1)

**Complex data types in Pig:**

1. **Tuple:** An ordered set of fields. Tuple is represented by braces.

Example: (1,2,3)

1. **Bag:** A set of tuples is called a bag. Bag is represented by curly braces.

Example: {(1,2),(3,4)}

1. **Map:** It is a set of key value pairs. Map is represented in square brackets.

Example: [key#value] where # is used to seperate key and value.

2)

**To interact with the shell in pig:**

The shell in apache pig is called grunt shell to interact with this grunt shell we need to use command “pig” or “pig –x local”.

3)

**Difference between Pig and map Reduce:**

**PIG** is a data flow language, the key focus of Pig is manage the flow of data from input source to output store. As part of managing this data flow it moves data feeding it to p1, taking output and feeding it to p2. The core features are preventing execution of subsequent stages if previous stage fails, manages temporary storage of data and most importantly compresses and rearranges processing steps for faster processing. While this can be done for any kind of processing tasks Pig is written specifically for managing data flow of Map reduce type of jobs. Most if not all jobs in a Pig are map reduce jobs or data movement jobs. Pig allows for custom functions to be added which can be used for processing in Pig, some default ones are like ordering, grouping, distinct, count etc.

**Map reduce** on the other hand is a data processing paradigm, it is a framework for application developers to write code in so that its easily scaled to PB of tasks, this creates a separation between the developer that writes the application vs the developer that scales the application. Not all applications can be migrated to Map reduce but good few can be including complex ones like k-means to simple ones like counting uniques in a dataset.

4)

**Difference between Pig and SQl:**

Usability is one difference between Pig and Sql.

Lets consider the scenario of ad-hoc queries. Data analyst/scientist want to analyze a particular trend on a particular day. He will also need to analyze a different trend another day as well. If he is going to use hive for this he will suffer for sure. He will be writing map-reduce jobs every time and modifying his queries. This is where Pig comes to rescue. He just needs to learn Pig Latin then deal directly with unstructured/semi-structured data.

5)

**Scalar Data Types:**

1. int: Signed 32-bit integer. Example: 10
2. long: Signed 64-bit integer. Example: 10l
3. float: 32-bit floating point. Example: 10.5F
4. double: 64-bit floating point. Example: 10.5e2
5. Arrays:
6. chararray: Character array (string) in Unicode UTF-8 format. Example: hello world.
7. bytearray: Byte array (blob)