Comandos executados no Hive e Impala, para a criação da bases final de climas.

drop table voos\_brasil\_labdata.voos\_historico\_tratada

CREATE TABLE IF NOT EXISTS voos\_brasil\_labdata.voos\_historico\_tratada

LOCATION '/user/cloudera/labdata/flights\_anac\_trat' as

select

a.icao

,a.num\_voo

,a.cod\_aut\_di

,a.cod\_tipo\_linha

,a.cod\_aero\_org

,b. descricao as aeroporto\_origem

,b.cidade as cidade\_origem

,b.uf as uf\_origem

,b.pais as pais\_origem

,b.continente as continente\_origem

,a.cod\_aero\_dest

,c.descricao as aeroporto\_dest

,c.cidade as cidade\_dest

,c.uf as uf\_dest

,c.pais as pais\_dest

,c.continente as continente\_dest

,case

when length(a.partida\_prevista)<3 then cast(

concat(

substr(a.partida\_real,7,4) , '-' ,

substr(a.partida\_real,4,2) , '-' ,

substr(a.partida\_real,1,2) , ' ' ,

substr(a.partida\_real,12,2) , ':' ,

substr(a.partida\_real,15,2) ,':00.000000'

) as timestamp)

else

cast(

concat(

substr(a.partida\_prevista,7,4) , '-' ,

substr(a.partida\_prevista,4,2) , '-' ,

substr(a.partida\_prevista,1,2) , ' ' ,

substr(a.partida\_prevista,12,2) , ':' ,

substr(a.partida\_prevista,15,2) ,':00.000000'

) as timestamp)

end as partida\_prevista

,cast(

concat(

substr(a.partida\_real,7,4) , '-' ,

substr(a.partida\_real,4,2) , '-' ,

substr(a.partida\_real,1,2) , ' ' ,

substr(a.partida\_real,12,2) , ':' ,

substr(a.partida\_real,15,2) ,':00.000000'

) as timestamp) as partida\_real

,case

when length(a.chegada\_prevista) < 3 then cast(

concat(

substr(a.chegada\_real,7,4) , '-' ,

substr(a.chegada\_real,4,2) , '-' ,

substr(a.chegada\_real,1,2) , ' ' ,

substr(a.chegada\_real,12,2) , ':' ,

substr(a.chegada\_real,15,2) ,':00.000000'

) as timestamp)

ELSE

cast(

concat(

substr(a.chegada\_prevista,7,4) , '-' ,

substr(a.chegada\_prevista,4,2) , '-' ,

substr(a.chegada\_prevista,1,2) , ' ' ,

substr(a.chegada\_prevista,12,2) , ':' ,

substr(a.chegada\_prevista,15,2) ,':00.000000'

) as timestamp)

end as chegada\_prevista

,cast(

concat(

substr(a.chegada\_real,7,4) , '-' ,

substr(a.chegada\_real,4,2) , '-' ,

substr(a.chegada\_real,1,2) , ' ' ,

substr(a.chegada\_real,12,2) , ':' ,

substr(a.chegada\_real,15,2) ,':00.000000'

) as timestamp) as chegada\_real

,a.situacao\_voo

,a.cod\_justificativa

from voos\_brasil\_labdata.voos\_historico a

left join voos\_brasil\_labdata.glossario\_aerodromos b

on a.cod\_aero\_org = b.sigla\_oacio

left join voos\_brasil\_labdata.glossario\_aerodromos c

on a.cod\_aero\_dest = c.sigla\_oacio

where b.pais = 'BRASIL'

and c.pais = 'BRASIL'

------------------------------------

drop table voos\_brasil\_labdata.voos\_historico\_tratada\_2

CREATE TABLE IF NOT EXISTS voos\_brasil\_labdata.voos\_historico\_tratada\_2

LOCATION '/user/cloudera/labdata/flights\_anac\_trat\_2' as

select

a.icao

,a.num\_voo

,a.cod\_aut\_di

,a.cod\_tipo\_linha

,a.cod\_aero\_org

,a.aeroporto\_origem

,a.cidade\_origem

,a.uf\_origem

,a.pais\_origem

,a.continente\_origem

,a.cod\_aero\_dest

,a.aeroporto\_dest

,a.cidade\_dest

,a.uf\_dest

,a.pais\_dest

,a.continente\_dest

,a.partida\_prevista as partida\_prevista\_completa

,date\_part('YEAR',a.partida\_prevista) as partida\_prevista\_ano

,date\_part('MONTH',a.partida\_prevista) as partida\_prevista\_mes

,date\_part('DAY',a.partida\_prevista) as partida\_prevista\_dia

,to\_date(a.chegada\_prevista) as partida\_prevista\_date

,cast(regexp\_replace(substr(cast(a.chegada\_prevista as varchar),12,5),':','') as int) as partida\_prevista\_hora

