

Setting up ml-survey-distinctCount-status datasource

The survey consumption report on Admin Dashboard is pointing to ml-survey-status therefore it showing incorrect values. Hence we are created new script `pyspark_sur_distinct_count_status.py` it will create new aggregated data source called `ml-survey-distinctCount-status` .

Config changes

1. `Pyspark_sur_distinct_count_status.py` script
2. `ml-survey-distinctCount-status` datasource
3. Changes in the `config.ini`
4. Backend json changes

`Pyspark_sur_distinct_count_status.py` script

1. This is a PySpark script that queries Mongo collections such as `surveySubmissions`, `solutions`, and `Programs`, processes the data, and then stores it in the druid `ml-survey-distinctCount-status` datasource.
2. This is the Batch Ingestion so every day triggers via cron job .
3. This script is configured in `run.sh` script as well .

config.ini

1.[DRUID]

ml_distinctCnt_survey_status_spec =

```
{"type":"index","spec":{"ioConfig":{"type":"index","inputSource":{"type":"local","baseDir":"local json file storage path", "filter" :
```

```
"ml_survey_distinctCount_status.json"},"inputFormat":{"type":"json"},"tuningConfig":{"type":"index","partitionsSpec":{"type":"dynamic"},"dataSchema":{"dataSource":"ml-surveydistinctCount-status","granularitySpec":{"type":"uniform","queryGranularity":"none","rollup":false,"segmentGranularity":"DAY"},"timestampSpec":{"column":"time_stamp","format":"auto"},"dimensionsSpec":{"dimensions":[{"type":"string","name":"program_name"}, {"type":"string","name":"program_id"}, {"type":"string","name":"survey_name"}, {"type":"string","name":"survey_id"}, {"type":"string","name":"submission_status"}, {"type":"string","name":"state_name"}, {"type":"string","name":"state_externalId"}, {"type":"string","name":"district_name"}, {"type":"string","name":"district_externalId"}, {"type":"string","name":"block_name"}, {"type":"string","name":"block_externalId"}, {"type":"string","name":"organisation_name"}, {"type":"string","name":"organisation_id"}, {"type":"string","name":"private_program"}, {"type":"string","name":"parent_channel"}, {"type":"long","name":"unique_users"}, {"type":"long","name":"unique_submissions"}, {"type":"string","name":"time_stamp"}]},"metricsSpec":[]}}}
```

2.[OUTPUT_DIR]

survey_distinctCount_status = "local json file storage path"

3.[COMMON]

survey_distinctCount_blob_path = "cloud json file storage path"

4.[LOGS]

survey_streaming_success_error : "logs storage path"

Backend Json

1 . ml_no_of_surveys_in_started_status_currently_sl.json :

https://github.com/shikshalokam/ml-analytics-service/blob/release-6.0.0/migrations/releases/6.0.0/config/backend/create/ml_no_of_surveys_in_started_status_currently_sl.json

2. ML_no_of_surveys_submitted_till_date_sl.json :

https://github.com/shikshalokam/ml-analytics-service/blob/release-6.0.0/migrations/releases/6.0.0/config/backend/create/ml_no_of_surveys_submitted_till_date_sl.json

