COMP5349 – Cloud Computing

Week 13: Course Review and Exam Info

Dr. Ying Zhou
School of Computer Science

COMP5349 Schedule in 2019

Week	Topic
Week 1	Cloud Computing Overview and Service Models
Week 2	Virtualization Technology
Week 3	Container Technology
Week 4	Map/Reduce Framework
Week 5	Spark Framework
Week 6	Distributed Execution: HDFS and YARN
Week 7	Distributed Execution: Spark
Week 8	Spark Data Frame
Week 9	Spark Machine Learning Library
Week 10	Cloud Storage and Databases Services
Week 11	Consistency in Cloud Storage and Database Service
Week 12	
Week 13	Course Review

The Big Picture

- Cloud Computing
  - ▶ Shared IT services for clients to rent from
  - ▶ On different levels (IaaS, PaaS, SaaS, FaaS, ….)
  - ▶ Made possible through web and data center technology
- Enabling Technologies
  - ▶ Virtualization
    - Used by all IaaS providers
  - ▶ Container
    - Can be used in various scenarios
      - IaaS customers can use container technology to deploy applications on VM
      - PaaS or SaaS
    - Active development to support Serverless computing
  - ▶ Key issues
    - Illusion of a *whole system* to every client
    - Performance isolation
    - Security and others

Analytics and BigData Services

- Basic Computational Model
  - ▶ Storage: distributed file systems (GFS, HDFS)
  - ▶ Programming Paradigm: MapReduce
  - ▶ Hadoop MapReduce as specific (open source) example
  - ▶ Map and Reduce phases
    - Each phase allows multiple tasks to run in parallel
    - Synchronization and shuffling happen between map and reduce phase
    - Map output key is used to reorganize intermediate result in reduce phase
  - ▶ An analytic workload may needs several map reduce phases
  - ▶ Simple localized fault tolerance mechanism depends on storage and I/O
- Other computation model
  - ▶ Spark
    - RDD based API
    - Data Frame based API
  - ▶ Main-memory based as compared with disk/batch based approach by MapReduce
- All based on functional programming paradigm

Cloud Storage and Database Services

■ Cloud Storage Services

▶ The cloud version of file system: GFS/HDFS, S3, EBS, etc

■ Cloud Database Services

▶ The cloud version of database: Bigtable, WAS, Dynamo, AWS Aurora

■ Common features

▶ Replication

▶ Partition

▶ Fault Tolerance

▶ Various consistency levels

▶ Various ways of handling read/write of the data

Cloud Storage/DB Services Consistency

- Many systems use customized algorithms for handling read/write

- Classic distributed system algorithm
  - Paxos
    - First phase only requests participants to make a promise, the actual value is proposed in the second phase
    - Participant is not requested to check the value proposed, instead, participant checks the proposal's sequence number
    - A leader is necessary to maintain the progress of the algorithm

- Paxos can be used in replicated environment to reach consensus
  - Run multiple Paxos, each is numbered and the value to be chosen represents an update command
  - Efficient mechanism to run infinite Paxos

Final Exam

- Online Open book two-hour exam
  - ▶ Conducted in Canvas
- Exam questions
  - ▶ Multiple choice questions and short answer questions
- Multiple choice question has single correct answer
- Short answer questions are structured into multiple parts.
- Type your answers in the text field provided in the exam paper
  - ▶ Label answer with question part
- The exam has a 100 points in total
- The exam has a 40% barrier
  - ▶ **You need to get at least 40 of 100 points in the final exam to pass this subject**

Final Exam Content

- Assessable:
  - ▶ Lecture content
    - All weeks
  - ▶ Tutorial material
    - All except week 1
  - ▶ Assignment
- **Big Data programming** may be assessed in various ways
  - ▶ Questions based on a short program
  - ▶ Design a workload by writing code/pseudo code

Thank You!