**BELK-Track-2 status :**

**18th march run :**

| **TABLE\_NAME** |  | **TOTAL\_RECORD\_COUNT\_HIVE** | **TOTAL\_RECORD\_COUNT\_SF** | **ROW\_COUNT\_ONLY\_IN\_HIVE** |  | **ROW\_COUNT\_ONLY\_IN\_SF** |  | **ROW\_COUNT\_DATA\_DISCREPANCY** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TRANSACTIONITEM |  | 331231 | 331231 | 0 |  | 0 |  | 331231 |
| C\_EDQ |  | 46242 | 46242 | 0 |  | 0 |  | 46241 |
| PRODUCT |  | 1068149 | 1068160 | 0 |  | 11 |  | 687 |
| C\_CUSTOMERCCTOKEN |  | 71828 | 71837 | 0 |  | 9 |  | 1 |
| FISCALCALENDAR |  | 10227 | 5866 | 6188 |  | 1827 |  | 4039 |
| ZIPDEMOGRAPHICS |  | 3542511 | 13677 | 3542511 |  | 13677 |  | 0 |
| HOLIDAYS |  | 712 | 1 | 711 |  | 0 |  | 1 |
| TRANSACTION |  | 108983 | 87739 | 23520 |  | 2276 |  | 13816 |
| CUSTOMERADDRESSXREF |  | 25850 | 0 | 25850 |  | 0 |  | 0 |
| C\_CUSTOMERCHANGEHISTORY |  | 741023 | 731379 | 741023 |  | 731379 |  | 0 |
| COUPON |  | 1508 | 0 | 1508 |  | 0 |  | 0 |
| TRANSACTIONPAYMENTMETHODXREF |  | 95940 | 88296 | 95940 |  | 88296 |  | 0 |
| CUSTOMER |  | 4218202 | 4800104 | 32548 |  | 614450 |  | 492485 |
| ADDRESS |  | 91486 | 91486 | 14088 |  | 14088 |  | 42961 |
| C\_CUSTOMERDEMOGRAPHIC |  | 14095 | 14088 | 7 |  | 0 |  | 2296 |
| EVENT |  | 32599245 | 32572446 | 32599245 |  | 32572446 |  | 0 |

**19 th march run :**

| **TABLE\_NAME** |  | **TOTAL\_RECORD\_COUNT\_HIVE** | **TOTAL\_RECORD\_COUNT\_SF** | **ROW\_COUNT\_ONLY\_IN\_HIVE** |  | **ROW\_COUNT\_ONLY\_IN\_SF** |  | **ROW\_COUNT\_DATA\_DISCREPANCY** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TRANSACTIONITEM** |  | 397185 | 397185 | 0 |  | 0 |  | 210682 |
| **PRODUCT** |  | 723998 | 724026 | 0 |  | 28 |  | 291 |
| C\_CUSTOMERCCTOKEN |  | 74737 | 74753 | 0 |  | 16 |  | 0 |
| FISCALCALENDAR |  | 10227 | 5866 | 6188 |  | 1827 |  | 4039 |
| ZIPDEMOGRAPHICS |  | 3542511 | 13677 | 3542511 |  | 13677 |  | 0 |
| HOLIDAYS |  | 712 | 1 | 711 |  | 0 |  | 1 |
| TRANSACTION |  | 118837 | 103116 | 16137 |  | 416 |  | 16599 |
| CUSTOMERADDRESSXREF |  | 228399 | 0 | 228399 |  | 0 |  | 0 |
| **C\_CUSTOMERCHANGEHISTORY** |  | 754428 | 745778 | 754428 |  | 745778 |  | 0 |
| C\_CUSTOMERDEMOGRAPHIC |  | 1777919 | 1777909 | 10 |  | 0 |  | 1551185 |
| COUPON |  | 2218 | 0 | 2218 |  | 0 |  | 0 |
| ADDRESS |  | 1920496 | 1924939 | 1757662 |  | 1762105 |  | 99935 |
| TRANSACTIONPAYMENTMETHODXREF |  | 122216 | 110735 | 122216 |  | 110735 |  | 0 |
| CUSTOMER |  | 6424702 | 7009691 | 29630 |  | 614619 |  | 19019 |
| EVENT |  | 67751594 | 33965021 | 67751594 |  | 33965021 |  | 0 |

