

# Leveraging Cloud Services for Large Scale Fuzzy Hashing

**DS7330** 

**Presenters:** 

Adam Alidra Leonardo Leal Filho Josh Mitchell

ssDeep Term Project Repo

### Introduction

What is fuzzy hashing?

A type of compression function for calculating the similarity between digital files

Use cases

loC's and malware analysis

#### How it works

Gather hashes and compare them to database(s) to see if the hash matches a known-bad sample

#### About ssDeep

It works by computing a fuzzy hash of each piece of data supplied to it and compares the hashes generating a similarity score ranging from 0 to 100

#### **Process**

- 1) Create hashes that correspond to files
- 2) Optimize chunks
- 3) Create a programmatic database on local machine (or using on-premise server) and schema
- 4) Query utilizing a matching mode technique
- 5) Spin up cloud resources (Azure Database MySQL) and replicate technique

You manage

Managed by vendor

## As-a-Service



Microsoft Dynamics 365

Office 365





**Traditional** on-premises

**Applications** 

Runtime

Middleware

Storage

Infrastructure (as a service)

**Applications** 

Data

**Runtime** 

Middleware

O/S

Virtualization

Servers

Storage

**Networking** 

**Platform** (as a service)

**Applications** 

Data

**Runtime** 

Middleware

O/S

Virtualization

Servers

Storage

**Networking** 

Software (as a service)

**Applications** 

Data

Runtime

Middleware

O/S

Managed by vendor

Virtualization

Servers

**Storage** 

**Networking** 

Data

O/S

Virtualization

**Servers** 

**Networking** 

You manage

Managed by vendor