EDH functional demo

1. 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. Demonstrate Role Based Access Controls on data
   5. Query the data using faceted text search
   6. Show audit logs of data access
   7. Show lineage of data (transformation history + any tags on file)
2. Slide of Cloudera Data Hub Edition and brief description of components (cake slide)
3. Slide showing HUE - describe its use and its function in demo
4. Slide – Data Ingestion into CDH (has Sqoop, Flume, Fuse/NFS, HDFS command line)
5. *Demo --Use Hue to show there are already 2 files in CDH*
6. *Demo --Use Data browser to show a table in RDBMS (keep window open)*
7. *Demo-- Sqoop data into CDH*
8. *Demo-- Use HUE to show there are now 3 file in CDH*
9. *Demo-- Use HUE to select \* from table that was from RDBMS (show data matched data browser)*
10. *Demo-- Use HUE Impala Browser to show tables in HDFS*
11. ---NEED discussion/demo of metadata on HDFS and show how to edit that metadata--
12. Slide Discussion of ETL & partners in that space
13. *Demo-- Pentaho to create file/table #4*
14. *Demo-- Use Hue to show there are 4 files in CDH*
15. *Demo—Use Hue to show there are 6 tables (4 table + 2 views on the base table)*
16. *Show a couple of SQL queries including some sort of join*
17. Slide Summary of what we have proven so far
    1. Easy to move data from outside Hadoop to inside Hadoop
    2. HDFS can treat data as file or tables or (as we will see documents)
    3. Some ISV icons for data analytics/BI partners
18. Slide on Cloudera Search
19. Slide on Faceted Search
20. *Demo of Cloudera Search on the Data*
21. Slide on Security – discuss 4 pillars
22. Slide – Authentication in Hadoop is Kerberos BUT Cloudera helped pioneer other standard authentication mechanisms. Hue is LDAP based. (no need to have Kerberos on clients). CDH also supports SAML used by many webbased apps for single sign on.
23. Slide on Sentry – discussion RBAC
24. *Demo--*
    1. *Logout – Log in as (user 1)*
    2. *You see 4 tables (3 + 1 view) – cannot see base table*
    3. *Select \* from view ( you see PII data)*
    4. *Logout – log in as (user 2)*
    5. *You see 4 tables (the 1 view is a different one based one your role)*
    6. *Select \* from the view (you don’t see PII data)*
    7. *Type Select \* on prior (user 1) view – you get no such table back*
25. Sentry limits data access to Tables, Views, Documents, and other objects based on users Roles.
26. Slide – Navigator – discussion of Audit and Lineage
27. *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.*
28. *Demo of Lineage for ETL job we ran*
29. *Demo of Lineage metadata tags on files*
30. Slide Summary of Security features (including Encryption and Key Management – not demoed)
31. Slide – moving data OUT of Hadoop
32. Demo – Sqoop data out of Hadoop – show data now in database
33. Slide -- summary