20th march run 10am

| **TABLE\_NAME** |  | **TOTAL\_RECORD\_COUNT\_HIVE** | **TOTAL\_RECORD\_COUNT\_SF** | **ROW\_COUNT\_ONLY\_IN\_HIVE** |  | **ROW\_COUNT\_ONLY\_IN\_SF** |  | **ROW\_COUNT\_DATA\_DISCREPANCY** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PRODUCT |  | 1371057 | 1371082 | 0 |  | 25 |  | 98 |
| FISCALCALENDAR |  | 10227 | 5866 | 6188 |  | 1827 |  | 4039 |
| ZIPDEMOGRAPHICS |  | 3542511 | 13677 | 3542511 |  | 13677 |  | 0 |
| HOLIDAYS |  | 712 | 1 | 711 |  | 0 |  | 1 |
| EVENT |  | 29041209 | 29063848 | 29041209 |  | 29063848 |  | 0 |
| TRANSACTIONITEM |  | 10920 | 0 | 10920 |  | 0 |  | 0 |
| C\_CUSTOMERDEMOGRAPHIC |  | 1882472 | 1882470 | 2 |  | 0 |  | 1756606 |
| CUSTOMERADDRESSXREF |  | 147291 | 0 | 147291 |  | 0 |  | 0 |
| ADDRESS |  | 1882685 | 1884821 | 1849913 |  | 1852049 |  | 12188 |
| CUSTOMER |  | 2079585 | 2665298 | 28772 |  | 614485 |  | 614 |

20 march run @5pm

Green are done

Yellow work in progress

Red not started

| **TABLE\_NAME** |  | **TOTAL\_RECORD\_COUNT\_HIVE** | **TOTAL\_RECORD\_COUNT\_SF** | **ROW\_COUNT\_ONLY\_IN\_HIVE** |  | **ROW\_COUNT\_ONLY\_IN\_SF** |  | **ROW\_COUNT\_DATA\_DISCREPANCY** | **Resaon** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PRODUCT |  | 1371057 | 1371082 | 0 |  | 25 |  | 31 | C\_ITEMAGEWEEKS only 1 column |
| FISCALCALENDAR |  | 10227 | 5866 | 6188 |  | 1827 |  | 4039 |  |
| ZIPDEMOGRAPHICS |  | 3542511 | 13677 | 3542511 |  | 13677 |  | 0 |  |
| HOLIDAYS |  | 712 | 1 | 711 |  | 0 |  | 1 |  |
| TRANSACTIONITEM |  | 10920 | 0 | 10920 |  | 0 |  | 0 |  |
| C\_CUSTOMERDEMOGRAPHIC |  | 1882472 | 1882470 | 2 |  | 0 |  | 1687 |  |
| CUSTOMERADDRESSXREF |  | 147291 | 0 | 147291 |  | 0 |  | 0 |  |
| ADDRESS |  | 1882685 | 1884821 | 1849913 |  | 1852049 |  | 12188 |  |
| CUSTOMER |  | 2079585 | 2665297 | 28772 |  | 614484 |  | 621 |  |
| EVENT |  | 29041209 | 29063848 | 29041209 |  | 29063848 |  | 0 | Done |

