

# **Failure Prediction in Cluster Systems**

**Muhammed Costan**

**September 2016**

**Dissertation submitted in partial fulfilment for the degree of  
Master of Science in Big Data**

**Computing Science and Mathematics  
University of Stirling**

## **Abstract**

Minimising the server downtimes in cluster systems by solving the problems as soon as possible is extremely important for companies. One of the most promising and cost efficient solutions is forecasting the failures before they happen to reduce their damages to the system. In this study, we aimed to create a model which is capable of predicting a specific type of failure in a big cluster of servers by using the predictive log analytics. To achieve this, several technologies have been used in harmony to benefit from different aspects of big data such as text clustering, machine learning, and parallel processing. Since the failure prediction problem can be represented as a sequential classification problem, Support Vector Machines (SVMs) has been chosen as a machine learning algorithm to train a prediction model. Different models were trained with SVMs by changing the values of the algorithm's parameters to find a model with an ideal accuracy rate.

**Keywords:** Log analytics, failure prediction, SVMs, machine learning, big data, classification