

## **MSc Project Report**

# **Serverless Computing Architecture for Fingerprint-based Audio Duplicate Identification**

Ishan Randeniya

2018

**Report submitted as part of the requirements for the degree of  
MSc Big Data Analytics at Robert Gordon University, Aberdeen, Scotland**

## Abstract

This research explores a distributed algorithm for similarity search in large audio database which has high scalability. Even though the database had been scaled, processing time will be an average time for all levels of scaling.

Depending on the background and the previous work, and the rational is to find the best matching audio and the comparison undertake to find the best matching sequential frequencies it is the most suited method that is using audio-fingerprints.

Instead of comparing the complete clips, the way that uses fingerprints and hashes more efficient due to the comparison is only for the particular hashes. Although it doesn't go through each files or hash.

The research describes a serverless architecture to identify audios based on the fingerprint-based approach. For the fingerprinting it has identified Min-Hash is the suitable hashing algorithm.

## Keywords

Audio Signal Processing, Audio Fingerprinting, Pattern Recognition, High-dimensional data, Distributed Computing, Serverless Architecture