21st March

| **TABLE\_NAME** |  | **TOTAL\_RECORD\_COUNT\_HIVE** | **TOTAL\_RECORD\_COUNT\_SF** | **ROW\_COUNT\_ONLY\_IN\_HIVE** |  | **ROW\_COUNT\_ONLY\_IN\_SF** |  | **ROW\_COUNT\_DATA\_DISCREPANCY** |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TRANSACTIONITEM |  | 729918 | 729918 | 0 |  | 0 |  | 423723 |  |
| PRODUCT |  | 1953259 | 1953275 | 0 |  | 16 |  | 24 |  |
| C\_CUSTOMERCCTOKEN |  | 131752 | 131779 | 0 |  | 27 |  | 0 |  |
| FISCALCALENDAR |  | 10227 | 5866 | 6188 |  | 1827 |  | 4039 |  |
| ZIPDEMOGRAPHICS |  | 3542511 | 13677 | 3542511 |  | 13677 |  | 0 |  |
| HOLIDAYS |  | 712 | 1 | 711 |  | 0 |  | 1 |  |
| TRANSACTION |  | 225869 | 209636 | 17071 |  | 838 |  | 37010 |  |
| C\_CUSTOMERCHANGEHISTORY |  | 1043099 | 1025967 | 17138 |  | 6 |  | 1174 |  |
| EVENT |  | 36055942 | 36667153 | 36055942 |  | 36667153 |  | 0 |  |
| C\_CUSTOMERDEMOGRAPHIC |  | 100 | 88 | 12 |  | 0 |  | 0 |  |
| COUPON |  | 4056 | 0 | 4056 |  | 0 |  | 0 | Main script is hardcoded with PUBLIC.coupon coupon instead of md5source |
| CUSTOMERADDRESSXREF |  | 22750 | 22837 | 1 |  | 88 |  | 0 |  |
| TRANSACTIONPAYMENTMETHODXREF |  | 241415 | 221750 | 241415 |  | 221750 |  | 0 |  |
| CUSTOMER |  | 4904537 | 5494813 | 23980 |  | 614256 |  | 23789 |  |
| ADDRESS |  | 169981 | 169981 | 88 |  | 88 |  | 108381 |  |

21 march –part 2

| **TABLE\_NAME** |  | **TOTAL\_RECORD\_COUNT\_HIVE** | **TOTAL\_RECORD\_COUNT\_SF** | **ROW\_COUNT\_ONLY\_IN\_HIVE** |  | **ROW\_COUNT\_ONLY\_IN\_SF** |  | **ROW\_COUNT\_DATA\_DISCREPANCY** |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ADDRESS |  | 169981 | 169981 | 0 |  | 0 |  | 8390 |  |
| TRANSACTIONITEM |  | 729918 | 729918 | 0 |  | 0 |  | 118 | C\_NONMERCHITEM,C\_CASHIER,C\_SUBTRANTYPE,C\_RECEIPTBARCODE,C\_SALESPERSON etc |
| PRODUCT |  | 1953259 | 1953275 | 0 |  | 16 |  | 24 | C\_ITEMAGEWEEKS only 1 column  See rounding ,celling floor |
| C\_CUSTOMERCCTOKEN |  | 131752 | 131779 | 0 |  | 27 |  | 0 |  |
| FISCALCALENDAR |  | 10227 | 5866 | 6188 |  | 1827 |  | 4039 |  |
| ZIPDEMOGRAPHICS |  | 3542511 | 13677 | 3542511 |  | 13677 |  | 0 |  |
| HOLIDAYS |  | 712 | 1 | 711 |  | 0 |  | 1 |  |
| TRANSACTION |  | 225869 | 225866 | 3 |  | 0 |  | 4 | Customer id mismatch  See with customer table |
| C\_CUSTOMERCHANGEHISTORY |  | 1043099 | 1025967 | 17138 |  | 6 |  | 1174 | Phone no mismatch |
| EVENT |  | 36055942 | 36667153 | 36055942 |  | 36667153 |  | 0 |  |
| C\_CUSTOMERDEMOGRAPHIC |  | 100 | 88 | 12 |  | 0 |  | 88 | Know changes, will apply in next run |
| COUPON |  | 4056 | 0 | 4056 |  | 0 |  | 0 | Main script is hardcoded with PUBLIC.coupon coupon instead of md5source  No custom script |
| CUSTOMERADDRESSXREF |  | 22750 | 22827 | 1 |  | 78 |  | 0 |  |
| TRANSACTIONPAYMENTMETHODXREF |  | 241415 | 221750 | 241415 |  | 221750 |  | 0 |  |
| CUSTOMER |  | 4904537 | 5494811 | 23980 |  | 614254 |  | 5169 | Need to look |
| ACCOUNT |  |  |  |  |  |  |  |  |  |
| C\_CUSTOMERCCTOKEN. |  |  |  |  |  |  |  |  |  |
| c\_customerchangehistory |  |  |  |  |  |  |  |  |  |
| C\_EDQ |  |  |  |  |  |  |  |  |  |
| c\_orderorigin |  |  |  |  |  |  |  |  |  |
| c\_pricetype |  |  |  |  |  |  |  |  |  |
| c\_restrictedstates |  |  |  |  |  |  |  |  | No data |
| c\_rfmrankpriority |  |  |  |  |  |  |  |  | No data |
| campaign |  |  |  |  |  |  |  |  |  |
| couponhistory |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

