# Introduction

The simplified process of taking raw data from source system to providing reports/kpis and requisite process for Data Quality Assurance is shown below:



1. Source System- provides data to DAAS to process and will require Data Quality Assurance at key stages, namely
   1. File arrival
      1. Currently in BiB, manual checks are done to see if there are .tmp files which are then rename to .txt to process. We recently discovered that sometimes the .tmp files are incomplete. The requirement will be for DAAS Operation to pick this file, move them to a temporary location and request for the replacement file from the source systems. Once the full file is received, the .tmp file could then be removed from the temporary holding location.
      2. The arrival statistics is done in the file ingestion stage.
   2. File Ingestion
      1. File loaded are catalog by filename, and number of rows in the file for reconciliation
      2. Trend is built form the catalog/metadata usually on a daily basis to check for consistent pattern of daily delivery. DAAS could take this further by adding hourly per day for early preemptive action.
      3. SMS Alerts are built around the trend to inform the team and based on the criticality of the situation, a manual communication/forwarding of the alert to the MTN ITS Management team is done
2. Processing on DAAS

As discussed, Saeed to help get the DAAS team to contribute on what to monitor on processing of Data on DAAS.

DaaS Processing Summary

Staging

* 1. BeeGFS
  2. Ingested into LLF (ODS)
  3. Tarred three days after into ARCHIVE
  4. Double tarred again into ARCHIVE/subfolder

Low Level Facts (LLF)

1. File ingestion happens on the from staging environment to LLF
2. Staged Data are cleaned (Kamanja->Flare [models] and then moved to LLF and Dimensions (type 2 dimensions)
   1. Duplicate check – how to monitor duplicate check metadata
   2. Rejected Data – how do we monitor rejected data metadata
3. Monitoring reports to reconcile daily, all data in STAGING have been moved to low level facts, any discrepancies are investigated and resolved by the team
   1. Data reconciliation/completeness metadata – Validation report table

Exploitation Layers– BSL Calculation for Customer Subject Areas

1. LLF are further processed into Daily Summaries, Weekly and monthly summaries
2. Monitoring here is by way of trending the daily rows summarized based on available meta data, and the actual KPI monitoring as covered in 3 below
3. Processing metadata required for monitoring and early detection of slow performance processes.
4. Reports/KPIs
   1. Static Daily KPIs (RGS, Revenue, Churn, Churn Returnee, Platform Revenue Reconciliation Variance etc.) are trend on daily basis – these need to be persistent and not change overtime once released.
   2. Any abnormal that cannot be resolved within available time for report delivery is communicated as delay in report while investigation continues.

Where the issue still could not be identified, MTN ITS Management (Yetunde) will be involved to make the call to release the report as is after providing her with all the monitoring reports. The report is then released with qualification that there may be issue with the report and investigation is ongoing. Usually problem traced to source system changing something without informing the team, that may require changes in the report soft rules.

* 1. Note that key KPIs are used to provide the required level of confidence to the business, Platform Revenue Reconciliation which is usually around 0.5% on the average. (I learnt from MTNN that this is current not in scope of DAAS).

1. Report Delivery Monitoring and Tracking
   1. All reports/KPIs by email, sms, and on Dashboards/BI tools are tracked on daily basis for delivery based on agreed SLA
   2. Data exports are also tracked in a similar manner.
   3. These are done for Daily and Monthly reports/KPIs and update sent to MTNN ITS management

# Source Systems

## File arrival

## Direct File from Source System

Files from different source systems arrives at designated file directories for ingestions. The files are first written as .tmp but renamed as .txt on completion.

