Demo flow -- items marked 5.1 require CDH 5.1 GA July 17

Use 3 or 4 files of with different information <all but 1 of them will be preloaded into CDH with all relevant metadata and tags preadded>

1. **Ingest**
2. Show a file in the windows/linux file system (Can we break this out into maybe two very different file types i.e. a CSV data file equating to a DB extract and another file equivalent to an Excel spreadsheet with mocked data within it that maps to the CSV file)
3. Use HDFS commands to put that file into HDFS (Though we may be limited given the cost of services on AWS it would be ideal to demonstrate ingesting data from a RDBMS to depict how would be users can offload archival data into HDFS using sqoop)
4. Use HDFS commands to show the contents of that file in HDFS – they will be the same
5. **ETL**
   1. Use ETL TOOL to 2 different files to same schema or similar action <need to look at our data to see what would be meaningful>
6. **Create/.Show metadata in Hive Metastore for file**
   1. Brief discussion of schema on read – slide? -
   2. Create Hive Schema for new table just uploaded
   3. Use Hue to show schema’s for the various tables
7. **Relational query** 
   1. Login into Hue as UserOne
   2. Use Hue to query the data with Impala (join 2 files on Tax Payer ID)
   3. <5.1> make a nice chart
8. **Text Search**
   1. Use Cloudera Manager to index the new file(s) just uploaded so it will be searchable
   2. Brief discussion on how/when data gets indexed, facets, and Solr
   3. Use Hue to Search the collection of documents <using a facet to limit the search by some attribute >
   4. <5.1> make nice chart/graph
9. **Security**
   1. Discussion of Sentry for Authorization
   2. Log into Hue <new tab> as UserTwo
   3. Try to query a table that you are not authorized to see
   4. Look at Hive Metastore notice that the table you are querying does not seem to exist
   5. Discuss navigator – slide? – (yes the navigator portion of the demo should demonstrate a 360 degree view of the individual and his/her presence in many of the datasets in the ecosystem)
   6. Open Audit logs – look at some recent actions in the log including unauthorized attempt to query a table

<continued on next page>

1. **Data lineage and tagging. <5.1>**
   1. Discussion on Lineage and Tagging <slides>
   2. Show metadata tags on files
   3. Add metadata tags to the file that was added
   4. Show lineage of previously ETLed file << need to take 2 file and use ETL tool to create a new one with some merged data>>