There are generally two types of model drift, data drift and performance drift. Data drift happens when the distribution of input data changes over time. Concept drift is when the relationship between the input and output changes. Thus, my pipeline needs to have the ability to collect data and store them, so that we can compare the new data in future when the historical data we have previously collected. The pipeline can store in database like PostgreSQL and AWS S3. This also ensures a baseline for the initial training data the model is trained on, especially on statistical properties like mean, variance and distribution.

The pipeline will first set up a data logging system to logs the raw inputs that are fed into my model. When the input data streams in, the pipeline will calculate the distribution of the input data. Then, it will use some statistical tests like Kolmogorov-Smirnov (KS), for continuous variables, and Chi-square test, for categorical features, to compare the distribution of the new input data against the historical training data. The pipeline can also use Jensen-Shannon Divergence to measure the distance between the distributions.

After my model makes some predictions based on the new input data, if there is any actual outcome or result of the new input data, the pipeline will log these actual outcomes as the ground truth labels. Then it will compare the predicted outcomes with the actual outcomes to calculate the latest performance metrics of my model. For every batch of input data, it will log the performance metrics of my model, if they are calculated. If the metrics deviate above a certain threshold or the residuals (difference between predicted and actual outcomes) are larger than certain thresholds, this means that the model is experiencing some concept drifts. If there are often no actual outcomes, some adaptive windowing techniques like Page-Hinkley can be used to test to detect changes in the mean prediction values and Drift Detection Method can monitor error rates. The distribution of the newly made predictions can also be compared against the distribution of prediction made in the past.