National Centre for Atmospheric Science
NATURAL ENVIRONMENT RESEARCH COUNCIL

PyActiveStorage

Maintained? yes | docs passing | Made with Python | Test passing | codecov 85%

Valeriu Predoi
David Hassell
Bryan Lawrence
External collab: StackHPC
Project: EXCALIBUR

https://github.com/NCAS-CMS/PyActiveStorage

Problem:

Problem at hand: the Polar Bear Conundrum

- Client needs some data, and they also need to apply some sort of reduction on it

Solution #1

- Client **downloads** data locally, applies reduction → polar bear dies

Solution #2

- Client asks (remote) Storage (nicely) to assemble their data, apply reduction, and give them the result of said reduction → polar bear lives

Solution:

Solution #2

- Client asks (remote) Storage (nicely) to assemble their data, apply reduction, and give them the result of said reduction → polar bear lives
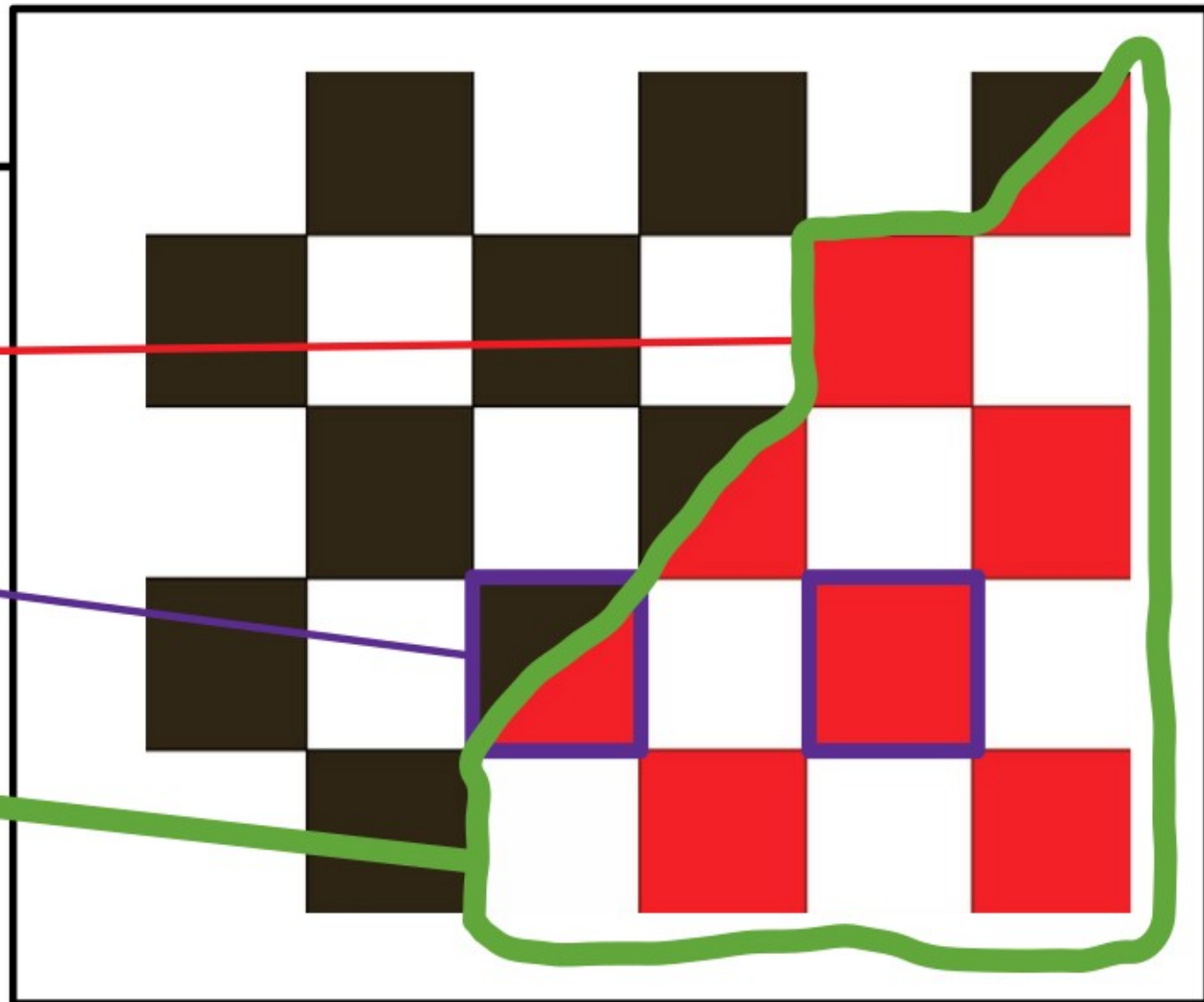
Mechanism

- Identify file that contains data of interest (uri, variable name); load said file on Storage
- Data is chunked and selection of interest overlaps a number of full or partial chunks
- We need to find:
  1. coords (offset, size) for each chunk that overlaps selection
  2. coords (offset, size) for each blob of data in the chunks that overlap selection
- Pass those coords to Storage to retrieve selection data
- Assemble selection data
- Transform assembled data (apply a method)

https://pyactivestorage.readthedocs.io/en/latest/