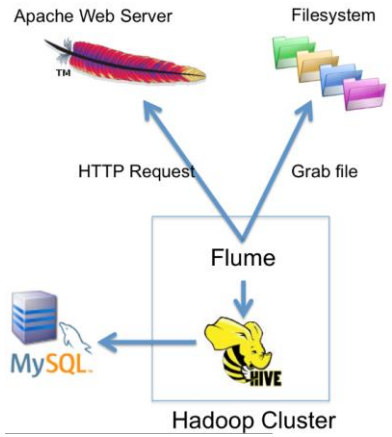
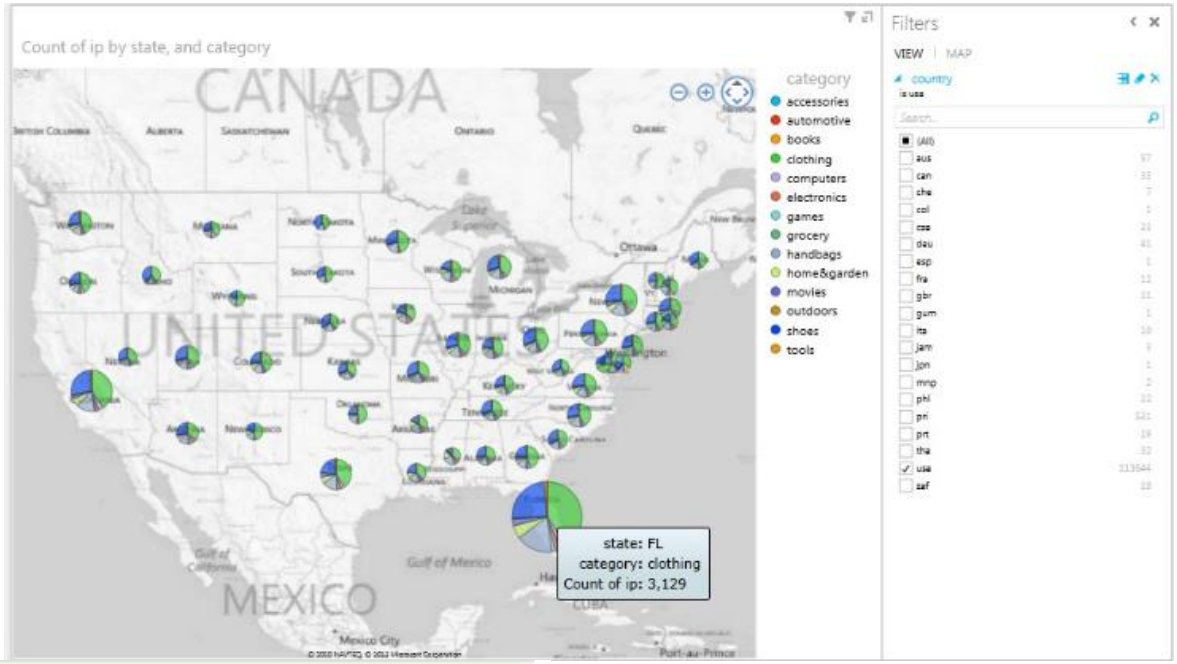
Hadoop - UseCases

BIS Academy

Clickstream Use Case

- Online retailer can optimize buying paths to improve its sales





Sensor Data – Examples

- To monitor machines or infrastructure such as ventilation equipment, bridges, energy meters, or airplane engines.
- The sensor data can be used for predictive analytics, to repair or replace these items before they break.

- To monitor natural phenomena such as meteorological patterns
- To monitor underground pressure during oil extraction
- To monitor patient vital statistics during recovery from a medical procedure.

Use Case - Customer Complaints Analysis

Industry:

- Retail

Data:

- Dataset, containing a few lakh observations with attributes like; CustomerId, Payment Mode, Product Details, Complaint, Location, Status of the complaint, etc.

Problem Statement:

- Analyze the data in the Hadoop ecosystem to:

1. Get the number of complaints filed under each product
2. Get the total number of complaints filed from a particular location
3. Get the list of complaints grouped by location which has no timely response

Tourism Data Analysis

Industry:

- Tourism

Data:

- The dataset comprises attributes like: City pair (combination of from and to), adults traveling, seniors traveling, children traveling, air booking price, car booking price, etc.

Problem Statement:

- Find the following insights from the data:

1. Top 20 destinations people frequently travel to: Based on given data we can find the most popular destinations where people travel frequently, based on the specific initial number of trips booked for a particular destination

2. Top 20 locations from where most of the trips start based on booked trip count

3. Top 20 high air-revenue destinations, i.e the 20 cities that generate high airline revenues for travel, so that the discount offers can be given to attract more bookings for these destinations.

Airline Data Analysis

Industry:

Aviation

Data:

- Dataset which contains the flight details of various airlines such as: Airport id, Name of the airport, Main city served by airport, Country or territory where airport is located, Code of Airport, Decimal degrees, Hours offset from UTC, Timezone, etc.

Problem Statement:

- Analyze the airlines' data to:

1. Find list of airports operating in the country
2. Find the list of airlines having zero stops
3. List of airlines operating with code share
4. Which country (or) territory has the highest number of airports
5. Find the list of active airlines in the United States

Analyze Loan Dataset

Industry:

- Banking and Finance

Data:

- Dataset which contains complete details of all the loans issued, including the current loan status (Current, Late, Fully Paid, etc.) and latest payment information.

Problem Statement:

- Find the number of cases per location and categorize the count with respect to reason for taking loan and display the average risk score.

Analyze Movie Ratings

Industry:

- Media

Data:

- Publicly available data from sites like rotten tomatoes, IMDB, etc.

Problem Statement:

- Analyze the movie ratings by different users to:

1. Get the user who has rated the most number of movies
2. Get the user who has rated the least number of movies
3. Get the count of total number of movies rated by user belonging to a specific occupation
4. Get the number of underage users

Analyze YouTube data

- **Industry:**

Social Media

Data:

- It is about the YouTube videos and contains attributes such as: VideoID, Uploader, Age, Category, Length, views, ratings, comments, etc.

Problem Statement:

- Identify the top 5 categories in which the most number of videos are uploaded, the top 10 rated videos, and the top 10 most viewed videos.

Thank You

Keerthiga Barathan