24 RUN 1ST

| **TABLE\_NAME** |  | **TOTAL\_RECORD\_COUNT\_HIVE** | **TOTAL\_RECORD\_COUNT\_SF** | **ROW\_COUNT\_ONLY\_IN\_HIVE** |  | **ROW\_COUNT\_ONLY\_IN\_SF** |  | **ROW\_COUNT\_DATA\_DISCREPANCY** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ADDRESS |  | 35 | 35 | 0 |  | 0 |  | 1 |
| PRODUCT |  | 1736624 | 1736648 | 0 |  | 24 |  | 42 |
| ZIPDEMOGRAPHICS |  | 3542511 | 13677 | 3542511 |  | 13677 |  | 0 |
| FISCALCALENDAR |  | 10227 | 5866 | 6188 |  | 1827 |  | 4039 |
| HOLIDAYS |  | 712 | 0 | 712 |  | 0 |  | 0 |
| TRANSACTIONITEM |  | 11058 | 0 | 11058 |  | 0 |  | 0 |
| EVENT |  | 28772933 | 0 | 28772933 |  | 0 |  | 0 |
| CUSTOMERADDRESSXREF |  | 2 | 30 | 2 |  | 30 |  | 0 |
| CUSTOMER |  | 203693 | 610539 | 203692 |  | 610538 |  | 1 |
| C\_CUSTOMERDEMOGRAPHIC |  | 37 | 35 | 2 |  | 0 |  | 0 |

PENDING ENTITY :

| **TABLE\_NAME** |  |
| --- | --- |
|  |  |
| TRANSACTION | Customer id mis |
|  |  |
| C\_CUSTOMERCHANGEHISTORY | Public schema is being used in MAIN\_500.sql script for c\_customerchangeHistory |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

Decide on a warehouse ?

25 th March

| **TABLE\_NAME** |  | **TOTAL\_RECORD\_COUNT\_HIVE** | **TOTAL\_RECORD\_COUNT\_SF** | **ROW\_COUNT\_ONLY\_IN\_HIVE** |  | **ROW\_COUNT\_ONLY\_IN\_SF** |  | **ROW\_COUNT\_DATA\_DISCREPANCY** |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ADDRESS |  | 136854 | 136854 | 0 |  | 0 |  | 7457 |  |
| TRANSACTIONITEM |  | 1406423 | 1406423 | 0 |  | 0 |  | 301 | Checked ,data difference is due to public is hardcoded instead of snapshot |
| PRODUCT |  | 1065263 | 1065307 | 0 |  | 44 |  | 3 | Due to null |
| C\_CUSTOMERCCTOKEN |  | 235507 | 235522 | 0 |  | 15 |  | 0 |  |
| ZIPDEMOGRAPHICS |  | 3542511 | 13677 | 3542511 |  | 13677 |  | 0 |  |
| FISCALCALENDAR |  | 10227 | 5866 | 6188 |  | 1827 |  | 4039 |  |
| HOLIDAYS |  | 712 | 0 | 712 |  | 0 |  | 0 |  |
| TRANSACTION |  | 423837 | 423833 | 4 |  | 0 |  | 4 |  |
| EVENT |  | 42591286 | 0 | 42591286 |  | 0 |  | 0 |  |
| C\_CUSTOMERCHANGEHISTORY |  | 1202873 | 1162242 | 40634 |  | 3 |  | 1286 | Public schema is being used in MAIN\_500.sql script for |
| COUPON |  | 2698 | 0 | 2698 |  | 0 |  | 0 |  |
| CUSTOMER |  | 5175311 | 5789329 | 5280 |  | 619298 |  | 7772 | OTS Changes - <https://acquiaps.atlassian.net/browse/BELK-331> |
| TRANSACTIONPAYMENTMETHODXREF |  | 462541 | 432979 | 462541 |  | 432979 |  | 0 |  |
| C\_CUSTOMERDEMOGRAPHIC |  | 57 | 39 | 18 |  | 0 |  | 0 |  |

