# Cloud Computing Systems
MIEI - 2016/2017
2nd test, 20 December 2016
duration: 2:00
*(You can answer in English or Portuguese)*

**1-** Consider the Map/Reduce programming model, like in Hadoop, for processing Web pages collected by a crawler.

**a)** Present the *pseudocode* of the mapper and reducer that allows you to find all the pages that link to a given page (by page consider its URL). The target page is in a global variable: `String target`. Assume that the map's input records have as key the page URL and as value this page HTML content. Assume also the method `value.getAllLinks()` that returns the List of all the URL links in the HTML value.

**b)** Can a Combiner accelerate this program? Describe how.

**2-** A big company is considering organizing and managing its data center as its own Private Cloud. In this data center, there are several software systems, heterogeneous regarding to OS versions, running services, software platforms and applications. Those systems include web sites, web services, databases, etc. each with varying loads along each week, month and year.

**a)** Present the advantages of this reorganization for this company. Justify.

**b)** Can you point two problems or disadvantages with the change? Justify.

**3-** Can we say that the term '*NoSQL database'* refers **just** to a database that doesn't offer any SQL interface? Justify.

**4-** S3 service (Simple Storage Service) offers scalability and reliability, allowing a client to archive large volumes of data, with high availability even in case of hardware faults or of heavy workloads. Explain how that characteristics are achieved by this service.

**5-** The words "Elastic" and "Elasticity" are commonly used regarding cloud resources and services (e.g. EC2 - Elastic Cloud Computing; EMR – Elastic Map Reduce). Explain the meaning of those words in this context and the importance of "elasticity" in cloud offerings.

**6-** The Zookeeper and Chubby services have an interface similar to the file systems and offer strong data consistency. Explain if these can be a good solution for an on-line shop (like Amazon) to save, update and share information about sales and goods stocks.

**7-** Consider an application where a server component must be available with low latency to all its clients, independent of its geographic localization (e.g. a massive multi-player game where the gamers register and pay a subscription with credit card). You are assuming some periods of heavy use by your clients.

**a)** Justify a public cloud based solution, considering its costs, workloads and availability.

**b)** Describe some eventual problems when saving personal data on a public cloud, assuming that the location of your data and services are unknown or outside your company's country. Discuss possible changes for the MMPG example above.