,a.partida\_real

,date\_part('YEAR',a.partida\_real ) as partida\_real\_ano

,date\_part('MONTH',a.partida\_real ) as partida\_real\_mes

,date\_part('DAY',a.partida\_real ) as partida\_real\_dia

,a.chegada\_prevista as chegada\_prevista\_completa

,date\_part('YEAR',a.chegada\_prevista ) as chegada\_prevista\_ano

,date\_part('MONTH',a.chegada\_prevista ) as chegada\_prevista\_mes

,date\_part('DAY',a.chegada\_prevista ) as chegada\_prevista\_dia

,to\_date(a.chegada\_prevista) as chegada\_prevista\_date

,cast(regexp\_replace(substr(cast(a.chegada\_prevista as varchar),12,5),':','') as int) as chegada\_prevista\_hora

,a.chegada\_real

,date\_part('YEAR',a.chegada\_real ) as chegada\_real\_ano

,date\_part('MONTH',a.chegada\_real ) as chegada\_real\_mes

,date\_part('DAY',a.chegada\_real ) as chegada\_real\_dia

,a.situacao\_voo

,a.cod\_justificativa

,case

when a.situacao\_voo = 'REALIZADO' and a.partida\_real > a.partida\_prevista then 1

else 0

end as flg\_atraso\_partida

,case

when a.situacao\_voo = 'REALIZADO' then (unix\_timestamp(partida\_real)-unix\_timestamp(partida\_prevista))/60

else 0

end as Diff\_Minutos\_partida

,case

when a.situacao\_voo = 'REALIZADO' and a.chegada\_real > a.chegada\_prevista then 1

else 0

end as flg\_atraso\_chegada

,case

when a.situacao\_voo = 'REALIZADO' then (unix\_timestamp(chegada\_real)-unix\_timestamp(chegada\_prevista))/60

