DA 320 Assignment 5 M. Blanco

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

**Problem 1:** Discuss why organizations are finding data warehouse can solve some of their data challenges without the need for complicated ETL, star schemas, and other conventions of enterprise data warehousing.

Thanks to Hadoop, organizations can scale horizontally as additional data is inserted into the system. MapReduce enables data processing close to the distributed data storage allowing for complex batch processing of data across the network.

**Problem 2:** Discuss why RDBMS’ are useful tools for asking questions about structured datasets.

Relational databases support Structured Query Language. SQL supports common database functions such as selecting the results of mathematical operations, joining query results from different relatable tables, and grouping results together by a value. Eventhoug SQL operation can be expensive, SQL is well understood by many people.

**Problem 3:** Compare HiveQL and SQL and discuss similarities and differences between them.

SQL is based on a relational database model whereas HiveQL is a combination of object-oriented programming with relational database concepts. SQL manipulates data stored in tables and changes table’s rows and columns. HiveQL is concerned about objects and its properties. SQL is concerned about the relationship that exists between tables while HiveQL is concerned about the relation between two objects. HiveQL does not support every possible type of SQL query. Transactions and materialized views are not supported. Hive allows for multiple insertion whereas SQL only allows for one. HiveQL does not support JOIN queries between two tables.

**Problem 4:**Explain why the MapReduce model is not well-designed to compare keys that do not match.

It is important to use the appropriate keys. The input and output keys must match up otherwise the MapReduce code will not work. The reduce-side join is a straightforward approach that takes advantage of that identical keys are sent to the same reducer. In the reduce-side join, the output key of Mapper must be join key so that they reach the same reducer.

MapReduce by examples 
Simplied flow (for developers) 
 



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Mapper Output | | | | |
| Mapper #1 | Mapper #2 | Mapper #3 | Mapper #4 | Mapper #5 |
| <It,1>  <is,1>  <been,1>  <a,1>  <long,1>  <day,1>  <without,1>  <you,1>  <Oh,1>  <That,1>  <I,1>  <will,1>  <be,1>  <standing,1>  <right,1>  <here,1> | <my,1>  <friend,1>  <I,1>  <will,1>  <tell,1>  <you,1>  <all,1>  <about,1>  <it,1>  <Talking,1>  <to,1>  <you,1>  <about,1>  <another,1>  <path,1> | <And,1>  <I,1>  <will,1>  <tell,1>  <you,1>  <all,1>  <about,1>  <it,1>  <When,1>  <I,1>  <see,1>  <you,1>  <again,1>  <I,1>  <know,1>  <we,1>  <loved,1>  <to,1>  <hit,1>  <the,1>  <road,1>  <and,1>  <laugh,1> | <when,1>  <I,1>  <see,1>  <you,1>  <again,1>  <All,1>  <the,1>  <planes,1>  <we,1>  <flew,1>  <But,1>  <something,1>  <told,1>  <me,1> | <We,1>  <have,1>  <come,1>  <a,1>  <long,1>  <way,1>  <from,1>  <where,1>  <we,1>  <began,1>  <Good,1>  <things,1>  <we,1>  <have,1>  <been,1>  <through,1>  <that,1>  <it,1>  <would,1>  <not,1>  <last,1> |

|  |  |
| --- | --- |
| Reducer Input | |
| Reducer #1 | Reducer #2 |
| <a,1>  <a,1>  <about,1>  <about,1>  <about,1>  <again,1>  <again,1>  <all,1>  <all,1>  <All,1>  <And,1>  <and,1>  <another,1>  <be,1>  <been,1>  <been,1>  <began,1>  <But,1>  <come,1>  <day,1>  <flew,1>  <friend,1>  <from,1>  <Good,1>  <have,1>  <have,1>  <here,1>  <hit,1>  <I,1>  <I,1>  <I,1>  <I,1>  <I,1>  <I,1>  <is,1>  <It,1>  <it,1>  <it,1>  <it,1>  <know,1>  <last,1>  <laugh,1>  <long,1>  <long,1>  <loved,1>  <me,1>  <my,1> | |  | | --- | | <not,1> | | <Oh,1> | | <path,1> | | <planes,1> | | <right,1> | | <road,1> | | <see,1> | | <see,1> | | <something,1> | | <standing,1> | | <Talking,1> | | <tell,1> | | <tell,1> | | <That,1> | | <that,1> | | <the,1> | | <the,1> | | <things,1> | | <through,1> | | <to,1> | | <to,1> | | <told,1> | | <way,1> | | <We,1> | | <we,1> | | <we,1> | | <we,1> | | <we,1> | | <When,1> | | <when,1> | | <where,1> | | <will,1> | | <will,1> | | <will,1> | | <without,1> | | <would,1> | | <you,1> | | <you,1> | | <you,1> | | <you,1> | | <you,1> | | <you,1> | |
| Reducer Output | |
| Reducer #1 | Reducer #2 |
| <a,2>  <about,3>  <again,2>  <all,2>  <All,1>  <And,1>  <and,1>  <another,1>  <be,1>  <been,2>  <began,1>  <But,1>  <come,1>  <day,1>  <flew,1>  <friend,1>  <from,1>  <Good,1>  <have,2>  <here,1>  <hit,1>  <I,6>  <is,1>  <It,1>  <it,3>  <know,1>  <last,1>  <laugh,1>  <long,2>  <loved,1>  <me,1>  <my,1> | <not,1>  <Oh,1>  <path,1>  <planes,1>  <right,1>  <road,1>  <see,2>  <something,1>  <standing,1>  <Talking,1>  <tell,2>  <That,1>  <that,1>  <the,2>  <things,1>  <through,1>  <to,2>  <told,1>  <way,1>  <We,1>  <we,4>  <When,1>  <when,1>  <where,1>  <will,3>  <without,1>  <would,1>  <you,6> |