

Containerized Solutions for astronomical surveys

A brief overview from DES/LSST

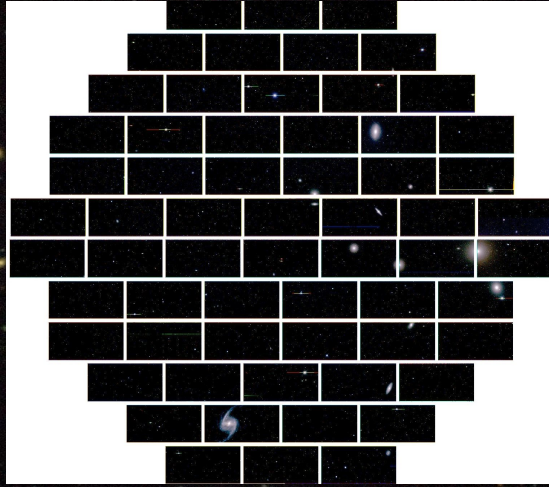
Matias Carrasco Kind (NCSA)

NDS Container Workshop, August 14th 2017, @ NCSA

Photometric Surveys



Telescope



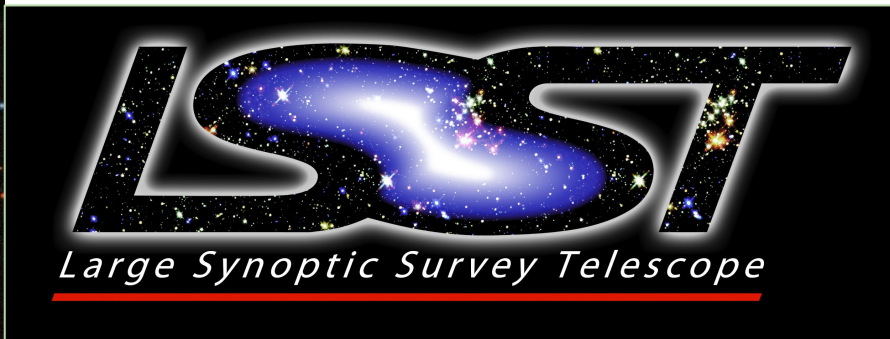
Images

IDX	RA	DEC	MAG_AUTO_G
0	9.55367	-52.858189	99
1	9.999752	-52.754332	22.4878
2	9.114992	-52.555158	25.3244
3	8.994403	-52.899365	24.9927
4	9.427335	-52.595066	23.6036
5	10.008397	-42.256504	24.9008
6	10.084474	-47.183702	24.6363
7	10.556631	-47.180639	23.4764
8	10.197952	-47.10095	26.3588
9	10.863236	-46.987299	21.7274
10	10.673803	-47.056676	23.6435
11	10.250745	-47.011011	24.6457
12	9.145812	-41.983799	22.9988
13	9.156004	-41.756442	20.9553

Catalogs

Collection of thousands of images in different filters. Raw Images are processed and combined and large catalogs are created with position and information of millions of astronomical sources.

