**Student Name:**

***Note:* \*\*\*\*** Be precise. Short correct answer is sufficient. \*\*\*\*

**Presentation 1 [CoHadoop & E3]**

[5 Points] Q1: Describe one disadvantage of CoHadoop that may occur if the handling of the colocated data is not performed carefully. *[2-3 sentences]*

[3 Points] Q2: Describe the difference between partitioning and colocation *[2-3 sentences]*? Are these two concepts independent of each other or one relay on the other *[2-3 sentences]*?

[2 Points] Q3: State whether each of the following is True or False:

* A Locator is assigned to a file automatically by CoHadoop at the loading time
* In CoHadoop, files having the same locator are not guaranteed to be stored in the same set of nodes

[5 Points] Q4: In E3 system, assume we have a file *F* containing 5 partitions, and we are searching for attribute *Name* = ‘ABC’, which exists 5 times in the file. Draw the best scenario (the distribution of ‘ABC’ across the partitions) in which an Inverted Index would perform the best.

[5 Points] Q5: The same as in Q4, but now draw the worst scenario in which the Inverted Index will be of no use. In this case, what other method E3 deploys (mention the Name of the method, and the content of the constructed file)

**Presentation 2 [HaLoop]**

[5 Points] Q6: What are the types of caches that HaLoop proposes? For each cache type, give a brief description on how/when it can save computations *[2-3 sentences each].*

[5 Points] Q7: For the following analytical techniques, *K-means clustering*, *Page Rank*, and *Naïve Bayes classifier*, mention which type (or types) of caches offered by Haloop can be used to speedup each technique? *[Just mention the cache types applicable to each technique without details]*

[5 Points] Q8: For each of the cache types supported by HaLoop, describe how Cache Reloading (i.e., Cache Recovery) is done in the case where the node having the cached data fails. *[2-3 sentences for each cache type].*