MELD File Transfer to Athena

Interface Control Document

Version 1.1 may 26, 2016

**Draft**

# SCOPE

This document describes the data flow of viewing events from MELD to ATHENA through data file transfer

1.1       System Identification  
1.1.1       MELD

MELD is the enterprise Hadoop environment processing set top box events into viewing events for the enterprise. MELD will be providing unfiltered STB viewing records to the Athena system.

1.1.2       Athena

Athena is the enterprise Advanced Advertising platform for Comcast. Athena monetized viewing data in support of advanced advertising initiatives.

# CONCEPT OF OPERATIONS

## Interface Overview

Viewing events will be gathered into data files for transfer to Athena. There are currently two workflows processing STB events: native and X1. These workflows operate independently sending data files to Athena. Data files will be identified by date-time and source. These files will be transferred hourly through the day every day.

Secure copy

MELD

Athena

## Data Transfer

Data files will be transferred using SCP protocol to the Athena edge node identified as: **vcpq2spc-ch2-a02s.sys.comcast.net** (10.54.231.135)

The user account **tsgmeldftp** will be used to authenticate to the Athena edge node. Additionally all data files will be gpg encryption using a previously provided key for [**jamie\_geyer1@cable.comcast.com**](mailto:jamie_geyer1@cable.comcast.com)**.** For each data file there will be an additional metadata file sent containing the md5 of the data file along with the viewing event count.

These data files will be stored on the Athena edge node in the following directory:

**~tsgmeldftp/ingest/<pdate>/daily/**

Where pdate (processing date ) is in the format of YYYY-MM-dd

The naming format for data files are:

athena\_<stype>\_yyyyMMddHHmm.gpg (datafile)

athena\_<stype>\_cnt\_yyyyMMddHHmm.txt (metadata file)

Where stype (system type) bali or X1

The metadata file will always be transferred after the data file to signal the completion of data transfer.

If there is a need to reprocess data due to any issue the reprocessed data files will be placed in the following directory:

**~tsgmeldftp/ingest/<pdate>/reprocessed/**

Conventions and naming standards will follow those files placed the in the hourly folder.

## TransactionS

Data and metadata files will be transferred on an hourly basis. It is anticipated that due to the multi-tenant nature of MELD, processing can be delayed on the system and there could be gaps between hourly files. Records that would be reported during those gaps will be contained in the next data file sent.

The naming format for data files are:

athena\_<stype>\_yyyyMMddHHmm.gpg (datafile)

athena\_<stype>\_cnt\_yyyyMMddHHmm.txt (metadata file)

Where stype (system type) bali or X1

The metadata file will always be transferred after the data file to signal the completion of data transfer.

The content of the ascii text meta data file will follow the following format:

<datafilename> : <md5sum>

count : <event record count>

The datafile content will be ascii text with fields delimted by tabs and records delimited by newlines

The format and order of the 188 fields is defined in the following table below:

