FireWorks and the Materials Project: Facilitating Scientific Workflows

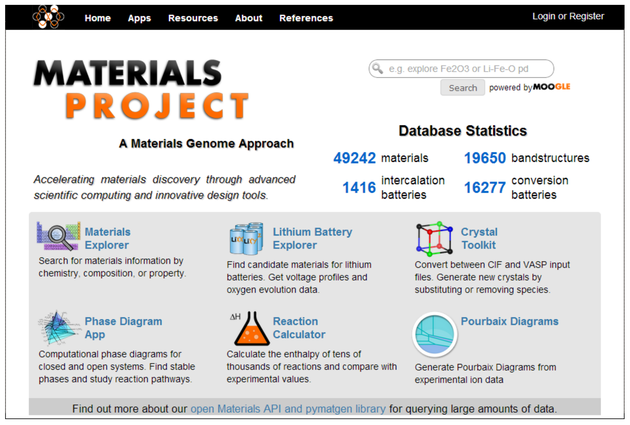
Morgan Hargrove[1, 2, 3], Dan Gunter[2, 3], Anubhav Jain[2, 4]
Louisiana State University[1], Lawrence Berkeley National Laboratory[2], Computational Research Division[3], Environmental Energy Technologies Division[4]

Introduction

As scientific research and experimentation move more towards a system of data-driven, distributed collaboration, the need for innovative new tools to facilitate these processes becomes readily apparent. In order to enable the creation of reliable, reproducible results under these conditions, scientists have begun developing scientific workflows to track the evolution of their data, results, and discoveries.

The Materials Project

Technological innovation is often enabled by materials advances. Yet, it takes an average of eighteen years to move new materials discoveries from the research laboratory to the market. The Materials Project aims to accelerate materials innovation and design by providing free access to a searchable, interactive database of computed materials properties spanning most known inorganic compounds.



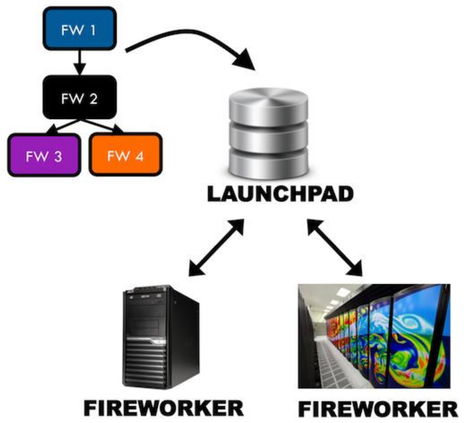
The current *materialsproject.org* release contains almost 50,000 different compounds, and additional materials are added monthly.

FireWorks

Computations for the Materials Project are currently driven by FireWorks, a high-throughput framework for defining, managing, and executing scientific workflows. FireWorks is a Python based, open-source* code that can be used to automate most types of calculations over arbitrary computing resources. FireWorks is intended to be a friendly workflow software that is easy to get started with but flexible enough to handle complicated use cases.

Main Components:

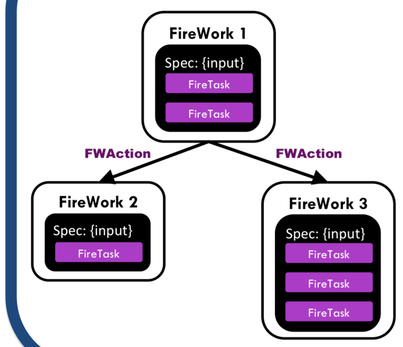
- LaunchPad (Server)
  - Manages workflows (can add, query the state of, or rerun workflows)
- FireWorkers (Workers)
  - Run jobs (request workflows from the LaunchPad, execute them, and send back information)



The basic infrastructure of FireWorks

* Documentation and installation instructions for FireWorks can be found at *http://pythonhosted.org/FireWorks*.

The FireWorks Workflow Model:



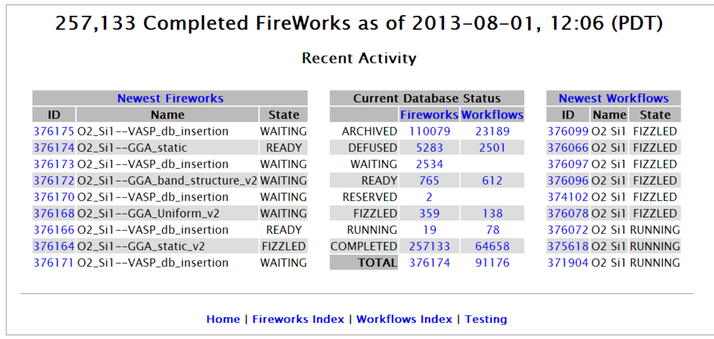
FireTask
: An atomic computing job

FireWork
: Contains all information needed to run a job (as a JSON specification: "spec")

Workflow
: A set of FireWorks with dependencies between them

BaseSite

BaseSite is a front-end user interface that serves as a performance dashboard for FireWorks. The goal of BaseSite is to facilitate the utilization of FireWorks in real-world scenarios, such as the Materials Project, by providing a simplistic, easy-to-navigate visualization of a FireWorks-based workflow system.



257,133 Completed FireWorks as of 2013-08-01, 12:06 (PDT)

Recent Activity

Newest Fireworks		
ID	Name	State
376175	O2_Si1--VASP_db_insertion	WAITING
376174	O2_Si1--GGA_static	READY
376173	O2_Si1--VASP_db_insertion	WAITING
376172	O2_Si1--GGA_band_structure_v2	WAITING
376170	O2_Si1--VASP_db_insertion	WAITING
376168	O2_Si1--GGA_Uniform_v2	WAITING
376166	O2_Si1--VASP_db_insertion	READY
376164	O2_Si1--GGA_static_v2	FIZZLED
376171	O2_Si1--VASP_db_insertion	WAITING

Current Database Status		
	Fireworks	Workflows
ARCHIVED	110079	23189
DEFUSED	5283	2501
WAITING	2534	
READY	765	612
RESERVED	2	
FIZZLED	359	138
RUNNING	19	78
COMPLETED	257133	64658
TOTAL	376174	91176

Newest Workflows		
ID	Name	State
376099	O2 Si1	FIZZLED
376066	O2 Si1	FIZZLED
376097	O2 Si1	FIZZLED
376096	O2 Si1	FIZZLED
374102	O2 Si1	FIZZLED
376078	O2 Si1	FIZZLED
376072	O2 Si1	RUNNING
375618	O2 Si1	RUNNING
371904	O2 Si1	RUNNING

Home | Fireworks Index | Workflows Index | Testing

The home page of BaseSite


```
{
    "name": "Mg1_O3_Si1--VASP_db_insertion",
    "fw_id": 354099,
    "state": "COMPLETED",
    "created_on": "2013-06-30T10:42:03.523022"
}
```


A summarized view of the spec for FireWork 354099

www.materialsproject.org team

MIT: G. Ceder, W. D. Richards, E. Chtykov
LBL: K. Persson, D. Bailey, M. Kocher, A. Jain, D. Gunter
NERSC: D. Skinner, S. Cholia, A. Greiner
UC Berkeley: M. Asta
UCL: G. Hautier
UKY: A. Dozier, M. Beck
NUS: S. Adams
UCSD: S. P. Ong

Acknowledgements

This work was supported in part by the TRUST Center (NSF award number CCF-0424422) with additional support from LBL and the DOE.