Photometric Surveys

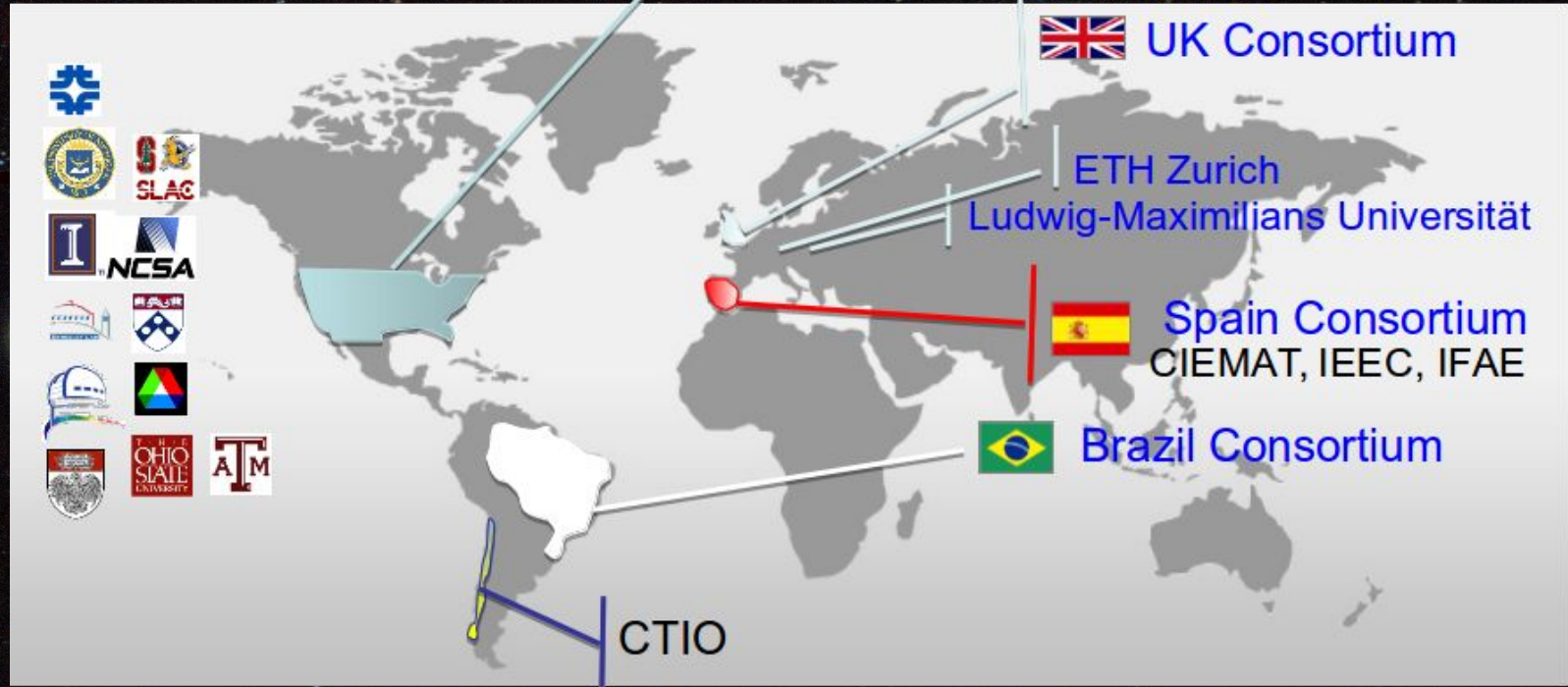


NCSA does Data Management for these 2

Dark Energy Survey (DES)

Started in 2003, DES is now an international collaboration of ~500 scientists from around the world

Fermilab, UIUC/NCSA, University of Chicago, LBNL, NOAO, University of Michigan, University of Pennsylvania, Argonne National Laboratory, Ohio State University, Santa-Cruz/SLAC/Stanford Consortium, Texas A&M



Dark Energy Survey facts

- 4 meters telescope, 520 Mpx camera
- 5 year survey, $\frac{1}{8}$ of the sky, Telescope in Chile, data @ NCSA, about to start 4th season
- Main Goal: To constrain the models of the Universe regarding Dark Energy and Dark Matter. Check Recent News!
- Many other Science Cases! (New dwarf planet, New galaxy satellites, Supernovae, etc)
- 1 - 3 TB of data per night, 1 PB of data
- Processing done at FermiGrid, Campus Cluster and Blue Waters
- Thousands of images and billions of rows, ~500 millions objects
- December to announce 1st Public Data Release
- NCSA provide means to access and interact with data → Containers

DES Labs : Collection of containerized tools for DES



DES Labs

<http://deslabs.ncsa.illinois.edu/>

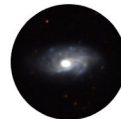
Easyaccess web



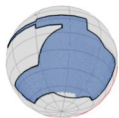
Jupyterhub + easyaccess



DES cutouts



Footprint



Easyaccess online



DESDM Services status



External Links

Science Server





NOAO Data Lab



CosmoHub



DES Labs : Cutout Service


 DARK ENERGY SURVEY Cutouts Services 



The content of these pages are restricted.

Use your DESDM DB credentials

Username

Password

 LOGIN

 DARK ENERGY SURVEY Cutouts Services 

My Jobs

Coadds Images

Single Epoch

API Help

Help

See the list of the REST API commands to interact with the cutout services from scripting tools such as curl, wget, python, etc.

