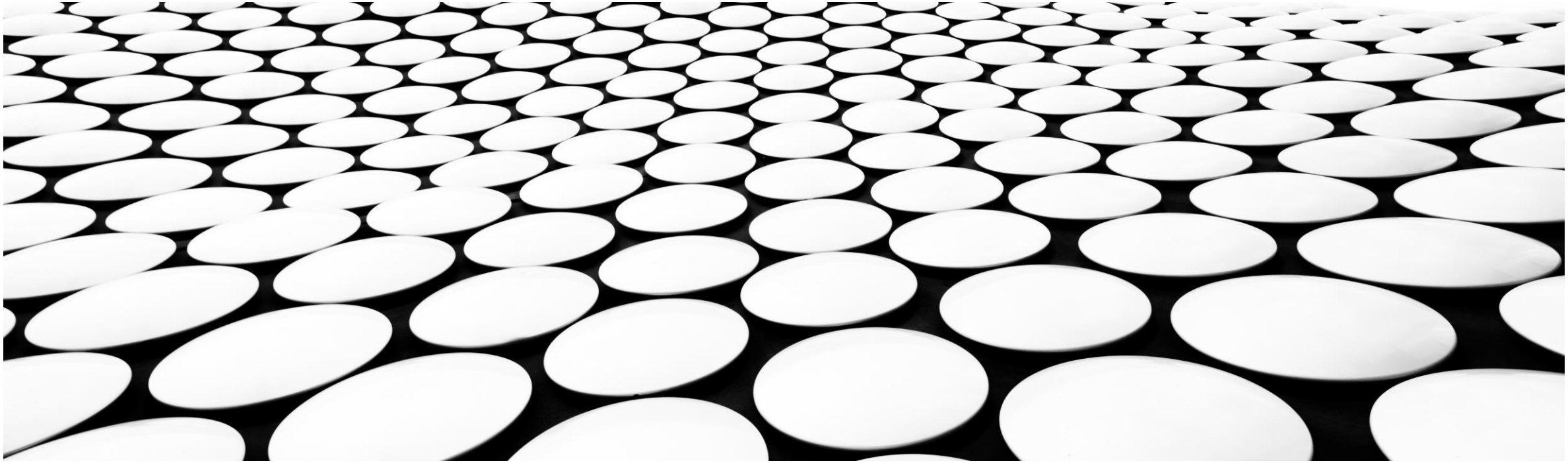
Scalability in distributed systems

COEN-317: Distributed Systems
Robert Bruce
Department of Computer Science and Engineering
Santa Clara University



Scalability: a definition

"Scalability is the ability of a system to expand to meet your business needs. You scale a system by adding extra hardware or by upgrading the existing hardware without changing much of the application" [1]

[1] https://learn.microsoft.com/en-us/biztalk/core/what-is-scalability

Facebook / Meta: Scribe

What is Scribe?

- A high-performance, buffered consumer-producer distribution model for aggregating huge volumes of service log files [1].

Scribe performance:

- "Scribe processes logs with an input rate that can exceed 2.5 terabytes [of data] per second and an output rate that can exceed 7 terabytes [of data] per second" [1].

Source code for scribe at https://github.com/facebookarchive/scribe

[1] https://engineering.fb.com/2019/10/07/data-infrastructure/scribe/

Introduction to Hadoop

What is Hadoop?

- Hadoop is a distributed framework for parallel processing of big data [1].

Hadoop has two components:

- A storage component: the Hadoop Filesystem (HDFS) [1].
- A processing component: YARN [1].

[1] https://www.edureka.co/blog/what-is-hadoop/

Hadoop file system (HDFS)

Hadoop File System (HDFS) [1]

- A distributed file structure comprised of many clusters spanning a massive number of machines.
- Each cluster is comprised of a *Namenode* and one or more *Datanodes* in a master/slave hierarchical tree structure.
- HFDS is hardware fault-tolerant through replication.
- HDFS is designed for enormous datasets.

[1] https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

Hadoop file system (HDFS)

Hadoop File System (HDFS)

Namenode:

- Contains HDFS metadata such as “permissions, modification and access times, namespace and disk space quotas” [1].
- “maintains the namespace tree and the mapping of file blocks to DataNodes” [1].

Datanode:

- Contains chunks of data typically 128 MB in size (user can also define size of each chunk) [1].
- Data chunks typically span multiple datanodes [1].

[1] K. Shvachko, H. Kuang, S. Radia, and R. Chansle, "The Hadoop Distributed File System," *Proceedings of the 26th Symposium on Mass Storage Systems and Technologies (MSST)*, 2010, pp. 1-10. doi:10.1109/MSST.2010.5496972

YARN (Yet Another Resource Negotiator)

Yarn (Yet Another Resource Negotiator) is “the brain of your Hadoop Ecosystem” [1]. It is comprised of two separate processes:

- **Resource manager**

“arbitrates resources among all the applications in the system” [2].

- **Node manager**

“responsible for containers, monitoring their resource usage (cpu, memory, disk, network) and reporting the same to the ResourceManager/Scheduler” [2]

[1] https://www.edureka.co/blog/hadoop-ecosystem
[2] https://hadoop.apache.org/docs/current/hadoop-yarn/hadoop-yarn-site/YARN.html

Apache Spark

What is Apache Spark?

- “A framework for real time data analytics in a distributed computing environment” [1].
- “executes in-memory computations to increase speed of data processing over Map-Reduce” [1].
- “100x faster than Hadoop for large scale data processing by exploiting in-memory computations and other optimizations” [1].
- Apache Spark homepage at: https://spark.apache.org/

[1] https://www.edureka.co/blog/hadoop-ecosystem

Apache Hive

What is Apache Hive?

- “a distributed, fault-tolerant data warehouse system that enables analytics at a massive scale and facilitates reading, writing, and managing petabytes of data residing in distributed storage using SQL" [1].
- Apache Hive homepage at: https://hive.apache.org/

[1] https://hive.apache.org/

Apache Hbase

What is Apache Hbase?

- "open-source, distributed, versioned, non-relational database modeled after Google's Bigtable: A Distributed Storage System for Structured Data" [1].
- Works in conjunction with Hadoop and HDFS [1].
- Apache Hbase home page at: https://hbase.apache.org/

[1] https://hbase.apache.org/

Apache ZooKeeper

What is Apache ZooKeeper?

- “A centralized service for maintaining configuration information, naming, providing distributed synchronization, and providing group services” [1].
- "ZooKeeper allows distributed processes to coordinate with each other through a shared hierarchical namespace which is organized similarly to a standard file system" [1].
- "The namespace consists of data registers - called znodes, in ZooKeeper parlance - and these are similar to files and directories" [1].
- "ZooKeeper data is kept in-memory, which means ZooKeeper can achieve high throughput and low latency numbers" [1].
- Apache ZooKeeper home page at: https://zookeeper.apache.org/

[1] https://zookeeper.apache.org/doc/current/zookeeperOver.html

Apache Oozie

What is Apache Oozie?

- “workflow scheduler system to manage Apache Hadoop jobs” [1].
- “jobs triggered by time (frequency) and data availability” [1].
- Apache Oozie homepage at: https://oozie.apache.org/

[1] https://oozie.apache.org/

Apache Flume

What is Apache Flume?

- “a distributed, reliable, and available system for efficiently collecting, aggregating and moving large amounts of log data from many different sources to a centralized data store.” [1].
- Apache Flume homepage at: https://flume.apache.org/

[1] https://flume.apache.org/