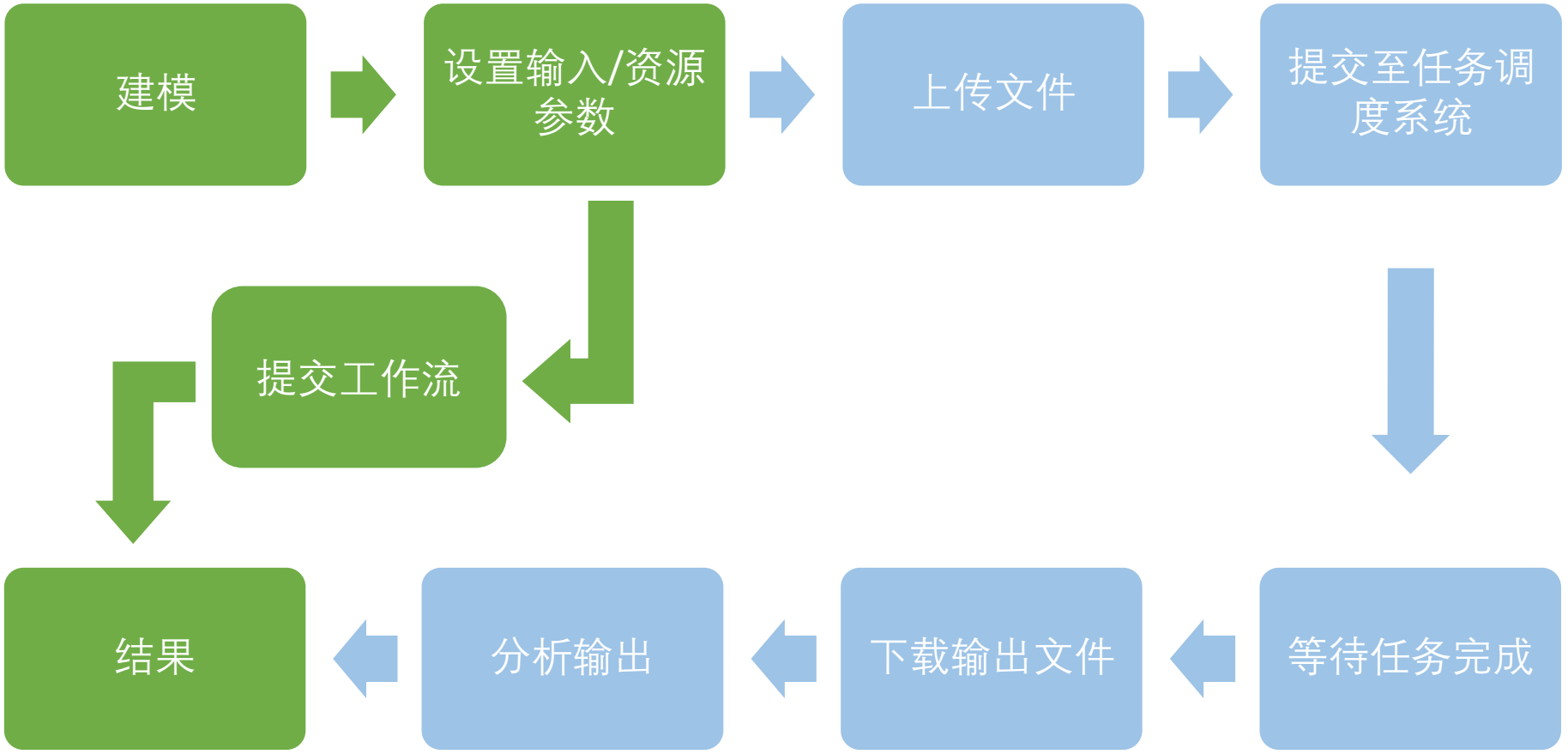
降低使用门槛

开源

使用工作流的好处


1
you
have coffee
copy files from previous simulation
edit 5 lines
REPORT
run simulation, analyze data
2
you
coffee...
forget coffee
copy files from previous simulation
edit 4 lines but forget LHFCALC=F
REPORT
run simulation, looks fine at first, in a month you discover it was wrong

对于计算流程的精简



输入文件的对比

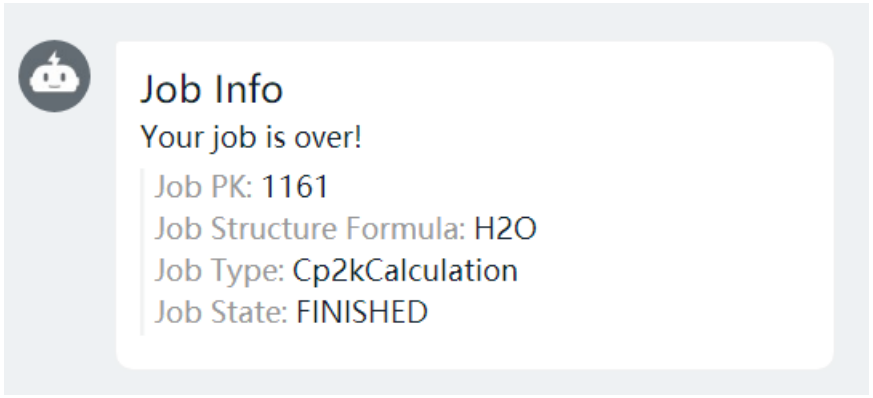
```yaml
workflow: NebSingleWorkChain
webhook: https://oapi.dingtalk.com/
resdir: results_yaml
structure:
  - is.xyz
  - fs.xyz
cell:
  - [10.239521, 0, 0]
  - [0, 11.352600, 0]
  - [0, 0, 27.891432]
metadata:
  kind_section:
    BASIS_SET: DZVP-MOLOPT-SR-GTH
    POTENTIAL: GTH-PBE
```

