Chapter 1 – Introduction

The growth of information technology (IT) micro, small, and medium-sized enterprises (MSMEs) and IT startups in Indonesia has greatly contributed to the country's economic development. However, these emerging companies often face challenges in optimizing their system's database, leading to difficult data access and huge computing resource usage. In this paper, we aim to address this problem by using probabilistic data structure filter to optimize query and comparing the efficiency of various filter types that could be used, including the Bloom filter and its combinations (classic, partitioned, counting), cuckoo filter, and xor filter. Our hypothesis is that using filter will provide significant decrease database query time for certain scenario, with Bloom Filter providing the best efficiency.

Previous studies have highlighted the importance of understanding the factors that influence e-commerce adoption by small and medium-sized enterprises (SMEs) in developing countries. In these studies, the adoption of e-commerce by SMEs in developing countries is still lagging that of large companies, yet it is crucial for the survival and success of these businesses in the current digital era [1]. This highlights the need for further research in this area, and our study aims to contribute to this by comparing the efficiency of various probabilistic data structure filter types to provide insights for IT MSMEs and IT startups in Indonesia.

In this study, we will be using a random generated transaction dataset with 64-bit index data strings, which will be used as the key that will be inserted into various filters. We will be using several techniques to measure the efficiency of the filters within three parameters: computing time, cost, and resources. By analyzing the data using these parameters, we will be able to determine which filter is the most efficient for IT MSMEs and IT startups in Indonesia.

The potential benefits of this research are not only solving the problems faced by emerging IT MSMEs and IT startups in Indonesia but also contributing to the country's economic factors. This research is also relevant to other papers on SMEs in the context of Indonesia, as it highlights the importance of understanding the factors that influence e-commerce adoption by small and medium-sized enterprises in developing countries, such as Indonesia. This research aims to contribute to the existing literature on IT industry adoption by SMEs in developing countries, and how it will benefit IT MSMEs and IT startups in Indonesia by providing them with insights on the most efficient filter to use in their systems.

The contribution of this research is as follows:
1. A comprehensive analysis of the Bloom filter and its combinations (classic, partitioned, counting) in comparison to cuckoo filter and xor filter in terms of efficiency.
2. A thorough examination of the trade-offs between the efficiency of the filters in terms of computing time, cost, and resources.
3. A novel approach to analyze the efficiency of various filter types through a random generated transaction dataset 64 bit index data (before hash), type strings, and 64 bit key (hashed).
4. An in-depth analysis of the efficiency of different filter types in the context of IT MSMEs and IT startups in Indonesia.

The § II provides a comprehensive review of the existing work on the topic of every data filters that's widely used while § III describes the research methods that were used in the study. It includes detailed information regarding the dataset used, technical specification of the software and hardware where the filters being implemented, and the specific parameters that were used to measure the efficiency of the filters. Then, the findings of the study, including a comparison of the

efficiency of the various filter types are reviewed in § IV. Finally, § V summarizes the main findings of the study and future recommendations of this research.