DA 320 Assignment 2 M. Blanco

**Part 1: Short Answers (use 3-5 lines to answer the following questions)**

**Problem 1:** Discuss the guiding principles of an IAAS storage solution.

IAAS architecture is based on various physical machines connected in a network cluster. Data from a single user can be stored in various machines. The IASS guiding principles are: cost effective by lowering infrastructure costs, redundancy, high availability and elasticity (scale up or down).

**Problem 2:** Discuss the practical use case on page 16 in the textbook.

In this practical case, a city’s decision makers had to choose the appropriate data format (CSV, XML, JSON) that would support sharing transit system data with the public. A device was installed on transit system buses to track position and speed generate large amounts of data.

**Problem 3:** Discuss the advantages and weakness of the XML data format and data conversion.

|  |  |
| --- | --- |
| **Advantages:** | **Weakness:** |
| * XML is stricter and has support for schemas and namespaces | * Bloated since each field name must be written out twice |
| * Ability to represent inheritance |  |
| * Parsing standards: DOM, SAX, STAX |  |
| * Querying standards: XQuery and XPath |  |
| * Transformation standards: XSLT |  |
| * Namespaces allow for sharing of standard structures |  |

**Problem 4:**Discuss the advantages and weakness of the JSON data format and data conversion.

|  |  |
| --- | --- |
| **Advantages:** | **Weakness:** |
| * Smaller in terms size | Lacks some key features: |
| * Additional structural information in the document | * Schema support |
| * Easier to parse and can more human readable | * Namespace support |
| * JSON compact and can be easily loaded/consumed in JavaScript |  |

**Problem 5:** Discuss a real world example of XML usage.

Many website use XML to render it’s website content. Here is a real-world example of a website using XML. In this case XML allows separation of information from presentation.



**Part 2: Research Paper (write a 1-2 page report)**

**Problem 6:** Cisco has proven that it is capable of moving data across fiber at 10 exabytes per second. But today’s limited technology prevents that in most cases. In almost every organization, the need for speed (bandwidth) has outgrown the growth in bandwidth. As more data gets processed and data variety expands, it becomes increasingly difficult to encounter efficient data transfers. Research and discuss or present information about Google’s data center project, and why it is important and ultimately need for big data. Make sure you consider ***at least*** the following questions:

* What exactly is Google’s data center project?
* What is the goal of the project?
* When was the project started?
* How many data centers are there in the US as of now? Worldwide?
* Do you think this project is a successful business model? Why?

Today, Google’s data centers have grown from a modest closet to impressive high-tech meccas. According to Urs Hölzle, Google’s eight employee starting data center “was a tiny room (7'x4', 2.5 sqm) and filled with about 30 PCs on shelves”. Google's incredible success is due to the company’s ability to build, organize, and operate a huge network of servers and fiber-optic cables efficiently and with great speed. Google business model has forced the company to build and operate its own data centers more cheaply than its competitors. The first built-from-scratch data center was in The Dalles, Oregon.

Today, Google relies upon 15 data centers to handle its business: eight locations in the United States, one in South America, four in Europe and two in Asia (Google Data Centers). In July 2016, research and advisory firm Gartner estimated that Google had 2.5 million servers and one can safely assume that this number continues to grow.

According to Internet Live Stats, Google data centers process an average of 40,000 search queries every second, resulting in 3.5 billion searches per day and 1.2 trillion searches per year (Google Search Statistics). In 1999, one year after Google was launched, the company processed 795.2 million searchers per year.

Google constructs its data centers vertically to boost its cloud capacity in response to the explosive growth in cloud computing. The shift is from a single-story design to four-story data centers. Data center design have evolved by the demand of cloud economics. Data centers are in rural locations with an abundant supply of inexpensive land and electricity which support for building bigger data center to house more servers.

After a reorganization in 2015, Google has dedicated its resources and focus to compete against AWS and Microsoft Azure platforms. This led to a major expansion of its data centers to support cloud services. At the Google Cloud Next conference, it was revealed that the company would invest $30 billion to expand data center locations (Finnegan).

# References

n.d. <https://datacentermurals.withgoogle.com/>.

Finnegan, Matthew. *Computer World UK*. 2017. <https://www.computerworlduk.com/galleries/it-vendors/google-cloud-next-announcements-3655877/>.

Ghiasi, Ali and Rich Baca. "Overview of Largest Data Centers." Ed. IEEE. n.d. <http://www.ieee802.org/3/bs/public/14\_05/ghiasi\_3bs\_01b\_0514.pdf>.

*Google Data Centers*. 16 01 2017. <https://en.wikipedia.org/wiki/Google\_Data\_Centers>.

*Google Search Statistics*. 16 01 2018. <http://www.internetlivestats.com/google-search-statistics/>.

**Part 3:** Save your file as ***DA320\_Assignment2\_XXXXX.docx (or .pdf)*** where ***XXXXX*** is the first five letters of your last name, and submit it online.