CREATING TRIGGERS FOR BASELINE MONITORING AND ANOMALY DETECTION
ANOMALLY DETECTION
WHAT IS ANOMALLY DETECTION

- Anomaly detection works by going through historical data and looking for values that are out of normal
- Works if the majority of data is considered «normal»
- Long term analytics – works with trends data
- Zabbix uses STL decomposition algorithm (Seasonal and Trend decomposition using Loess)
Decomposition using STL algorithm is a way to split a single time series sequence into three other sequences:
DEVIATIONS

Deviation is a measure of data variability.

How “far” values are from average?

Zabbix has capabilities to determine deviation in multiple ways.
FIND ANOMALY RATE WITH TRENDSTL

Data to work with: 28d, start to analyze starting from previous hour, use ‘1d’ to seek anomalies, weekdays matters (Season: 7d). Deviations – how many is considered as anomaly (default: 3)
OUTPUT OF TRENDSTL FUNCTION

- a decimal value between 0 and 1
- (the number of anomaly values in detect period) / (total number of values in detect period).

Example 1: Context is 7d, detect anomalies in last 1d. If one is detected, then reported value is:

\[ \frac{1}{24} = 1 \text{ anomaly detected within last 24h} \]

Example 2: Context is 7d, detect anomalies for last 2d. If one is detected, then reported value is:

\[ \frac{1}{48} = 1 \text{ anomaly detected within last 48h} \]
DEVIATION ALGORITHMS

- **mad (default) – «median absolute deviation»**
  A robust measure of the variability of a univariate sample of quantitative data.

- **stddevpop – «population standard deviation»**
  Looks at the square root of the variance of the set of numbers.

- **stddevsamp – «sample standard deviation»**
  Average distance of the observed data from the expected values
CONCEPT OF SEASON DATA

How the service has been used:

- All days are the same (24h scale)
- All Thursdays are the same (7d scale)
- 8h working day in a 24/7 factory. There are 3 sessions in the level of 1d.
- First day of each month
CONCLUSION

- Get trend values from the period
- Decompose values, get reminder
- Calculate deviation for values in remainder
- Select values with deviation and compare with threshold
02

BASELINE MONITORING
WHAT IS BASELINE MONITORING?

- Baseline is a value derived from an average over multiple calendar periods of the same length.
BASELINE PROPERTIES

- Periods and seasons.
- Average from past calendar period:
  - E.g., every Monday of the past 4 weeks
  - Monday is a period, week is a season
- Periods are linked to current time:
  - If today is Wednesday, then periods are Tuesdays
PERIODS VS SEASONS

Tuesday  Tuesday  Tuesday  Tuesday  Tuesday

season 1 week  season 1 week  season 1 week  season 1 week

seasons 4 weeks

now

Period 1 day Tuesday

current week
2 WAYS TO CALCULATE BASELINE

- ‘baselinewma’ - Calculates the baseline by averaging data from the same timeframe in multiple equal time periods (‘seasons’) using the weighted moving average (WMA) algorithm.

- ‘baselinedev’ - Returns the number of deviations (by stddevpop algorithm) between the last data period and the same data periods in preceding seasons.
BASELINE WEIGHTED MOVING AVERAGE

Check if CPU usage is 2x higher than WMA on the same weekdays over last 5 days (exclusive)
COUNT OF DEVIATIONS

More than 3 deviations detected in the last 8h, by using input periods from 12 weeks.
Thank you