Wiki Operacional PSafe

Total Basic Metrics Country

Basic metrics calculated with country dimension Also exports data to Impala DB.

- dau INT,
- dni INT,
- wau INT,
- mau INT,
- last30 INT,
- acu INT,
- deu int,
- country STRING,
- date TIMESTAMP *(rounded to day, e.g. 2014-12-31 00:00:00)*

Script Location in HDFS:

`/user/bi.user/scripts/mobile_operational_reports/statistics_by_country.hql`

Workflow:

`Total_Basic_Metrics_Country`. Part of `Total_Daily` coordinator

Sql destination:

`StageDb.exp_factbasicmetrics`

Script body:

```
ADD jar hdfs://PSBIHNN001/user/bi.user/scripts/csv-serde.jar;

SET hive.exec.dynamic.partition.mode=nonstrict;
SET hive.exec.max.dynamic.partitions = 100000;
SET hbase.client.scanner.caching = 1000;
SET hive.auto.convert.join = false;

set CURRENT_TIMEZONE='GMT-03:00';

set LAST_REPORT_DATE='${LAST_REPORT_DATE_PARAM}';
set CURRENT_REPORT_DATE=date_sub(${hiveconf:LAST_REPORT_DATE}, 1);
set DAU_REPORT_DATE=date_sub(${hiveconf:LAST_REPORT_DATE}, 1);
set WAU_REPORT_DATE=date_sub(${hiveconf:LAST_REPORT_DATE}, 8);
set MAU_REPORT_DATE=concat(year(${hiveconf:CURRENT_REPORT_DATE}), '-', lpad(cast(month(${hiveconf:CURRENT_REPORT_DATE}) as STRING), 2, '0'), '-01');
set 30AGO_REPORT_DATE=date_sub(${hiveconf:LAST_REPORT_DATE}, 31);

drop table if exists tmp_db.tmp_dau_country;
drop table if exists tmp_db.tmp_wau_country;
drop table if exists tmp_db.tmp_mau_country;
drop table if exists tmp_db.tmp_30ago_country;
drop table if exists tmp_db.tmp_dni_country;
drop table if exists tmp_db.tmp_acu_country;
drop table if exists tmp_db.tmp_dni_country;
drop table if exists tmp_db.tmp_acu_country;
drop table if exists tmp_db.tmp_deu_country;

CREATE TABLE IF NOT EXISTS exp_db.exp_factbasicmetrics
(
  dau INT,
  dni INT,
  wau INT,
  mau INT,
  last30 INT,
  acu INT,
  deu int,
  country STRING,
  date TIMESTAMP
)
PARTITIONED BY (year INT, month INT, day INT)
ROW FORMAT SERDE 'com.bizo.hive.serde.csv.CSVSerde'
WITH serdeproperties (
   "separatorChar" = ",",
   "quoteChar"     = "'",
   "escapeChar"    = "\\"
)
STORED AS TEXTFILE;

CREATE TABLE IF NOT EXISTS impala_db.total_basic_metrics
(
  dau INT,
  dni INT,
  wau INT,
  mau INT,
  last30 INT,
  acu INT,
  deu int,
  country STRING,
  date TIMESTAMP
)
PARTITIONED BY (year INT, month INT, day INT)
STORED AS RCFILE;

CREATE TABLE IF NOT EXISTS tmp_db.tmp_returning_country
(
  date TIMESTAMP,
  country STRING,
  count INT
)
STORED AS TEXTFILE;
```

```
CREATE TABLE IF NOT EXISTS tmp_db.tmp_dau_country
(
  date TIMESTAMP,
  country STRING,
  count INT
)
STORED AS TEXTFILE;

CREATE TABLE IF NOT EXISTS tmp_db.tmp_wau_country
(
  date TIMESTAMP,
  country STRING,
  count INT
)
STORED AS TEXTFILE;

CREATE TABLE IF NOT EXISTS tmp_db.tmp_mau_country
(
  date TIMESTAMP,
  country STRING,
  count INT
)
STORED AS TEXTFILE;

CREATE TABLE IF NOT EXISTS tmp_db.tmp_30ago_country
(
  date TIMESTAMP,
  country STRING,
  count INT
)
STORED AS TEXTFILE;

CREATE TABLE IF NOT EXISTS tmp_db.tmp_dni_country
(
  date TIMESTAMP,
  country STRING,
  count INT
)
STORED AS TEXTFILE;

CREATE TABLE IF NOT EXISTS tmp_db.tmp_acu_country
(
  date TIMESTAMP,