|  |  |  |
| --- | --- | --- |
| Position | Name | Type |
| 1 | di\_device\_id | string |
| 2 | di\_account\_num | string |
| 3 | di\_encoding\_cd | string |
| 4 | geo\_id | string |
| 5 | geo\_type\_cd | string |
| 6 | time\_zone | string |
| 7 | controller\_id | string |
| 8 | start\_timestamp\_local | timestamp |
| 9 | end\_timestamp\_local | timestamp |
| 10 | start\_timestamp\_utc | timestamp |
| 11 | end\_timestamp\_utc | timestamp |
| 12 | duration\_seconds | decimal(38,10) |
| 13 | device\_platform\_type\_cd | string |
| 14 | device\_type\_cd | string |
| 15 | media\_platform\_type\_code | string |
| 16 | service\_type\_cd | string |
| 17 | tuner\_status\_cd | string |
| 18 | channel\_number | string |
| 19 | source\_id | string |
| 20 | source\_type\_cd | string |
| 21 | source\_type\_code | string |
| 22 | start\_offset\_seconds | decimal(38,10) |
| 23 | end\_offset\_seconds | decimal(38,10) |
| 24 | record\_timestamp\_local | timestamp |
| 25 | record\_timestamp\_utc | timestamp |
| 26 | pre\_initiation\_context\_cd | string |
| 27 | initiation\_cd | string |
| 28 | termination\_cd | string |
| 29 | start\_event\_sequence | string |
| 30 | end\_event\_sequence | string |
| 31 | status | string |
| 32 | unique\_session\_id | string |
| 33 | viewing\_end\_timestamp\_local | timestamp |
| 34 | viewing\_end\_timestamp\_utc | timestamp |
| 35 | viewing\_duration\_seconds | decimal(38,10) |
| 36 | program\_viewing\_duration\_seconds | decimal(38,10) |
| 37 | viewing\_cap\_rule\_id | string |
| 38 | station\_id | string |
| 39 | station\_digicableid | string |
| 40 | station\_title | string |
| 41 | station\_callsign | string |
| 42 | station\_secondarydigicableids | string |
| 43 | station\_onscreencallsign | string |
| 44 | station\_hd | string |
| 45 | station\_language | string |
| 46 | station\_otachannelnumber | int |
| 47 | station\_shortname | string |
| 48 | station\_timezone | string |
| 49 | station\_stationtype | string |
| 50 | station\_dmaregionid | string |
| 51 | station\_tagids\_title\_type | string |
| 52 | station\_merlin\_station\_key | string |
| 53 | channel\_id | string |
| 54 | channel\_channelnumber | string |
| 55 | channel\_locationid | string |
| 56 | channel\_title | string |
| 57 | channel\_digicableid | int |
| 58 | channel\_callsign | string |
| 59 | channel\_onscreencallsign | string |
| 60 | channel\_hd | string |
| 61 | channel\_language | string |
| 62 | channel\_otachannelnumber | int |
| 63 | channel\_shortname | string |
| 64 | channel\_timezone | string |
| 65 | channel\_stationtype | string |
| 66 | channel\_merlin\_channel\_key | string |
| 67 | company\_id | string |
| 68 | company\_title | string |
| 69 | company\_distributor | string |
| 70 | company\_musicdistributor | string |
| 71 | company\_displayname | string |
| 72 | company\_studio | string |
| 73 | company\_label | string |
| 74 | company\_description | string |
| 75 | company\_networktype | string |
| 76 | company\_tagids\_title\_type | string |
| 77 | company\_merlin\_company\_key | string |
| 78 | stationcompany\_associationtype | string |
| 79 | unique\_tune\_id | string |
| 80 | listing\_id | string |
| 81 | listing\_starttime | timestamp |
| 82 | listing\_endtime | timestamp |
| 83 | listing\_airingtype | string |
| 84 | listing\_audiotype | string |
| 85 | listing\_captiontype | string |
| 86 | listing\_contentratingscheme | string |
| 87 | listing\_contentrating | string |
| 88 | listing\_contentsubratings | string |
| 89 | listing\_hdlevel | string |
| 90 | listing\_seriesid | string |
| 91 | listing\_showingtype | string |
| 92 | listing\_subjecttoblackout | string |
| 93 | listing\_subtitled | string |
| 94 | listing\_sap | string |
| 95 | listing\_dvrprogramid | string |
| 96 | listing\_dvrseriesid | string |
| 97 | listing\_descriptivevideoservice | string |
| 98 | listing\_merlin\_listing\_key | string |
| 99 | program\_id | string |
| 100 | program\_title | string |
| 101 | program\_shorttitle | string |
| 102 | program\_mediumtitle | string |
| 103 | program\_longtitle | string |
| 104 | program\_year | int |
| 105 | program\_shortsynopsis | string |
| 106 | program\_mediumsynopsis | string |
| 107 | program\_longsynopsis | string |
| 108 | program\_runtime | int |
| 109 | program\_type | string |
| 110 | program\_language | string |
| 111 | program\_contentratingscheme | string |
| 112 | program\_contentrating | string |
| 113 | program\_contentsubratings | string |
| 114 | program\_starrating | int |
| 115 | program\_category | string |
| 116 | program\_firstruncompanyid | string |
| 117 | program\_adult | string |
| 118 | program\_local | string |
| 119 | program\_listbytitle | string |
| 120 | program\_sportssubtitle | string |
| 121 | program\_releasedate | string |
| 122 | program\_seriesid | string |
| 123 | program\_tvseasonid | string |
| 124 | program\_originalairdate | string |
| 125 | program\_tvseasonepisodenumber | int |
| 126 | program\_seriesepisodenumber | int |
| 127 | program\_tvseasonnumber | int |
| 128 | program\_episodetitle | string |
| 129 | program\_productionepisodenumber | string |
| 130 | program\_tagids\_title\_type | string |
| 131 | program\_merlin\_program\_key | string |
| 132 | adm\_customer\_account\_id | int |
| 133 | adm\_upgrade\_cdv\_score | double |
| 134 | adm\_upgrade\_video\_score | double |
| 135 | adm\_upgrade\_hsd\_score | double |
| 136 | adm\_customer\_segment\_id | int |
| 137 | adm\_customer\_segment\_name | string |
| 138 | adm\_active\_clv | double |
| 139 | adm\_active\_clv\_nat\_decile | string |
| 140 | adm\_active\_clv\_div\_decile | string |
| 141 | adm\_star\_segment | string |
| 142 | adm\_nine\_box\_segment | string |
| 143 | adm\_division | string |
| 144 | adm\_region | string |
| 145 | adm\_city | string |
| 146 | adm\_state | string |
| 147 | adm\_zip | string |
| 148 | adm\_product\_mix | string |
| 149 | adm\_hsd\_tier\_name | string |
| 150 | adm\_video\_tier\_name | string |
| 151 | adm\_cdv\_tier\_name | string |
| 152 | adm\_xh\_tier\_name | string |
| 153 | adm\_mrc | double |
| 154 | adm\_ebill | int |
| 155 | adm\_account\_status | string |
| 156 | adm\_mrc\_net\_offer\_amt | double |
| 157 | adm\_blastplus\_ind | tinyint |
| 158 | adm\_nternetplus\_ind | tinyint |
| 159 | adm\_x1\_platform | tinyint |
| 160 | adm\_customer\_type | string |
| 161 | adm\_tenure\_by\_pkey | int |
| 162 | adm\_offer\_name | string |
| 163 | adm\_offer\_months\_duration | int |
| 164 | adm\_contract\_term\_name | string |
| 165 | adm\_contract\_start\_date | string |
| 166 | adm\_contract\_term\_period | int |
| 167 | adm\_dma | string |
| 168 | vodasset\_vod\_asset\_id | string |
| 169 | vodasset\_provider\_id | string |
| 170 | vodasset\_title\_asset\_id | string |
| 171 | vodasset\_product | string |
| 172 | vodasset\_title | string |
| 173 | vodasset\_title\_description | string |
| 174 | vodasset\_actors\_display | string |
| 175 | vodasset\_display\_run\_time | string |
| 176 | vodasset\_episode\_id | string |
| 177 | vodasset\_episode\_name | string |
| 178 | vodasset\_licensing\_window\_start\_date | timestamp |
| 179 | vodasset\_licensing\_window\_end\_date | timestamp |
| 180 | vodasset\_mpaa\_or\_tv\_rating | string |
| 181 | vodasset\_genre\_display | string |
| 182 | vodasset\_provider\_name | string |
| 183 | vodasset\_suggested\_price | decimal(38,2) |
| 184 | meld\_audit\_ts | string |
| 185 | src\_flag | string |
| 186 | local\_day\_id | string |
| 187 | box\_type | string |
| 188 | media\_platform\_type\_cd | string |

