ZABBIX '25
CONFERENCE
GERMANY

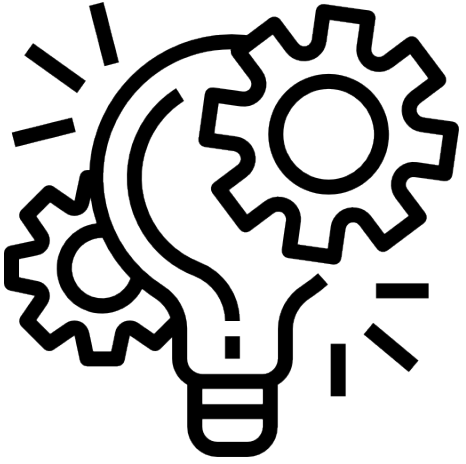
AI-based Anomaly Detection in Zabbix using DataForge

Wolfgang Alper

CEO
IntelliTrend GmbH

Zabbix meets AI

The idea of anomaly detection

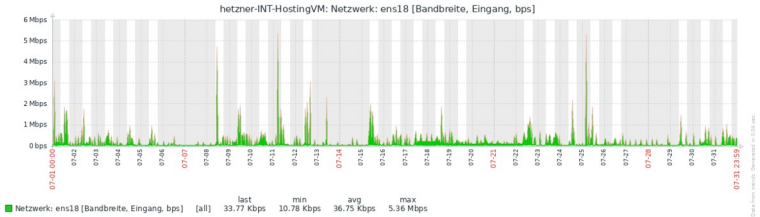


ZABBIX
PREMIUM PARTNER

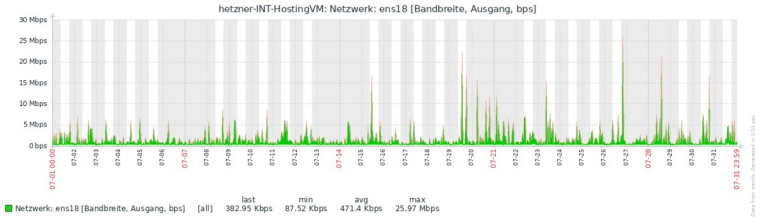
IntelliTrend
IT-Services

The idea of anomaly detection

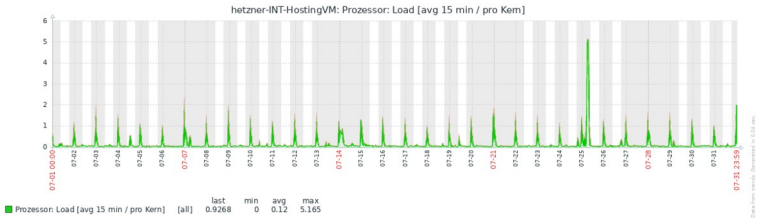
Network In



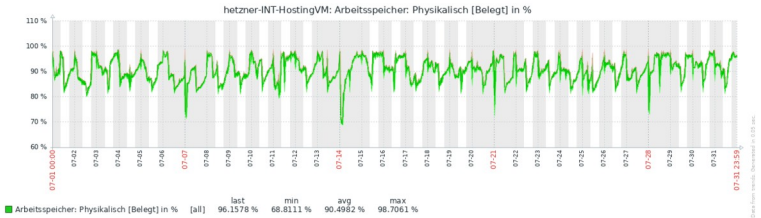
Network Out



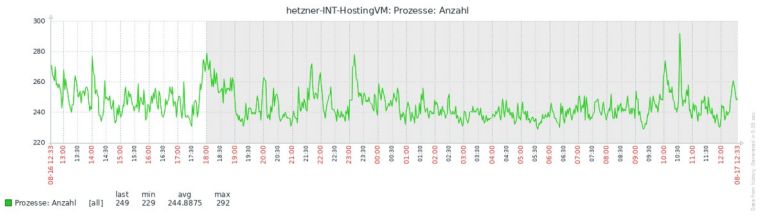
CPU Load



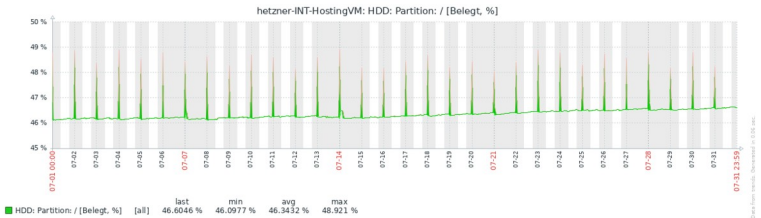
Memory Used



Processes Count



Free Disk Space



Generic metrics of a system

The idea of anomaly detection

Place individual metrics in a system-specific context to detect anomalies

- Instead of looking to one metric at a time, have a system that looks at multiple metrics at once
- Instead of using simple trigger functions, look at the data as a whole over a period of time
- Instead of using triggers with static conditions, let the system “learn” the specific characteristics over time with variable conditions
- Allow this concept to be used across multiple hosts

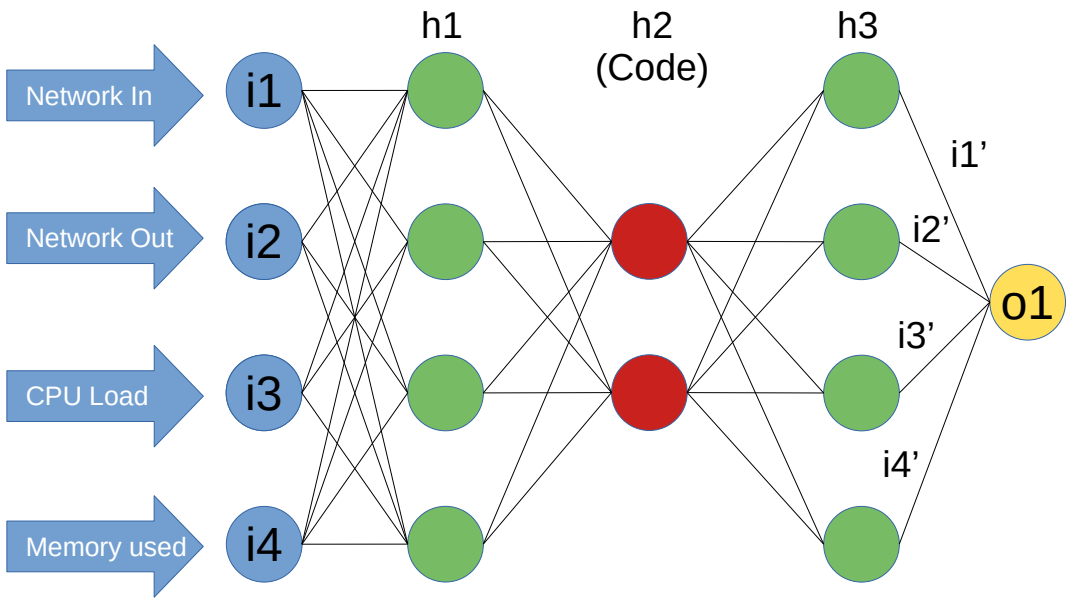
The idea of anomaly detection

Examples

- Monitoring server rooms:
  Pay attention not only to the temperature, but also relate it to the power consumption of the systems, air conditioning, etc.
- Server utilization:
  Pay attention not only to CPU utilization, but also to memory usage, the number of users, network traffic, etc.

Do this with series of values over time, not only with single values per metric

Long short-term memory based Autoencoder



- Incoming data is connected to i1 to i4 (ix)
- Hidden layer “h1” acts as an encoder to extract the features of that data
- Hidden layer “h2” is much smaller and extracts the essence of these features
- Hidden layer “h3” acts as an decoder to reconstruct the original data
- If there is a problem, hidden layer “h3” will cause a data reconstruction error
- This error is calculated as an anomaly using a “loss function” in “o1”

Simplified schematic representation of an Autoencoder neural network

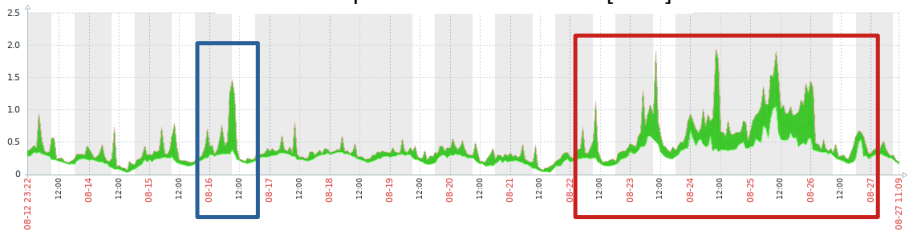
ZABBIX PREMIUM PARTNER

IntelliTrend IT-Services

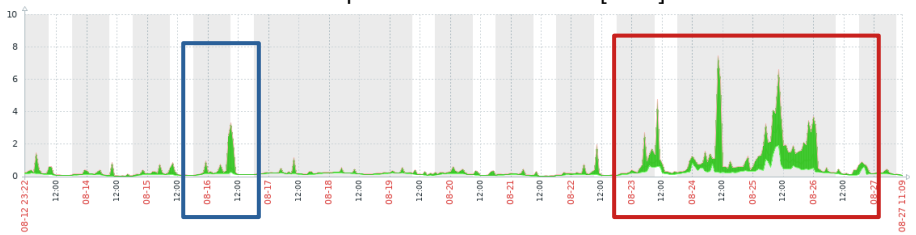
Using a model in production

Reconstruction errors represent the probabilities for anomalies based on their loss function

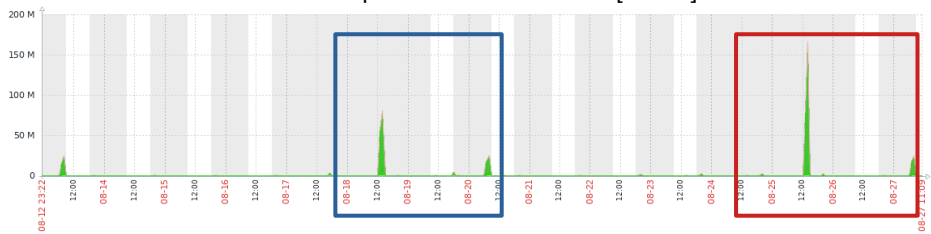
AI-Webshop: Reconstruction Error [MAE]



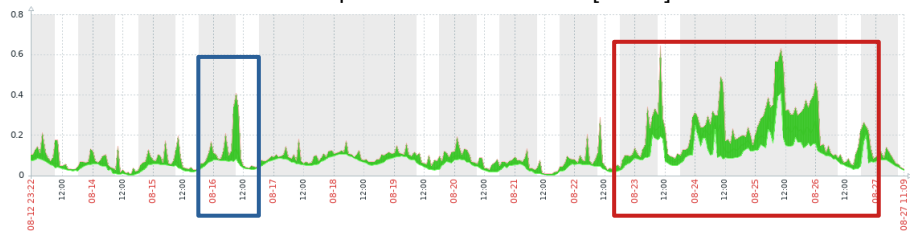
AI-Webshop: Reconstruction Error [MSE]



AI-Webshop: Reconstruction Error [MAPE]



AI-Webshop: Reconstruction Error [MSLE]



Example trigger expressions in Zabbix to alert an anomaly:

```
avg(/AI-Webshop/1.loss.mae,600)  >={$MIN_ANOMALY_MAE}
last(/AI-Webshop/m1.loss.mape)   >={$MIN_ANOMALY_MAPE}
min(/AI-Webshop/1.loss.mse,300)  >={$MIN_ANOMALY_MSE}
min(/AI-Webshop/1.loss.msle,300)>={$MIN_ANOMALY_MSLE}
```

Available loss functions to indicate an anomaly:

