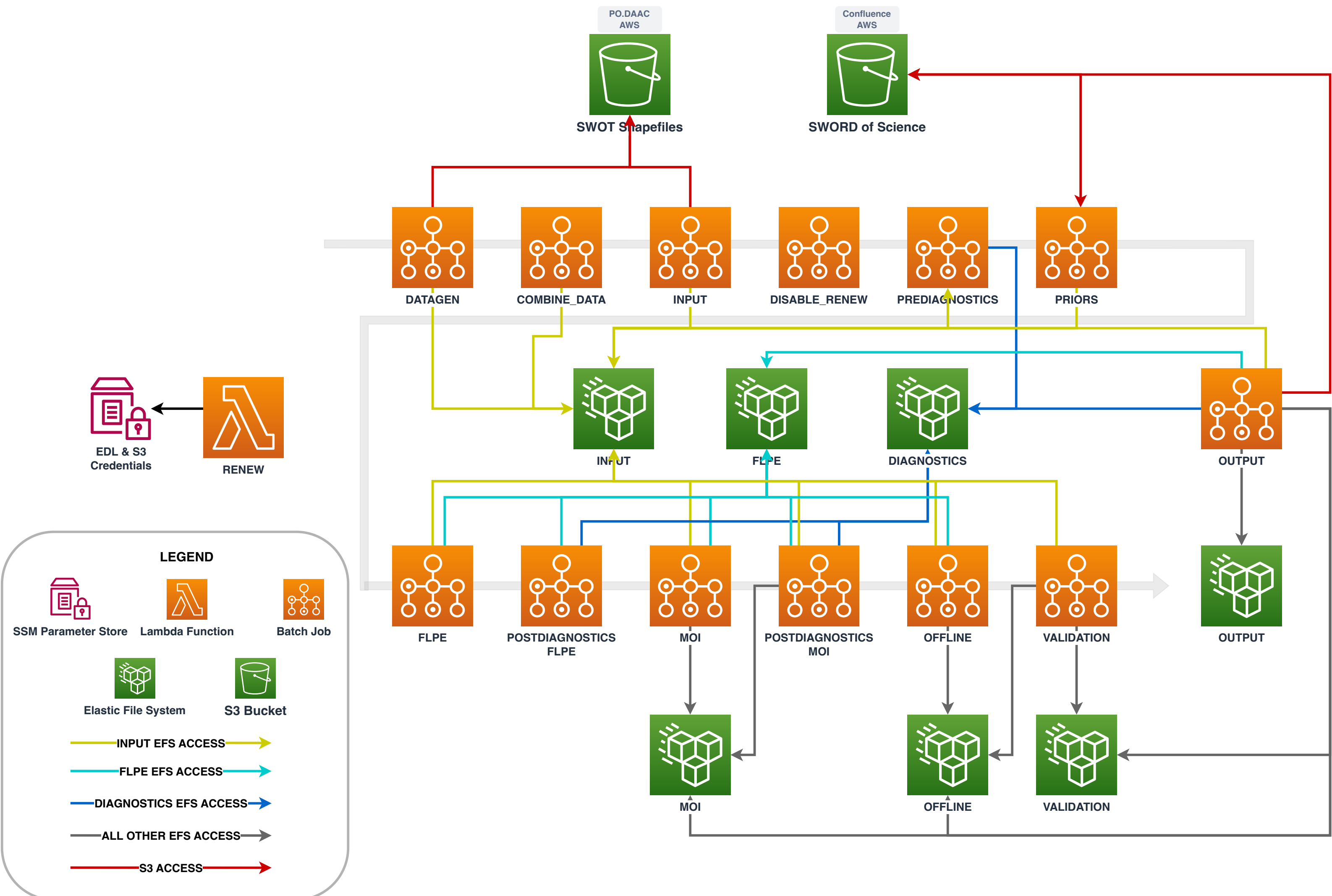


Confluence AWS Data Flow



Description

This diagram represents the data flow through Confluence in AWS. The legend describes which service is associated with each icon type. The light gray arrow indicates the order of execution for each module.

The renew Lambda function retrieves temporary S3 credentials for in-region direct access to the SWOT shapefiles hosted in a PO.DAAC S3 bucket. It stores these credentials in the SSM parameter store. It executes every 50 minutes to prevent the expiration of the temporary credentials. This gets enabled by the confluence-aws program and executes outside of the AWS Batch jobs.

The data flow begins and executes as follows:

1. The datagen module takes SWOT shapefiles and determines the reach and nodes to run on for each module. It produces JSON files for each continent and stores them in the input Elastic File System (EFS).
2. The combine\_data module combines the continental datagen produced JSON files into global JSON files used as input to the other modules and stores the global JSON files in the input EFS.
3. The input module takes SWOT shapefiles and transforms them into time series data stored in NetCDF files in the input EFS.
4. The disable\_renew module then disables the renew Lambda function as access to the PO.DAAC SWOT S3 bucket is no longer needed.
5. The prediagnostics module runs which takes the SWOT NetCDF files created by the input module and stored in the input EFS and runs various diagnostic operations on the data. It updates the NetCDF files in the input EFS and stores the results of the flagging diagnostic operations in the diagnostics EFS.
6. The priors module runs to produce various gaged priors as well as global priors generated from geoBAM. The priors module creates a new version of the SOS which it downloads from the SOS S3 bucket and stores in the input EFS. It preserves any overwritten data in a separate SOS group.
7. The FLPE modules execute taking the input-created NetCDF files as input and outputting NetCDF files to the FLPE EFS.
8. The postdiagnostics module runs to compare previous executions of the workflow with the current execution for the FLPE modules. The input is taken from the FLPE EFS and output is stored in the diagnostics EFS.
9. The MOI module runs and uses the FLPE-generated NETCDF files stored in the FLPE EFS as well as data from the input EFS. It stores output in the MOI EFS.
10. The postdiagnostic module runs to compare previous executions of the workflow with the current execution for the MOI module. The input is taken from the input, FLPE, and MOI EFS and output is stored in the diagnostics EFS.
11. The offline module runs to produce discharge values and uses input data from the input and FLPE EFS and stores the results in the offline EFS.
12. The validation module executes and validates the entire execution and takes input from input EFS and the offline EFS. It stores output in the validation EFS.
13. The output module executes and takes input from every EFS to create an updated SOS result file that contains the collocated results from each module organized by continent. It stores the new SOS result file in the SOS S3 bucket.