1. apm系统：

a. druid.io：时序数据库；

b. hbase（phoenix查询）：非关系型分布式数据库；

2. 用户行为分析系统：

a. hive：基于hadoop的数据仓库工具；

b. impala：基于hive的大数据实时分析引擎。

3. HBase如何实现模糊查询?

```
try {
    HTable table = new HTable(conf, tablename);
    Scan s = new Scan();
    //查询rowkey包括xx的行
    Filter filter = new RowFilter(CompareFilter.CompareOp.EQUAL, new SubstringComparator("
    s.setFilter(filter);
    ResultScanner rs = table.getScanner(s);
    for (Result r : rs) {
       KeyValue[] kv = r.raw();
        for (int i = 0; i < kv.length; i++) {
            System.out.print(new String(kv[i].getRow()) + "  ");
            System.out.print(new String(kv[i].getFamily()) + ":");
            System.out.print(new String(kv[i].getQualifier()) + "  ");
            System.out.print(kv[i].getTimestamp() + "  ");
            System.out.println(new String(kv[i].getValue()));
        }
    }
} catch (IOException e) {

}
```

4. map/reduce过程，如何用map/reduce实现两个数据源的联合统计

a. 简单概括的说，MapReduce是将一个大作业拆分为多个小作业的框架

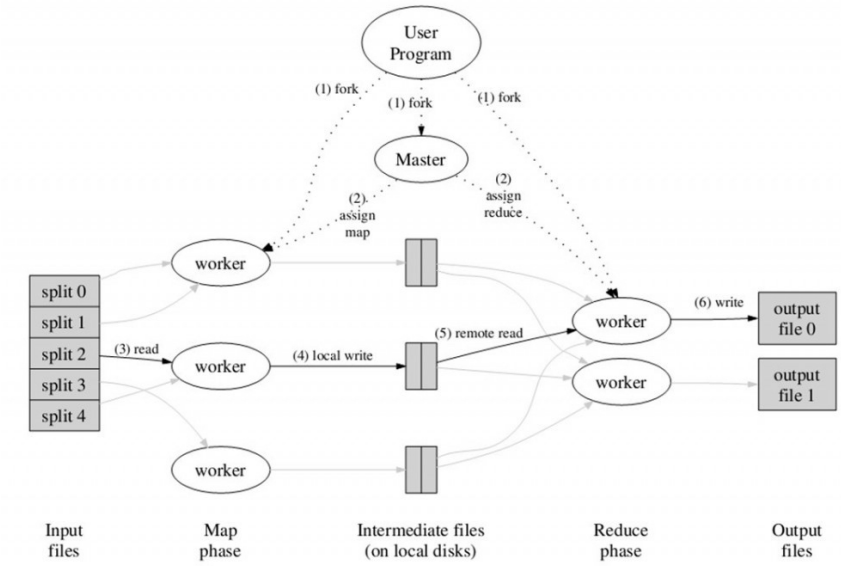

Figure 1: Execution overview

i. 第一阶段是准备阶段，包括1、2，主角是MapReduce库，完成拆分作业和拷贝用户程序等任务；
ii. 第二阶段是运行阶段，包括3、4、5、6，主角是用户定义的map和reduce函数，每个小作业都独立运行着；
iii. 第三阶段是扫尾阶段，这时作业已经完成，作业结果被放在输出文件里，就看用户想怎么处理这些输出了。