## SECURITY and Integrity

The security of the data will be enforced using a gpg encrypted file using the Athena supplied key. Additionally, the data will be transferred using a secure transfer protocol. Access to the data is limited to those with access to the protected private key. It is MELD’s assertion that Athena personnel will be safeguarding the private key using industry accepted standards.

# Qualification Methods

There are four methods provided to ensure the data from MELD can be validated

1. Metadata file has been received – signifying a complete data file transfer
2. MD5 contained in metadata file matches that of the corresponding data file
3. Data file can be decrypted using protected private key
4. Record count contained in metadata file matches the record count in the data file

Failures to transfer files will be alarmed and retried at regular intervals. In the event of repeated failures. Athena tier 1 support will be engaged to resolve the issue.

# Notes

# Appendices

# ApprovalS

|  |  |  |
| --- | --- | --- |
| Name | Team | Date |
| Michelle Vagnati | Advanced Advertising |  |
| Greg Miller | Enterprise Business Intelligence |  |
|  |  |  |
|  |  |  |

# Record of Changes

|  |  |  |
| --- | --- | --- |
| Change | Date | Description |
| V1.0 | May 25, 2016 | Initial Draft |
| v1.1 | May 26,2016 | Removed 3 duplicate fields S2.3 (local\_day\_id, box\_type and media\_platform\_type\_cd) |
|  |  |  |