**NGWEE SIMON CHIDI**

Week 5: Daily Morning Challenge,Day 1: Tuesday 21st January 2020

**Question 1: Give an overview of what you understand as the MapReduce technology. Explain how it is similar to the Split-Apply-Combine technology created by Hadley Wickham.**

MapReduce is a processing technique and a program model for distributed computing based on java. The MapReduce algorithm contains two important tasks, namely Map and Reduce. Map takes a set of data and converts it into another set of data, where individual elements are broken down into tuples (key/value pairs). Secondly, reduce task, which takes the output from a map as an input and combines those data tuples into a smaller set of tuples. As the sequence of the name MapReduce implies, the reduce task is always performed after the map job.

The major advantage of MapReduce is that it is easy to scale data processing over multiple computing nodes. Under the MapReduce model, the data processing primitives are called mappers and reducers. Decomposing a data processing application into *mappers* and *reducers* is sometimes nontrivial. But, once we write an application in the MapReduce form, scaling the application to run over hundreds, thousands, or even tens of thousands of machines in a cluster is merely a configuration change. This simple scalability is what has attracted many programmers to use the MapReduce model.

The whole process goes through four phases of execution namely, splitting, mapping, shuffling, and reducing.

**Input Splits:** An input to a MapReduce job is divided into fixed-size pieces called **input splits**Input split is a chunk of the input that is consumed by a single map

**Mapping**: This is the very first phase in the execution of map-reduce program. In this phase data in each split is passed to a mapping function to produce output values. In our example, a job of mapping phase is to count a number of occurrences of each word from input splits (more details about input-split is given below) and prepare a list in the form of <word, frequency>

**Shuffling**: This phase consumes the output of Mapping phase. Its task is to consolidate the relevant records from Mapping phase output. In our example, the same words are clubed together along with their respective frequency.

**Reducing**: In this phase, output values from the Shuffling phase are aggregated. This phase combines values from Shuffling phase and returns a single output value. In short, this phase summarizes the complete dataset.

The split-apply-combine strategy is similar to the map-reduce strategy for processing large data, recently popularized by Google. In map-reduce, the map step corresponds to split and apply, and reduce corresponds to combine, although the types of reductions are much richer than those performed for data analysis. Map-reduce is designed for a highly parallel environment, where work is done by hundreds or thousands of independent computers, and for a wider range of data processing needs than just data analysis.

**Question 2: Briefly explain three effective method for field research**

Field research is defined as a [qualitative method](https://www.questionpro.com/blog/qualitative-research-methods/) of [data collection](https://www.questionpro.com/blog/qualitative-data/) that aims to observe, interact and understand people while they are in a natural environment. Field research, field studies, or fieldwork is the [collection](https://en.wikipedia.org/wiki/Empirical_research) of [raw data](https://en.wikipedia.org/wiki/Raw_data) outside a [laboratory](https://en.wikipedia.org/wiki/Laboratory), [library](https://en.wikipedia.org/wiki/Library), or [workplace](https://en.wikipedia.org/wiki/Workplace) setting. The approaches and methods used in field research vary across [disciplines](https://en.wikipedia.org/wiki/Branches_of_science)

* Case Study

A case study research is an in-depth analysis of a person, situation or event. This method may look difficult to operate, however, it is one of the simplest ways of conducting research as it involves a deep dive and thorough understanding the data collection methods and inferring the data.

* Participant Observation

In this method of field research, the researcher is deeply involved in the research process, not just purely as an observer, but also as a participant. This method too is conducted in a natural environment but the only difference is the researcher gets involved in the discussions and can mould the direction of the discussions. In this method, researchers live in a comfortable environment with the participants of the research, to make them comfortable and open up to in-depth discussions.

* Direct Observation

In this method, the data is collected via an [observational method](https://www.questionpro.com/blog/qualitative-observation/) or subjects in a natural environment. In this method, the behavior or outcome of situation is not interfered in any way by the researcher. The advantage of direct observation is that it offers contextual data on people, situations, interactions and the surroundings. This method of field research is widely used in a public setting or environment but not in a private environment as it raises an ethical dilemma.