else 0

end as Diff\_Minutos\_chegada

from voos\_brasil\_labdata.voos\_historico\_tratada a

------------------------------------------------------------------

--------------------------------------

--drop table voos\_brasil\_labdata.voos\_historico\_tratada\_3

CREATE TABLE IF NOT EXISTS voos\_brasil\_labdata.voos\_historico\_tratada\_3

LOCATION '/user/cloudera/labdata/flights\_anac\_trat\_3' as

select

a.icao

,a.num\_voo

,a.cod\_aut\_di

,a.cod\_tipo\_linha

,a.cod\_aero\_org

,a.aeroporto\_origem

,a.cidade\_origem

,a.uf\_origem

,a.pais\_origem

,a.continente\_origem

,b.icao icao\_origem

,b.altitudeicao altitudeicao\_origem

,b.direcaovento direcaovento\_origem

,b.pressaoatmestacao pressaoatmestacao\_origem

,b.tempbulboseco tempbulboseco\_origem

,b.tempbulboumido tempbulboumido\_origem

,b.umidaderelativa umidaderelativa\_origem

,b.velocidadeventonebulosidade velocidadeventonebulosidade\_origem

,b.descricao\_cod descricao\_cod\_origem

,a.cod\_aero\_dest

,a.aeroporto\_dest

,a.cidade\_dest

,a.uf\_dest

,a.pais\_dest

,a.continente\_dest

--CHEGADA

,a.chegada\_prevista\_completa

,a.chegada\_prevista\_ano

,a.chegada\_prevista\_mes

,a.chegada\_prevista\_dia

,a.chegada\_prevista\_date

,a.chegada\_prevista\_hora

,a.chegada\_real

,a.chegada\_real\_ano

,a.chegada\_real\_mes

,a.chegada\_real\_dia

,c.icao icao\_dest

,c.altitudeicao altitudeicao\_dest

,c.direcaovento direcaovento\_dest

,c.pressaoatmestacao pressaoatmestacao\_dest

,c.tempbulboseco tempbulboseco\_dest

,c.tempbulboumido tempbulboumido\_dest

,c.umidaderelativa umidaderelativa\_dest

,c.velocidadeventonebulosidade velocidadeventonebulosidade\_dest

,c.descricao\_cod as descricao\_cod\_dest

--PARTIDA

,a.partida\_prevista\_completa

,a.partida\_prevista\_ano

,a.partida\_prevista\_mes

,a.partida\_prevista\_dia

,a.partida\_prevista\_date

,a.partida\_prevista\_hora

,a.partida\_real

,UPPER(situacao\_voo) AS situacao\_voo

,a.cod\_justificativa

,a.flg\_atraso\_partida

,Diff\_Minutos\_partida

,a.flg\_atraso\_chegada

,Diff\_Minutos\_chegada

from

voos\_brasil\_labdata.voos\_historico\_tratada\_2 a

inner join

voos\_brasil\_labdata.clima\_historico\_tratada b

on a.cod\_aero\_org = b.icao

and a.partida\_prevista\_date = b.dataref

and (case

when a.partida\_prevista\_hora between 601 and 1500 then 1200

when a.partida\_prevista\_hora between 1501 and 2100 then 1800

when a.partida\_prevista\_hora > 2100 or a.partida\_prevista\_hora < 601 then 0

else null

end) = b.hora

inner join

voos\_brasil\_labdata.clima\_historico\_tratada c

on a.cod\_aero\_dest = c.icao

and a.chegada\_prevista\_date = c.dataref

and (case

when a.chegada\_prevista\_hora between 601 and 1500 then 1200

when a.chegada\_prevista\_hora between 1501 and 2100 then 1800

when a.chegada\_prevista\_hora > 2100 or a.chegada\_prevista\_hora < 601 then 0

else null

end) = c.hora

------------------------------------------

--drop table voos\_brasil\_labdata.voos\_historico\_tratada\_4

CREATE TABLE voos\_brasil\_labdata.voos\_historico\_tratada\_4

LOCATION '/user/cloudera/labdata/flights\_anac\_trat\_4' AS

SELECT voos.\*,

CASE WHEN feriado\_federal\_partida.codigo\_tipo is not null THEN 1 ELSE 0

END as feriado\_federal\_partida,

CASE WHEN feriado\_federal\_chegada.codigo\_tipo is not null THEN 1 ELSE 0

END as feriado\_federal\_chegada,

CASE WHEN feriado\_estadual\_partida.codigo\_tipo is not null THEN 1 ELSE 0

END as feriado\_estadual\_partida,

CASE WHEN feriado\_estadual\_chegada.codigo\_tipo is not null THEN 1 ELSE 0

END as feriado\_estadual\_chegada,

CASE WHEN feriado\_municipal\_partida.codigo\_tipo is not null THEN 1 ELSE 0

END as feriado\_municipal\_partida,

CASE WHEN feriado\_municipal\_chegada.codigo\_tipo is not null THEN 1 ELSE 0

END as feriado\_municipal\_chegada,

CASE WHEN feriado\_facultativo\_partida.codigo\_tipo is not null THEN 1 ELSE 0

END as feriado\_facultativo\_partida,

CASE WHEN feriado\_facultativo\_chegada.codigo\_tipo is not null THEN 1 ELSE 0

END as feriado\_facultativo\_chegada,

CASE WHEN trim(feriado\_facultativo\_partida.descricao) = 'Carnaval' THEN 1 ELSE 0

END as carnaval\_partida,

CASE WHEN trim(feriado\_facultativo\_chegada.descricao) = 'Carnaval' THEN 1 ELSE 0

END as carnaval\_chegada,

dia\_da\_semana\_partida.codigo\_dia as dia\_da\_semana\_partida,

dia\_da\_semana\_chegada.codigo\_dia as dia\_da\_semana\_chegada,

CASE WHEN dia\_da\_semana\_partida.codigo\_dia IN (4,5,6) THEN 1 ELSE 0

END as final\_de\_semana\_partida,

CASE WHEN dia\_da\_semana\_chegada.codigo\_dia IN (4,5,6) THEN 1 ELSE 0

END as final\_de\_semana\_chegada

FROM voos\_brasil\_labdata.voos\_historico\_tratada\_3 voos

LEFT JOIN voos\_brasil\_labdata.calendario\_feriado\_tratado feriado\_federal\_partida

ON feriado\_federal\_partida.codigo\_tipo = 1

AND voos.partida\_prevista\_date = feriado\_federal\_partida.data

LEFT JOIN voos\_brasil\_labdata.calendario\_feriado\_tratado feriado\_federal\_chegada

ON feriado\_federal\_chegada.codigo\_tipo = 1

AND voos.chegada\_prevista\_date = feriado\_federal\_chegada.data

LEFT JOIN voos\_brasil\_labdata.calendario\_feriado\_tratado feriado\_estadual\_partida