Create Authentication Token

Check Token Status

Submit Job

Get Job Results

Get Jobs List

Delete Jobs

```

{
  "status": "ok",
  "message": "List of jobs returned",
  "list_jobs": [{"job1": "job2", ...}
  "creation_time": [{"time1", "time2", ...}
  "job_status": [{"SUCCESS", "PENDING", ...}
  "job_type": [{"coadd", "single", ...}
}

```

Success:

Errors: 400: no token, 403: Token doesn't exists or expired

Example: curl https://descut.cosmology.illinois.edu/api/jobs/token=mytoken&list_jobs

Python example:

```

import requests
req = requests.get('https://descut.cosmology.illinois.edu/api/jobs/?token=1222...&list_jobs')
print(req.text)
{'list_jobs': ['16b...', '0100...', 'b056...'], 'message': 'List of jobs returned', 'status': 'ok'}

```

Delete Jobs: Delete Job by its Id

URL: /api/jobs/?

Method: DELETE

URL Params: token=mytoken&jobid=abc...

Success: 200: content: {"message": "Job abc... was deleted", "status": "ok"}

Errors: 400: no token, 403: Token does not exists or it expired

Example: curl -X DELETE -G https://descut.cosmology.illinois.edu/api/jobs/ -d "token=c17...&jobid=9839..."

Python example:

```

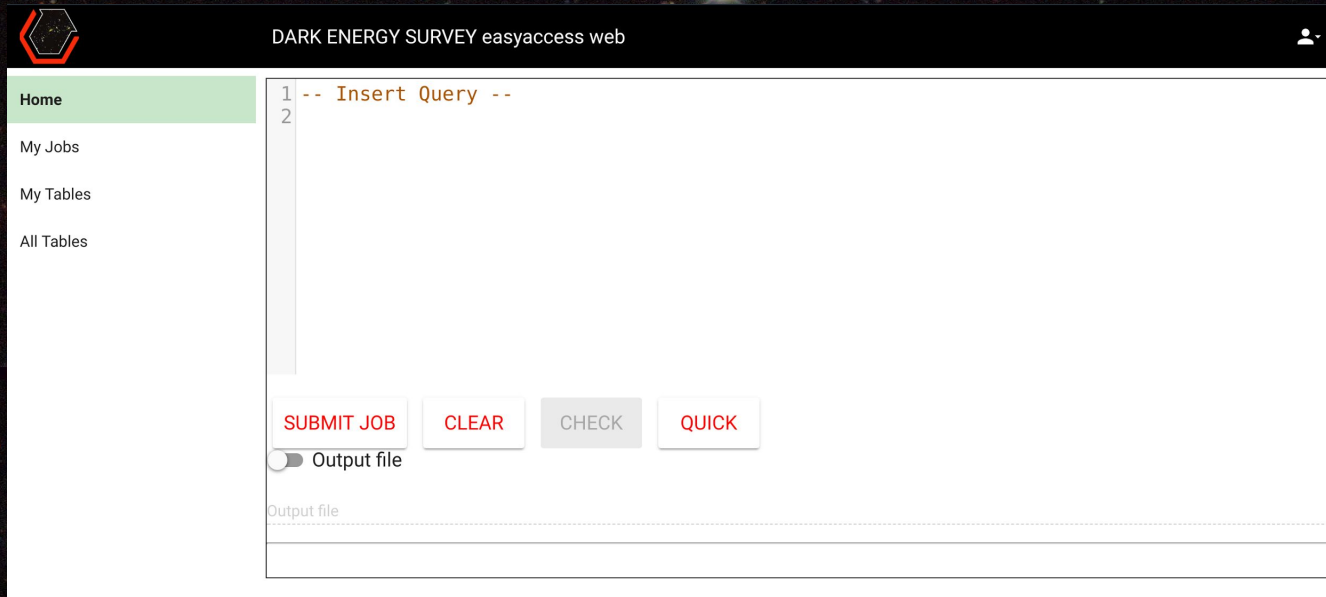
import requests
req = requests.delete('https://descut.cosmology.illinois.edu/api/jobs/?token=1222...&jobid=abc...')
print(req.text)
{'message': 'Job abc... was deleted', 'status': 'ok'}

```

Input a position or a list of positions and return a cutout (all filters, all exposures) around that position for further analysis, direct access to FS

