EDH functional demo

1. Slide of demo components besides Cloudera Data Hub edition
2. Slide showing HUE - describe its use and its function in demo
3. Slide of Cloudera Data Hub Edition and brief description of components (cake slide)
4. Slide – Data Ingestion into CDH (has Sqoop, Flume, Fuse/NFS, HDFS command line)
5. Slide of demo flow
   1. Put data into CDH from
      1. File
      2. RDBMS
   2. Transform data using ETL tool
   3. Query data using SQL
   4. Query the data using faceted text search
   5. Demonstrate Role Based Access Controls on data
   6. Show audit logs of data access
   7. Show lineage of data (transformation history + any tags on file)
6. Slides – Load data – ingest, metadata
7. *Demo --Use Hue to show there are already 2 files in CDH*
8. *Demo – Use Hue to show the Metadata repository –*
9. *Use Data browser to show a table in RDBMS (keep window open)*

* *“This is the current state of the system 2 file, mapped to relational format”*
* *“Now we can load data into HDFS”*

1. *Demo— ingest from DBMS*
   1. *use Data browser to show table in RDBMS*
   2. *Sqoop data into CDH*
2. *Demo – ingest from file*
   1. *Show excel file*
   2. *Use Hue to put CSV file into CDH (process shows how Metadata is being mapped)*

* *“Now we can see that the data is in HDFS*

1. *Demo-- Use HUE to show there are now 4 files in CDH*

* *“The Relational Metadata is there too”.*

1. *Demo—Use HUE to show the new data in Metastore – field names, datatypes,*
2. *Demo-- Use HUE Impala Browser to show tables in HDFS <<this may be overkill>>*
3. *Demo-- Use HUE to select \* from table that was from RDBMS (show data matched data browser) OR from table that was in Excel*
4. Slide Discussion of ETL & partners in that space
5. *Demo-- Pentaho to create file/table #4*
6. *Demo-- Use Hue to show there are 4 files in CDH << where do they views come in??>>*
7. Slide Summary of what we have proven so far
   * Easy to move data from outside Hadoop to inside Hadoop
   * HDFS can treat data as file, tables, record, documents(documents coming)
   * Easy to clean/transform data
8. Slide on Impala
9. *Demo of Impala*
   1. *Demo—Use Hue to show there are 6 tables (4 table + 2 views on the base table)*
   2. *Show an SQL queries including some sort of join*
10. Slide on Cloudera Search
11. *Demo of Cloudera Search on the Data*
    1. *Make sure to highlight both text and facets*
12. Slide on Security – discuss 4 pillars
13. Slide on Sentry – discussion RBAC
14. *Demo--*
    * *Logout – Log in as (user 1)*
    * *You see 4 tables (3 + 1 view) – cannot see base table*
    * *Select \* from view ( you see PII data)*
    * *Logout – log in as (user 2)*
    * *You see 4 tables (the 1 view is a different one based one your role)*
    * *Select \* from the view (you don’t see PII data)*
    * *Type Select \* on prior (user 1) view – you get no such table back*
15. Sentry limits data access to Tables, Views, Documents, and other objects based on users Roles.
16. Slide – Navigator – discussion of Audit and Lineage
17. *Demo Audit logs for what we did today – point out audit except for user’s attempt to query a table that they are not authorized to see/access.*
18. *Demo of Lineage for ETL job we ran*
19. *Demo of Lineage metadata tags on files and user properties*
20. Slide Summary of Security features (including Encryption and Key Management – not demoed)
21. Slide – moving data OUT of Hadoop
22. Demo – Sqoop data out of Hadoop – show data now in database
23. Slide -- summary