PR list :

<https://github.com/AgilOneClientProjects/PolarisEAP/pull/2971/files>

1. **Entity** : C\_EDQ

**Field**: c\_storeId

**Actual** : 0025

**Expected** : 25

**Old code** : CAST(COALESCE(sparse.c\_storeId,dw.c\_storeId) AS STRING) AS c\_storeId

**New code** : CAST(CAST(COALESCE(sparse.c\_storeId,dw.c\_storeId) AS INT) AS STRING) AS c\_storeId

**Error Tag** : Leading\_zero

**Reason** :

*CAST(CAST(COALESCE(sparse.c\_storeId,dw.c\_storeId) AS INT) AS STRING) AS c\_storeId:*

* 1. First, COALESCE(sparse.c\_storeId,dw.c\_storeId) takes the first non-null value from sparse.c\_storeId and dw.c\_storeId.
  2. Then, CAST(... AS INT) attempts to cast this non-null value to an integer. If the value cannot be cast to an integer (e.g., if it's a non-numeric string), this step would result in an error.
  3. Finally, CAST(... AS STRING) casts the integer to a string. This is useful if you want to ensure that the value is a valid integer before converting it to a string, possibly to remove any leading zeros or to validate the data.

CAST(COALESCE(sparse.c\_storeId,dw.c\_storeId) AS STRING) AS c\_storeId:

* 1. Again, COALESCE(sparse.c\_storeId,dw.c\_storeId) takes the first non-null value.
  2. CAST(... AS STRING) directly casts the non-null value to a string without any intermediate casting to an integer. This means that the original format of the non-null value is preserved in the string representation, including any leading zeros or non-numeric characters.

1. **Entity** : PRODUCT

**Field**: C\_SUPPLIERDESC,C\_VENDORPARENTDESC,C\_SUPPLIERSTANDARDDESC,NAME,C\_STYLECOLORDESC

**Actual** : "LIGHTNING BUG\\\\WE CONNOR:WE CONNOR"

**Expected** : C\_SUPPLIERDESC",

"LIGHTNING BUG\\\\\\\\WE CONNOR:WE CONNOR

**Old code** : CAST(GET\_IGNORE\_CASE(raw.recordcontent,’C\_SUPPLIERDESC’) AS STRING)

**New code** : CAST(GET\_IGNORE\_CASE(raw.recordcontent,’C\_SUPPLIERDESC’) AS STRING)

**Error Tag** : Extra slash(/)

**Reason** :

Casting to string remove extra \\ this is not happening in hive

this is what we use in conformed to udm mapping , So we can ignore it

1. **Entity** : Customer

**Field**: C\_BESTCUSTFOREMAIL

**Actual(SF)** : Y

**Expected(HIVE)** : NULL

**Code\_change** : No

**Old code** :

**New code** :

**Error Tag** : Y\_and\_null

Action : mark it as ignore

**Reason** :

BI changes :

| SELECT  o.ID  ,CASE  WHEN o.Email IS NOT NULL AND o.rnk = 1 THEN 'Y'  ELSE NULL  END AS c\_bestCustForEmail  ,o.c\_erroneousName |
| --- |

1. **Entity** : C\_CUSTOMERCCTOKEN

**Field**: C\_PREFERREDCREDITCARD,C\_CCTYPE

**Actual(SF)** : null,null

**Expected(HIVE)** : true,BELK\_PRIVATE\_LABEL

**Code\_change** :

**Old code** :

**New code** :

**Error Tag** : null\_value\_overide\_in\_sf

**Action** :

**Reason/Findings** :

**Query** :

select \* from udm\_pv\_10169\_c\_customercctoken where sourcec\_customercctokennumber='816923945243007615172422';