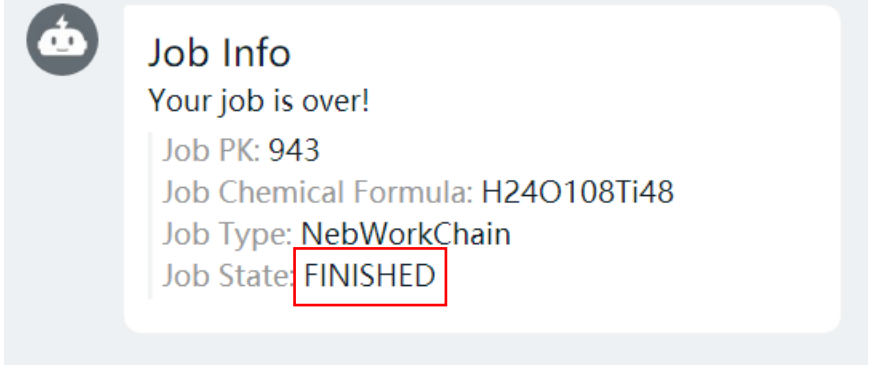
VS

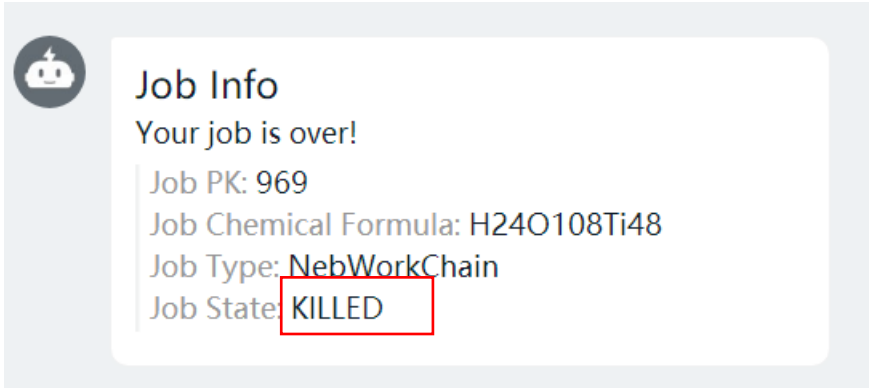
```
&FORCE_EVAL
  METHOD Quickstep
  &DFT
    BASIS_SET_FILE_NAME BASIS_MOLOPT
    POTENTIAL_FILE_NAME GTH_POTENTIALS
    MULTIPLICITY 1
    &MGRID
      CUTOFF 400
    &END MGRID
    &QS
      EPS_DEFAULT 1.0E-13
      EXTRAPOLATION ASPC
      EXTRAPOLATION_ORDER 2
    &END QS
    &SCF
      SCF_GUESS RESTART
      EPS_SCF 3.0E-7
      MAX_SCF 50
      &OUTER_SCF
         EPS_SCF 3.0E-7
         MAX_SCF 10
      &END
     &OT
         MINIMIZER DIIS
         PRECONDITIONER FULL_SINGLE_INVERSE
         ENERGY_GAP 0.1
      &END
    &END SCF
    &XC
      &XC_FUNCTIONAL PBE
      &END XC_FUNCTIONAL
       &vdW_POTENTIAL
         DISPERSION_FUNCTIONAL PAIR_POTENTIAL
         &PAIR_POTENTIAL
            TYPE DFTD3
            PARAMETER_FILE_NAME dftd3.dat
            REFERENCE_FUNCTIONAL PBE
         &END PAIR_POTENTIAL
      &END vdW_POTENTIAL
    &END XC
  &END DFT
```

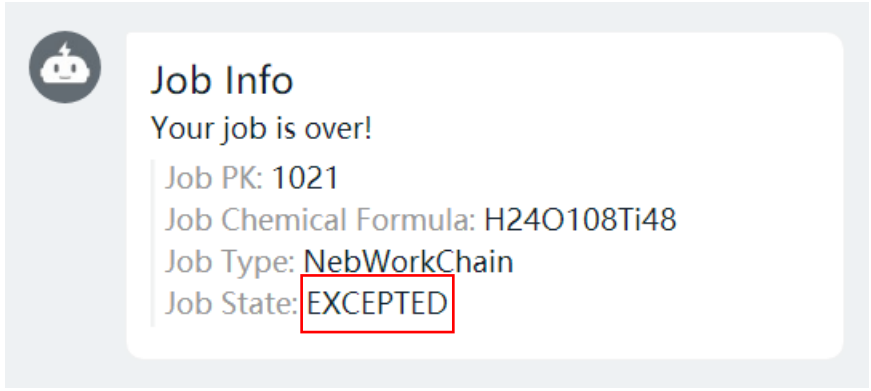
.....

推送


Job Info
Your job is over!
Job PK: 1161
Job Structure Formula: H2O
Job Type: Cp2kCalculation
Job State: FINISHED

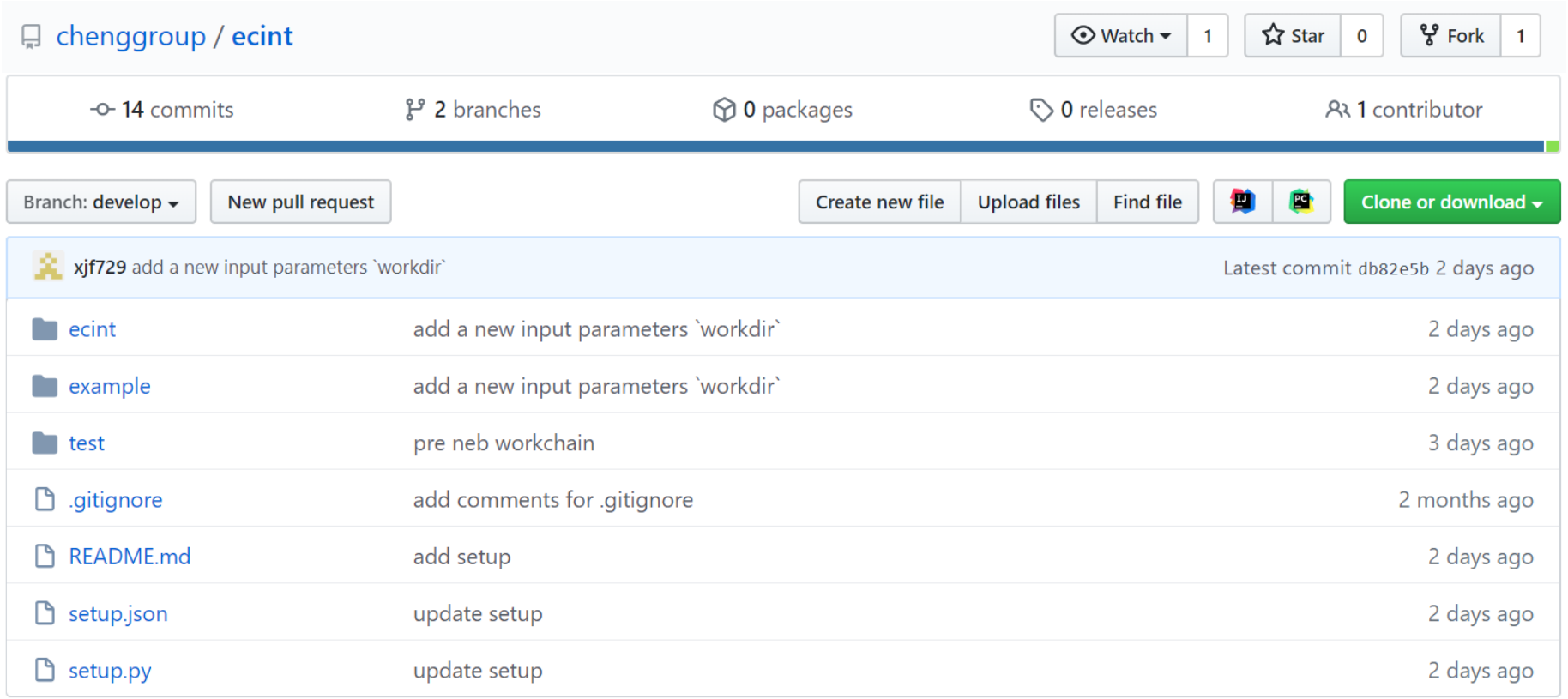

Job Info
Your job is over!
Job PK: 943
Job Chemical Formula: H24O108Ti48
Job Type: NebWorkChain
Job State: FINISHED


