Questions

1. Does it matter if we use varchar or string? N

2. Do I have to do any static or dynamic partitioning of the data? Y

2a. Partition branch by state dynamically

2b. sqoop branch as a file so it can be loaded into hive tables

3. Time table to have timeid as date or string? if date then would be incremental lastmodified if string then incremental append? mAKE AS DATE

IF DATE TYPE THEN CAN’T BE WITHOUT DASHES

4. Automating the Process with Oozie: Delete all folders from where the sqoop jobs loaded data into? For example /Credit\_Card\_System/CDW\_SAPP\_BRANCH/ Directory? From the ETL portion? Y

DON'T USE INCREMENTAL APPEND OR LAST MODIFIED HERE ONLY IN NEXT SECTION

5. Automating the Process with Oozie: Tables in Hive should be dropped from the ETL portion? Y

6. Since we are deleting folders and Hive tables, for the sqoop queries in the workflow, can they just be standard queries and not incremental last modified or append? DON'T USE INCREMENTAL APPEND OR LAST MODIFIED HERE ONLY IN NEXT SECTION

7. How important are comments beyond the Java portion? N

8. workflow.xml action naming convection? n

9. Do we have to use the sqoop metastore? run metastore for the optimizing part for the oozie job

10. What do you mean by if SSN is invalid? If null? Abort the session? Abort the sqoop job? convert to number

11. How important is it to make HIVE table with NOT NULL? no

12. Branch phone number is (XXX)XXX-XXXX, complete

13. Check about dynamic partitioning from external table to external table ok.

14. Style .txt. files

15. Process Optimization Module

16. OOzie Coordinator Schedule

17. Incorporate the job –exec for all jobs in a new workflow.

18. Make bash script to create all jobs

19. Mapping says branch zip to be int but should be string for hive visualization