**Output** :

C\_preferredcreditcard→true

C\_CCTYPE→BELK\_PRIVATE\_LABEL

**Query** :

select \* from udm\_s\_10169\_c\_customercctoken\_222133 where sourcec\_customercctokennumber='816923945243007615172422';

**Output** :

No data

**Query** : Data in dw\_snapshot

**Output:**

-- C\_PREFERREDCREDITCARD->true

-- C\_CCTYPE-->BELK\_PRIVATE\_LABEL

1. **Entity** : TRANSACTIONITEM

**Field**: DATEMODIFIED

**Actual(SF)** :"2025-03-18 06:15:32.720000000"

**Expected(HIVE)** : “2025-03-17 13:41:06.000000000"

**Code\_change** : Not required

**Old code** :

**New code** :

**Error Tag** : DATE\_IGNORE\_LIST

**Action** : Add this field to ignore list

**Reason/Findings** : Add datemodified field to ignore list as both hive and sf run at different time

1. **Entity** : C\_CUSTOMERCHANGEHISTORY

**Field**: ID

**Actual(SF)** :"2025-03-18 11:00:52.37800055ac0-e2a1-11ed-bd4b-0242ac110004"

**Expected(HIVE)** : “174229565237800055ac0-e2a1-11ed-bd4b-0242ac110004"

**Code\_change** : required

**Old code** :CONCAT(sparse.RowCreated,sparse.Sourcec\_customerchangehistoryNumber) AS ID

**New code** : CONCAT(TO\_NUMBER\_CUSTOM(sparse.RowCreated),sparse.Sourcec\_customerchangehistoryNumber) AS ID

**Error Tag** : DATE\_EPOC

**Action** : Change date to epoc

**Reason/Findings** : In hive date are stored in epoc format so id contain epoc

1. **Entity** : PRODUCT

**Field**: C\_ITEMAGEWEEKS

**Actual(SF)** :"488"

**Expected(HIVE)** : “489"

**sourceproductnumber\_EXAMPLE** : [469492381,230548178]

**Code\_change** : NO

**Old code**

**New code** :

**Error Tag** :

**Action** : For now ignoring

**Reason/Findings** :

C\_ITEMAGEWEEKS have 1 unit extra in hive than sf

No reference of code is found where we alter C\_ITEMAGEWEEKS.

1. **Entity** : Customer

**Field**: C\_ERRONEOUSNAME,C\_BESTCUSTFOREMAIL

**Actual(SF)** :"N"

**Expected(HIVE)** : “Y"

**ID\_EXAMPLE** : [FTP\_CSV\_6\_C0629102000084,230548178,FTP\_CSV\_6\_20220411057045739]

**Code\_change** :

**Old code** :

**New code** :

**Error Tag** :

**Action** :Skip

**Reason/Findings** : BI is changing sf\_bi\_customersummary\_PRE\_110.sql

1. **Entity** : Event

**Field**: NA

**Actual(SF)** :

**Expected(HIVE)** :

**Code\_change** :

**Old code** :

**New code** :

**Error Tag** :

**Action** : No

**Reason/Findings** :

In EVENT we took following as ID

[

md5(concat\_ws('|',coalesce(SOURCESYSTEMID, '|'),

coalesce(SOURCECUSTOMERNUMBER, '|'),

coalesce(SOURCEMESSAGENUMBER, '|'),

COALESCE(EVENTTIMESTAMP, 0::TIMESTAMP),

coalesce(SOURCEPRODUCTNUMBER, '|'),

coalesce(REALTIMEREQUESTNUMBER, '|'),

coalesce(MESSAGEID, '|'),

coalesce(EMAIL, '|'),

coalesce(BATCH, '|'),

coalesce(TYPE, '|'),

coalesce(LinkName, '|'),

coalesce(URL, '|'),

coalesce(Domain, '|')

,coalesce(c\_sfmcbatchid, '|')

)) as ID

]