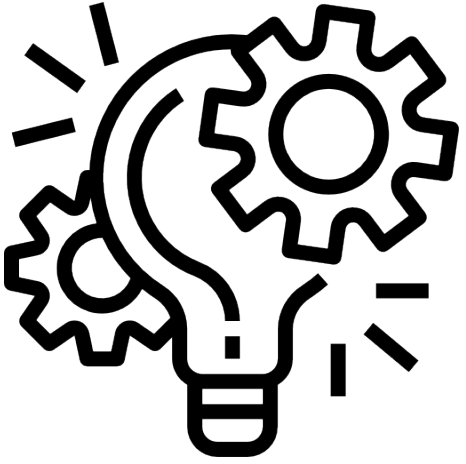
MAE	- **M**ean **A**bsolute **E**rror
MAPE	- **M**ean **A**bsolute **P**ercentage **E**rror
MSE	- **M**ean **S**quared **E**rror
MSLE	- **M**ean **S**quared **L**ogarithmic **E**rror

ZABBIX PREMIUM PARTNER

IntelliTrend IT-Services

AI-based anomaly detection

Integration with Zabbix

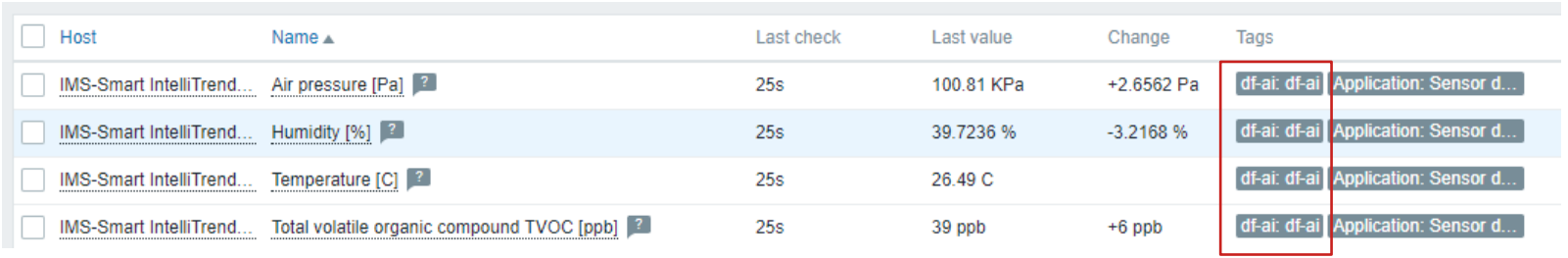


ZABBIX
PREMIUM PARTNER

IntelliTrend
IT-Services

Data extraction from Zabbix

Use of Tags to identify the items that should be used for training and later for inference



Host	Name ▲	Last check	Last value	Change	Tags
IMS-Smart IntelliTrend...	Air pressure [Pa]	25s	100.81 KPa	+2.6562 Pa	df-ai: df-ai, Application: Sensor d...
IMS-Smart IntelliTrend...	Humidity [%]	25s	39.7236 %	-3.2168 %	df-ai: df-ai, Application: Sensor d...
IMS-Smart IntelliTrend...	Temperature [C]	25s	26.49 C		df-ai: df-ai, Application: Sensor d...
IMS-Smart IntelliTrend...	Total volatile organic compound TVOC [ppb]	25s	39 ppb	+6 ppb	df-ai: df-ai, Application: Sensor d...

API Request

```
{
    "jsonrpc": "2.0",
    "method": "history.get",
    "params": {
        "output": "extend",
        "history": 0,
        "itemids": "912257",
        "sortfield": "clock",
        "sortorder": "DESC",
        "limit": 10
    },
    "id": 1
}
```

API Response

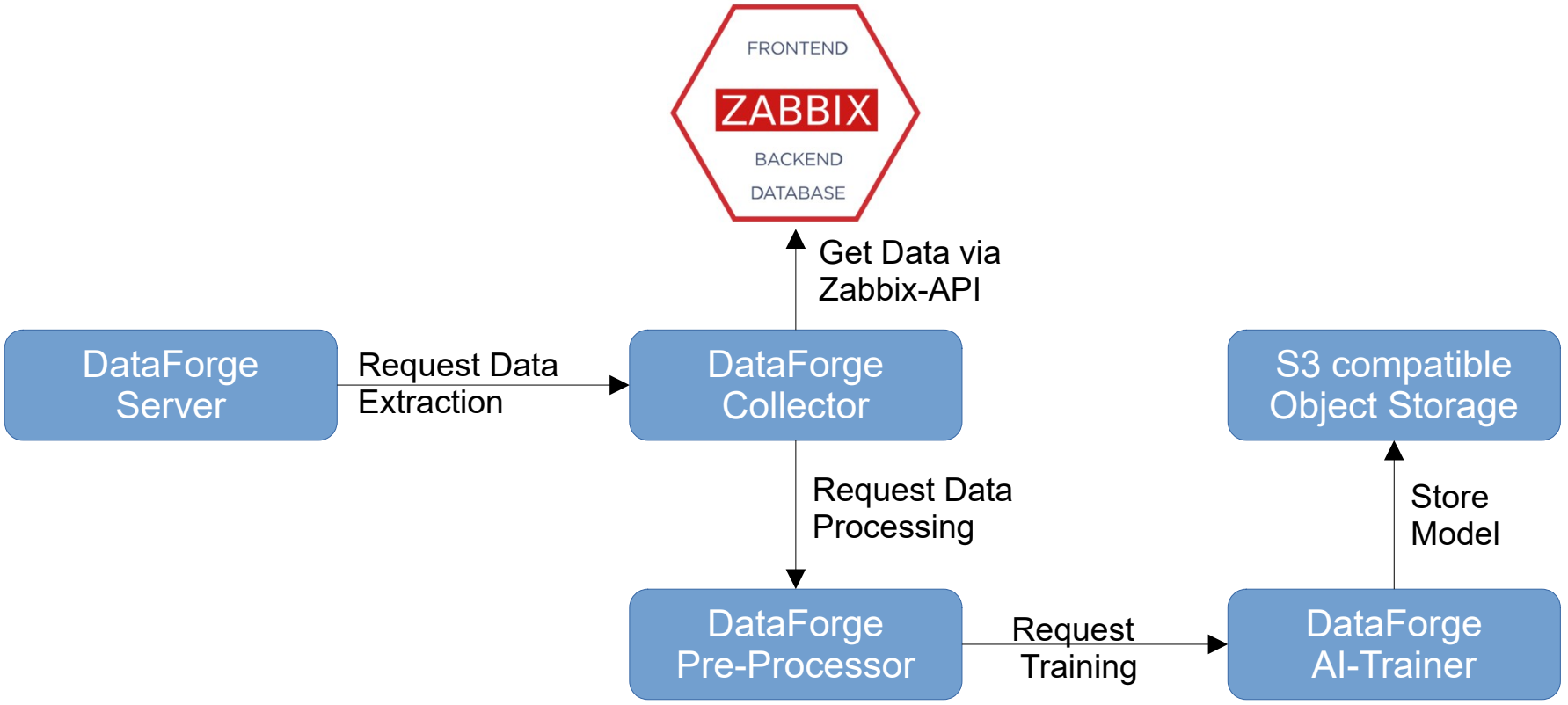
```
{
  "jsonrpc": "2.0",
  "result": [
    {
      "itemid": "912257",
      "clock": "1724588945",
      "value": "26.31999969",
      "ns": "227934509"
    },
    {
      "itemid": "912257",
      "clock": "1724588885",
      "value": "26.32999992",
      "ns": "87916196"
    },
```

