Use OpenCellID Data To Estimate GSM/UMTS/LTE Coverage

Practical Exam

Goal

OpenCellID.com provides regulary exports of worldwide cell data:

- https://www.opencellid.org
- Latest Full Dump and Diffs: https://www.opencellid.org/downloads.php

Full Database

- cell_towers.csv.gz
  Updated: 2019-11-10 (937MB)

Differential

- OCID-diff-cell-export-2019-11-10-T000000.csv.gz (856KB)
- OCID-diff-cell-export-2019-11-09-T000000.csv.gz (1827KB)
- OCID-diff-cell-export-2019-11-08-T000000.csv.gz (1904KB)
- OCID-diff-cell-export-2019-11-07-T000000.csv.gz (1902KB)
- OCID-diff-cell-export-2019-11-06-T000000.csv.gz (1873KB)
- OCID-diff-cell-export-2019-11-05-T000000.csv.gz (2006KB)
- OCID-diff-cell-export-2019-11-04-T000000.csv.gz (689KB)

```
radio,mcc,net,area,cell,unit,lon,lat,range,samples,changeable,created,updated,averageSignal
UMTS,262,2,801,86355,0,13.285512,52.522202,1000,7,1,1282569574,1300155341,0
GSM,262,2,801,1795,0,13.276907,52.525714,5716,9,1,1282569574,1300155341,0
GSM,262,2,801,1794,0,13.285064,52.524,6280,13,1,1282569574,1300796207,0
UMTS,262,2,801,211250,0,13.285446,52.521744,1000,3,1,1282569574,1299466955,0
UMTS,262,2,801,86353,0,13.293457,52.521515,1000,2,1,1282569574,1291380444,0
UMTS,262,2,801,86357,0,13.289106,52.53273,2400,3,1,1282569574,1298860769,0
UMTS,262,3,1107,83603,0,13.349675,52.497575,3102,222,1,1282672189,1300710809,0
GSM,262,2,776,867,0,13.349711,52.497367,1000,214,1,1282672189,1301575206,0
GSM,262,3,1107,13971,0,13.349743,52.497437,1000,212,1,1282672189,1300710809,0
GSM,262,3,1107,355,0,13.34963,52.497378,1000,198,1,1282672189,1300710809,0
UMTS,262,3,1107,329299,0,13.349223,52.497519,3041,186,1,1282672189,1299860879,0
```

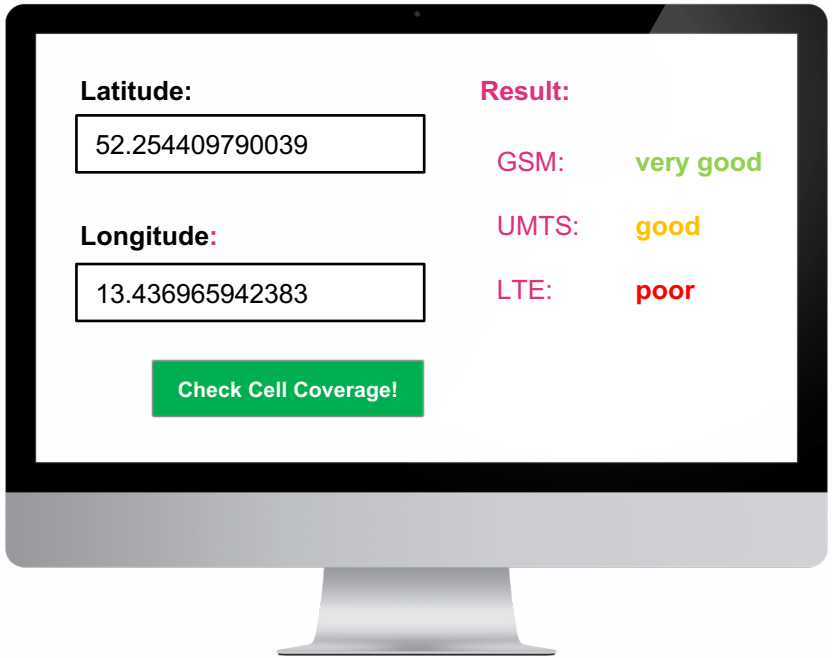
cell_towers.csv

Goal

We want to make use of this data to estimate the coverage of GSM, UMTS and LTE for a certain place (*latitude, longitude*).

