**database** pollyana.ferreira@capgemini.com**;**

**show** **table** "marcos.nagato@capgemini.com".flights**;**

**CREATE** **SET** **TABLE** "pollyana.ferreira@capgemini.com".flights ,**NO** **FALLBACK** ,

**NO** **BEFORE** **JOURNAL**,

**NO** **AFTER** **JOURNAL**,

**CHECKSUM** = **DEFAULT**,

**DEFAULT** **MERGEBLOCKRATIO**

(

ano INTEGER COMPRESS 2014 ,

mes INTEGER COMPRESS (1 ,2 ,3 ,4 ,5 ,6 ,7 ,8 ,9 ,10 ,11 ,12 ),

dia INTEGER COMPRESS (1 ,2 ,3 ,4 ,5 ,6 ,7 ,8 ,9 ,10 ,11 ,12 ,13 ,14 ,15 ,16 ,17 ,18 ,19 ,20 ,21 ,22 ,23 ,24 ,25 ,26 ,27 ,28 ,29 ,30 ,31 ),

dep\_time VARCHAR(4) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

dep\_delay VARCHAR(3) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

arr\_time VARCHAR(4) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

arr\_delay VARCHAR(3) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

carrier VARCHAR(2) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

tailnum VARCHAR(6) CHARACTER **SET** LATIN **NOT** CASESPECIFIC COMPRESS 'NA',

flight VARCHAR(4) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

origin VARCHAR(3) CHARACTER **SET** LATIN **NOT** CASESPECIFIC COMPRESS ('SEA','PDX'),

dest VARCHAR(3) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

air\_time VARCHAR(3) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

distance INTEGER,

hora VARCHAR(2) CHARACTER **SET** LATIN **NOT** CASESPECIFIC,

minuto VARCHAR(2) CHARACTER **SET** LATIN **NOT** CASESPECIFIC)

**PRIMARY** **INDEX** ( flight )**;**

*--Tarefas*

*--Considere o dataset flights.csv para realizar as seguintes tarefas:*

*--1. Crie a coluna qa\_year\_month\_day e aponte inconsistências das colunas ano, mes, dia de acordo com as regras abaixo:*

**ALTER** **TABLE** flights

**ADD** qa\_year\_month\_day char(2)**;**

**select** \* **from** flights**;**

*--MY : Indica que está com dado faltante no ano.*

**update** flights **set** qa\_year\_month\_day = 'MY'

**where** (ano **is** **Null**

**or** trim(ano) = ''

)**;**

*--MM : Indica que está com dado faltante no mes.*

**update** flights **set** qa\_year\_month\_day = 'MM'

**where** (mes **is** **Null**

**or** trim(mes) = ''

**or** length(trim(mes)) = 0

)**;**

*--MD : Indica que está com dado faltante no dia.*

**update** flights **set** qa\_year\_month\_day = 'MD'

**where** (dia **is** **Null**

**or** trim(dia) = ''

**or** length(trim(dia)) = 0

)**;**

*--IY : Indica que o valor excede o intervalo [1950,+∞) no ano.*

**select** \* **from** flights**;**

**update** flights **set** qa\_year\_month\_day = **null;**

**update** flights **set** qa\_year\_month\_day = 'IY'

**where** ano < 1950

**;**

*--IM : Indica que o valor excede o intervalo [1, 12] no mês.*

**update** flights **set** qa\_year\_month\_day = 'IM'

**where** (length(trim(mes)) < 1

**or** length(trim(mes)) > 12

)**;**

*--ID : Indica que o valor excede o intervalo esperado para o dia conforme o mês (para fevereiro usar o limite de 29 dias).*

**update** flights **set** qa\_year\_month\_day = 'ID'

**where** mes **in**(1,3,5,7,8,10,12) **and** dia **not** between 1 and 31**;**

**update** flights **set** qa\_year\_month\_day = 'ID' **where** mes **in**(2) **and** dia **not** between 1 and 29**;**

**update** flights **set** qa\_year\_month\_day = 'ID' **where** mes **in**(4,6,9,11) **and** dia **not** between 1 and 30**;**

*--2. Crie a coluna qa\_hour\_minute e aponte inconsistencias das colunas hora e minuto de acordo com as regras abaixo:*