ON feriado\_estadual\_partida.codigo\_tipo = 2

AND lower(voos.uf\_origem) = lower(feriado\_estadual\_partida.uf)

AND voos.partida\_prevista\_date = feriado\_estadual\_partida.data

LEFT JOIN voos\_brasil\_labdata.calendario\_feriado\_tratado feriado\_estadual\_chegada

ON feriado\_estadual\_partida.codigo\_tipo = 2

AND lower(voos.uf\_dest) = lower(feriado\_estadual\_chegada.uf)

AND voos.chegada\_prevista\_date = feriado\_estadual\_chegada.data

LEFT JOIN voos\_brasil\_labdata.calendario\_feriado\_tratado feriado\_municipal\_partida

ON feriado\_municipal\_partida.codigo\_tipo = 3

AND lower(voos.uf\_dest) = lower(feriado\_municipal\_partida.uf)

AND lower(voos.cidade\_dest) = lower(feriado\_municipal\_partida.cidade)

AND voos.partida\_prevista\_date = feriado\_municipal\_partida.data

LEFT JOIN voos\_brasil\_labdata.calendario\_feriado\_tratado feriado\_municipal\_chegada

ON feriado\_municipal\_chegada.codigo\_tipo = 3

AND lower(voos.uf\_dest) = lower(feriado\_municipal\_chegada.uf)

AND lower(voos.cidade\_dest) = lower(feriado\_municipal\_chegada.cidade)

AND voos.chegada\_prevista\_date = feriado\_municipal\_chegada.data

LEFT JOIN voos\_brasil\_labdata.calendario\_feriado\_tratado feriado\_facultativo\_partida

ON feriado\_facultativo\_partida.codigo\_tipo = 4

AND voos.partida\_prevista\_date = feriado\_facultativo\_partida.data

LEFT JOIN voos\_brasil\_labdata.calendario\_feriado\_tratado feriado\_facultativo\_chegada

ON feriado\_facultativo\_chegada.codigo\_tipo = 4

AND voos.chegada\_prevista\_date = feriado\_facultativo\_chegada.data

INNER JOIN voos\_brasil\_labdata.calendario\_dia\_da\_semana dia\_da\_semana\_partida

ON voos.partida\_prevista\_date = dia\_da\_semana\_partida.data

INNER JOIN voos\_brasil\_labdata.calendario\_dia\_da\_semana dia\_da\_semana\_chegada

ON voos.chegada\_prevista\_date = dia\_da\_semana\_chegada.data;

------------------------------------------

--drop table voos\_brasil\_labdata.voos\_historico\_tratada\_5

CREATE TABLE voos\_brasil\_labdata.voos\_historico\_tratada\_5

LOCATION '/user/cloudera/labdata/flights\_anac\_trat\_5' AS

SELECT \*,CASE WHEN altitudeicao\_origem is null then 750 else altitudeicao\_origem end as altitudeicao\_origem\_moda

,CASE WHEN direcaovento\_origem is null then 14 else direcaovento\_origem end as direcaovento\_origem\_moda

,CASE WHEN pressaoatmestacao\_origem is null then 926.7 else pressaoatmestacao\_origem end as pressaoatmestacao\_origem\_moda

,CASE WHEN tempbulboseco\_origem is null then 23.3 else tempbulboseco\_origem end as tempbulboseco\_origem\_moda

,CASE WHEN tempbulboumido\_origem is null then 20.1 else tempbulboumido\_origem end as tempbulboumido\_origem\_moda

,CASE WHEN umidaderelativa\_origem is null then 82 else umidaderelativa\_origem end as umidaderelativa\_origem\_moda

,CASE WHEN velocidadeventonebulosidade\_origem is null then 1 else velocidadeventonebulosidade\_origem end as velocidadeventonebulosidade\_origem\_moda

,CASE WHEN altitudeicao\_dest is null then 750 else altitudeicao\_dest end as altitudeicao\_dest\_moda

,CASE WHEN direcaovento\_dest is null then 14 else direcaovento\_dest end as direcaovento\_dest\_moda

,CASE WHEN pressaoatmestacao\_dest is null then 926.7 else pressaoatmestacao\_dest end as pressaoatmestacao\_dest\_moda

,CASE WHEN tempbulboseco\_dest is null then 23.2 else tempbulboseco\_dest end as tempbulboseco\_dest\_moda