1. **Entity** : c\_customerDemographic

**Field**: C\_ISBESTDEMOGRAPHIC

**Actual(SF)** : Y/N

**Expected(HIVE)** : NULL

**Code\_change** :

**Old code** :

**New code** :

**ID\_EXAMPLE** : []

**Error Tag** :

**Action** : Skip

**Reason/Findings** : BI is changing sf\_bi\_customersummary\_PRE\_106.sql

1. **Entity** : Customer

**Field**: EMAILSTATUS

**Actual(SF)** : v

**Expected(HIVE)** : u

**Code\_change** :

**Old code** :

**New code** :

**ID\_EXAMPLE** : ['FTP\_CSV\_1\_553538cccc5f93cf9079f7dd8bc22ef06b1472db32fffcb8f846ab30fcce1423,FTP\_CSV\_1\_12bdc8adca223a402fa5999b8843835026a2600a3605e8765d72f4c2bc92a279]

**Error Tag** :

**Action** : Skip

**Reason/Findings** : Mellisa might be changing this

1. **Entity** : C\_CUSTOMERDEMOGRAPHIC

**Field**: BATCH etc

**Actual(SF)** :

**Expected(HIVE)** :

**Code\_change** :

**Old code** :

**New code** :

**ID\_EXAMPLE** : [20220411004202446",

"20220411008561124"]

**Error Tag** :

**Action** : Need to verify

**Reason/Findings** : DW\_HIVE\_INC is not updated. Seems to be a timing issue.

1. **Entity** : CUSTOMER

**Field**: C\_NUMERICCUSTOMERNUMBER etc

**Actual(SF)** :

**Expected(HIVE)** :

**Code\_change** :

**Old code** :

**New code** :

**ID\_EXAMPLE** : [KFK\_0\_7288007",

"KFK\_0\_48899308"]

**Error Tag** :

**Action** : "filters": "sourcesystemid <> 'KFK\_0'"

**Reason/Findings** :

Ignoring KFK\_0

1. **Entity** : Customer

**Field**: c\_syf% columns

**Actual(SF)** : NULL

**Expected(HIVE)** : NON-NULL values

**Code\_change** : No

**Old code** :

**New code** :

**Error Tag** : OTS Changes - <https://acquiaps.atlassian.net/browse/BELK-331>

Action : mark it as ignore

1. **Entity** : c\_CustomerChangeHistory

**Field**: c\_primaryphone, c\_mobilephone

**Actual(SF)** : NULL

**Expected(HIVE)** : NON-NULL values

**Code\_change** : No

**Old code** :

**New code** :

**Error Tag** : Public schema is being used in MAIN\_500.sql script for c\_customerchangeHistory

Action : mark it as ignore

1. **Entity** : transactionitem

**Field**: custom attributes

**Actual(SF)** : NULL

**Expected(HIVE)** : NON-NULL values

**Code\_change** : No

**Old code** :

**New code** :

**Error Tag** : Deduping is happening twice, once in TI\_PRE\_115.sql and then in MAIN\_500.sql. Therefore, it’s picking values from DW\_SNAPSHOT if NULL is found for any attribute.

Action : mark it as ignore

General Guideline :

