

# Neobi API Pipeline

Tue, February 25

---

Cliford Rojas

EnhancelT

1785 The Exchange SE

Suite. 200

Atlanta, GA 30339

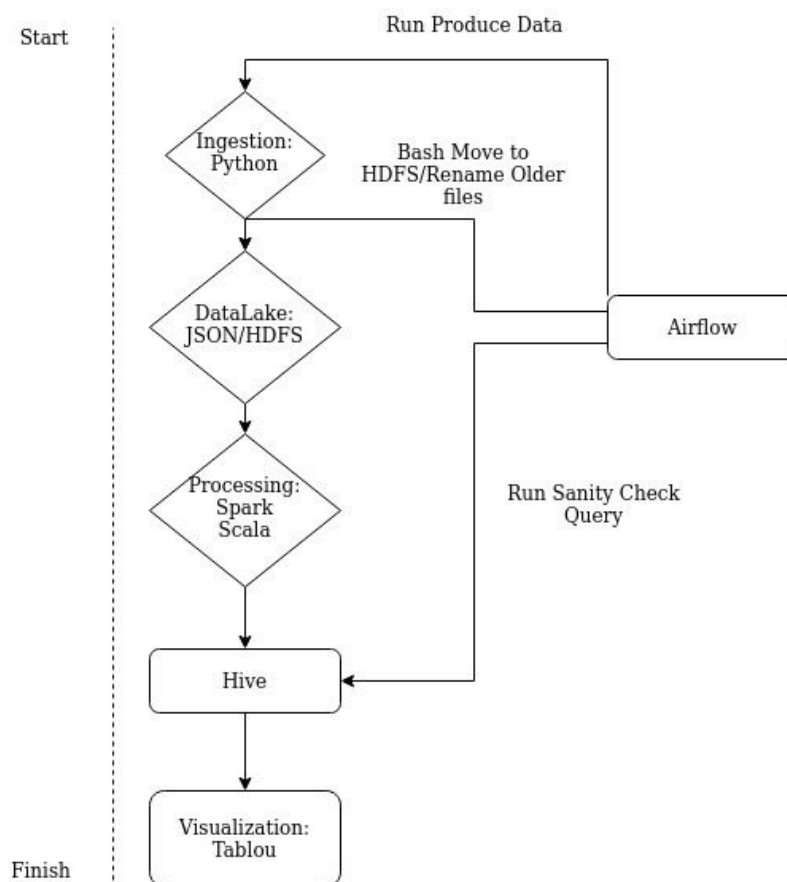
## Overview

The cannabis industry is a newly risen industry with states like colorado and california leading the way for other states. The industry is still in its infancy and is quickly growing from sales of marijuana for medical purposes to using hemp as a substitute for traditional plastic and paper use cases.

## Goals

1. Use marijuana data to establish a pricing guide using canada's pricing as a means to estimate the United States market.
2. Gather distributor/producer information to potentially get a list of known growers that can provide the quantities needed to build a business in some US states.

## Pipeline Overview



## Timeline

	Monday	Tuesday	Wednesday	Thursday	Friday
Week 1	Ingestion: Raw Data Local JSON ~100 Records	Igestion: Raw-Data Analysis of Producers/C annabis Products ~1000 Records	Igestion: Write to HDFS ~10,000 Records	Processing: Remove unwanted products.	Processing: Remove unwanted information Picture,Link s,..etc
Week 2	Data warehouse: Wire cleaned data into hive	Data warehouse: Troublesho ot Data import	Data Warehouse: Create Queries for analysis.	Basic Visualizatio n: Demo data	Automation Run Airflow with bash commands.
Week 3	Troublesho ot Automation Check stages	Moderate Visualizatio n: Create more graphs and show valuable information.	Safety net	Safety net	Pass to the next team.