SEMINAR 4

- What is the article about?

  The main purpose of this article is to describe how Pinterest split its data across many MySQL servers in 2012. They used this sharding, which is a database architecture. Pinterest is using this method in order to store its data.

- What requirements and design philosophies influenced the final solution?

  - ☐ Requirements :

    Scalability and stability (their system needed to be stable and scale to the moon;

    Accessibility;

    Data consistency;

  - ☐ Design philosophies:

    To not use database joins, foreign keys, or indexes to gather all data

    Support load balancing for their data

    All data needed to be replicated

    Generate universally unique IDs (UUID) for all of their objects.

- What are the relation between a MySQL instance, a database, and a shard?

  The relation between those components can be explained in the following way:

  They started with some servers running each MySQL instance. Each MySQL server is master-master replicated onto a backup host in case the primary fails. Their production servers only read/write to the master, and each instance can have multiple databases, which explains the relationship between instances of MySQL and a database. Each database is uniquely named and is a shard of their data.

- What was ZooKeeper used for?

  The usage of ZooKeeper was to send config to services that maintain the MySQL shard. The config only changes when they needed to move shards around or replace a host. If there was a case that the master dies, they decided to promote the slave and then bring up a new slave. They stored the shard configuration table in ZooKeeper.