The current file based source systems are the following:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Data Source** | **Description** | **Landing Directory** | **Directory for Incomplete File (tmp)** | **Final Storage Object** |
| bundle4u\_gprs |  |  |  |  |
| bundle4u\_voice |  |  |  |  |
| cb\_serv\_mast\_view |  |  |  |  |
| cs5\_air\_adj\_da |  |  |  |  |
| cs5\_air\_adj\_ma |  |  |  |  |
| cs5\_air\_refill\_ac |  |  |  |  |
| cs5\_air\_refill\_da |  |  |  |  |
| cs5\_air\_refill\_ma |  |  |  |  |
| cs5\_ccn\_gprs\_ac |  |  |  |  |
| cs5\_ccn\_gprs\_da |  |  |  |  |
| cs5\_ccn\_gprs\_ma |  |  |  |  |
| cs5\_ccn\_sms\_ac |  |  |  |  |
| cs5\_ccn\_sms\_da |  |  |  |  |
| cs5\_ccn\_sms\_ma |  |  |  |  |
| cs5\_ccn\_sms\_ma\_v |  |  |  |  |
| cs5\_ccn\_voice\_ac |  |  |  |  |
| cs5\_ccn\_voice\_da |  |  |  |  |
| cs5\_ccn\_voice\_ma |  |  |  |  |
| cs5\_ccn\_voice\_ma\_v |  |  |  |  |
| cs5\_sdp\_acc\_adj\_ac |  |  |  |  |
| cs5\_sdp\_acc\_adj\_da |  |  |  |  |
| cs5\_sdp\_acc\_adj\_ma |  |  |  |  |
| cs5\_sdp\_acc\_adj\_ma\_tmp |  |  |  |  |
| cug\_access\_fees |  |  |  |  |
| fin\_log |  |  |  |  |
| ggsn\_cdr |  |  |  |  |
| maps2g |  |  |  |  |
| maps3g |  |  |  |  |
| maps4g |  |  |  |  |
| mobile\_money |  |  |  |  |
| msc\_cdr |  |  |  |  |
| sdp\_dmp\_ma |  |  |  |  |
| sgsn\_cdr |  |  |  |  |

Required Quality Check Reports. Note that each one of the columns will be chatted on a separate time series graph with additional graph to show day of the week series.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Data Source** | **No of Files Received** | **No of Records Received** | **No of File Left in Landing** | **Number of Incomplete File (tmp)** | **Number of Records loaded for the day in the Final Objects** |
| bundle4u\_gprs |  |  |  |  |  |
| bundle4u\_voice |  |  |  |  |  |
| cb\_serv\_mast\_view |  |  |  |  |  |
| cs5\_air\_adj\_da |  |  |  |  |  |
| cs5\_air\_adj\_ma |  |  |  |  |  |
| cs5\_air\_refill\_ac |  |  |  |  |  |
| cs5\_air\_refill\_da |  |  |  |  |  |
| cs5\_air\_refill\_ma |  |  |  |  |  |
| cs5\_ccn\_gprs\_ac |  |  |  |  |  |
| cs5\_ccn\_gprs\_da |  |  |  |  |  |
| cs5\_ccn\_gprs\_ma |  |  |  |  |  |
| cs5\_ccn\_sms\_ac |  |  |  |  |  |
| cs5\_ccn\_sms\_da |  |  |  |  |  |
| cs5\_ccn\_sms\_ma |  |  |  |  |  |
| cs5\_ccn\_sms\_ma\_v |  |  |  |  |  |
| cs5\_ccn\_voice\_ac |  |  |  |  |  |
| cs5\_ccn\_voice\_da |  |  |  |  |  |
| cs5\_ccn\_voice\_ma |  |  |  |  |  |
| cs5\_ccn\_voice\_ma\_v |  |  |  |  |  |
| cs5\_sdp\_acc\_adj\_ac |  |  |  |  |  |
| cs5\_sdp\_acc\_adj\_da |  |  |  |  |  |
| cs5\_sdp\_acc\_adj\_ma |  |  |  |  |  |
| cs5\_sdp\_acc\_adj\_ma\_tmp |  |  |  |  |  |
| cug\_access\_fees |  |  |  |  |  |
| fin\_log |  |  |  |  |  |
| ggsn\_cdr |  |  |  |  |  |
| maps2g |  |  |  |  |  |
| maps3g |  |  |  |  |  |
| maps4g |  |  |  |  |  |
| mobile\_money |  |  |  |  |  |
| msc\_cdr |  |  |  |  |  |
| sdp\_dmp\_ma |  |  |  |  |  |
| sgsn\_cdr |  |  |  |  |  |

## Data from Database (RDBMS) Source System

dim\_bill\_type

dim\_ dmc\_dump\_all

## File Ingestion

critique