1：技能

模板1：  
顺序：

1. 熟练使用Linux常用的操作命令，熟悉shell脚本编程；
2. 熟悉MapReduce的原理和流程，并能用其API开发应用程序；
3. 熟悉Spark的原理和流程，并能用Java或Scala语言进行Spark应用程序开发，熟悉Spark Streaming组件，并能用其开发实时计算系统，了解Spark源码；
4. 熟悉Hive和HBase的系统架构，能使用Hive进行海量数据的统计分析，能根据需求实际HBase表，能对Hive、HBase进行调优；
5. 理解机器学习的思想，熟悉python编程，熟悉mahout，熟练使用SparkMLlib，了解逻辑回归，朴素贝叶斯，决策树，随机森林，K-Means算法；
6. 熟悉常见的开源日志收集框架flume，Kafka消息队列，并能使用其实现不同场景的日志收集；
7. 熟练掌握JavaSE和JavaEE相关知识，具备良好的面向对象编程思想；
8. 熟练使用常见的JavaEE开源框架;

模板2：按照类型分开写：

**大数据开发方面技能：**

1、熟练使用Linux常用的操作命令，了解shell脚本编程

2、掌握hadoop体系架构各种组件，熟悉MapReduce的工作架构，以及HDFS架构，Yarn架构

3、熟悉Hadoop的调优

4、熟悉Hive和HBase架构，能用Hive，Hbase进行海量数据的统计分析和高效的查询，熟悉Hive、HBase的调优

5、熟练flume和sqoop技术。能用Flume做到高效的数据采集，用Sqoop在hadoop和关系型数据库之间传递数据

6、精通Spark体系架构，包括SparkCore，SparkSQL，SparkStreaming，SparkMLlib，读过spark的部分核心源代码

7、熟悉Spark的调优

8、熟悉Storm集群架构及其工作原理，能够使用Flume+kafka+Storm收集数据并进行清洗和实时统计分析

9、了解常用的机器学习算法(例如逻辑回归、朴素贝叶斯)，熟练运用Mahout，SparkMllib中提供的机器学习算法

10、熟悉常用的推荐算法（例如KNN、基于内容的推荐、基于模型的推荐），对推荐系统有较深的理解

11、了解Zookeeper，Redis，ElasticSearch

**Java开发方面技能：**

1. 掌握JavaSE和JavaEE相关知识，并且熟悉常用的Java设计模式(例如工厂、单例、适配器等设计模式)
2. 掌握常见的JavaEE开源框架，如Struts2、Spring MVC、Spring、Hibernate、Mybatis
3. 掌握MySQL数据库，熟练编写SQL的增删改查语句
4. 深入理解java虚拟机

模板三：

|  |  |
| --- | --- |
| Spark | 精通Spark运行机制，研读部分Spark源码，熟悉资源调度和任务调度流程;  熟练使用Spark Streaming和Spark Core开发应用程序，对SparkSQL有一定的了解和使用;  熟练使用Spark Streaming+Kafka实时流式处理框架;  擅长Spark性能调优，如内存调优、shuffle调优、并行度调优、数据倾斜等 |
| Hadoop | 熟练搭建及维护Hadoop集群;  熟悉MapReduce的运行原理和流程，能对其进行优化；  熟练使用CDH部署搭建Hadoop集群及其辅助组件。 |
| 机器学习 | 了解预测（线性回归，梯度下降法）、分类（逻辑回归、朴素贝叶斯、聚类、决策树、协同过滤、神经网络等机器学习常用算法，能使用Mathout和SparkMLib等技术栈进行简单的处理分析。 |
| Hive | 熟悉Hive的架构原理和搭建，能够使用HQL进行海量数据分析，了解Hive的优化，熟练使用Hive对数据进行预处理和分析 |
| Storm | 熟悉Storm的系统架构及其工作原理，并能使用API开发应用程序；  熟悉Storm+Kafka实时流处理框架。 |
| 计算机语言 | 熟练JAVA编程，可以使用Scala语言进行Spark开发；  熟悉Python语言；  了解R语言，能够读懂R语言编写开发的代码。 |
| 数据库 | 熟练MySql、Oracle的使用；  熟练HBase体系架构与集群搭建，实现HBase的存储与查询并进行优化；  熟练Redis的使用，并利用Redis实现诸多业务需求。 |
| 其他组件 | 熟悉Zookeeper的架构，能应用其为集群提供协同服务；  熟悉Flume、Kafka等开源日志收集框架，能够使用Sqoop做数据迁移;  能阅读英文文献及相关文档. |