Uses: MongoDB, Redis, Celery and Web Framework on different containers. Prototype version for Kubernetes

DES Labs : easyaccess Web client



The screenshot shows the 'DARK ENERGY SURVEY easyaccess web' interface. On the left is a sidebar with a logo and navigation links: 'Home' (highlighted), 'My Jobs', 'My Tables', and 'All Tables'. The main area contains a query editor with line numbers 1 and 2, and the text '-- Insert Query --'. Below the editor are four buttons: 'SUBMIT JOB' (red text), 'CLEAR' (red text), 'CHECK' (grey), and 'QUICK' (red text). A toggle switch labeled 'Output file' is currently turned off. Below the toggle is a text input field labeled 'Output file'.

SQL client that runs easyaccess in the backend. Allows to submit Jobs, Table Schema and soon to be linked to the cutout Server

Managed by Kubernetes, uses MySQL, Redis, Celery, and others

DES Labs : jupyterHub + easyaccess



Sign In

Username (Use DES DB credentials here):

mcarras2|

Password:

Sign In

- Managed by Kubernetes, Persistent storage through NFS
- Multi-user jupyter notebook sessions
- Access with DB credentials
- Easyaccess and other tools installed
- Spark Cluster and Batch Submission in alpha stage
- Local JS9 (Next gen. DS9)
- Open to DES collaborators

DES Labs

- Footprint exploration and Image Archive
- DESDM Service Status Managed with Kubernetes, uses InfluxDB and Grafana
- DES Release Pages
- Others...