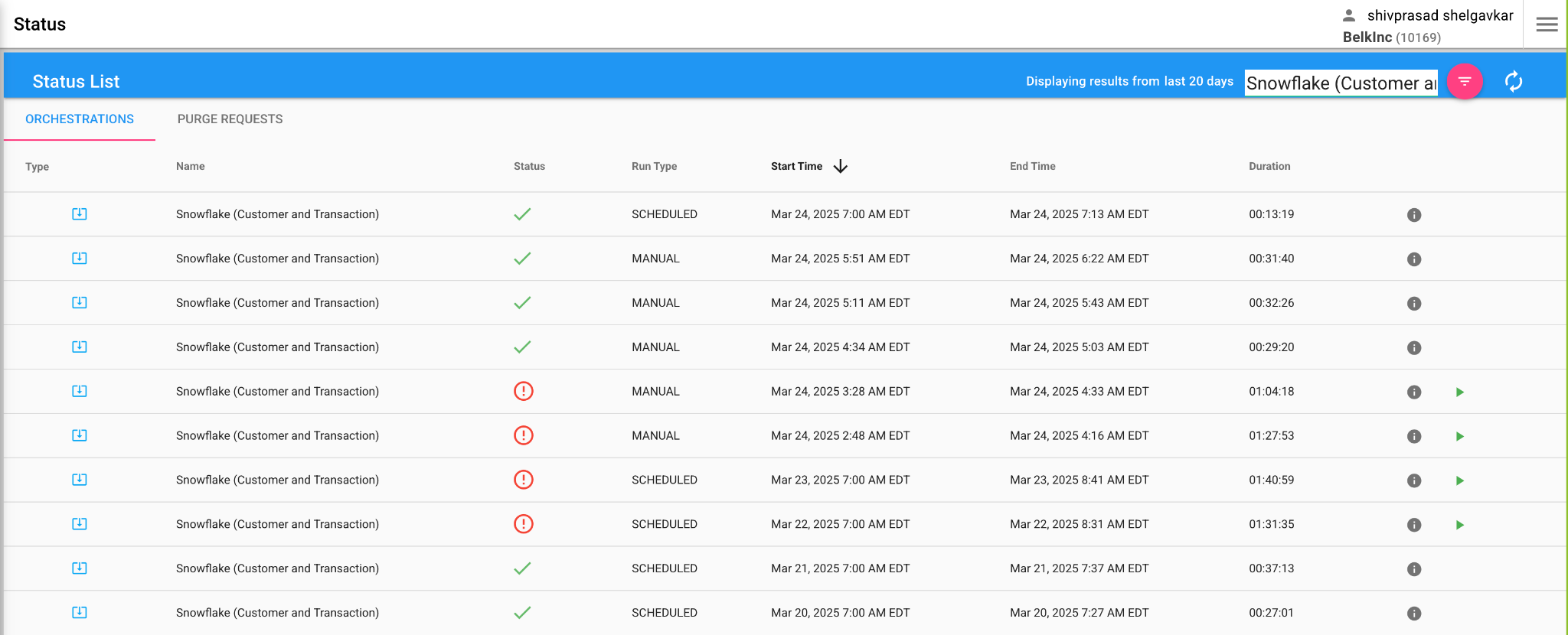
1. REGEXP\_REPLACE should not be replaced with REGEXP\_SUBSTR
2. THEN REGEXP\_REPLACE(COALESCE(SourceCustomerNumber, ''), '^FTP\_CSV\_**\\d\_**', '') was not working on sf added ‘+’

THEN REGEXP\_REPLACE(COALESCE(SourceCustomerNumber, ''), '^FTP\_CSV\_**\\d+**\_', '') ’

1. Check if any OTS are run on snowflake and not on Parque (OTS Changes - <https://acquiaps.atlassian.net/browse/BELK-331>)
2. Check if Public schema is being used in any script instead of dw\_snapshot

Workflow Failure

1. AIF Workflow for transaction failed on 22nd and 22rd March, 2025.



This is brief about the process of data validation on prod.

We start with *no knowledge* about features and offerings provided to a tenant.

* We do not analyse the script before starting to validate the data, because not all scripts will have issues. We focus on scripts which show some discrepancies.
* A lot of times even after successful data validation we don't know what a script is trying to achieve.
* We execute dw workflow in shadow mode.
* The application compares results from the Hive pipeline with that of the shadow pipeline on Snowflake.
* A lot of times due to some missed steps we do not get accurate results in shadow mode.
* After running shadow mode, the application creates metrics about the deviation of the data in two pipelines.
* Many times Snowflake pipeline is correct in its own form where the hive generates bad records, they are ignored in Hive because they are minor but they do cause our application to report them as errors.
* Now in each of the reported discrepancies we first check the discrepancy manually by querying the database.
* Next we analyse the script and try to find where the impacted column is being modified.
* Then we try to understand why there is a discrepancy in that column.
* There could be multiple reasons for it, wrong function, wrong syntax, missed scripts, modified by bi workflow or in worst case some unhandled data value.
* After fixing the issue, we execute the scripts again to generate the whole table again.
* We compare the new table so generated.
* If there are more discrepancies, we repeat the process.