Job Info
Your job is over!
Job PK: 969
Job Chemical Formula: H24O108Ti48
Job Type: NebWorkChain
Job State: KILLED


Job Info
Your job is over!
Job PK: 1021
Job Chemical Formula: H24O108Ti48
Job Type: NebWorkChain
Job State: EXCEPTED

ECINT

A package based on AiiDA and ASE
Developed by chenggroup



*Install this package by pip

开源工作流框架

AiiDA

Group Leader : Nicola Marzari (EPFL)

后端稳定，数据库 (PostgreSQL) 查询模式更为丰富，通过插件来对接不同的计算软件，可拓展性强，但是部署较为麻烦

有着众多第三方开发者，维护各种插件，不过也导致各种插件质量层次不齐

atomate

Group Leader: Gerbrand Ceder (UCB)

基于 pymatgen, custodian 和 FireWorks，数据库 (MongoDB) 查询性能高

对 VASP/LAMMPS 支持较好，并对其有着丰富的错误处理以及后处理工具

pyiron

Group Leader: Jörg Neugebauer (MPI)

使用 Jupyter notebooks，用户交互性好

容易部署，不需要特意部署数据库

signac

Group Leader : Sharon Glotzer (UM)

批量任务的提交以及批量数据的处理

并未提供与其他计算软件的接口，因此对于不同计算软件的接口通常需要用户自行编写，门槛较高

为何选用 AiiDA

	CP2K	VASP	QE	LAMMPS
AiiDA	✓	✓	✓	○
atomate	✗	✓	✗	✓
pyiron	✗	✓	✗	✓
signac	✗	✗	✗	✗

✗ 不支持　　✓ 相对好的支持　　○ 部分支持

对于 CP2K 工作流: AiiDA

AiiDA 工作流主要组成

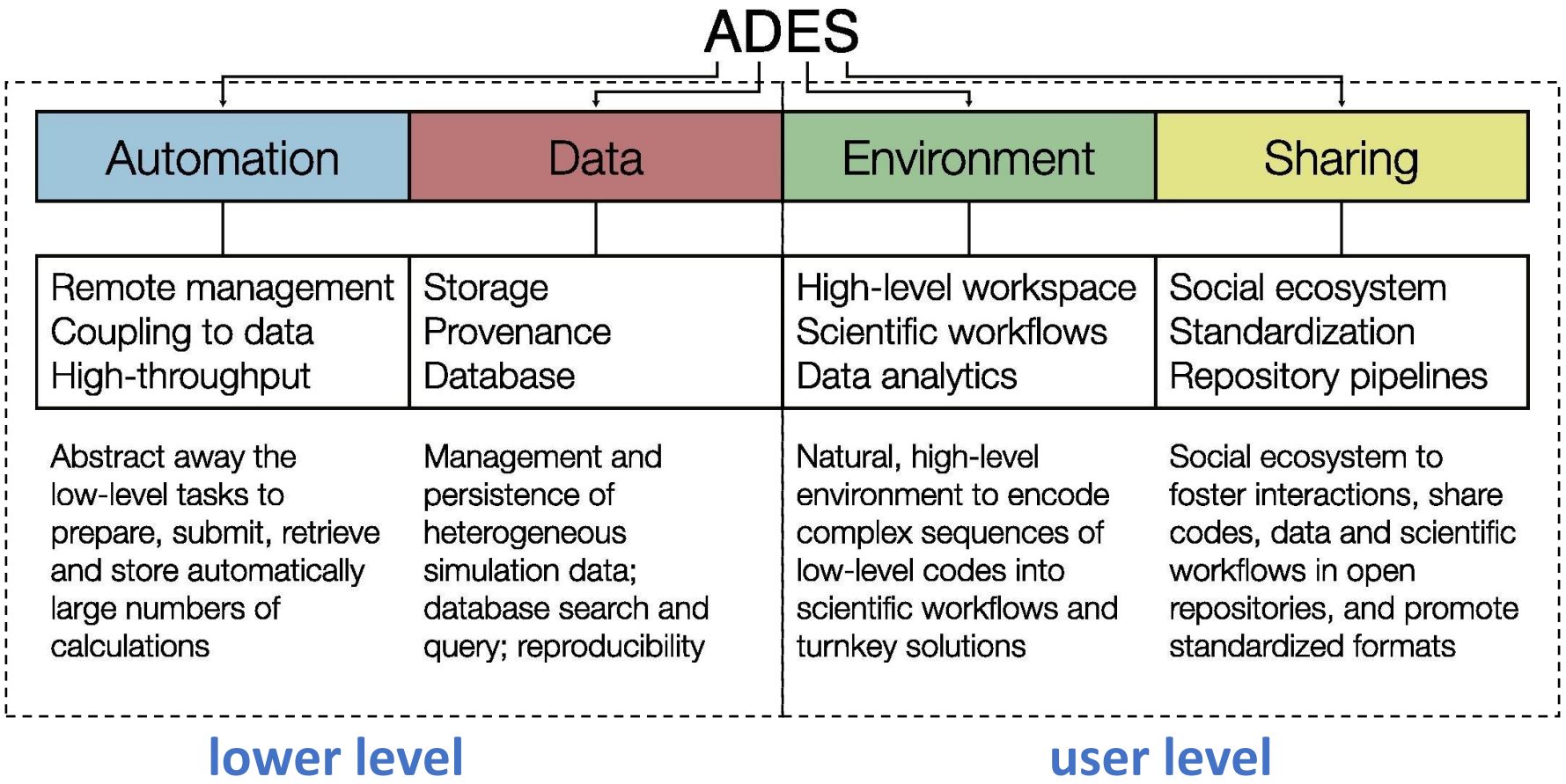
```python
class EnergyWorkChain(WorkChain):
    @classmethod
    def define(cls, spec):
        super(EnergyWorkChain, cls).define(spec)

        # Input parameters
        spec.input('input_files.structure_file', valid_type=str, non_db=True)
        spec.input('input_files.config_file', valid_type=str, default='energy.json', required=False, non_db=True)
        spec.input('input_files.kind_section_file', valid_type=str, default='DZVPBLYP', required=False, non_db=True)
        spec.input('input_files.machine_file', valid_type=str, default='machine.json', non_db=True)
        spec.input('parameters', required=False)

        # Outline
        spec.outline(
            cls.submit_workchain,
            cls.inspect_workchain,
            cls.get_result,
        )

        # Output parameters
        spec.output('output_parameters', valid_type=Dict, required=True, help='the results of the calculation')
```

