Pokemon Data Analysis by APACHE HIVE

==========

Input Dataset:

---

https://drive.google.com/open?id=0ByJLBTmJojjzZTJQM2UzN2J6aUU

Project Description:

Pokémon Go is a free-to-play, location-based augmented reality game developed by Niantic for iOS and Android devices. It was released only in July 2016 and only in selected countries. You can download Pokémon for free of cost and start playing. You can also use PokéCoins to purchase Pokéballs, the in-game item you need to be able to catch Pokémon.

Pokemonid_Number: Unique Pokémon id .

Name: name of the Pokémon.

Type 1: Primary property of a Pokémon.

Type 2: Extended property of the same Pokémon.

A Pokémon may be one or both the types. For instance, Charmander is a Fire type, while Bulbasaur is both a Grass type as well as a Poison type. With the current 18-type system, there are 324 possible ways to assign these types to Pokémon, along with 171 unique combinations. As of Generation VI, 133 different type combinations have been used.

Total: the sum of all character points of a Pokémon (HP, Attack, Defense, Sp. Atk, Sp. Def, and Speed).

HP (Hit Points): Pokémon Hit Points, which is a value that determines how much damage a Pokémon can receive. When a Pokémon's HP is down to '0', the Pokémon will faint. HP is the most frequently affected stat of them all, as a depleting HP is a key factor in winning a battle.

Attack: the Attack stat.

Defense: the Defense stat.

Sp. Atk: Pokémon's Special Attack stat.

Sp. Def: Pokémon's Special Defense stat.

Speed: the speed stat of a Pokémon.

```
[cloudera@quickstart ~]$ hive
hive> show databases ;
hive> CREATE DATABASE IF NOT EXISTS pokemondatabase;
hive> USE pokemondatabase;
hive> CREATE TABLE IF NOT EXISTS pokemon (
    pokemonid int, pokemonname string, type1 string, type2 String, total Int, hp Int,
attack Int, defense Int, specialattack Int,   specialdefense Int, speed Int)
    ROW FORMAT DELIMITED
    FIELDS TERMINATED BY ','
    LINES TERMINATED BY '\n'
    tblproperties("skip.header.line.count"="1");
hive> LOAD DATA LOCAL INPATH '/home/cloudera/Downloads/Pokemon.csv' INTO
TABLE pokemon;
hive> SELECT * FROM pokemon;
hive> DESCRIBE pokemon;
hive> DESCRIBE FORMATTED pokemon;
```

Problem Statement 1:

Find out the average HP (Hit points) of all the Pokémon,

```
hive> SELECT avg(HP) FROM pokemon ;
```

Problem Statement 2:

Create and insert values of existing table 'pokemon' into a new table 'pokemon1', with an additional column 'power_rate' to find the count of 'powerful' and 'moderate' from the table 'pokemon1'.

if the HP is greater than the average HP, then it is powerful, and if the HP is less than the average, then it is Moderate and a neutral condition is considered as powerless..

```
hive> create table pokemon11 as select *, IF(HP>69.25875, 'powerful',
IF(HP<69.25875, 'moderate','powerless')) AS power_rate from pokemon;

hive> SELECT * FROM pokemon11;

hive> SELECT COUNT(*) FROM pokemon11;

hive> SELECT power_rate, COUNT(power_rate) FROM pokemon11 GROUP BY
power_rate;
    HAVING (power_rate =='powerful' OR power_rate =='moderate');
```

Problem Statement 3:

Find the count of powerful pokemons.

```
hive> SELECT power_rate, COUNT(power_rate) FROM pokemon11
    GROUP BY power_rate HAVING (power_rate =='powerful');
```

Problem Statement 4:

Find out the top 10 Pokémons according to their HP's.

```
hive> SELECT pokemonid, pokemonname, hp FROM pokemon ORDER by hp
DESC limit 10;
```

Problem statement 5:

Find out the top 10 Pokémons based on their Attack stat, using the below query.

```
hive> SELECT pokemonid, pokemonname, attack FROM pokemon ORDER by
attack DESC limit 10;
```

Problem statement 6:

Find out the top 10 Pokémons based on their Defense stat, using the below query.

```
hive> SELECT pokemonid, pokemonname, defense FROM pokemon ORDER by
defense DESC limit 10;
```

Problem statement 7:

Find out the top 10 Pokémons based on their total power.

```
hive> SELECT pokemonid, pokemonname, total FROM pokemon ORDER by total
DESC limit 10;
```

Problem statement 8:

Find out the top 10 Pokémons having a drastic change in their attack and sp.attack, using the below query.

```
hive> SELECT pokemonid, pokemonname, (attack-specialattack) AS
difference_in_attack
FROM pokemon ORDER BY difference_in_attack DESC LIMIT 10;
```

Problem statement 9:

Find out the top 10 Pokémons having a drastic change in their defense and sp.defense, using the below query.

```
Hive> SELECT pokemonid, pokemonname, (defense-specialdefense) AS
difference_in_defense
FROM pokemon ORDER BY difference_in_defense DESC LIMIT 10;
```

Problem statement 10:

Find out the top 10 fastest Pokémons, using the below query.

```
Hive> SELECT pokemonid, pokemonname, speed from pokemon
ORDER BY speed DESC LIMIT 10;
```