

Highlight

Note

Data Drift Schedules and Alerts

When you define a data monitor, you specify a schedule on which it should run. Additionally, you can specify a threshold for the rate of data drift and an operator email address for notifications if this threshold is exceeded.

Configuring Data Drift Monitor Schedules

Data drift monitoring works by running a comparison at scheduled **frequency**, and calculating data drift metrics for the features in the dataset that you want to monitor. You can define a schedule to run every **Day**, **Week**, or **Month**.

For dataset monitors, you can specify a **latency**, indicating the number of hours to allow for new data to be collected and added to the target dataset. For deployed model data drift monitors, you can specify a **schedule_start** time value to indicate when the data drift run should start (if omitted, the run will start at the current time).

Configuring Alerts

Data drift is measured using a calculated *magnitude* of change in the statistical distribution of feature values over time. You can expect some natural random variation between the baseline and target datasets, but you should monitor for large changes that might indicate significant data drift.

You can define a **threshold** for data drift magnitude above which you want to be notified, and configure alert notifications by email.

```
alert_email AlertConfiguration('data_scientists@contoso.com')
monitor = DataDriftDetector.create_from_datasets(ws, 'dataset-drift-detector',
                                                baseline_data_set, target_data_set,
                                                compute_target=cpu_cluster,
                                                frequency='Week', latency=2,
                                                drift_threshold=.3,
                                                alert_configuration=alert_email)
```