*--MH : Indica que está com dado faltante na hora.*

**ALTER** **TABLE** flights

**ADD** qa\_hour\_minute CHAR(2)**;**

*--MM : Indica que está com dado faltante no minuto. ------34 alterçoes*

**update** flights **set** qa\_hour\_minute = 'MM'

**where** (minuto **is** **Null**

**or** trim(minuto) = ''

**or** to\_number(minuto) **is** **null**

)**;**

*--IH : Indica que o valor excede o intervalo [0, 23] na hora.*

**update** flights **set** qa\_hour\_minute = 'IH'

**where** **not** (length(trim(hora)) >= 0

**or** **not** length(trim(hora)) <= 23

)**;**

*--IM : Indica que o valor excede o intervalo [0, 59] no minuto.*

**update** flights **set** qa\_hour\_minute = 'IH'

**where** **not** (length(trim(minuto)) >= 0

**or** **not** length(trim(minuto)) <= 59

)**;**

*--3. Crie a coluna qa\_dep\_arr\_time e aponte inconsistências da coluna dep\_time e arr\_time de acordo com as regras abaixo:*

**ALTER** **TABLE** flights

**ADD** qa\_dep\_arr\_time CHAR(2)**;**

*--MD : Indica que está com dado faltante no dep\_time . ---- 34 alteraçoes*

**update** flights **set** qa\_dep\_arr\_time = 'MD'

**where** (dep\_time **is** **Null**

**or** trim(dep\_time) = ''

**or** to\_number(dep\_time) **is** **null**

)**;**

**select** \* **from** flights**;**

*--MA : Indica que está com dado faltante no arr\_time . --- 37 alterações*

**update** flights **set** qa\_dep\_arr\_time = 'MA'

**where** (arr\_time **is** **Null**

**or** trim(arr\_time) = ''

**or** to\_number(arr\_time) **is** **null**

)**;**

*--FD : Indica que não respeita o formato esperado no dep\_time (3 ou 4 dígitos).*

**update** flights **set** qa\_dep\_arr\_time = 'FD'

**where** length(dep\_time) **not** **in** (3,4)

**or** (to\_number(dep\_time) **is** **null**

**and** dep\_time <> 'NA')**;** *-- 84*

*--FA : Indica que não respeita o formato esperado no arr\_time (3 ou 4 dígitos). ---15 alteraçoes*

**update** flights **set** qa\_dep\_arr\_time = 'FA'

**where** length(arr\_time) **not** **in** (3,4)

**or** (to\_number(arr\_time) **is** **null**

**and** arr\_time <> 'NA')**;** *-- 108*

*--ID : Indica que o dep\_time não é um horário válido (HHMM ou HMM).*

**update** flights **set** qa\_dep\_arr\_time = 'ID'

**where** length(dep\_time) = 4

**and** (to\_number(substr(dep\_time,1,2)) > 23

**or** to\_number(substr(dep\_time,3,2)) > 59)**;**

*-- 1*

**update** flights **set** qa\_dep\_arr\_time = 'ID'

**where** length(dep\_time) = 3

**and** to\_number(substr(dep\_time,2,2)) > 59**;**

*--IA : Indica que o arr\_time não é um horário válido (HHMM ou HMM).*

**update** flights **set** qa\_dep\_arr\_time = 'IA'

**where** length(arr\_time) = 4

**and** (to\_number(substr(arr\_time,1,2)) > 23

**or** to\_number(substr(arr\_time,3,2)) > 59)**;** *-- 1*

**update** flights **set** qa\_dep\_arr\_time = 'IA'

**where** length(arr\_time) = 3

**and** to\_number(substr(arr\_time,2,2)) > 59**;**

*--4. Crie a coluna qa\_dep\_arr\_delay e aponte inconsistências da coluna dep\_delay e arr\_delay de acordo com as regras abaixo:*