JavaScript



Celery



openstack™



kubernetes



polymer

ORACLE®



MEMSQL



docker



redis

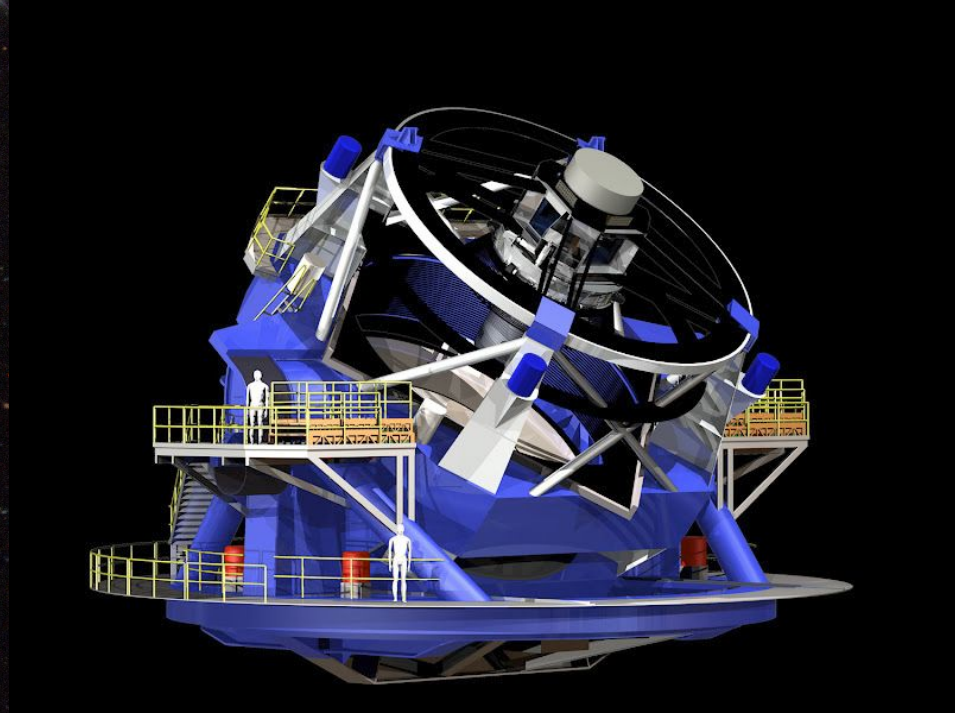


InfluxDB

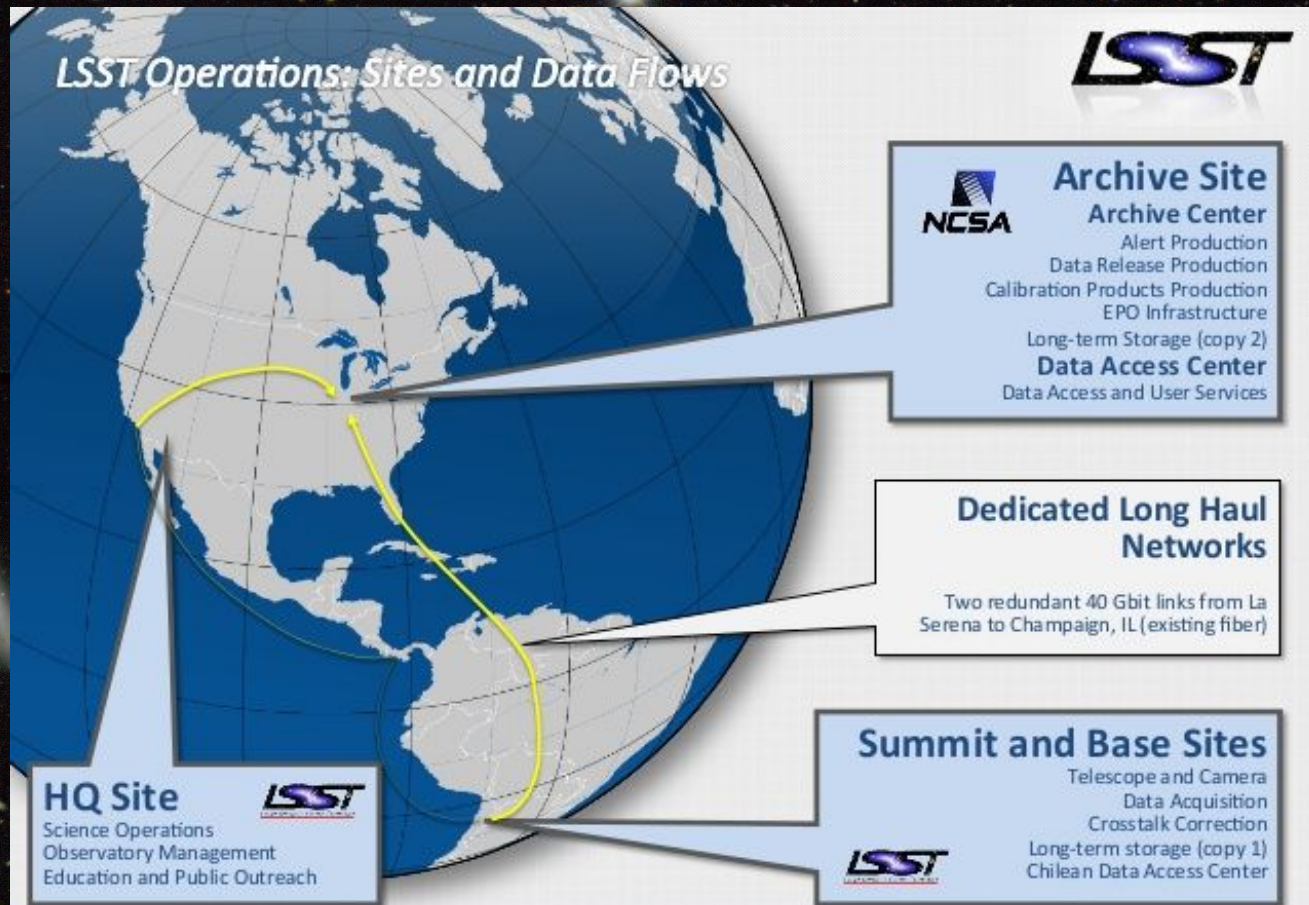


Grafana

Large Synoptic Survey Telescope (LSST)



Large Synoptic Survey Telescope facts



Large Synoptic Survey Telescope facts

- 8.4 Meters Telescope, 3.2 Gpx camera
- 10 year survey, $\frac{1}{2}$ of the sky, Telescope in Chile, data @ NCSA, starts 2020
- Several Goals: Dark Energy/Matter, Transients (Bright/positions), Milky Way
- Public Access since the beginning
- 15 TB per night, 0.5 Exabytes in images, 15 PB in Database
- 37 Billions objects, 5.5 Million 6.4 GB images
- Processing done at NCSA LSST Cluster
- DAC in Chile and NCSA
- Planned to containerized scientific pipelines, DB, Interfaces