Define the input parameters

Define the steps of the workflow

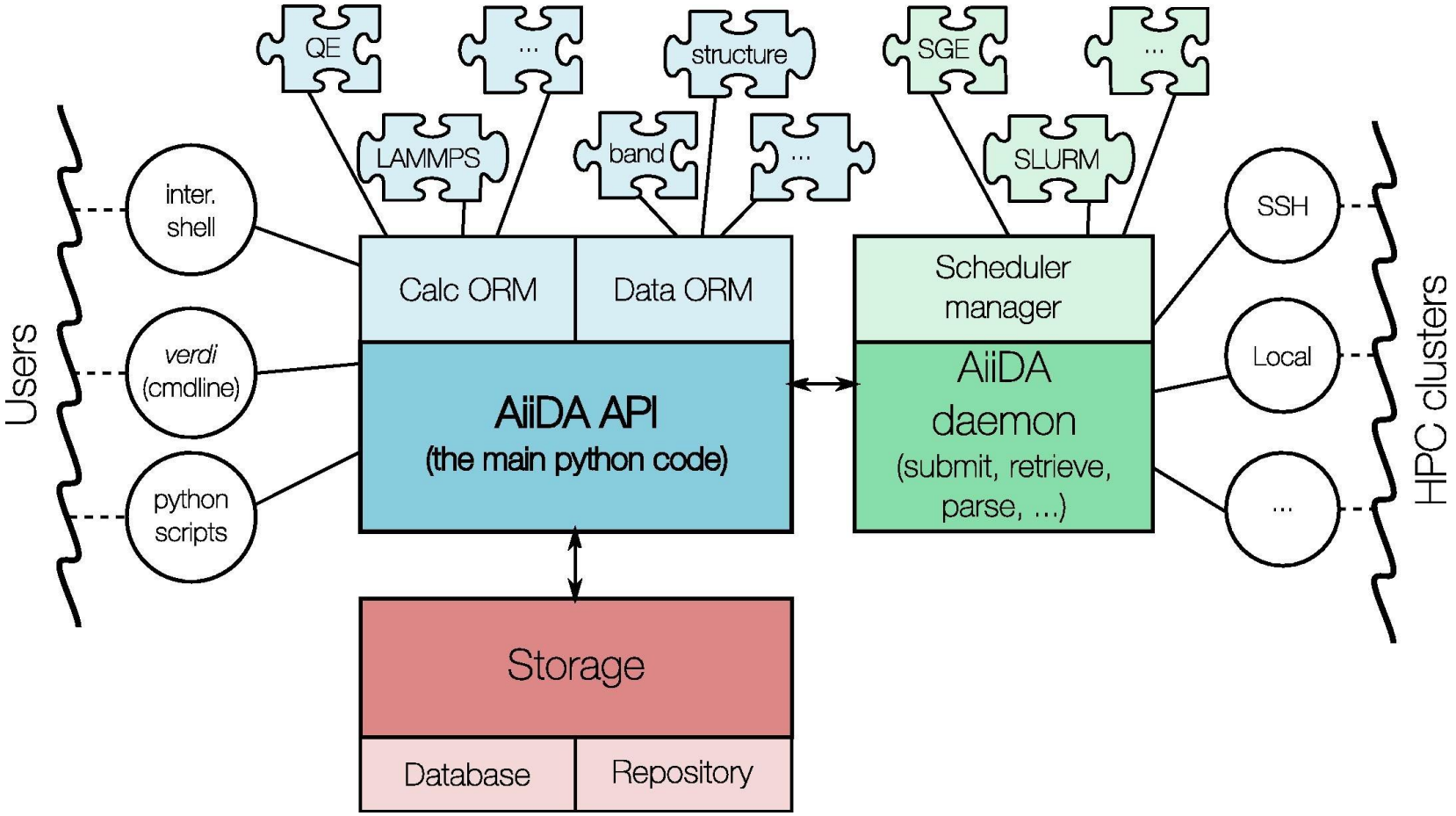
Define the output parameters

AiiDA—Framework Requirements



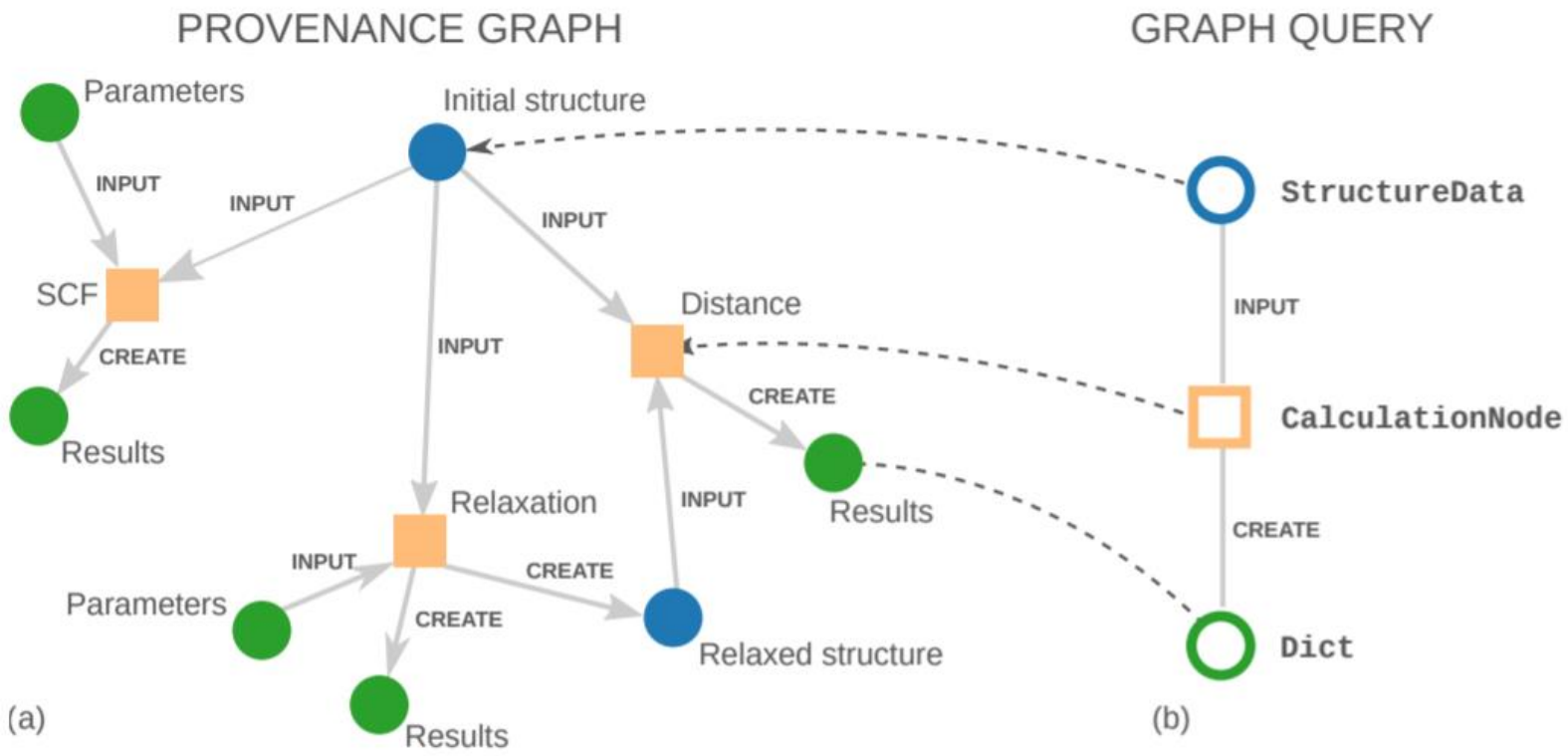
Pizzi G, Cepellotti A, Sabatini R, et al. *Computational Materials Science*, 2016, **111**: 218-230.