Workflow:

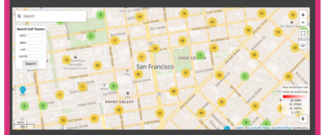
- **Gather data** from OpenCellID.com
- **Save raw data** (*CSV files*) to HDFS (partitioned by radio, e.g. *GSM*, *UMTS*, *LTE*…)
- **Optimize**, **reduce** and **clean raw data** and save it to **final** directory on HDFS
- **Export** address data **to end-user database** (e.g. MySQL, MongoDB…)
- Provide a simple **HTML Frontend** which is able to:
  - read from end-user database
  - process user input (Latitude, Longitude…)
  - checks against OpenCellID data in end-user database
  - Display result (GSM, LTE and UMTS coverage)
- The whole data workflow **must be implemented** within an ETL **workflow tool** (e.g. **Pentaho Data Integration** or **Airflow**) and **run automatically**



Dataflow: 1. Get Cell Data

Get **Initial data** cell_towers.csv.gz
as well as **Diffs** (OCID-diff-cell-export-2019-11-08-T000000.csv.gz) for
every future execution of ETL Workflow
(Airflow or Pentaho Data Integration)

1



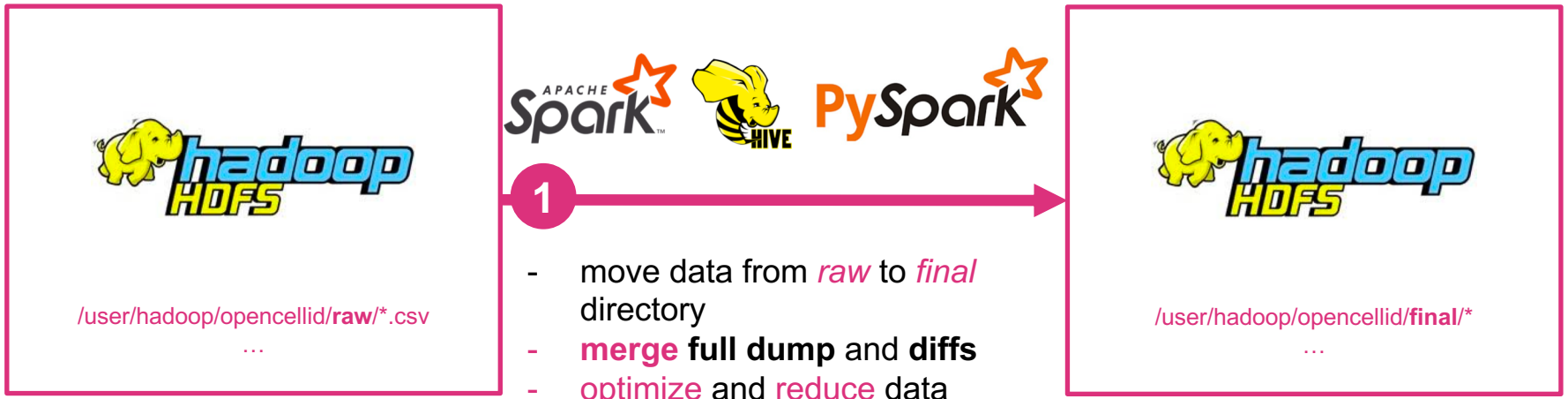
https://www.opencellid.org

2

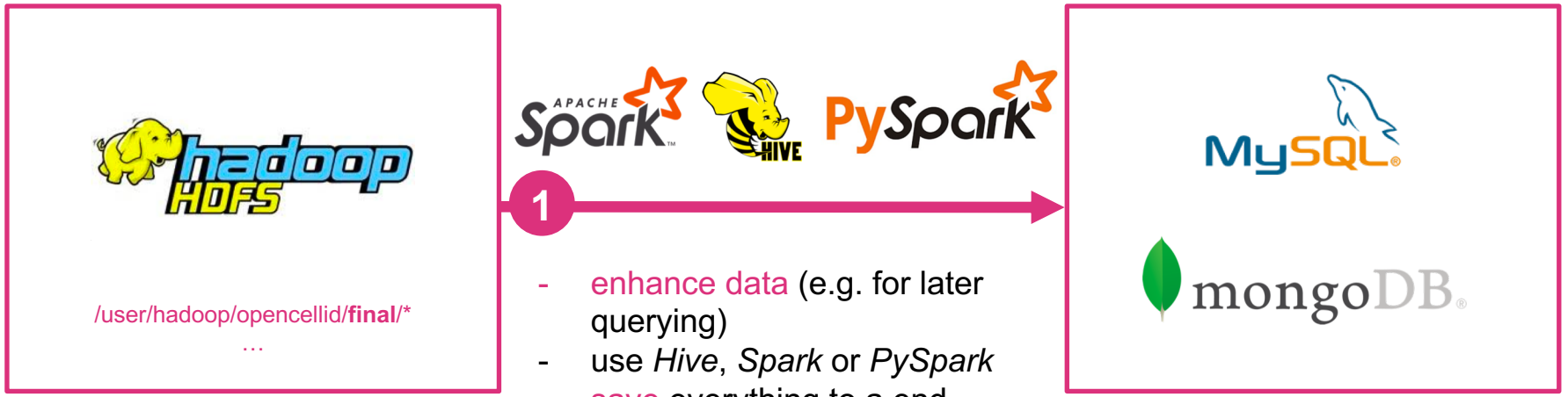
hadoop HDFS

/user/hadoop/opencellid/**raw**/*.csv
…

Dataflow: 2. Raw To Final Transfer

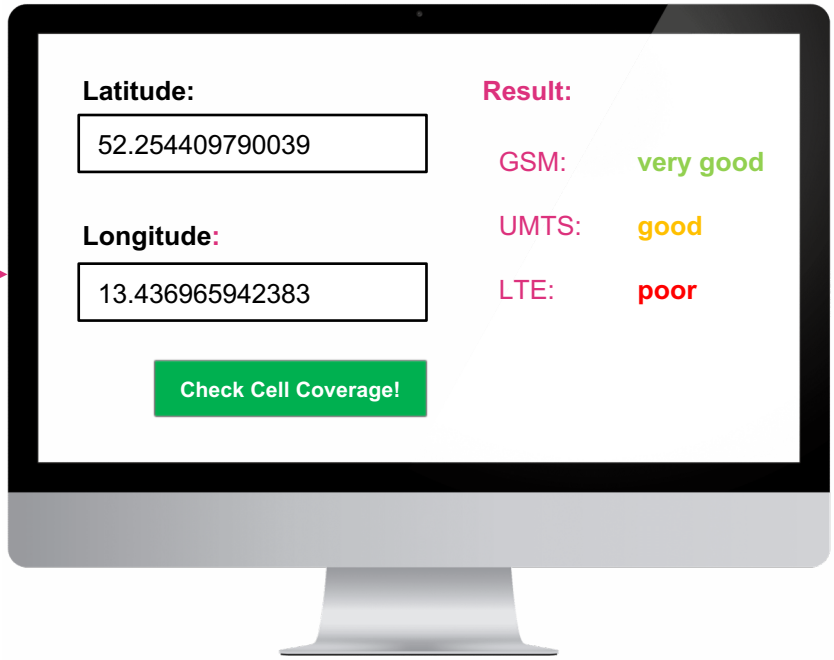


Dataflow: 3. Enhance Data And Save Results



Dataflow: 4. Provide Simple Web Interface

1



- Provide a simple **HTML Frontend** which is able to:
  - read from end-user database
  - process user input (Latitude, Longitude…)
  - checks against OpenCellID data in end-user database
  - Display result (GSM, LTE and UMTS coverage)