DS 5110 Big Data Systems Term Project

Team 1: Harold Haugen, George Shoriz, Zach Lisman, Dan Anthony

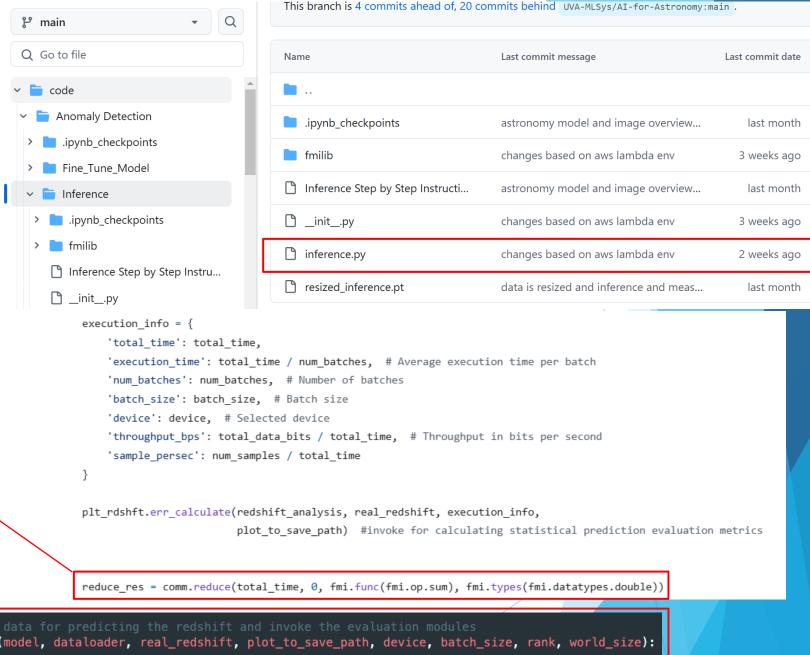
11.19.2024

Prior Work (Steps #1-3)

- ▶ 1: Step Function
 - ► Facilitates orchestration of FMI lambda functions
 - Measure execution time and cost
- 2: Rendezvous Server
 - Establishes communication between AWS lambda functions
- 3: Al for Astronomy Inference: Run Inference.py file
 - resized_images
- Additional: Review Machine Learning

Recent Work

- Review fork of GitHub Repo with edits from Mills
- Integration of Cosmic Al with AWS/Lambda
 - Make edits to inference.py and Plot_Redshift.py for when it is run
 - Passes in rank & world_size
 - When run, it calls engine function which calls inference function
 - Inference function
 - Reduction operation
 - Pass in function type (sum)
 - Works with execution info (as doubles)
 - Added arguments for Rank & World_Size

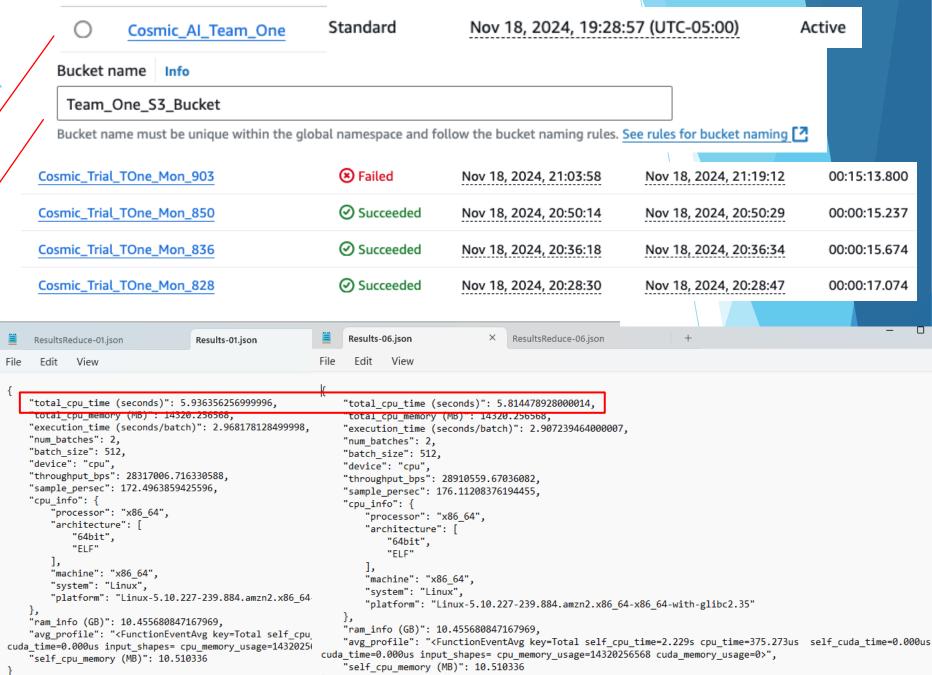


```
#Iterate over data for predicting the redshift and invoke the evaluation modules def inference(model, dataloader, real_redshift, plot_to_save_path, device, batch_size, rank, world_size):

parser.add_argument('--rank', type=int, **environ_or_required('RANK', required=False))
parser.add_argument('--world_size', type=int, **environ_or_required('WORLD_SIZE', required=False))
```

Recent Work

- Copied/created our own:
 - State Machine (copied IAM roles)
 - S3 Bucket (and copied files)
- First Trial for Experiments
 - Ran Step Function
 - Output to Cosmic Al
 - Changed Inference.py
 'Bucket =' to point to our
 S3 Bucket
 - Try Different World Sizes
 - World Size 1: ~6 second
 - World Size 6: ~6 Second
 - World Size 16 & 128: Did not finish



Future Work (Deliverables)

- Try restarting rendesvous server and review video from last week
- Experiment with partitioning and run inference with varied parameters and visualize results via charts or graphs
- Record 5-10 Minute video presenting term project
 - Introduction
 - Data
 - Experimental Design
 - Beyond original specification
 - Results
 - Testing
 - Conclusions (Systems & ML)
- Update README to reflect presentation and final outcomes
- Upload any files to GitHub

Code

Presentations/Updates