Context
Cloud environments perform continuous logging of various metrics of and around the applications deployed on them. This data is generally stored as a so-called time-series data. Hereby, a vast amount of information describing hardware and software behavior is accumulated over time. To provide better understanding, mitigate errors and improve capacity provisioning and orchestration of cloud applications, an analysis of such time series is required.

This does not only include the ability to correctly predict future behavior in timely manner, but also recognize behavioral anomalies in the data, first retrospectively, and later as they occur.

Scope
In the scope of this work, techniques for analysis and anomaly detection in time-series data shall be explored. A concept shall be presented for correctly performing validation tests and evaluation of anomaly detection in time series data. Possible approaches for detecting anomalies in time series shall be implemented (the scope is not set on any specific approaches).

Requirements and Comments
If this thesis achieves the desirable outcome, its results are to be integrated in the RECAP research project (https://recap-project.eu/) which is released under an OpenSource license. For that reason, we appreciate if you are ready to OpenSource your results.

If you are interested in this or similar theses, please contact Mark Leznik either by mail or directly in his office.

mail: mark.leznik@uni-ulm.de
office: Uni West, 43.2.213
Cloud environments perform continuous logging of various metrics of and around the applications deployed on them. This data is generally stored as a so-called time-series data. Hereby, a vast amount of information describing hardware and software behavior is accumulated over time. To provide better understanding, mitigate errors and improve capacity provisioning and orchestration of cloud applications, an analysis of such time series is required.

However, the large amount of sensors and metrics logged leads to a very high dimensionality of the data at hand. This issue can be tackled using any given dimensionality reduction algorithms. Additionally, a mapping of the data into latent space is possible. This is not only beneficial for later analysis (machine learning, clustering) of the data, but also provides possibility for an easier visualization of the data.

Scope
In the scope of this work, techniques for dimensionality reduction and mapping into latent space of time-series data shall be explored. A concept shall be presented for correctly performing and evaluating the efficiency of both procedures.

Requirements and Comments
If this thesis achieves the desirable outcome, its results are to be integrated in the RECAP research project (https://recap-project.eu/) which is released under an OpenSource license. For that reason, we appreciate if you are ready to OpenSource your results.

If you are interested in this or similar theses, please contact Mark Leznik either by mail or directly in his office.

mail: mark.leznik@uni-ulm.de
office: Uni West, 43.2.213

Faculty of Engineering and Computer Science
Institute of Information Resource Management
Context
Time-series data analysis including machine learning and/or other statistical analysis of any given time-series data at hand requires upfront knowledge concerning the data. Questions such as „predictability“ can be answered in advance without performing any machine learning tasks on the data. Hence, a set of time-series quality measurement metrics would offer information on a data set in advance. This additionally could conceptually be coupled with classification and anomaly detection techniques to further improve the data analysis process.

Scope
In the scope of this work, a toolset of quality measure metrics/criteria for time-series data shall be defined and implemented in accordance with the current state-of-the-art. Further, either real-time analysis or classification and anomaly detection using the data may be explored.

Requirements and Comments
If this thesis achieves the desirable outcome, its results are to be integrated in the RECAP research project (https://recap-project.eu/) which is released under an OpenSource license. For that reason, we appreciate if you are ready to OpenSource your results.
Context
An autoencoder is a type of artificial neural network used to learn efficient data coding in an unsupervised manner. The aim of an autoencoder is to learn a representation (encoding) for a set of data by training the network to ignore signal “noise”. Along with the reduction side, a reconstructing side is learnt, where the autoencoder tries to generate from the reduced encoding a representation as close as possible to its original input, hence its name.

For several purposes (such as simulation, evaluation and anonymization) it is necessary to derive synthetical data from a given real input.

Scope
In the scope of this work, an autoencoder able to realistically mimic real input data logged from cloud monitoring solutions shall be developed. A concept to evaluate the correctness and plausibility of the generated data shall be developed. Additionally, a comparison do state of the art solution shall be performed.

Requirements and Comments
If this thesis achieves the desirable outcome, its results are to be integrated in the RECAP research project (https://recap-project.eu/) which is released under an OpenSource license. For that reason, we appreciate if you are ready to OpenSource your results.

If you are interested in this or similar theses, please contact Mark Leznik either by mail or directly in his office.

mail: mark.leznik@uni-ulm.de
office: Uni West, 43.2.213

Faculty of Engineering and Computer Science
Institute of Information Resource Management
Context
Synthetical data generation using several methods (classic statistics, GANs) are necessary in the context of data analysis for several purposes such as simulation, evaluation and anonymization. Hereby, not only a synthetization in terms of similar data reproduction with an obfuscation effect is of interest, but also data generation in a “what-if” scenario. In other words, synthetic data generation based on input data and additional parameters (i.e. how would the data look with more users?). Such data can then be used not only in evaluation scenarios where the real data might be too sensitive, but also find its way into prediction applications, where the validation can be performed using such data.

Scope
In the scope of this work, an approach for generating synthetic data shall be developed, with the possibility of adding a user defined bias in the process of the generation. A fitting validation approach shall be implemented for the evaluation of the diffused data.

Requirements and Comments
If this thesis achieves the desirable outcome, its results are to be integrated in the RECAP research project (https://recap-project.eu/) which is released under an OpenSource license. For that reason, we appreciate if you are ready to OpenSource your results.

If you are interested in this or similar theses, please contact Mark Leznik either by mail or directly in his office.

mail: mark.leznik@uni-ulm.de
office: Uni West, 43.2.213

Faculty of Engineering and Computer Science
Institute of Information Resource Management
Context
Time-series data analysis, specifically anomaly detection (collective anomalies) are still a challenge to recognize for machine learning and classical statistical approaches. The use of CNN and similar deep learning based classification methods has proven to be successful in image based classification. In this work, the possibility of using classification techniques to recognize anomalies based on a labeled dataset shall be explored.

Scope
In the scope of this work, based on the creation of so-called density maps (optionally any other more suitable visualization), a classification approach for recognizing time series anomalies shall be developed. The validation of the work is to be performed using an open source labeled time series dataset.

Requirements and Comments
If this thesis achieves the desirable outcome, its results are to be integrated in the RECAP research project (https://recap-project.eu/) which is released under an OpenSource license. For that reason, we appreciate if you are ready to OpenSource your results.
Context
Auto machine learning approaches for data analysis are a promising up and coming research field for several proposes ranging from image analysis to speech recognition. Auto machine learning algorithms based on the minimum nescience principle have also shown to be successful for time series prediction and classification. The algorithms include Decision Trees, Multilayer Perceptrons, and Grammatical Evolution.

Scope
In the scope of this work, the possibility of using an AutoML framework for time series analysis shall be explored. A comparison to more established approaches during the evaluation is order hereby.

Requirements and Comments
If this thesis achieves the desirable outcome, its results are to be integrated in the RECAP research project (https://recap-project.eu/) which is released under an OpenSource license. For that reason, we appreciate if you are ready to OpenSource your results.

If you are interested in this or similar theses, please contact Mark Leznik either by mail or directly in his office.

mail: mark.leznik@uni-ulm.de
office: Uni West, 43.2.213

Faculty of Engineering and Computer Science
Institute of Information Resource Management