GeoNotebook

An extension to the Jupyter Notebook for exploratory geospatial analysis

Christopher Kotfila, Aashish Chaudhary - Kitware Inc.
Petr Votava - NASA Ames

JupyterCon 2017
What is GeoNotebook?

- Python & Javascript extensions to Jupyter Notebook
  - Interactive map connected to python execution environment via RPC.
  - Integrated tile server for rendering geospatial data onto the map.
- Supports interactive exploration and subsetting of geospatial data (especially raster data).
Geospatial Data Types

Vector

Raster

Author: David DiBiase, The Nature of Geographic Information
What is a Tile Server?

- Provides API endpoints for downscaled images
- Serves more detailed images at lower ‘zoom’
- Both Raster & Vector data can be tiled
Demonstration
Technical Objectives

- No assumptions about relationship between data access and data visualisation.
  - Data access always at native resolution
- Sensible defaults; target “traditional” Jupyter user.
- Support custom visualisation/data sources via third party python packages
Technical Stack

- **GDAL** - file I/O with custom VRT generation
- **Mapnik** - styling and orchestration of downsampling
- **KTile** - TileStache fork for serving tiles
- **Jupyter Notebook** - plugins + custom python kernel
- **GeoJS/OpenLayers** - interactive presentation of map
Current Limitations

- Complex deployment of tile rendering stack (Mapnik/GDAL); docker image available
- Re-rendering transformations of data subsets (e.g. rendering in-memory data).
- Interactive elements break some of the implicit promises about reproducibility of notebooks.
Development Context

● NASA ESTO - AIST 2014 Grant
  ○ Public cloud tools for extending analysis capabilities.
  ○ Bring analysis to the data, rather than data to the analyst

● NASA Earth Exchange (NEX) @ NASA Ames
  ○ Portal for collaborative sharing network for researchers
  ○ Large scale data processing pipelines for enabling ‘science-quality’ satellite data products.
Call to Action

- Originally developed as a demonstration of what was possible for Q/A on NEX data processing pipelines.
- Still early days
  - Apache 2.0 licensed software seeks interested community of users & contributors.
  - Cathedral and the Bazaar - “plausible promise”
Many Thanks!

Contributing projects:
GDAL, Mapnik, TileStache, Jupyter Notebook, NumPy, Fiona, Shapely, Rasterio, Pytest and many others.

Contributing humans:
Mike Little†, Raman Nemani†, Petr Votava, Andrew Michaelis†, Jon Beazley‡, Doruk Ozturk‡, Dan Lamanna‡

† NASA, ‡ Kitware Inc.
Questions!

Github: [OpenGeoscience/geonotebook](https://github.com/OpenGeoscience/geonotebook)
Gitter: [gitter.im/OpenGeoscience/geonotebook](https://gitter.im/OpenGeoscience/geonotebook)
RTD: [geonotebook.readthedocs.io](https://geonotebook.readthedocs.io)
Email: chris.kotfila@kitware.com

docker run --net=host -v /path/to/notebooks:/notebooks geonotebook/geonotebook:latest
Additional Slides
Reproducible, automated provisioning of Q/A instance & cloud infrastructure, delivered on demand via Ansible.
Subset with a vector

Investigate

Do on the fly processing
Nascent Geotrellis integration

Geotrellis & GeoPySpark

Deep Learning for Semantic Segmentation of Aerial Imagery