**ALTER** **TABLE** flights

**ADD** qa\_dep\_arr\_delay Char(2)**;**

**select** \* **from** flights**;**

*--MD : Indica que está com dado faltante no dep\_delay.34*

**update** flights **set** qa\_dep\_arr\_delay = 'MD'

**where** (dep\_delay **is** **null**

**or** trim(dep\_delay) = ''

**or** dep\_delay = 'NA'**;**

)**;**

*--MA : Indica que está com dado faltante no arr\_delay.50*

**update** flights **set** qa\_dep\_arr\_delay = 'MA'

**where** (arr\_delay **is** **null**

**or** trim(arr\_delay) = ''

**or** arr\_delay = 'NA'**;** *-- 48*

)**;**

*--FD : Indica que há caracteres no campo dep\_delay. 34*

**update** flights **set** qa\_dep\_arr\_delay = **null;**

**update** flights **set** qa\_dep\_arr\_delay = 'FD'

**where** (dep\_delay) **is** **null**

**or** trim(dep\_delay) = ''

**or** dep\_delay = 'NA'

;

*--FA : Indica que há caracteres no campo arr\_delay -- 50*

**update** flights **set** qa\_dep\_arr\_delay = 'FA'

**where** (arr\_delay) **is** **null**

**or** trim(arr\_delay) = ''

**or** arr\_delay = 'NA'

**;**

*--5. Crie a coluna qa\_carrier e aponte inconsistências da coluna carrier de acordo com as regras abaixo:*

**ALTER** **TABLE** flights

**ADD** qa\_carrier Char(2)**;**

**select** \* **from** flights**;**

*--M : Indica que está com dado faltante.*

**update** flights **set** qa\_carrier = 'M'

**where** carrier **is** **null**

**or** length(trim(carrier)) = ''

**;**

*--F : Indica que não respeita o formato esperado (2 caracteres alfanuméricos).*

**update** flights **set** qa\_carrier = 'F'

**where** length(carrier) = 1**;**

**;**

*--6. Crie a coluna qa\_tailnum e aponte inconsistências da coluna tailnum de acordo com as regras abaixo:*

**ALTER** **TABLE** flights

**ADD** qa\_tailnum Char(2)**;**

**update** flights **set** qa\_tailnum = **null;**

*--M : Indica que está com dado faltante. --353*

**update** flights **set** qa\_tailnum = 'M'

**where** (tailnum **is** **Null**

**or** trim(tailnum) = ''

**or** tailnum = 'NA'**;**

*--R : Indica que a aeronave está ausente no dataset planes.*

**update** flights **set** qa\_tailnum = 'R'

**where** tailnum **not** **in**

(**select** tailnum **from** planes)

**;** *-- 353*

*--7. Crie a coluna qa\_flight e aponte inconsistências da coluna flight de acordo com as regras abaixo:*

**ALTER** **TABLE** flights

**ADD** qa\_flight Char(2)**;**

*--M : Indica que está com dado faltante.*

**update** flights **set** qa\_flight = 'M'

**where** flight **is** **Null**

**or** trim(flight ) = ''

**;**

*--F : Indica que o campo não está com formato numérico*

**update** flights **set** qa\_flight = 'F'

**where** to\_number(flight) **is** **null;**

*--8. Crie a coluna qa\_origin\_dest e aponte inconsistências da coluna origin, dest de acordo com as regras abaixo:*

**ALTER** **TABLE** flights

**ADD** qa\_origin\_dest Char(2)**;**

*--MO : Indica que está com dado faltante no origin.*

*--MD : Indica que está com dado faltante no dest.*

*--RO : Indica que o aeroporto origem está ausente no dataset airports.*

*--RD : Indica que o aeroporto destino está ausente no dataset airports.*

*--9. Crie a coluna qa\_air\_time e aponte inconsistencias da coluna air\_time de acordo com as regras abaixo:*

*-- M : Indica que está com dado faltante.*

*--I : Indica que o valor excede o intervalo [20, 500].*

*--F :Indica que o valor não é numérico.*

*--10. Crie a coluna qa\_distance e aponte inconsistências da coluna distance de acordo com as regras abaixo:*

*--M : Indica que está com dado faltante.*

*--I : Indica que o valor excede o intervalo [50, 3000].*

*--11. Crie a coluna qa\_distance\_airtime e aponte inconsistências entre as colunas distance e air\_time de acordo com as regras abaixo:*

*--M : Indica que está com distance ou air\_time faltante.*

*--TL : Indica que a viagem é longa de acordo com a condição: air\_time >= distance × 0.1 + 30.*

*--TS : Indica que a viagem é curta de acordo com a condição: air\_time <= distance × 0.1 + 10.*

*--TR : Indica que a viagem é normal caso as duas anteriores não sejam verdade.*