

Data Monetization and Customer Experience (CX) Optimization using telco data assets: Amazon CloudFormation stack details

Following key resources will be created in the region where you deploy the stack:

- Amazon VPC with
 - 2 public subnets
 - 2 private subnets
 - associated network configuration
- Amazon S3 bucket to host
 - directory structure and sample data files.
 - to host AWS Glue temporary files, if not already exists.
 - **Dictionaries**
 - **DPI Dictionary** (sample provided). You can choose to build your own dictionary or license it from a 3rd party. A few probe vendors package this dictionary in their offering. URL and User Agent (UA) attributes under HTTP traffic can also be leveraged to provide granular level of content classification and categorization. Although, HTTP traffic is small portion of overall traffic nowadays, effectiveness of classification based on URL/UA becomes significantly stronger when combined with classification and categorization based on DPI detected application/website.

DPI dictionary sample used in this blog closely follows IAB (Interactive Advertising Bureau) two tier categorization. Carriers, depending on the market they operate, can adopt multi-tier categorization dictionary with customized IAB framework suited for their marketing requirements. Sensitive content categories in each market must be masked for consumers with Privacy labels.
 - **Device Dictionary** (sample provided). You can license it from GSMA, a 3rd party or can generate from your in-house device management solution. Device attributes are not included in the provided sample. Device dictionaries usually comes with hundreds of device attribute as a function of device type. These attributes play very important role in several use cases like targeting subscribers to adopt contactless payment methods via their handset with NFV capable handsets or prioritizing 4G/5G RAN (Radio Access Network) roll out by targeting areas with high value subs and 4G/5G capable handsets.
- IAM Roles for
 - AWS Glue
 - Amazon Redshift
 - AWS Lambda
- Amazon Redshift cluster, associated security groups and schema **cmdm**, tables with attributes descriptions and procedures.

Telco CX and Data Monetization CF Stack Details

#	Amazon Redshift Objects	Type	Description
1	f_raw_dpi	Table	Stores filtered, transformed and enriched raw dpi data
2	f_hourly_dpi	Table	Stores aggregated hourly dpi data
3	f_daily_dpi	Table	Stores aggregated daily dpi data
4	d_customer_demographics	Table	Stores customer demographics attributes.
5	d_device	Table	Stores basic device dictionary data.
6	d_dpi_dictionary	Table	Stores dpi dictionary data
7	d_location	Table	Stores 3G and 4G normalized cell/site configuration data.
8	d_operator_plmn	Table	Stores operator plmn (MCC + MNC) to country and carrier mapping data
9	d_tac	Table	Stores IMEI TAC (Type Allocation Code) data to device dictionary mapping data
10	d_tariff_plan	Table	Stores customer and subscription mapping data
11	d_tariff_plan_desc	Table	Stores tariff plan details data
12	d_audit	Table	Stores basic audit data (read rows, filtered rows, db loaded rows) for all fact processing and aggregation.
13	f_hourly_dpi	Procedure	Amazon Redshift procedure for dpi data hourly aggregations.
14	f_daily_dpi	Procedure	Amazon Redshift procedure for dpi data daily aggregations. This procedure reads data from dpi raw table but can also read data from hourly table.

Note: Amazon Redshift DDLs are kept at basic level for this walkthrough.

- Amazon Secret Manager with Amazon Redshift credentials
- AWS Lambda functions for
 - AWS Glue catalog table attribute descriptions update
 - hourly DPI data aggregation in Amazon Redshift
 - daily DPI data aggregation in Amazon Redshift
- AWS Glue
 - Catalog Database

#	Database	Description
1	cemdm	This database contains all AWS Glue cataloged tables related to customer experience management and data monetization

- Amazon Redshift connection (Data Warehouse)

Telco CX and Data Monetization CF Stack Details

#	Connection	Description
1	redshift- <i><cloudformation stack name></i>	This connection is used in Glue ETL Jobs to connect to Amazon Redshift

- Crawlers (discover data stored on Amazon S3)

#	Crawler	Source (s3:// <i><Bucket Name>/</i>)	Description
1	DimensionDataCrawler	dimensions/	Catalog all dimensional data
2	DPIRawDataCrawler	dpi/raw/incoming/	Catalog incoming raw dpi data
3	DPIProcessedDataCrawler	dpi/processed/	Catalog all processed datasets.

- Jobs (transform data stored on Amazon S3 and load it in Amazon Redshift tables)

#	Job	Source (s3:// <i><Bucket Name>/..</i>)	Target: Amazon Redshift table (schema: cemdm)	Job Bookmark	Description
1	d_crm_demographics	dimensions/crm_demographics/	d_customer_demographics	Enable	Loads transformed and enriched subscriber demographics data.
2	d_device	dimensions/device/	d_device	Enable	Loads device dictionary data.
3	d_dpi_dictionary	dimensions/dpi_dictionary/	d_dpi_dictionary	Enable	Loads dpi dictionary with media category and sub-category data.
4	d_location	dimensions/network_cm_3g/ dimensions/network_cm_4g/	d_location	Enable	Load transformed and enriched Network 3G and 4G Configuration Management cell/site data.
6	d_operator_plmn	dimensions/operator_plmn/	d_operator_plmn	Enable	Loads operator/carrier plmn data.
7	d_tac	dimensions/tac/	d_tac	Enable	Loads transformed IMEI tac to device dictionary mapping data

Telco CX and Data Monetization CF Stack Details

8	d_tariff_plan	dimensions/tariff_plan/	d_tariff_plan	Enable	Loads transformed tariff plan customer mapping data
9	d_tariff_plan_desc	dimensions/tariff_plan_desc/	d_tariff_plan_desc	Enable	Loads transformed tariff plan description data
10	raw_dpi_incoming	dpi/raw/incoming/	f_raw_dpi_d_audit	Disable	Loads PS probe dpi data and moves files to raw/dpi/archive/s3 folder after processing the files. This job takes in job parameters like home plmn and 3-digit country code mcc values. The job also does basic auditing to capture total number of records read from source, filtered and loaded in Amazon Redshift.

AWS Glue tracks data that has already been processed during a previous run of an ETL job by persisting state information from the job run. This persisted state information is called a *job bookmark*. Read more about AWS Glue job bookmarks [here](#).