Using the Zabbix-API to get univariate time-series data for a given item

ZABBIX PREMIUM PARTNER

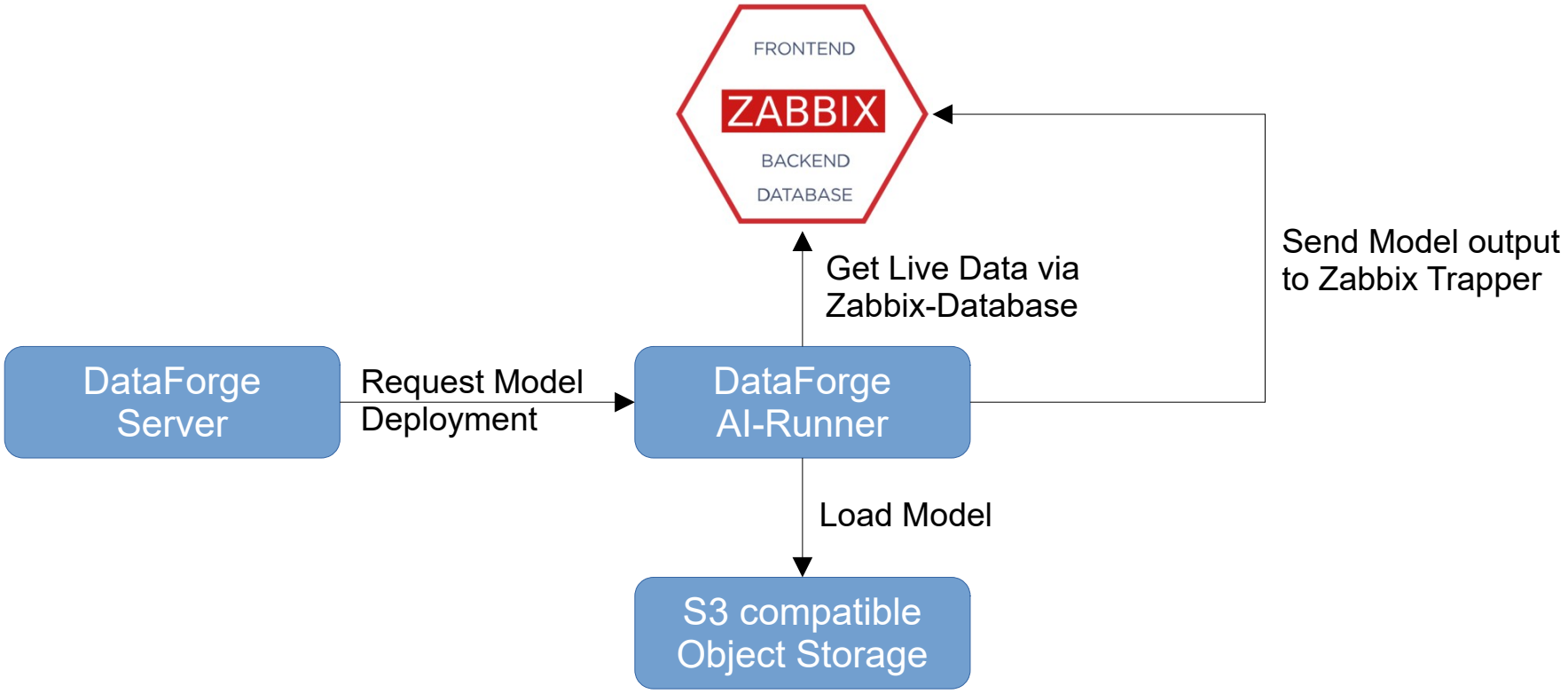
IntelliTrend IT-Services

Integration with Zabbix - Training



No changes to Zabbix Source code required

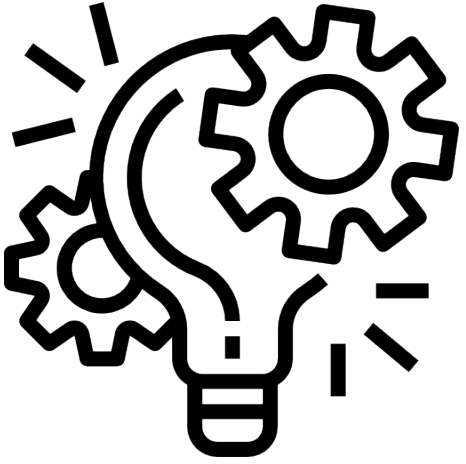
Integration with Zabbix - Inference



No changes to Zabbix Source code required

AI-based anomaly detection

Detecting simple anomalies



ZABBIX
PREMIUM PARTNER

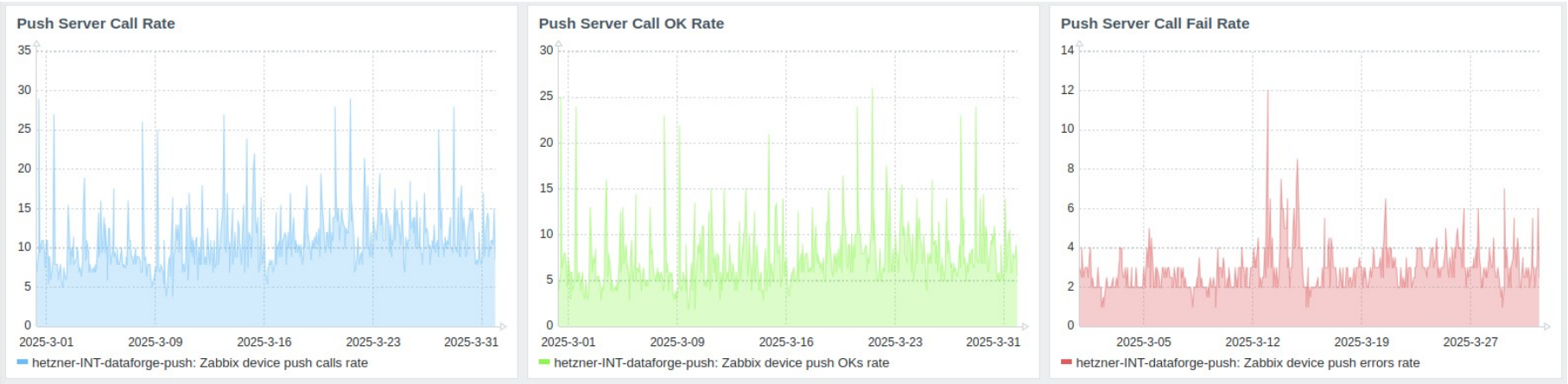
IntelliTrend
IT-Services

Detecting simple anomalies

These are the three most important metrics of our Push Notification Server:

- “Call Rate”, “Call OK Rate” and “Call Fail Rate”
- These rates indicate the number of attempted, successful and failed push notification requests over time

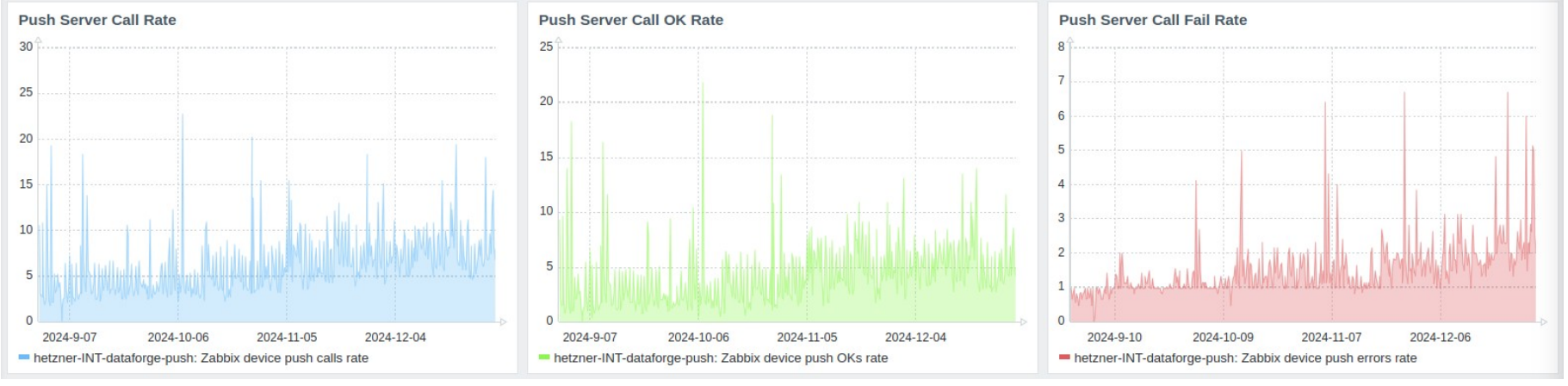
The failure rate of around 20%-30% is based on users who reinstall / uninstall the app, but leave the mediatype configuration as is on their Zabbix user account



Detecting simple anomalies

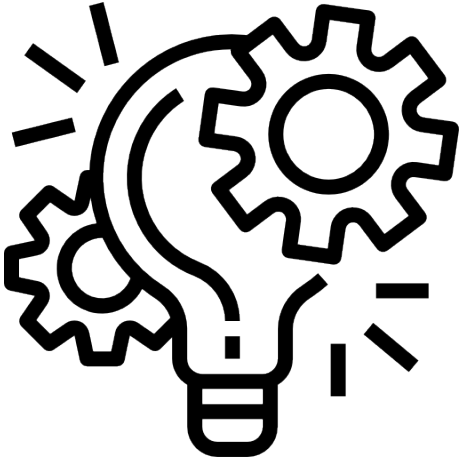
Writing a trigger for these metrics is non-trivial:

- Large spikes can be expected when a customer has a large number of issues that trigger many push notifications
- Looking at a longer time interval all metrics are trending upwards as the application gains users



AI-based anomaly detection

How can we solve this with AI?

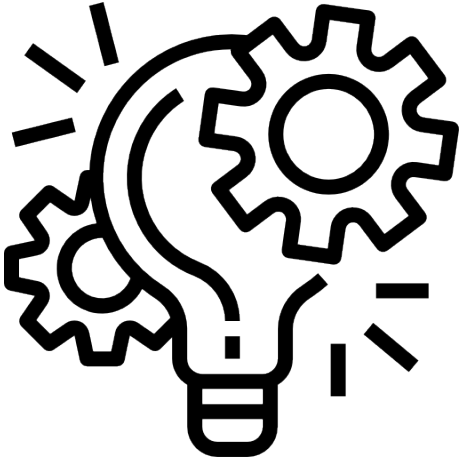


From data to model

- Create a training dataset with history data from Zabbix
- Create one or more test datasets with history data from Zabbix
- Create a model configuration
- Create / Train a model
- Test the model
- Assign the the model to an AI-Runner

AI-based anomaly detection

Datasets



Tagging Items for Data Extraction

- Mark metrics that AI should use by assigning the tag “df-ai” to relevant items in Zabbix
- These tags will be used to extract the data from Zabbix for:
  - Training the model
  - Testing the model
  - Using the model in production (Inference)

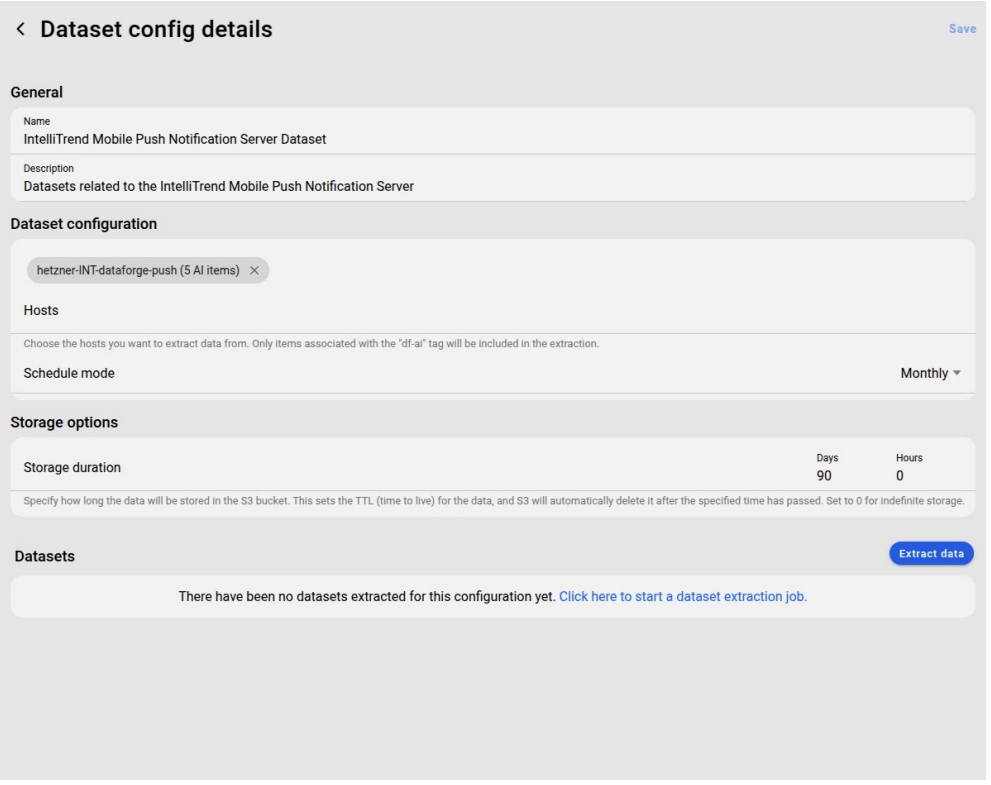
Name ▲	Triggers	Key	Interval	History	Trends	Type	Status	Tags
DataForge Push Server: Push server raw data: Zabbix device push calls rate	Triggers 2	dfpush.zabbix_device_push_calls_rate		90d	365d	Dependent item	Enabled	df-ai
DataForge Push Server: Push server raw data: Zabbix device push errors/calls absolute ratio	Triggers 2	dfpush.zabbix_device_push_errors_calls_ratio		90d	365d	Dependent item	Enabled	df-ai
DataForge Push Server: Zabbix device push errors/calls recent ratio	Triggers 2	dfpush.zabbix_device_push_errors_calls_recent_ratio	1m	90d	365d	Calculated	Enabled	df-ai
DataForge Push Server: Push server raw data: Zabbix device push errors rate		dfpush.zabbix_device_push_errors_rate		90d	365d	Dependent item	Enabled	df-ai

ZABBIX PREMIUM PARTNER

IntelliTrend IT-Services

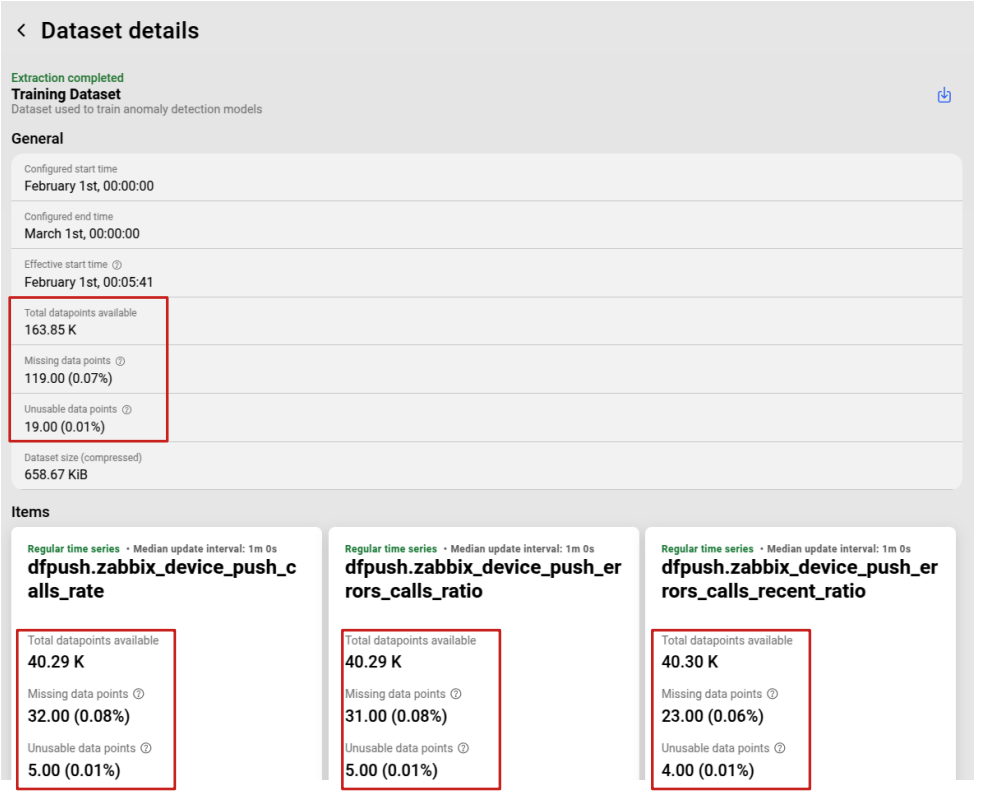
Creating a dataset

- Add host(s) with tagged items
- Choose scheduling mode for automatic update
- Set storage duration for house keeping
- Datasets needs to be created for:
  - Training
  - Testing



