Eoin Falconer (13331016)

# Distributed Systems

## Ceph Summary

- Ceph maximally separates data from metadata management, allowing them to scale independently and this separation relies on CRUSH (Con- trolled Replication Under Scalable Hashing), a data distribution function that generates a pseudo-random distribution, allowing clients to calculate object locations instead of looking them up.
- I liked the failure detection in Ceph as OSDs can self report and each OSD reports to the system about it peers, allowing OSDs to step in as primaries quickly.
- Ceph also addresses that the MDS cluster is important in the performance of the file system as it accounts for over half of the work load and addresses this using object inode numbers which are distributed to all the OSDs in the system.
- My favourite aspect of Ceph is that it uses it's knowledge of metadata popularity to provide a wide distribution of hot spots without losing its root directory, the contents of frequently read directories are distributed across many nodes and the nodes that are heavily written to have their contents hashed across the cluster evenly, so that the updates are quickly referenced across nodes.
- In addition, the OSD cluster map is completely distributed, this means that any of the OSDs can independently calculate the location of any object and each OSD also has a weighting to indicate how much data it should be assigned based on it's placement in the map.