,CASE WHEN tempbulboumido\_dest is null then 20.1 else tempbulboumido\_dest end as tempbulboumido\_dest\_moda

,CASE WHEN umidaderelativa\_dest is null then 82 else umidaderelativa\_dest end as umidaderelativa\_dest\_moda

,CASE WHEN velocidadeventonebulosidade\_dest is null then 0 else velocidadeventonebulosidade\_dest end as velocidadeventonebulosidade\_dest\_moda

,CASE WHEN altitudeicao\_origem is null then 406 else altitudeicao\_origem end as altitudeicao\_origem\_media

,CASE WHEN direcaovento\_origem is null then 12 else direcaovento\_origem end as direcaovento\_origem\_media

,CASE WHEN pressaoatmestacao\_origem is null then 957.25 else pressaoatmestacao\_origem end as pressaoatmestacao\_origem\_media

,CASE WHEN tempbulboseco\_origem is null then 23.55 else tempbulboseco\_origem end as tempbulboseco\_origem\_media

,CASE WHEN tempbulboumido\_origem is null then 19.4 else tempbulboumido\_origem end as tempbulboumido\_origem\_media

,CASE WHEN umidaderelativa\_origem is null then 71.95 else umidaderelativa\_origem end as umidaderelativa\_origem\_media

,CASE WHEN velocidadeventonebulosidade\_origem is null then 1.58 else velocidadeventonebulosidade\_origem end as velocidadeventonebulosidade\_origem\_media

,CASE WHEN altitudeicao\_dest is null then 405.9 else altitudeicao\_dest end as altitudeicao\_dest\_media

,CASE WHEN direcaovento\_dest is null then 13.9 else direcaovento\_dest end as direcaovento\_dest\_media

,CASE WHEN pressaoatmestacao\_dest is null then 957.28 else pressaoatmestacao\_dest end as pressaoatmestacao\_dest\_media

,CASE WHEN tempbulboseco\_dest is null then 23.6 else tempbulboseco\_dest end as tempbulboseco\_dest\_media

,CASE WHEN tempbulboumido\_dest is null then 19.41 else tempbulboumido\_dest end as tempbulboumido\_dest\_media

,CASE WHEN umidaderelativa\_dest is null then 19.41 else umidaderelativa\_dest end as umidaderelativa\_dest\_media

,CASE WHEN velocidadeventonebulosidade\_dest is null then 1.58 else velocidadeventonebulosidade\_dest end as velocidadeventonebulosidade\_dest\_media

,CASE WHEN altitudeicao\_origem is null then 0 else altitudeicao\_origem end as altitudeicao\_origem\_zero

,CASE WHEN direcaovento\_origem is null then 0 else direcaovento\_origem end as direcaovento\_origem\_zero

,CASE WHEN pressaoatmestacao\_origem is null then 0 else pressaoatmestacao\_origem end as pressaoatmestacao\_origem\_zero

,CASE WHEN tempbulboseco\_origem is null then 0 else tempbulboseco\_origem end as tempbulboseco\_origem\_zero

,CASE WHEN tempbulboumido\_origem is null then 0 else tempbulboumido\_origem end as tempbulboumido\_origem\_zero

,CASE WHEN umidaderelativa\_origem is null then 0 else umidaderelativa\_origem end as umidaderelativa\_origem\_zero

,CASE WHEN velocidadeventonebulosidade\_origem is null then 0 else velocidadeventonebulosidade\_origem end as velocidadeventonebulosidade\_origem\_zero

,CASE WHEN altitudeicao\_dest is null then 0 else altitudeicao\_dest end as altitudeicao\_dest\_zero

,CASE WHEN direcaovento\_dest is null then 0 else direcaovento\_dest end as direcaovento\_dest\_zero

,CASE WHEN pressaoatmestacao\_dest is null then 0 else pressaoatmestacao\_dest end as pressaoatmestacao\_dest\_zero

,CASE WHEN tempbulboseco\_dest is null then 0 else tempbulboseco\_dest end as tempbulboseco\_dest\_zero

,CASE WHEN tempbulboumido\_dest is null then 0 else tempbulboumido\_dest end as tempbulboumido\_dest\_zero

,CASE WHEN umidaderelativa\_dest is null then 0 else umidaderelativa\_dest end as umidaderelativa\_dest\_zero

,CASE WHEN velocidadeventonebulosidade\_dest is null then 0 else velocidadeventonebulosidade\_dest end as velocidadeventonebulosidade\_dest\_zero

FROM voos\_brasil\_labdata.voos\_historico\_tratada\_4 ;