HADOOP Data Ingestion

**Section1)**

**FTP ( Data movement from FTP to HADOOP Landing)**

Landing

HDFS



FTP Server

**FTP code to transfer Files**

SCRIPT\_DIR="/dc01/iap1/ftphome/sap/"

. ${SCRIPT\_DIR}/transactions

cur\_date=`date +'%m%d%Y'`

ftp\_file=${FTP\_FILENAME\_PREFIX}"\*"${cur\_date}".csv"

log="${LOGDIR}/ftp\_${cur\_date}.log"

host=`hostname`

rc=0

boj=`date`

cd ${SRCDIR}

#If not deployed correctly will give the following system failure

if [ "$?" -ne "0" ]; then

echo " Invalid Source File Path $SRCDIR " | mailx -s "IMPORTANT!!TransactionProcess $cur\_date FTP Failed" $SUPPORT\_EMAIL

exit 1

fi

################### FTP block Start ######################

exec 1>$log 2>&1

echo "---------------------------------------------------"

echo " Begin FTP Session "

echo "---------------------------------------------------"

echo "open $FTP\_SERVER

quote USER $FTP\_USER

quote PASS $FTP\_PWD

cd $FTP\_DIR

verbose

prompt

mget $ftp\_file

close

quit" |$FTP -n

echo "---------------------------------------------------"

echo "Target FTP Location is : $SRCDIR "

echo " End FTP Session "

echo "---------------------------------------------------"

################## FTP block ends ########################

# Check the log file for FTP status

foo=`grep -i "Transfer complete" $log`

if [ $? -eq 0 ] ; then

status="Successful"

rc=0

else

no\_files=`grep -i "No such file or directory" $log`

echo "NO\_FILES = $no\_files"

if [ $? -eq 0 ]; then

rc=1

status="No files to FTP"

else

rc=2

status="FTP Process Failed"

fi

fi

echo "FTP Program Completed" | mailx -s "FTP Program completed `date`" ${SUPPORT\_EMAIL}

**Section 2) Transfer files from Landing to HDFS**

Hadoop\_server=

Name\_node=

Port=

srcFilename=

targetdir=

archiveDir

hrun=`hadoop fs –put $srcfilename $targetdir `

if [ $? -eq 0 ] ; then

status="Successful"

rc=0

else

status="put process failed"

fi

echo "FTP Program Completed" | mailx -s "FTP Program completed with status $status `date`" ${SUPPORT\_EMAIL}

gzip ${srcfilename }

mv ${srcfilename }.gz ${archiveDir}