  country STRING,
  count INT
)
STORED AS TEXTFILE;

CREATE TABLE IF NOT EXISTS tmp_db.tmp_deu_country
(
  date TIMESTAMP,
  country STRING,
  count INT
)
STORED AS TEXTFILE;


-- prepare profile table as file, cuz it is faster to read from it
drop table if exists tmp_db.total_profile_rcfile;
CREATE TABLE tmp_db.total_profile_rcfile STORED AS RCFILE AS
select * from ods_db.total_profile;

-- DEU
INSERT OVERWRITE TABLE tmp_db.tmp_deu_country
SELECT date,country,count(1) from
(select distinct
    concat(to_date(from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE})), ' 00:00:00') date,
    country,
    imei
  FROM
    staging_db.mobile_operational_data LATERAL VIEW explode(features) ft AS feature_id, feature_cnt

WHERE concat(year, '-', month, '-', day) >= to_date(${hiveconf:DAU_REPORT_DATE})
    AND concat(year, '-', month, '-', day) <= to_date(${hiveconf:LAST_REPORT_DATE})
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) >= ${hiveconf:DAU_REPORT_DATE}
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) < ${hiveconf:LAST_REPORT_DATE}
 and feature_id in ('1000','3010','6000','6008','33005','40001','45001','45002','45003','45004','45005','45006','45007','45008',
'45021','45032','45034','45037','45040','45043','45047','45051','45054','45057','45060','45063','55000','55001')
) s1
GROUP by date, country;

-- DAU

INSERT OVERWRITE TABLE tmp_db.tmp_dau_country
SELECT concat(to_date(${hiveconf:CURRENT_REPORT_DATE}), ' ', '00:00:00') date, country, count(1) count
FROM (SELECT a.imei, b.country FROM ods_db.mobile_operational_user_activity a
      JOIN tmp_db.total_profile_rcfile b ON a.imei = b.user_id
WHERE concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) >= to_date(${hiveconf:DAU_REPORT_DATE})
    AND concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) <= to_date(${hiveconf:LAST_REPORT_DATE})
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) >= ${hiveconf:DAU_REPORT_DATE}
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) < ${hiveconf:LAST_REPORT_DATE}
GROUP BY a.imei, b.country) d
GROUP BY country;

-- DNI

INSERT OVERWRITE TABLE tmp_db.tmp_dni_country
SELECT concat(to_date(${hiveconf:CURRENT_REPORT_DATE}), ' ', '00:00:00') date, country, count(1) count
FROM tmp_db.total_profile_rcfile
WHERE from_utc_timestamp(first_install, ${hiveconf:CURRENT_TIMEZONE}) >= ${hiveconf:DAU_REPORT_DATE}
    AND from_utc_timestamp(first_install, ${hiveconf:CURRENT_TIMEZONE}) < ${hiveconf:LAST_REPORT_DATE}
GROUP BY country;

-- AC-U

INSERT OVERWRITE TABLE tmp_db.tmp_acu_country
SELECT concat(to_date(${hiveconf:CURRENT_REPORT_DATE}), ' ', '00:00:00') date, country, count(1) count
FROM tmp_db.total_profile_rcfile
WHERE from_utc_timestamp(first_install, ${hiveconf:CURRENT_TIMEZONE}) < ${hiveconf:LAST_REPORT_DATE}
GROUP BY country;
```

```
-- WAU

INSERT OVERWRITE TABLE tmp_db.tmp_wau_country
SELECT concat(to_date(${hiveconf:CURRENT_REPORT_DATE}), ' ', '00:00:00') date, country, count(1) count
FROM (SELECT a.imei, b.country FROM ods_db.mobile_operational_user_activity a
      JOIN tmp_db.total_profile_rcfile b ON a.imei = b.user_id