AiiDA—Architect



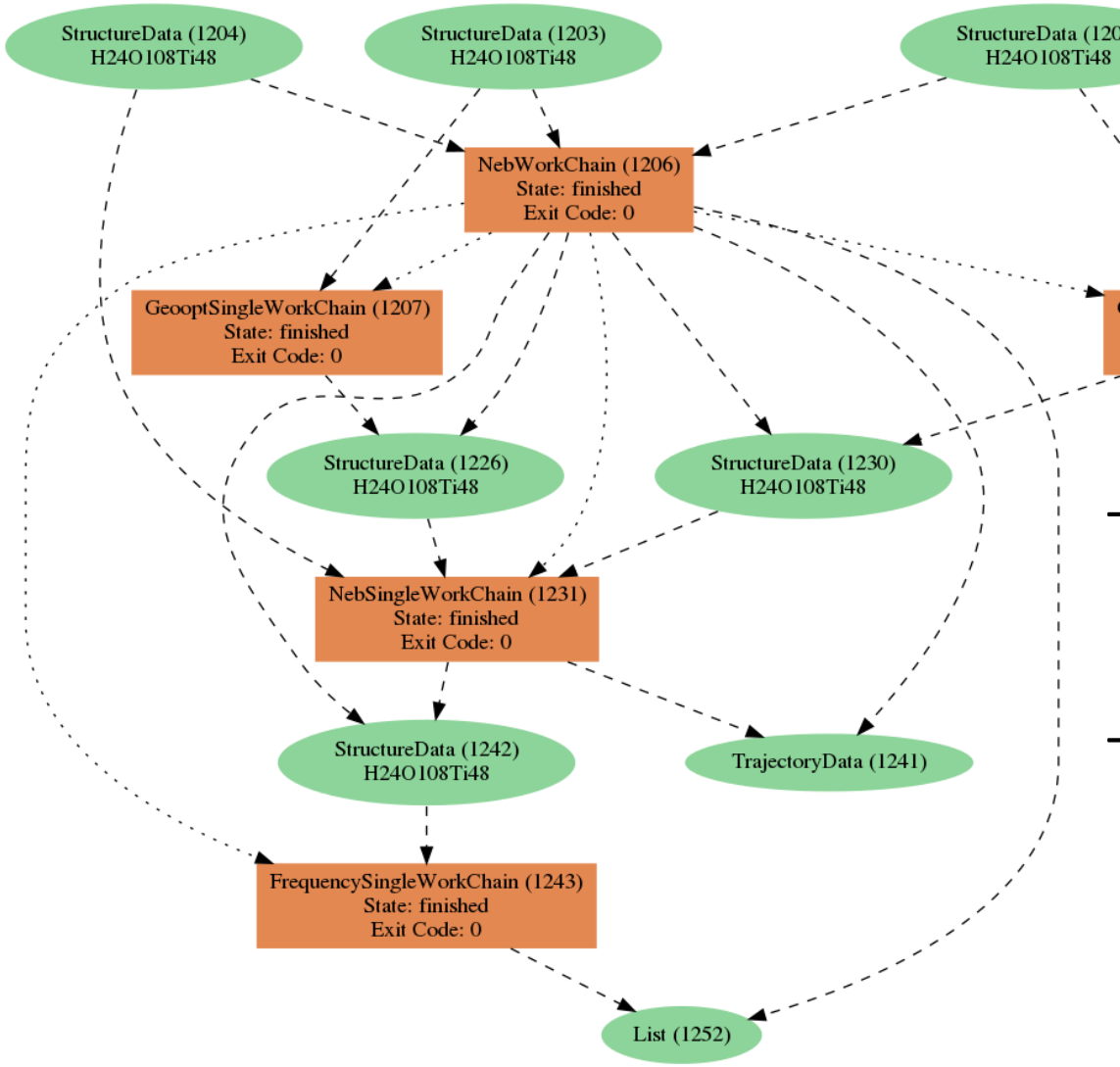
Pizzi G, Cepellotti A, Sabatini R, et al. *Computational Materials Science*, 2016, **111**: 218-230.

