

# HDFS-APIs

## Hadoop Distributed File System (HDFS) APIs in perl, python, ruby and php

The Hadoop Distributed File System is written in Java. An application that wants to store/fetch data to/from HDFS can use the Java API. This means that applications that are not written in Java cannot access HDFS in an elegant manner.

Thrift is a software framework for scalable cross-language services development. It combines a powerful software stack with a code generation engine to build services that work efficiently and seamlessly between C++, Java, Python, PHP, and Ruby.

This project exposes HDFS APIs using the Thrift software stack. This allows applications written in a myriad of languages to access HDFS elegantly.

### The Application Programming Interface (API)

The HDFS API that is exposed through Thrift can be found in [hdfs/src/contrib/thriftfs/it/hadoopfs.thrift](#). The [perl](#), [python](#), [ruby](#), and [php](#) APIs can be found at [src/contrib/thriftfs/gen-\\*](#) directories.

### Compilation

The compilation process creates a server `org.apache.hadoop.thriftfs.HadoopThriftServer` that implements the Thrift interface defined in `it/hadoopfs.thrift`.

The thrift compiler is used to generate API stubs in python, php, ruby, cocoa, etc. The generated code is checked into the directories `gen-*`. The generated java API is checked into `lib/hadoopthriftapi.jar`.

There is a sample python script `hdfs.py` in the `scripts` directory. This python script, when invoked, creates a [HadoopThriftServer](#) in the background, and then communicates with HDFS using the API. This script is for demonstration purposes only.