WHERE concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) >= to_date(${hiveconf:WAU_REPORT_DATE})
    AND concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) <= to_date(${hiveconf:LAST_REPORT_DATE})
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) >= ${hiveconf:WAU_REPORT_DATE}
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) < ${hiveconf:LAST_REPORT_DATE}
GROUP BY a.imei, b.country) d
GROUP BY country;

-- 30ago

INSERT OVERWRITE TABLE tmp_db.tmp_30ago_country
SELECT concat(to_date(${hiveconf:CURRENT_REPORT_DATE}), ' ', '00:00:00') date, country, count(1) count
FROM (SELECT a.imei, b.country FROM ods_db.mobile_operational_user_activity a
      JOIN tmp_db.total_profile_rcfile b ON a.imei = b.user_id
WHERE concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) >= to_date(${hiveconf:30AGO_REPORT_DATE})
    AND concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) <= to_date(${hiveconf:LAST_REPORT_DATE})
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) >= ${hiveconf:30AGO_REPORT_DATE}
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) < ${hiveconf:LAST_REPORT_DATE}
GROUP BY a.imei, b.country) d
GROUP BY country;

-- MAU (month to date)

INSERT OVERWRITE TABLE tmp_db.tmp_mau_country
SELECT concat(to_date(${hiveconf:CURRENT_REPORT_DATE}), ' ', '00:00:00') date, country, count(1) count
FROM (SELECT a.imei, b.country FROM ods_db.mobile_operational_user_activity a
      JOIN tmp_db.total_profile_rcfile b ON a.imei = b.user_id
WHERE concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) >= to_date(${hiveconf:MAU_REPORT_DATE})
    AND concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) <= to_date(${hiveconf:LAST_REPORT_DATE})
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) >= ${hiveconf:MAU_REPORT_DATE}
    AND from_utc_timestamp(time, ${hiveconf:CURRENT_TIMEZONE}) < ${hiveconf:LAST_REPORT_DATE}
GROUP BY a.imei, b.country) d
GROUP BY country;


-- JOINING TABLES INTO ONE

INSERT OVERWRITE TABLE impala_db.total_basic_metrics PARTITION (year, month, day)
SELECT dau.count dau, dni.count dni, wau.count wau, mau.count mau, last30.count last30, acu.count acu, deu.count deu, dau.country, dau.date,
year(dau.date), month(dau.date), day(dau.date)
FROM tmp_db.tmp_dau_country dau
  JOIN tmp_db.tmp_wau_country wau ON dau.country = wau.country
  JOIN tmp_db.tmp_mau_country mau ON dau.country = mau.country
  JOIN tmp_db.tmp_30ago_country last30 ON dau.country = last30.country
  JOIN tmp_db.tmp_dni_country dni ON dau.country = dni.country
  JOIN tmp_db.tmp_acu_country acu ON dau.country = acu.country
 JOIN tmp_db.tmp_deu_country deu ON dau.country = deu.country
GROUP BY dau.count, dni.count, wau.count, mau.count, last30.count, acu.count, deu.count, dau.country, dau.date, year(dau.date), month(dau.date), day(dau.date);

-- EXPORT TO CSV TABLE
INSERT OVERWRITE TABLE exp_db.exp_factbasicmetrics PARTITION (year, month, day)
SELECT dau, dni, wau, mau, last30, acu, deu, country, date, year, month, day
FROM impala_db.total_basic_metrics WHERE concat(year, '-', lpad(cast(month as string), 2, '0'), '-', lpad(cast(day as string), 2, '0')) = to_date(${hiveconf:CURRENT_REPORT
```