AiiDA—Provenance Graph



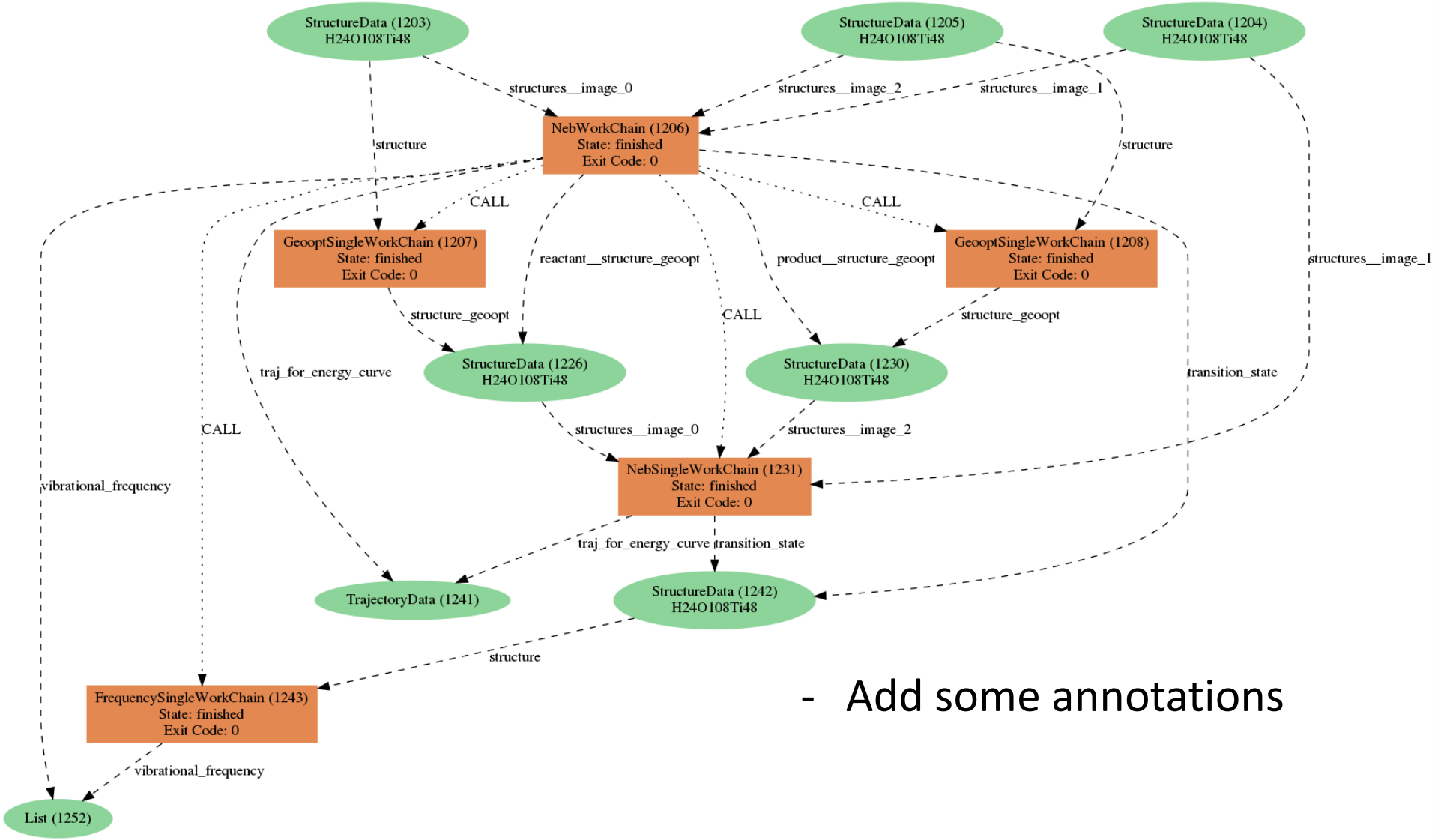
Huber S, Zoupanos S, Uhrin M, et al. *arXiv preprint arXiv:2003*.12476, 2020.

Provenance Graph—minimal



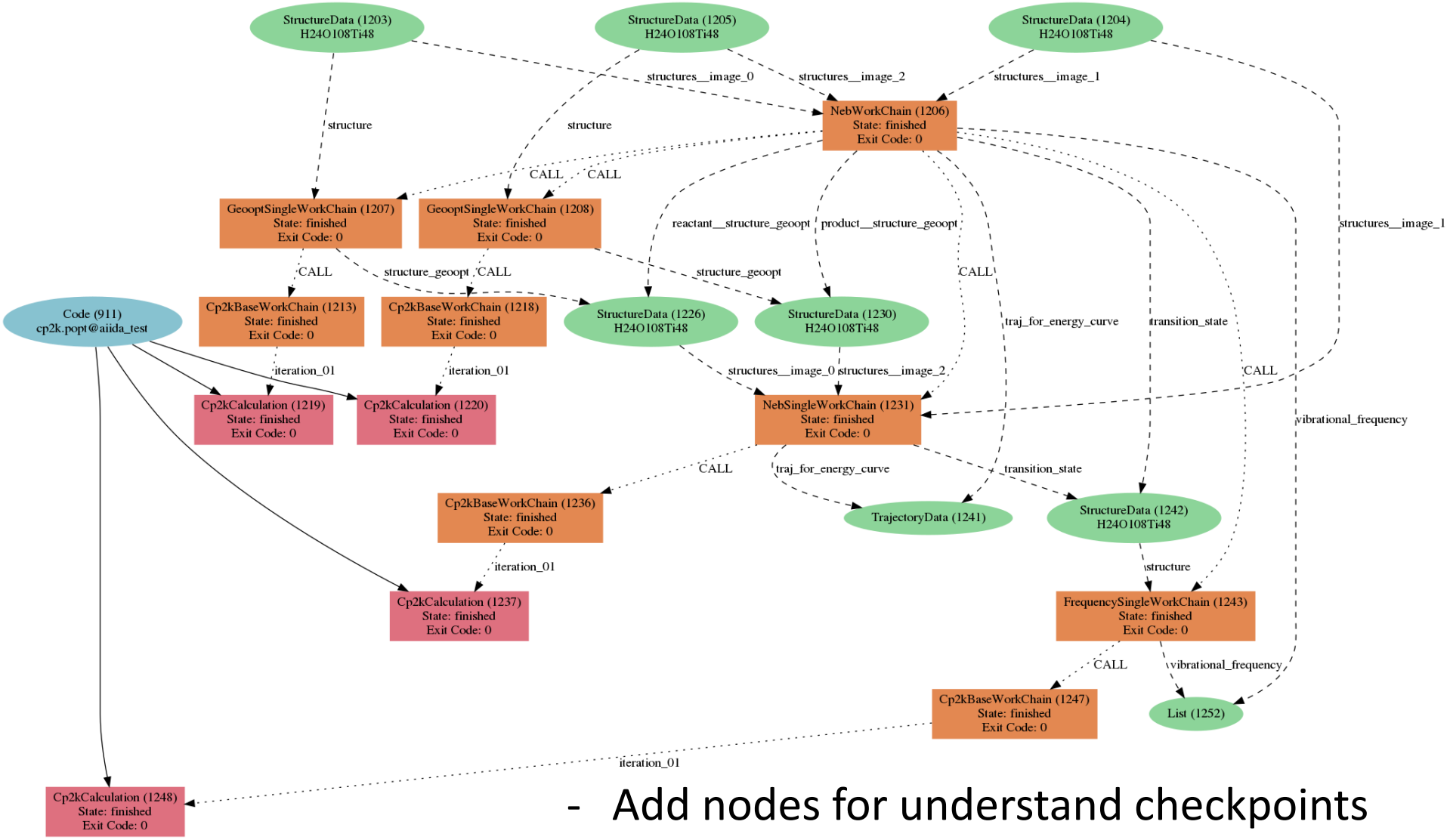
- Only main input and output
- WorkChain: core WorkChain and first called WorkChain