Creating a dataset

- After the data extraction process finished, dataset details are updated
- The analysis is helpful to decide whether the dataset will be useful in a training or in a test
- Datasets can be be downloaded, decompressed and converted to a variety of formats for further processing in 3th party applications



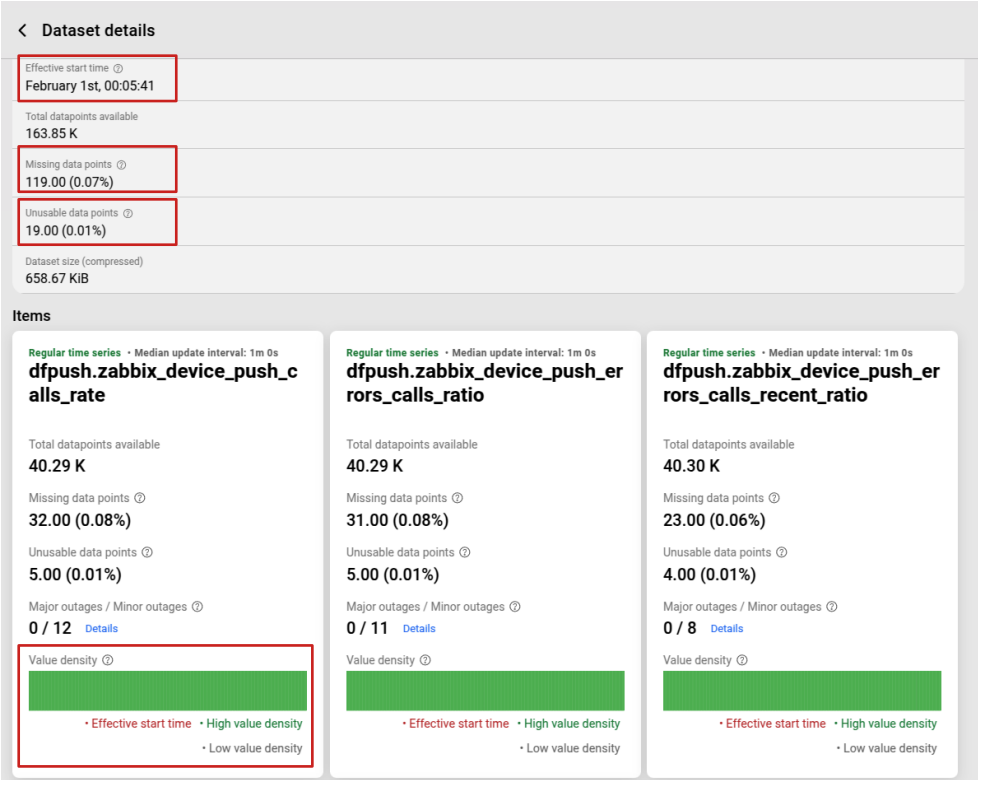
ZABBIX PREMIUM PARTNER

IntelliTrend IT-Services

Dataset – Quality of data

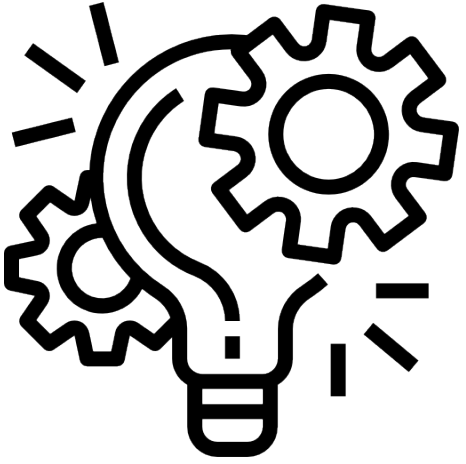
Analysis of the dataset:

- This dataset has a high value density and looks healthy
- Very few datapoints are missing or unusable
- All of the time series were detected as having a consistent update interval
- There are no major outages and an evenly distributed value density
- The effective start time of the dataset (the time at which each time series had at least one value) is only 5m41s after the configured start time

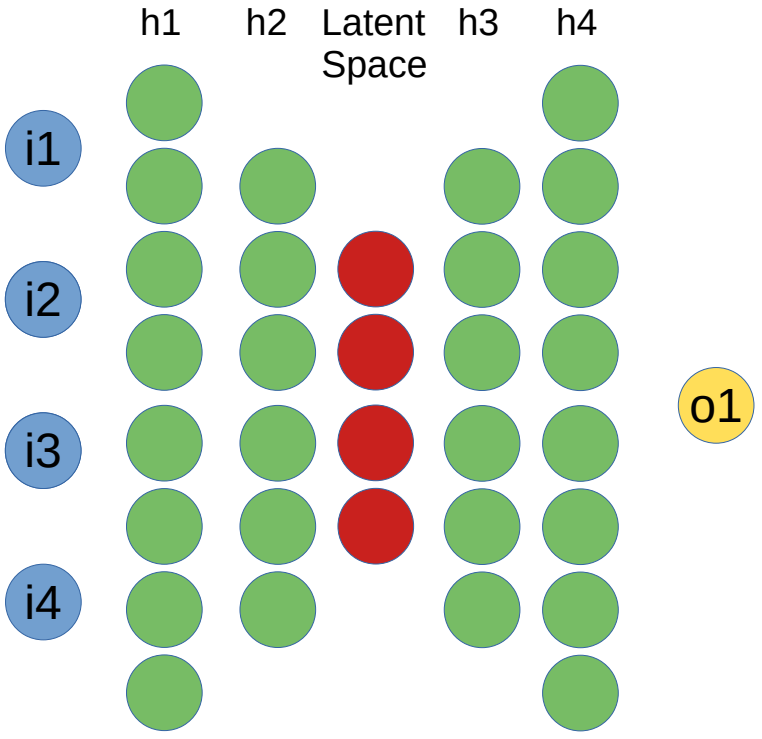


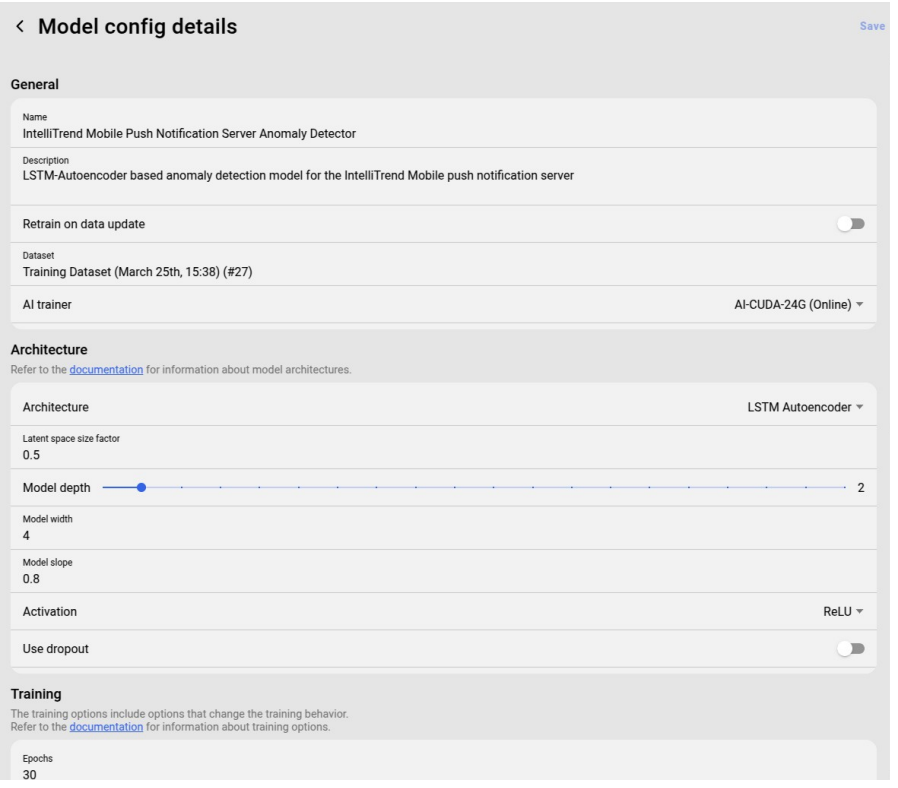
AI-based anomaly detection

Creating a Model



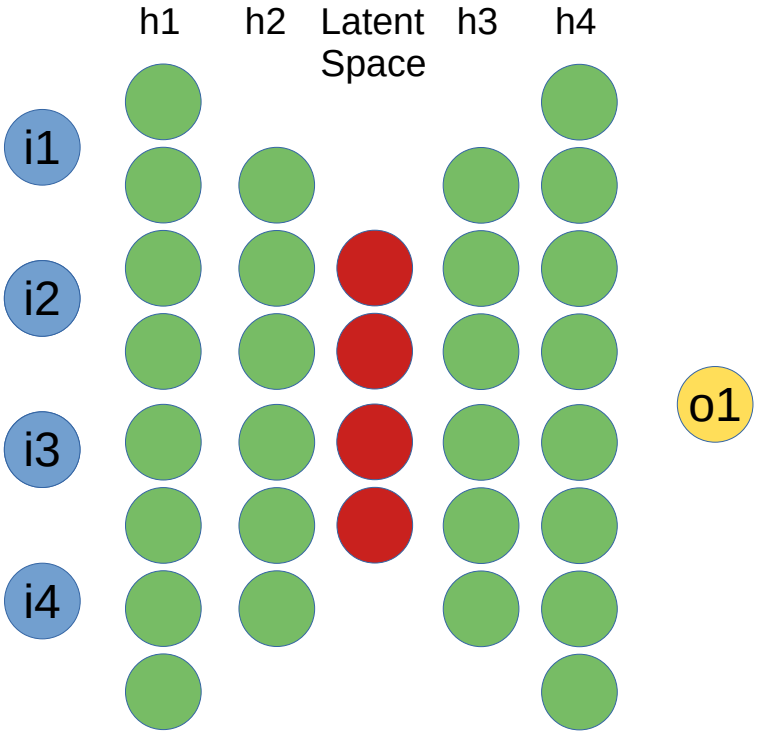
Model configuration

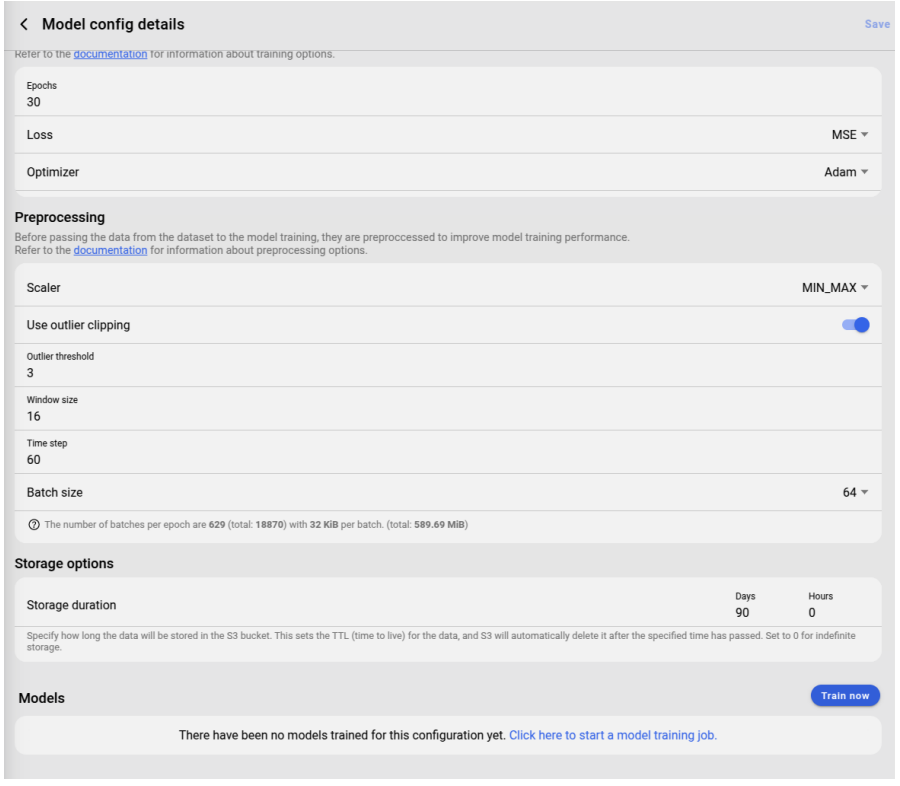




Architecture defines model type, number of layers / neurons, activation function etc.

