Local File Storage Indexing

Project aims to create an index of whole file system on your computer. It includes Metadata related to files. Includes in-file data into metadata. In-file data includes text data of PDFs, Word, Notepads, Images, etc. Text data will be extracted from images using Optical Character Recognition (OCR). All the data will be saved in the local file storage for security purpose. Data will be analyzed based on type of files, number of files, type of data included on those files. Big Data technology will be used to store all the above mentioned and generated data. Dataset will be created by a python program created for this specific project.