Provenance Graph—minimal



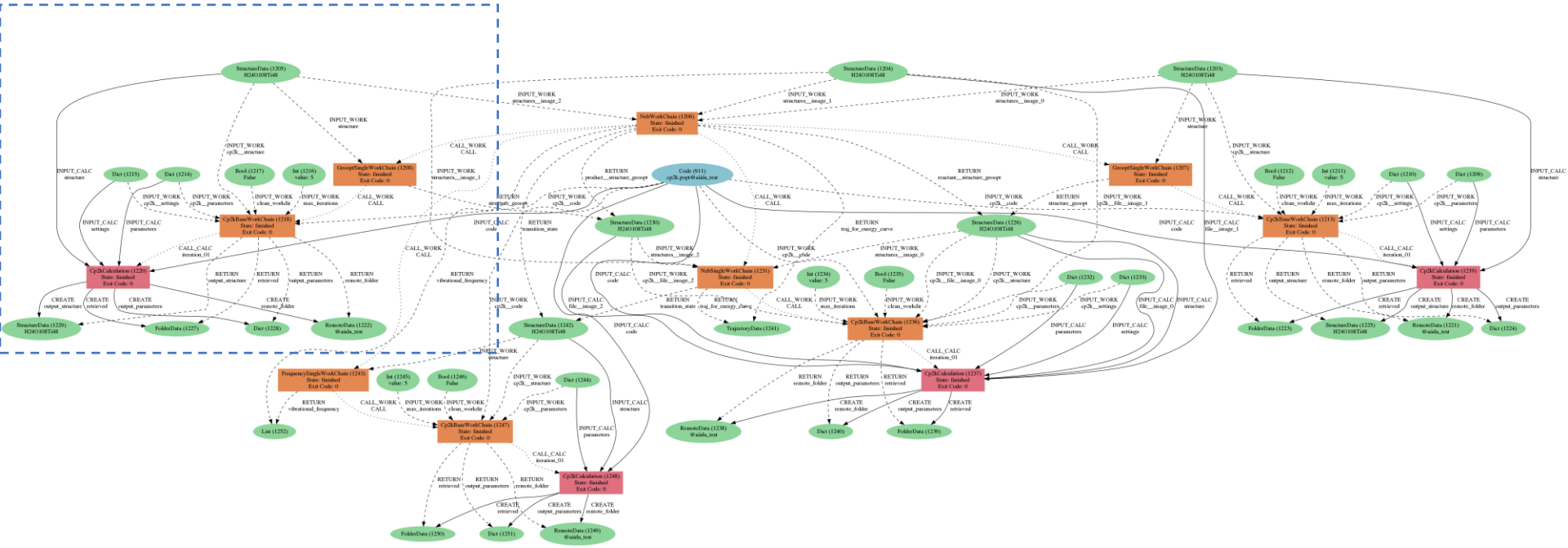
- Add some annotations

Provenance Graph—medium



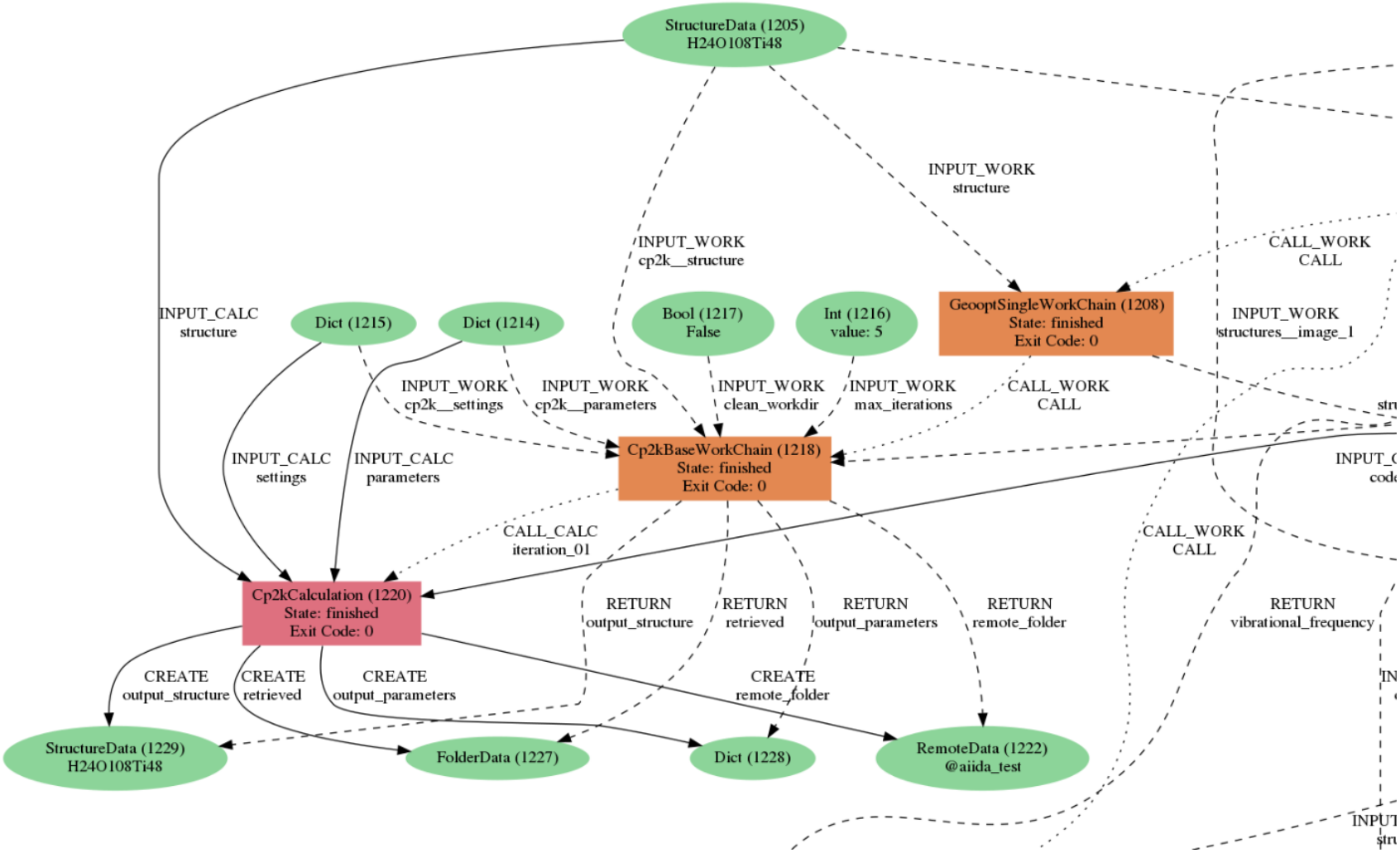
- Add nodes for understand checkpoints

Provenance Graph—high

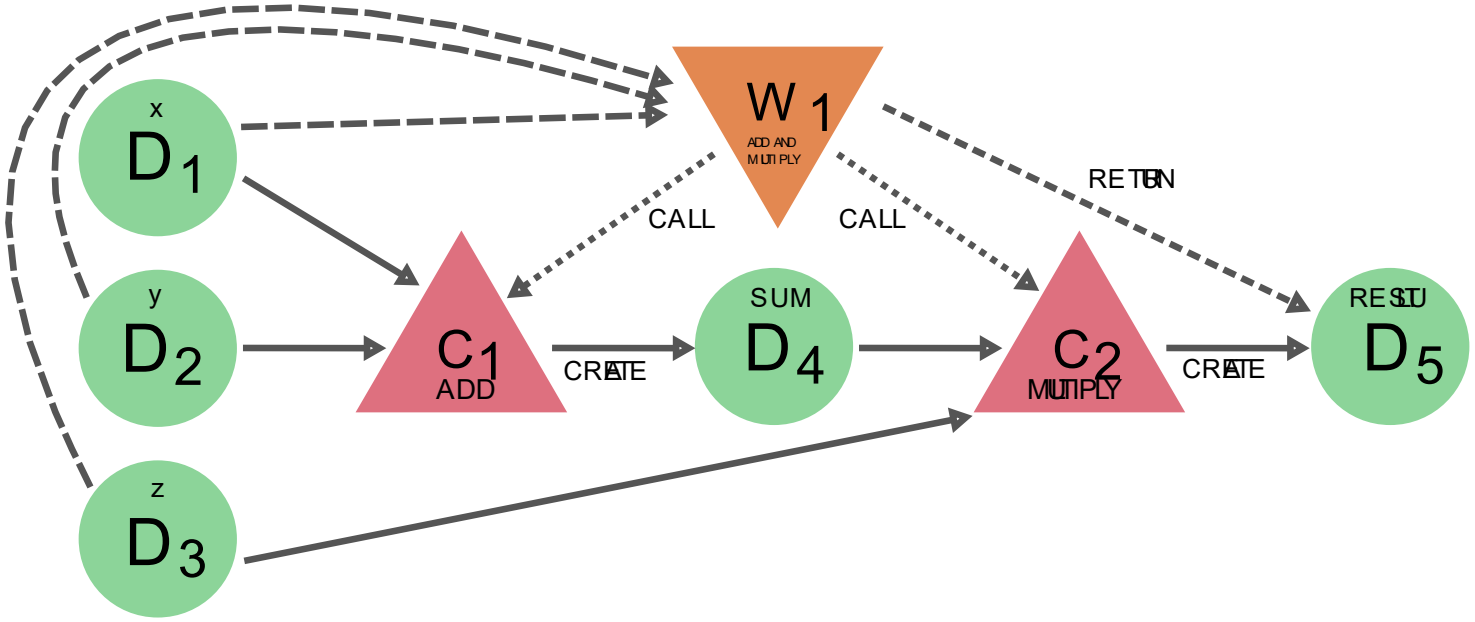


- All annotations, both “label” and “type”
- All relevant node

Provenance Graph—high(partial)

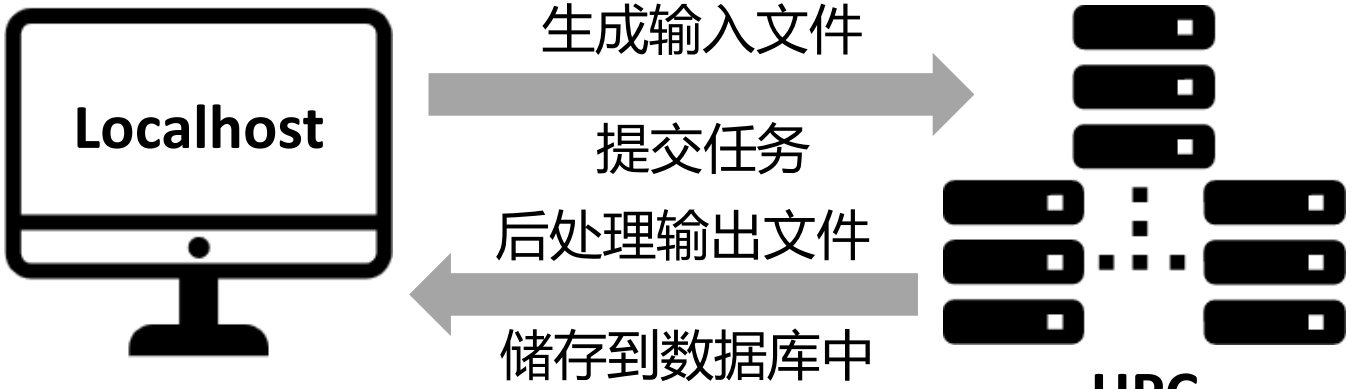


AiiDA—Checkpoints



Uhrin M, Huber S P, Yu J, et al. *arXiv preprint arXiv*:2007.10312, 2020.

部署



Host: hydrogen
User: chenglab
IP: 10.24.3.144
Port: 8099

HPC