Model configuration

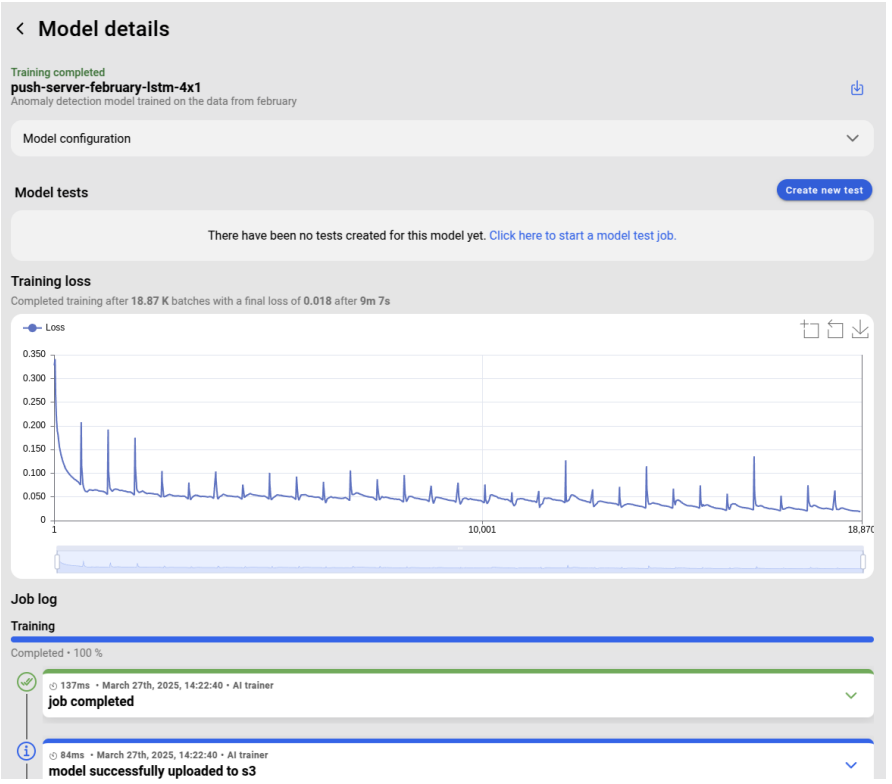




Preprocessing defines operations on the dataset for training including batch size

Model creation

- Model is created based on model configuration and training dataset
- The number of selected metrics (items) effected the number of neurons
- Model details show the training loss during the training process



ZABBIX PREMIUM PARTNER

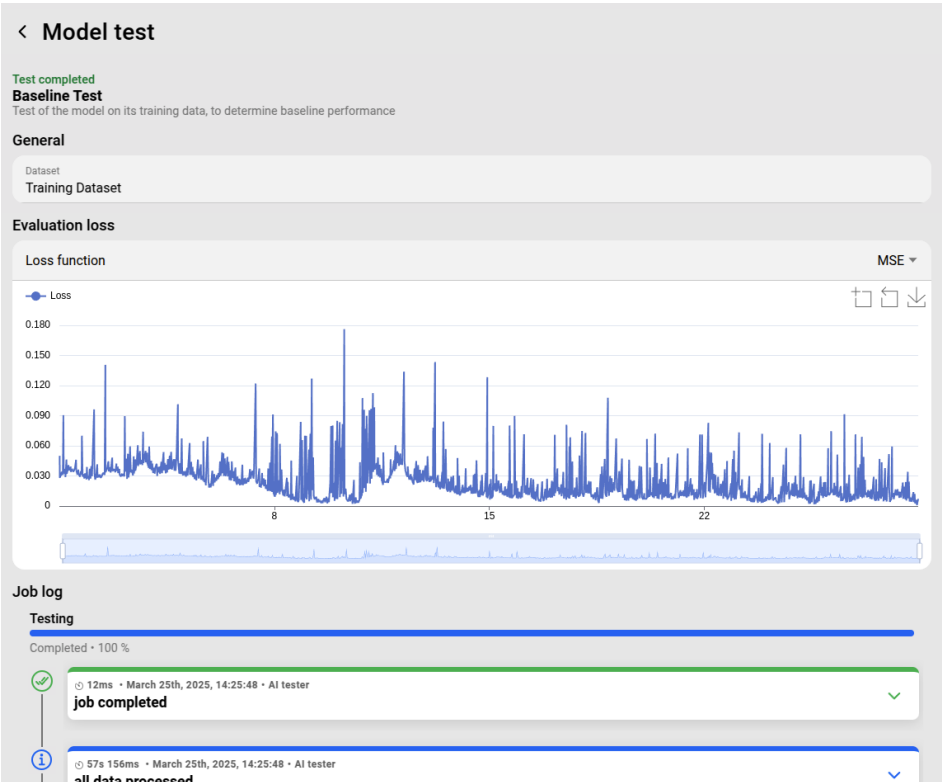
IntelliTrend IT-Services

Model test

- After the model is created, it can be tested using datasets
- Typically there is a baseline test and also validation tests
- Any number of datasets can be created from Zabbix history data
- Different loss functions can be used

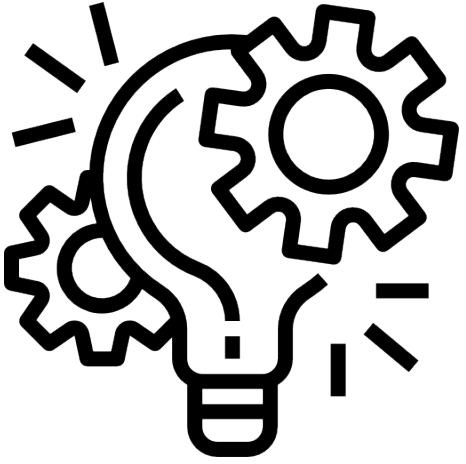
Available loss functions to indicate an anomaly:

MAE	- **M**ean **A**bsolute **E**rror
MAPE	- **M**ean **A**bsolute **P**ercentage **E**rror
MSE	- **M**ean **S**quared **E**rror
MSLE	- **M**ean **S**quared **L**ogarithmic **E**rror



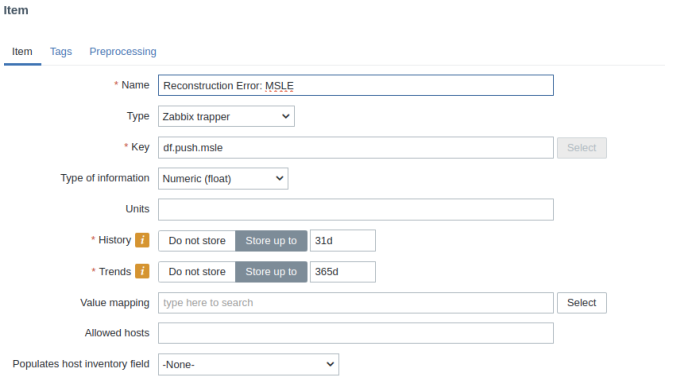
AI-based anomaly detection

Using the model



Creating output items

- The result of the model evaluation is sent to Zabbix using items
- Item type must be trapper with type numeric (float)
- The item keys must use a consistent prefix and then be suffixed with the appropriate loss metric like “df.push.msle”



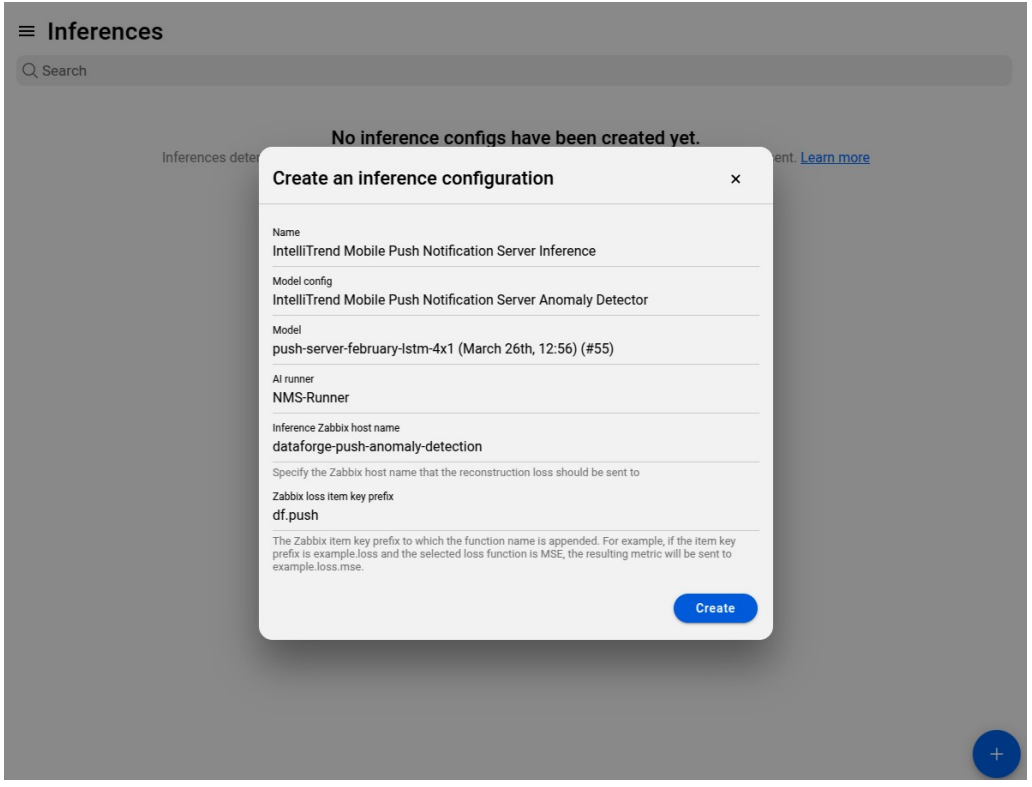
Name ▲	Triggers	Key	Interval	History	Trends	Type	Status
Reconstruction Error: MAE		df.push.mae		31d	365d	Zabbix trapper	Enabled
Reconstruction Error: MAPE		df.push.mape		31d	365d	Zabbix trapper	Enabled
Reconstruction Error: MSE		df.push.mse		31d	365d	Zabbix trapper	Enabled
Reconstruction Error: MSLE		df.push.msle		31d	365d	Zabbix trapper	Enabled

ZABBIX PREMIUM PARTNER

IntelliTrend IT-Services

Deploying the model

- The process of actually using a model is called inference
- The inference configration defines the model, AI-Runner and the host in Zabbix to receive the model output