LSST Container Solutions



Internet

LSST Science Platform



Portal

JupyterLab



Web APIs



Data Releases



Alert Streams



User Databases



User Files



User Computing



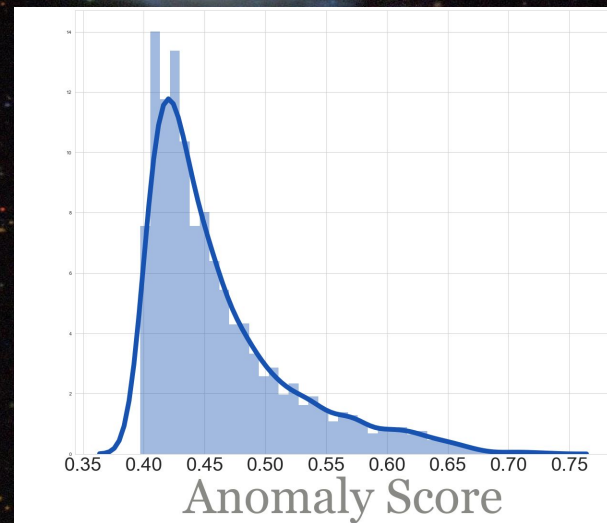
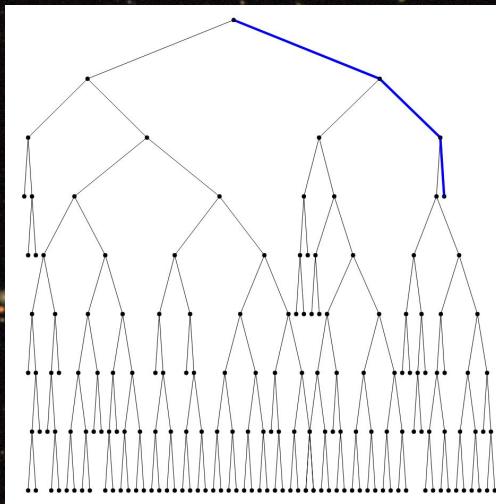
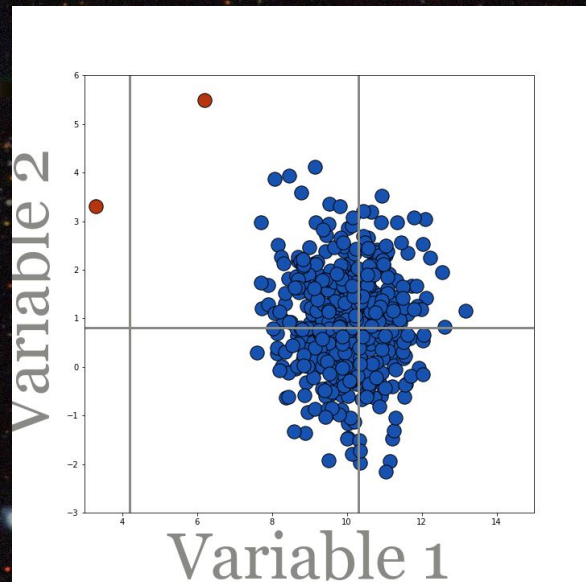
Software Tools

LSST Container Solutions (+ Kubernetes)

- Running Scientific Processing Pipelines (Being tested, IN2P3)
- QA Framework and System (Under Development)
- HTCondor and Spark Cluster (Being tested at bare metal K8S install)
- User interfaces (Under Development)
- JupyterHub (Currently on GKE, moving to Nebula)
- LSST Documentation and Forum (Currently on AWS, eventually NCSA)
- Other Services
- Security, GPFS and Developer/Production current issues
- Local Docker Registry

Isolation Forest (Anomaly Detection)

- Used for anomaly detection in static and stream mode
- Managed by Kubernetes, Front-end, Spark back, Accesible from Jupyter notebook
- Elastic Computing



Credit: Sahand Hariri

Conclusions

- Moving forward for containerized applications
- DES mostly uses containers for data access and user interfaces via cluster and container management
- LSST will go beyond that and will do processing, production, user interfaces, HPC in a multiple DAC schema
- Common needs, solutions and synergy @ NCSA

Thanks!

Questions?

Matias Carrasco Kind
mcarras2@illinois.edu
github.com/mgckind
matias-ck.com