Disco takes on the lofty task of attempting to be a substitute for a Hadoop Distributed File System in a light weight, Python implemented method for polling data. While Hadoop is the industry leader in distributed filesystems, Disco offers simplified coding, concepts, and implementation, appealing to a wide range of data users. The libraries involved are still in the works but deployment is very simple, especially in a Python-centric organization/SAN environment. The access is simplified and diverse as it can leverage many different access protocols. APIs exist to simplify. It otherwise is similar to Hadoop’s MapReduce but instead of consisting of <key,value> pairs it follows more a database infrastructure. Disco uses an opensource distribution that can be easily found on GitHub, as mentioned in the references.

Big data analytics is dramatically shifting over to Python, despite lack of computing efficiency, because of its robust libraries and simplified programming styles. From this, it was only a matter of time until the data storage and distribution took on a Python implementation for a distributed file system. The intention is that the analytics and storage will still be able to be offloaded to other compute powers instead of simply on one local machine. The benefit of ease for programmers can now be enjoyed, potentially without needing to learn the intricacies of SQL/NoSQL and other, more complicated, programming languages.

The Disco project is relatively new and young but has high promise, especially in Python implemented environments. The concept of a distributed file system is that, in big data analytics, many times the data is too large to be both stored and analyzed on a single machine. While super-computers are one way around needing such a program, cheaper implementations, using existing hardware and smaller storage clusters is viable and valuable. Time will tell whether Disco will be a valuable competitor for a Hadoop Distributed File System.

References

Disco Project (2014). Retrieved from: https://github.com/discoproject/disco

Disco Release notes(2014): Retrieved from: https://disco.readthedocs.io/en/latest/releases.html