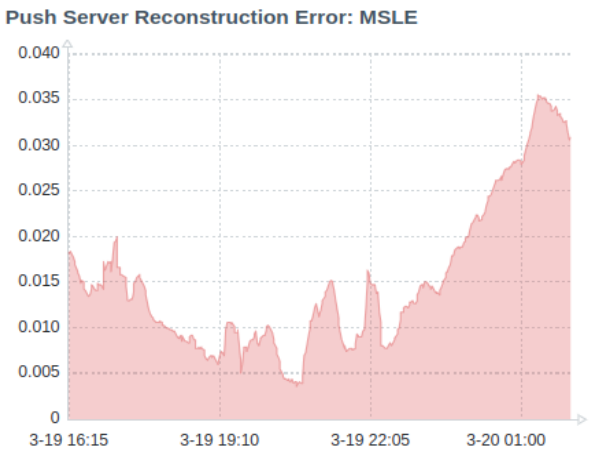


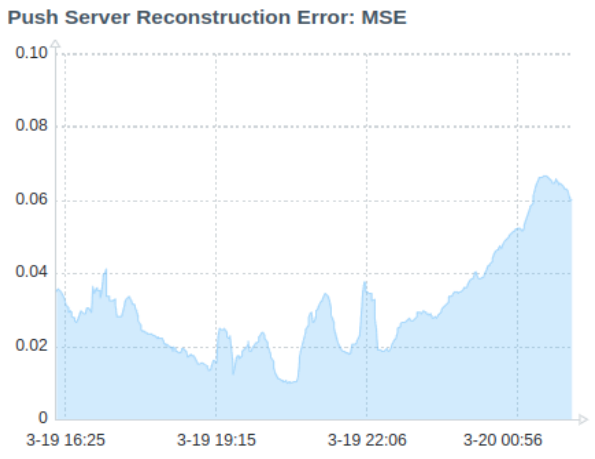
ZABBIX PREMIUM PARTNER

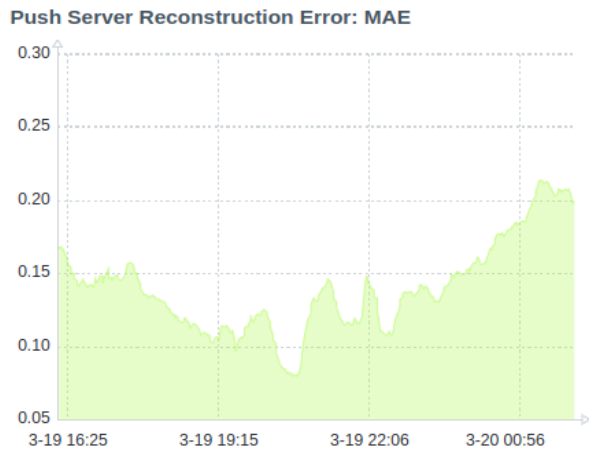
IntelliTrend IT-Services

Deploying the model

- After deploying the model, the output for each loss function shows up as item in Zabbix
- These items can be used with any trigger function in Zabbix to detect an anomaly



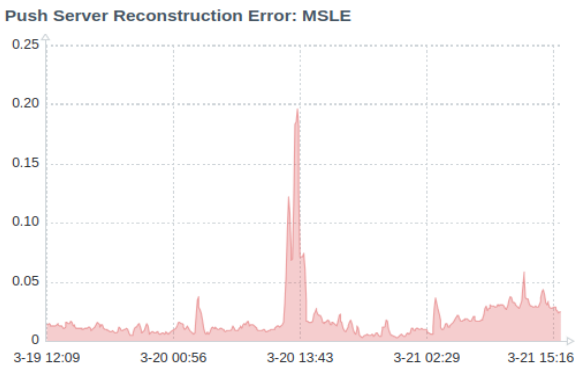


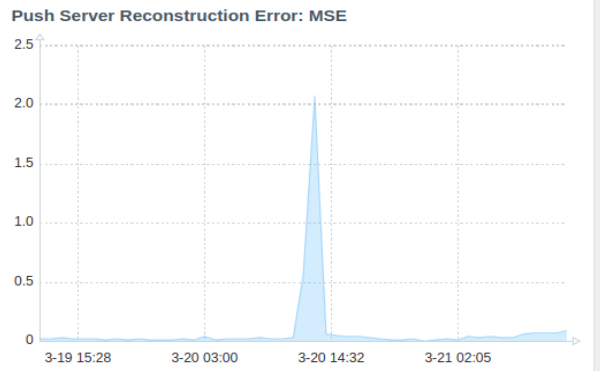


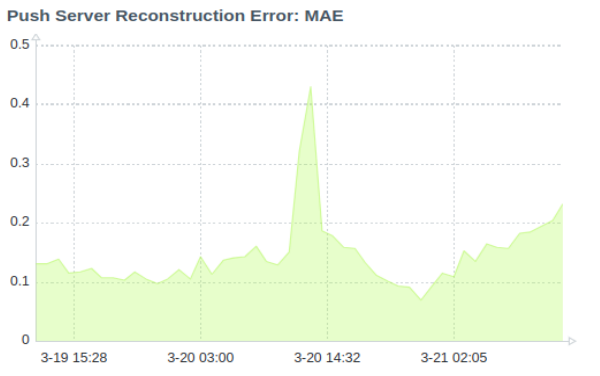
Detecting an anomaly

When an anomaly is detected, the values of the loss functions change significantly

For example MSE (Mean Squared Error) is more sensible than MAE (Mean absolute Error)

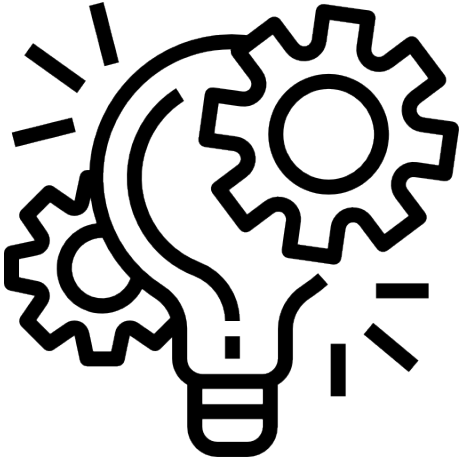






AI-based anomaly detection

Lets look at more examples

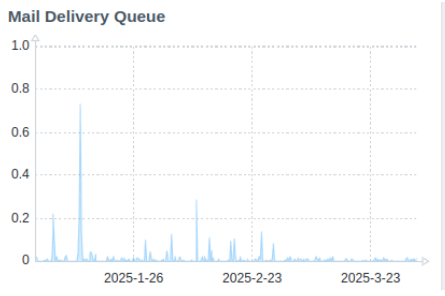


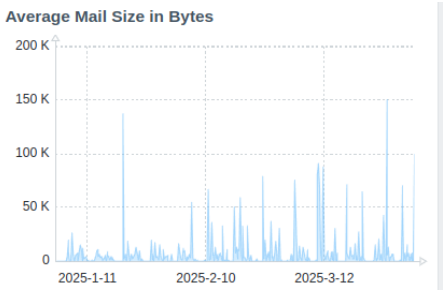
ZABBIX
PREMIUM PARTNER

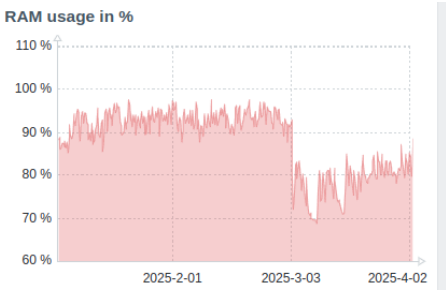
IntelliTrend
IT-Services

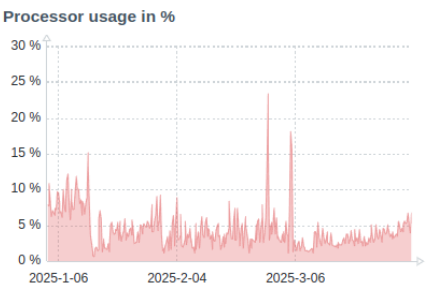
Example Exchange Server

- Below some KPI’s from a Microsoft Exchange Server
- The graphs show the mail queue, average mail size in bytes and %ram and %cpu usage
- There are 17 additional items that will be used for training for a total of 21 items









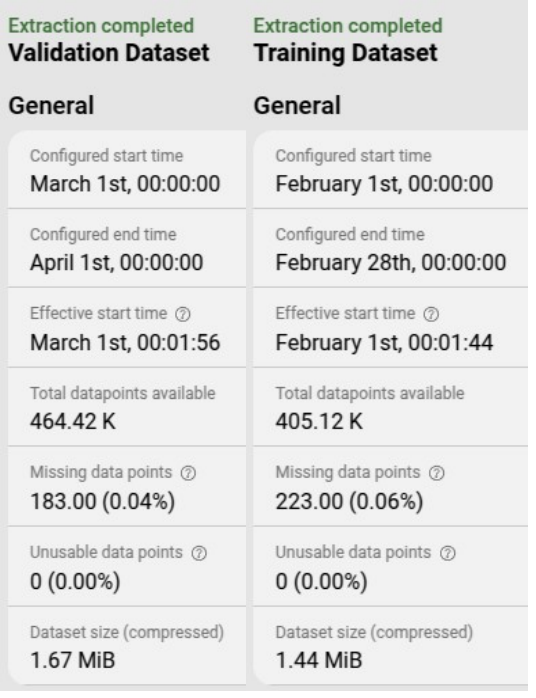
Example Exchange Server

Name ▲	Last check	Last value	Change	Tags
Arbeitsspeicher: Physikalisch [Belegt] in %	1m 47s	87.7376 %	+0.4971 %	df-ai
Arbeitsspeicher: Swap [Belegt] in %	1m 47s	41.9535 %	-0.011 %	df-ai
Exchange: MSExchange-Datenbank [Information Store / Protokoll: Generierte Bytes/s]	53s	0		df-ai
Exchange: MSExchange-Datenbank [Information Store / Protokoll: Schreiben Bytes/s]	16s	0		df-ai
Exchange: MSExchange-Datenbank [Information Store I/O / Datenbanklesevorgänge/s]	1m 30s	0		df-ai
Exchange: MSExchange-Datenbank [Information Store I/O / Datenbankschreibvorgänge/s]	1m 39s	0.9874	-0.000143	df-ai
Exchange: MSExchangeTransport [SMTP-Empfang(_total) / Aktuelle Verbindungen]	53s	2		df-ai
Exchange: MSExchangeTransport [SMTP-Empfang(_total) / Empfangene Bytes/s]	16s	41373.7277	-116681.7052	df-ai
Exchange: MSExchangeTransport [SMTP-Empfang(_total) / Empfangene Nachrichten / Delta]	1m 30s	7	-4	df-ai
Exchange: MSExchangeTransport [SMTP-Empfang(_total) / Empfangene Nachrichten/s]	4s	0		df-ai
Exchange: MSExchangeTransport [SmtpSend(_total) / Aktuelle Verbindungen]	1m 42s	0		df-ai
Exchange: MSExchangeTransport [SmtpSend(_total) / Gesendete Bytes/s]	53s	408635.0496	+378003.4473	df-ai
Exchange: MSExchangeTransport [SmtpSend(_total) / Gesendete Nachrichten / Delta]	4s	14	-2	df-ai
Exchange: MSExchangeTransport [SmtpSend(_total) / Gesendete Nachrichten/s]	11s	1.9698	+0.000057	df-ai
Exchange: MSExchangeTransport [Zustellungswarteschlange / Warteschlangen extern]	1m 30s	0		df-ai
Exchange: MSExchangeTransport [Zustellungswarteschlange / Warteschlangen intern]	53s	0		df-ai
Exchange: MSExchange [Client Type / Messages opened/sec]	1m 42s	2.9536	+2.9536	df-ai
Exchange: MSExchange [IS Store / Messages opened/sec]	1m 41s	0	-0.9906	df-ai
Exchange: MSExchange [MapiHttp Emsmdb / aktive Benutzer / Anzahl]	1m 42s	2	+1	df-ai
Netzwerk: Ping [Antwortzeit]	1m 31s	0.53ms	+0.054ms	df-ai
Netzwerk: Ping [Status]	1m 31s	Up (1)		df-ai

