Big Data Lab Answer Sheet.

Please complete this answer sheet and turn it in at the beginning of class on the due date posted in LEARN.

|  |  |
| --- | --- |
| **Part I:** | |
| **Q** | **A** |
| **1** | Metadada (Variables, Columns) -> Data (Variables used, rows) |
| **2** | Cells as ‘Rows’ with cell position defined in data slots |
| **3** | Metadata (Categories,State) -> Data -> Metadata -> Data -> … |
| **4** | CSV (Notepad++) – Human Readable |
| **Part II:** | |
| **Q** | **A** |
| **5** | Count word reuse on one input line |
| **6** | Sentence (newline delimited) |
| **7** | Distributed Computing; larger chunks = slower runs |
| **8** | Characters instead of Words (space delimited -> character delimited) |
| **9** | For each(word in sentence){  Read word from input array  Add to total for word X into current shard  } |
| **10** | Shards are input; Fetch total of words from each shard and print result |
| **11** | Reducer splits each line with a tab character and adds 1 to count of word |
| **12** | Alice 81.6666666667  Bob 68.0  Carol 67.0  Dave 78.0  Eve 63.6666666667 |
| **13** |  |