

Bdd :

Namespace	Name	State	OPEN	OPENING	CLOSED	CLOSING	OFFLINE	FAILED	SPLIT	Other	Description
default	location	ENABLED	1	0	0	0	0	0	0	0	'location', {NAME => 'crime_info'}
practice	crimes	ENABLED	1	0	0	0	0	0	0	0	'practice:crimes', {NAME => 'crime_info'}, {NAME => 'location'}

On GitHub , you will find the docker setup and python notebooks.

Here is a document report summarizing:

- Modeling choices
- Screenshots of HBase Shell queries
- Sample retrieved data
- EDA results

## Modeling choices

The data were firstly presented on this form (different columns):

['DR\_NO', 'Date Rptd', 'DATE OCC', 'TIME OCC', 'AREA', 'AREA NAME', 'Rpt Dist No', 'Part 1-2', 'Crm Cd', 'Crm Cd Desc', 'Mocodes', 'Vict Age', 'Vict Sex', 'Vict Descent', 'Premis Cd', 'Premis Desc', 'Weapon Used Cd', 'Weapon Desc', 'Status', 'Status Desc', 'Crm Cd 1', 'Crm Cd 2', 'Crm Cd 3', 'Crm Cd 4', 'LOCATION', 'Cross Street', 'LAT', 'LON']

I choose this kind of model for my data:

```
catalog = {'location': ['location', 'cross_street', 'lat', 'lon', 'area_name', 'premis_desc'],  
          'crime_info': ['crm_cd_desc', 'weapon_desc', 'status_desc', 'part_1-2', 'vict_age', 'vict_sex',  
                        'vict_descent', 'time_occ', 'mocodes']}
```

I separated the columns into two logical column families to reflect domain concepts and optimize access:

Location summarize geographical and physical details of the crime and crime\_info specifies details of crimes like weapon used, victim demographics, and crime classification.

## Screenshots of HBase Shell queries

There were several issues using the hbase shell not allowing me to have answer via hbase. So I answered the question in python. The answers are directly in the notebook.