Tag items for data extraction

Example Exchange Server

- The training dataset contains 405K datapoints obtained from 21 items over a duration of one month
- The validation dataset contains 464K datapoints obtained from 21 items over a duration of one month
- During the training process with 100 epochs, the dataset will be expanded to 4.8GiB

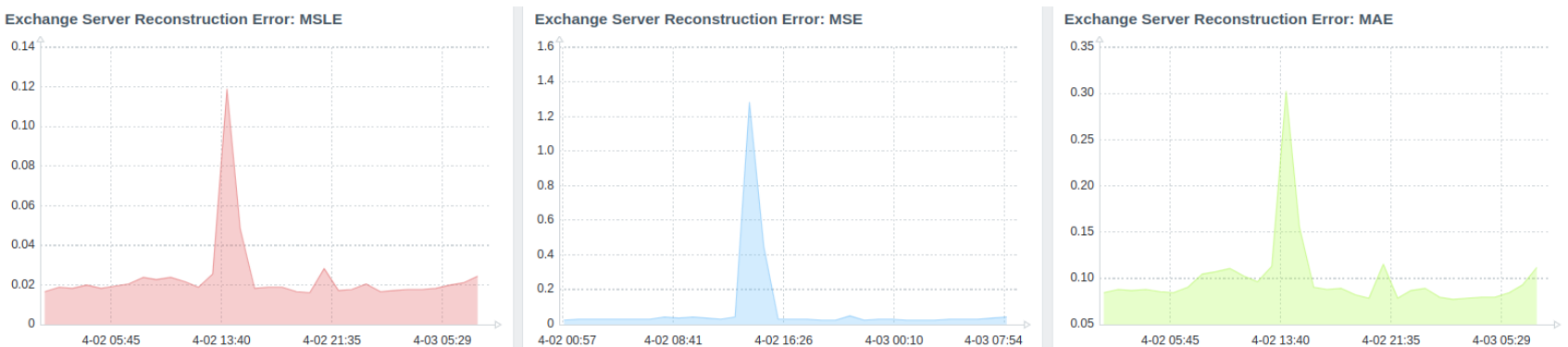


Extraction completed **Validation Dataset**	Extraction completed **Training Dataset**
General	**General**
Configured start time March 1st, 00:00:00	Configured start time February 1st, 00:00:00
Configured end time April 1st, 00:00:00	Configured end time February 28th, 00:00:00
Effective start time March 1st, 00:01:56	Effective start time February 1st, 00:01:44
Total datapoints available 464.42 K	Total datapoints available 405.12 K
Missing data points 183.00 (0.04%)	Missing data points 223.00 (0.06%)
Unusable data points 0 (0.00%)	Unusable data points 0 (0.00%)
Dataset size (compressed) 1.67 MiB	Dataset size (compressed) 1.44 MiB

ZABBIX PREMIUM PARTNER

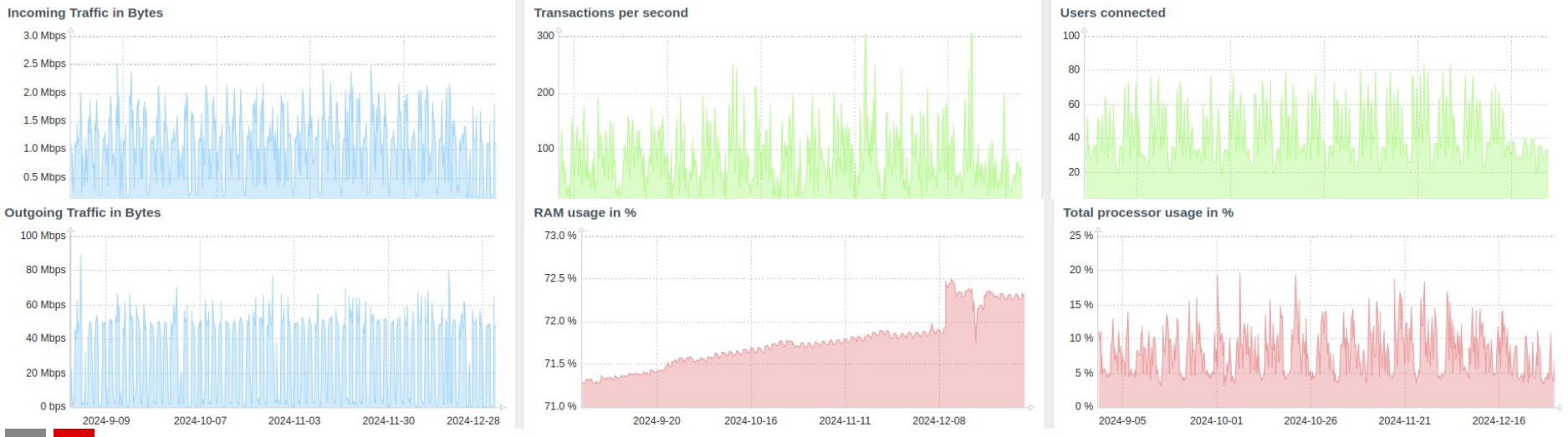
Example Exchange Server

- We now have created one trigger that monitors a total of 21 related items
- If any of these items shows an unexpected behavior over time or in relation to other items, this trigger will fire
- To get the root cause, further investigation is needed
- If required, specific triggers can be added



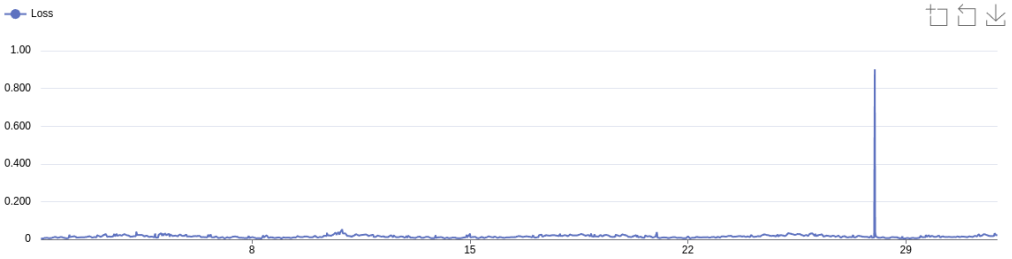
Example MS-SQL Server

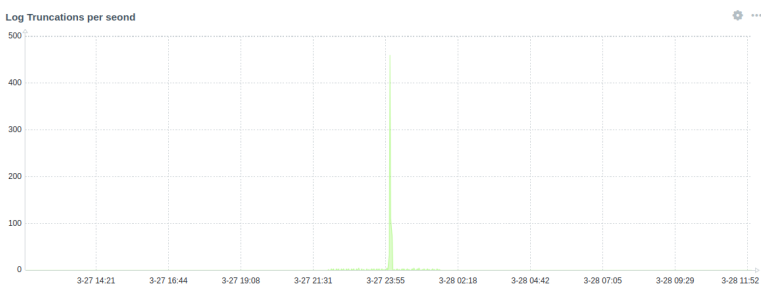
- Below some KPI’s from a Microsoft SQL Server
- There are 56 additional items that will be used for training for a total of 62 items

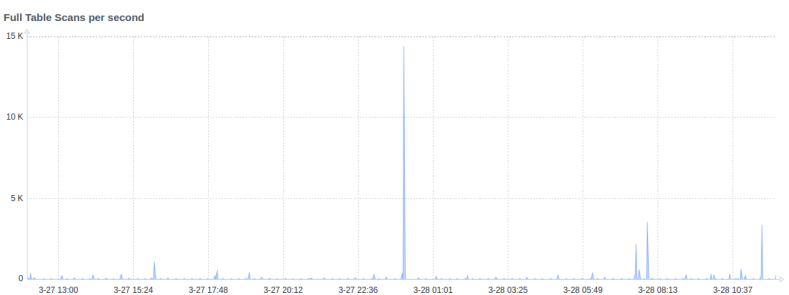


Example MS-SQL Server

- At 3-27 the model detected an anomaly
- After reviewing the metrics, the culprit was found quickly
- The number of full table scans per second spiked to almost 15.000 while the number of log truncations also spiked at around 480

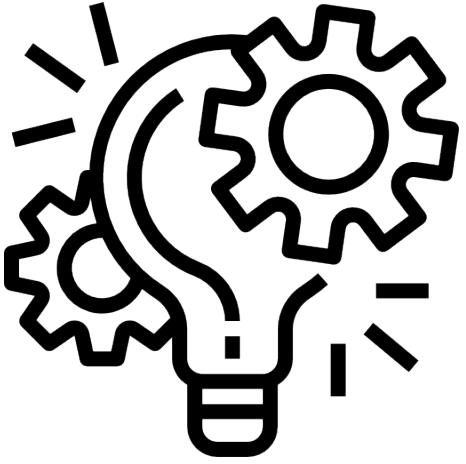






AI-based anomaly detection

Combine AI with regular trigger



ZABBIX
PREMIUM PARTNER

IntelliTrend
IT-Services

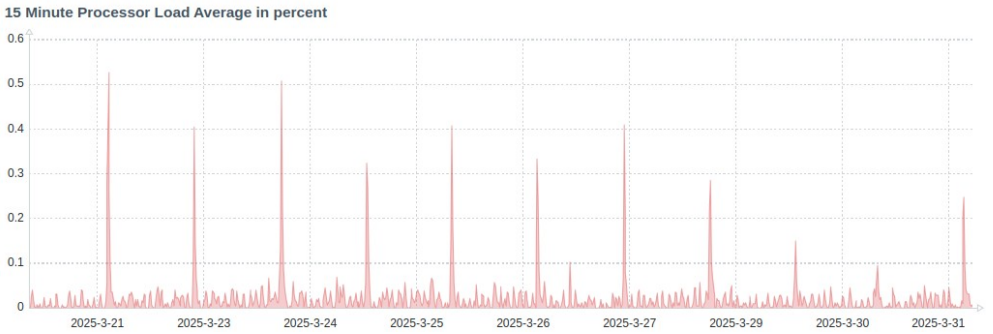
Combine AI with regular trigger

Challenge

- Every evening at around 21:00, Zabbix sends an alert
- Reason: Processor load gets high

Possible solutions

- Schedule periodic maintenance
- Use time() function
- … combine with anomaly detection