User: aiida_test
IP: 121.192.191.52
Port: 6666

ecint
aiida-core
aiida-cp2k
…

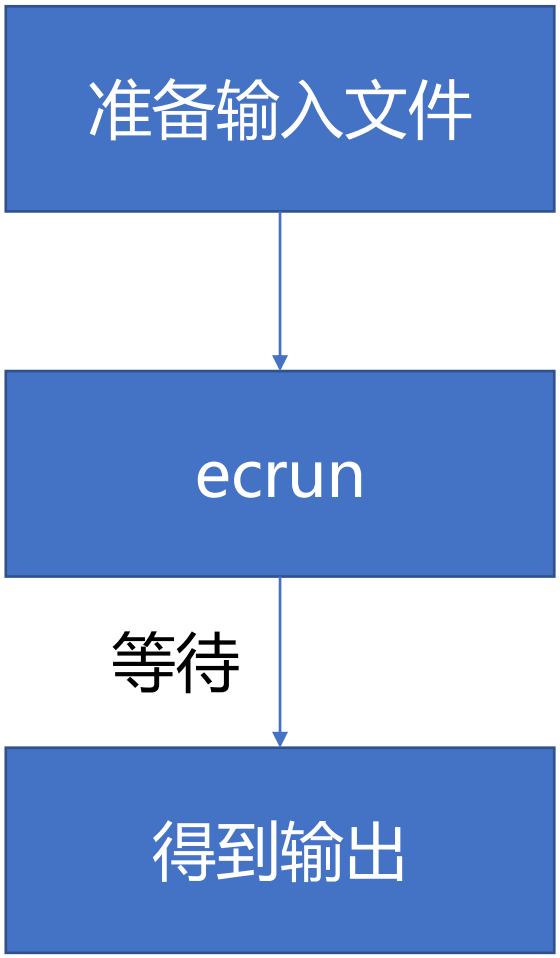
RabbitMQ:
message broker

PostgreSQL:
database

cp2k 6.1
[设置cp2k环境变量]

使用

https://chenggroup.github.io/wiki/ecint/user




```
{
  "workflow": "EnergySingleWorkChain",
  "webhook": "https://oapi.dingtalk.com/
  "resdir": "results_json",
  "structure": "Cu13CO.vasp"
}
```

```
$ ls
Cu13CO.vasp  energy.json  energy.yaml
$ ecrun energy.json
START SUBMIT...
END SUBMIT
```

```
$ cd results_json
$ ls
coords.xyz  results.dat
```

厦門大學
XIAMEN UNIVERSITY

Q&A