Importing a File

Unlike the upload function, which is a push from the client to the server, the import function is a parallelized reader and pulls information from the server from a location specified by the client. The path is a server-side path. This is a fast, scalable, highly optimized way to read data. H2O pulls the data from a data store and initiates the data transfer as a read operation.

Refer to the Supported File Formats topic to ensure that you are using a supported file type.

Note: When parsing a data file containing timestamps that do not include a timezone, the timestamps will be interpreted as UTC (GMT). You can override the parsing timezone using the following:

- R: `h2o.setTimezone("America/Los Angeles")`
- Python: `h2o.cluster().timezone = "America/Los Angeles"`

R | Python

```
# To import airlines file from H2O's package:
library(h2o)
h2o.init()
iris_path <- "https://s3.amazonaws.com/h2o-airlines-unpacked/allyears2k.csv"
iris <- h2o.importFile(path = iris_path)

# To import from S3:
library(h2o)
h2o.init()
airlines_path <- "https://s3.amazonaws.com/h2o-airlines-unpacked/allyears2k.csv"
airlines <- h2o.importFile(path = airlines_path)

# To import from HDFS, you must include the node name:
library(h2o)
h2o.init()
airlines_path <- "hdfs://node-1:/user/smalldata/airlines/allyears2k_headers.zip"
airlines <- h2o.importFile(path = airlines_path)
```