ZABBIX PREMIUM PARTNER

IntelliTrend IT-Services

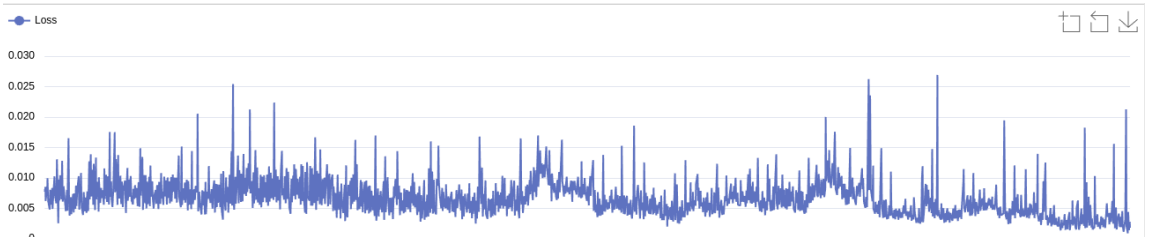
Combine AI with regular trigger

Mark items, create dataset, train model

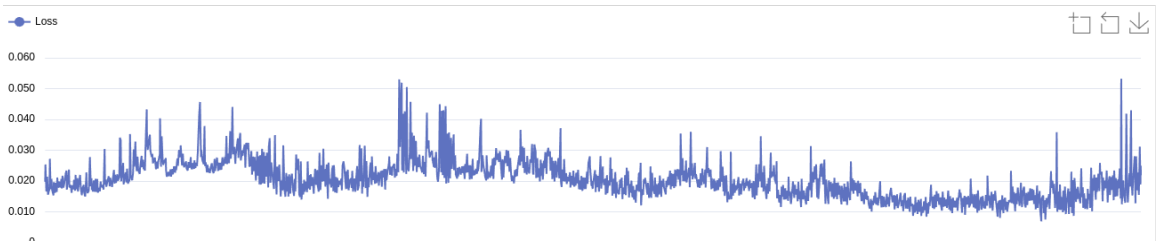
Name ▲	Triggers	Key	Interval	History	Trends	Type	Status	Tags	Info
TPL: OS_Linux [Server, All] [Basic]: Arbeitsspeicher: Physikalisch [Belegt] in %	Triggers 2	vm.memory.size[pused]	2m	30d	365d	Zabbix agent	Enabled	df-ai Application: Arbeitssp…	
TPL: OS_Linux [Server, All] [Basic]: Arbeitsspeicher: Physikalisch [FREI] in %		vm.memory.size[pavailable]	2m	30d	365d	Zabbix agent	Enabled	df-ai Application: Arbeitssp…	
TPL: OS_Linux [Server, All] [Basic]: Arbeitsspeicher: Swap [Belegt] in %	Triggers 2	system.swap.size[,pused]	2m	30d	365d	Zabbix agent	Enabled	df-ai Application: Arbeitssp…	
TPL: OS_Linux [Server, All] [Basic]: Arbeitsspeicher: Swap [Frei] in %		system.swap.size[,pfree]	2m	30d	365d	Zabbix agent	Enabled	df-ai Application: Arbeitssp…	
TPL: Netzwerk Ping [All] [basic]: Netzwerk: Ping [Antwortzeit]		icmppingsec	2m	30d	365d	Simple check	Enabled	df-ai Application: Netzwerk	
TPL: Netzwerk Ping [All] [basic]: Netzwerk: Ping [Status]	Triggers 1	icmpping	2m	30d	365d	Simple check	Enabled	df-ai Application: Netzwerk	
TPL: OS_Linux [Server, All] [Basic]: Prozesse: Anzahl [aktiv]		proc.num[,,run]	2m	30d	365d	Zabbix agent	Enabled	df-ai Application: Prozesse	
TPL: OS_Linux [Server, All] [Basic]: Prozessor: Auslastung [Gesamt] in %	Triggers 2	system.cpu.util[,total]	2m	30d	365d	Calculated	Enabled	df-ai Application: Prozessor	
TPL: OS_Linux [Server, All] [Basic]: Prozessor: Context switches [/sec]		system.cpu.switches	2m	30d	365d	Zabbix agent	Enabled	df-ai Application: Prozessor	
TPL: OS_Linux [Server, All] [Basic]: Prozessor: Interrupts [/sec]		system.cpu.intr	2m	30d	365d	Zabbix agent	Enabled	df-ai Application: Prozessor	

Combine AI with regular trigger

Baseline test



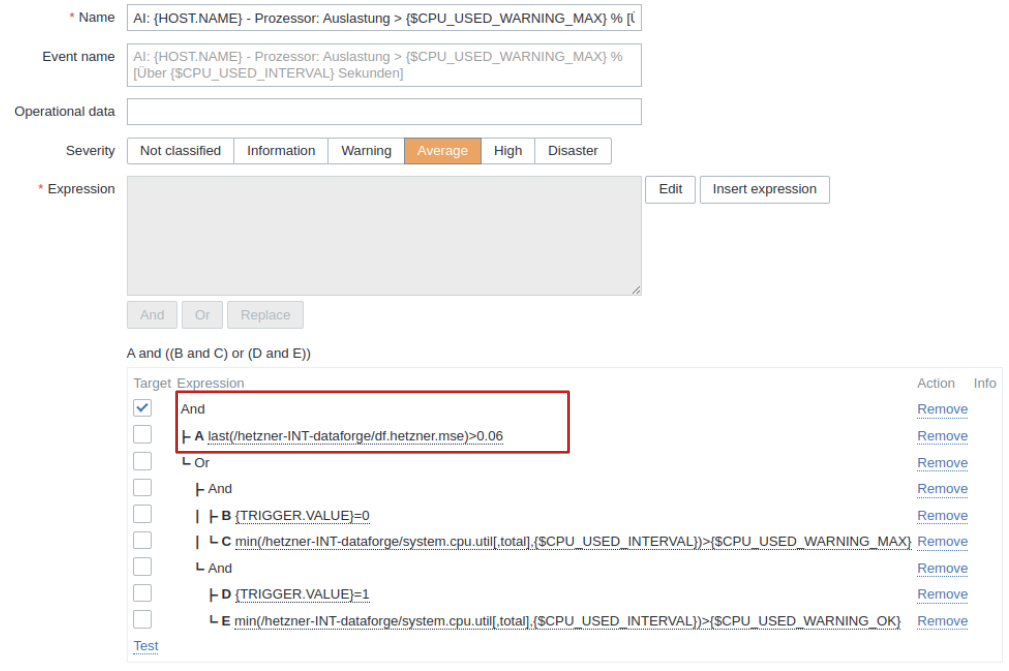
Validation test



Training looks good,
no anomaly at 21:00
during backup

Combine AI with regular trigger

Reconfigure the trigger – only fire if there is also an anomaly



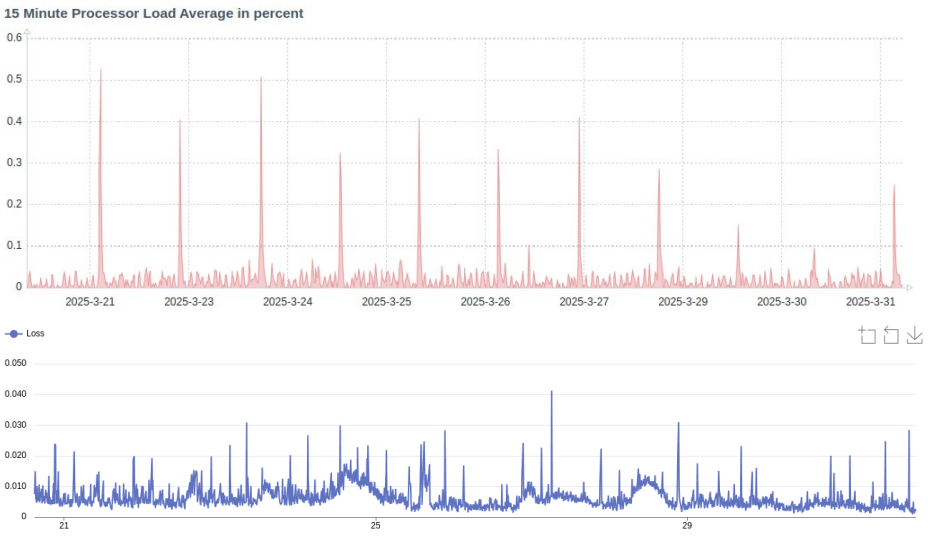
ZABBIX PREMIUM PARTNER

IntelliTrend IT-Services

Combine AI with regular trigger

It works – no alarm sent during the backup at 21.00

Average	OK	AI: {HOST.NAME} - Prozessor: Auslastung > {$CPU_USED_WARNING_MAX} % [Über {$CPU_USED_INTERVAL} Sekunden]	**last(**/hetzner-INT-dataforge/df.hetzner.mse**)**>0.1 and(({TRIGGER.VALUE}=0 and **min(**/hetzner-INT-dataforge/system.cpu.util[,total], {$CPU_USED_INTERVAL}**)**>{$CPU_USED_WARNING_MAX}) or ({TRIGGER.VALUE}=1 and **min(**/hetzner-INT-dataforge/system.cpu.util[,total], {$CPU_USED_INTERVAL}**)**>{$CPU_USED_WARNING_OK}))	Ena
Average	PROBLEM	{HOST.NAME} - Prozessor: Auslastung > {$CPU_USED_WARNING_MAX} % [Über {$CPU_USED_INTERVAL} Sekunden]	({TRIGGER.VALUE}=0 and **min(**/hetzner-INT-dataforge/system.cpu.util[,total],{$CPU_USED_INTERVAL}**)**>{$CPU_USED_WARNING_MAX}) or ({TRIGGER.VALUE}=1 and **min(**/hetzner-INT-dataforge/system.cpu.util[,total],{$CPU_USED_INTERVAL}**)**>{$CPU_USED_WARNING_OK})	Ena

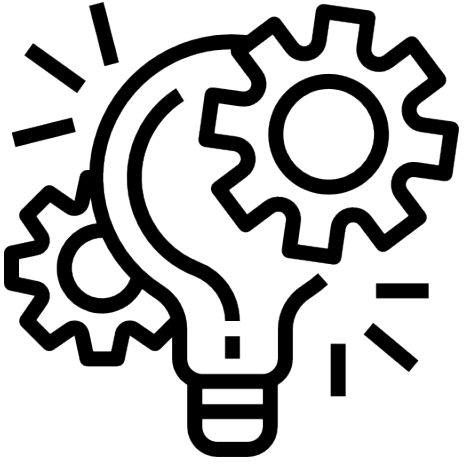


Combine AI with regular trigger

But – It sents alarms when the load is at an unusal time like here at 11:18

AI-based anomaly detection

Hardware requirements



ZABBIX
PREMIUM PARTNER

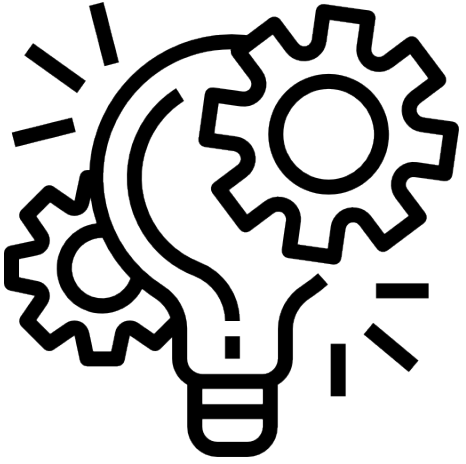
Hardware requirements

Hardware requirements for these autoencoder models are quite moderate. As an example we will use the 100 epoch MS-SQL server model with 62 items and one month of data:

- NVIDIA RTX-4090: Training process used 1.3GiB of DRAM and 458MiB of VRAM with the GPU hovering around 60% utilization and the CPU loading 2 cores for the training process.
  Training took 6m 32s or 3.92s per epoch using this setup.
- AMD R9 7950X3D CPU: Training process used 533MiB of DRAM with the CPU loading 8 cores for the training process.
  Training took 20m 42s or 12.42s per epoch using this setup.
- AMD R7 3700U CPU: Training process used 510MiB of DRAM with the CPU loading 4 cores for the training process.
  Training took 8h 53m or 5m 32s per epoch using this setup.
- Inference does not require a GPU. For example an AMD R9 7950X3D CPU can handle upto 560 NVPS. The MS-SQL server model with 62 items equals to 0.51 NVPS.

AI-based anomaly detection

Summary



Summary

AI-based anomaly detection systems can:

- ... learn simple and complex threshold values
- … learn temporal dependencies (Certain values are expected at certain times)
- … learn value dependencies (Certain values must maintain a mathematical relationship between each other)
- … learn dependencies across multiple systems and services
- … improve monitoring setups by looking into anomalies across many items at once
- … express the magnitude of an anomaly using the value of reconstruction error instead of just true/false

ZABBIX '25
CONFERENCE
GERMANY

Thank you

Wolfgang Alper

CEO
IntelliTrend GmbH