[Course Title] Cloud Computing and Big Data Systems

[Course Code] COMP4651

[No. of Credits] 3

[Any pre-/co-requisites] Object-oriented programming, Data structures

Name: Song Guo

Email: songguo@cse.ust.hk

Course Description

Big data systems, including Cloud Computing and parallel data processing frameworks, emerge as enabling technologies in managing and mining the massive amount of data across hundreds or even thousands of commodity servers in datacentres. This course exposes students to both the theory and hands-on experience of this new technology. By walking through a number of hands-on labs and assignments, students are expected to gain first-hand experience programming on real world clusters in cloud.

List of Topics:

- Basic concepts of Cloud Computing and production Cloud services
- Virtualization: virtual machine and container
- MapReduce: the de facto datacentre-scale programming abstraction and its open source implementation of Hadoop
- Big data processing: predictive analytics, descriptive analytics, graph analytics, text analytics. etc.
- Spark: a new generation parallel processing framework and its infrastructure, programming model, cluster deployment, tuning and debugging
- The state-of-the-art research topics in Cloud systems, including workload management, resource allocation and scheduling.

Assessments:

[List specific assessed tasks, exams, quizzes, their weightage]

Assessment Task	Contribution to Overall Course grade (%)
In-class Quizzes and Participation	10%
Mid-Term	20%
Homework and Project	40%
Final examination	30%

Required Texts and Materials

Since Cloud computing and big data systems are emerging technologies under heavy development, there is no official textbook. The followings books are good references to learn Hadoop and Spark programming:

- T. White, "Hadoop: The Definitive Guide Links to an external site.," 4th Eds, O'Reilly, 2015.
- 2. B. Chambers and M. Zaharia, "Spark: The Definitive Guide -- Big Data Processing Made Simple Links to an external site.," O'Reilly, 2018.

[Optional] Additional Resources

In addition to the reference books, some course materials come from seminal papers published in recent years' top conferences, which will be released as the course develops.