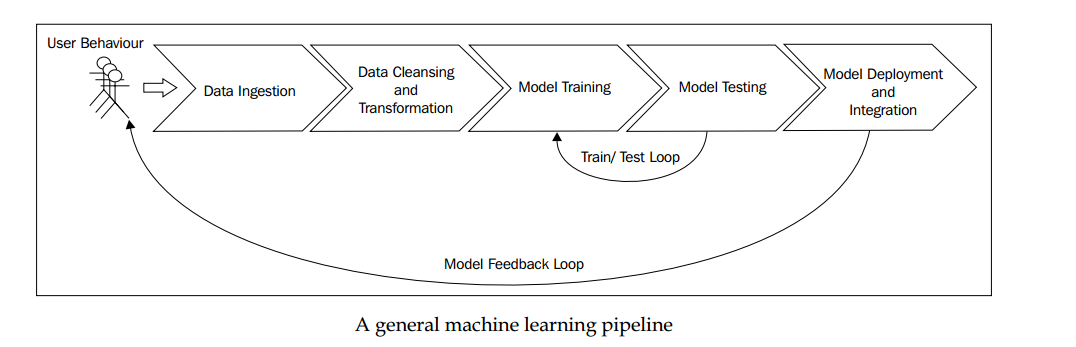
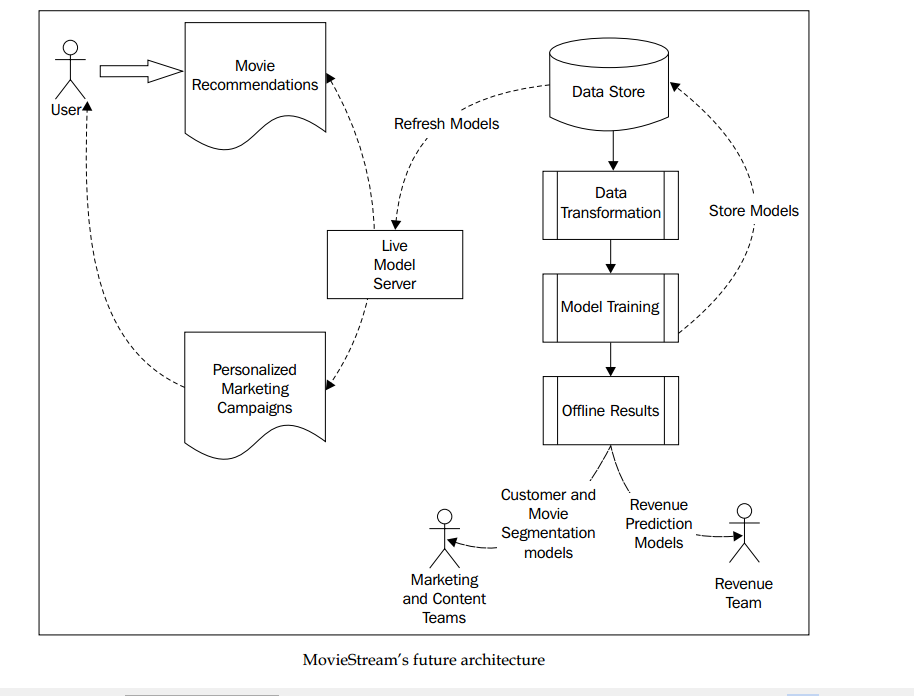
**Unit02—Spark 工作流程概述**

1. Spark 真实机器学习的工作流程：
2. 获取数据，比如从前端的日志,后台的controller 收集等获得数据
3. 数据清洗 转换，这里可能会从在有些不要的，有些不全，等等需要处理好 为下一步做准备
4. 现在可以根据需求去做些模型设计，也就是采用什么算法来做什么功能
5. 拿出部分数据进行训练，对模型进行训练并测试，调整模型
6. 测试通过后，部署上线，具体流程图如下：



1. 根据书上所说，会涉及以下几个方面
2. 静态的建模，会使用spark MLIBS里的工具
3. **会使用SPARK SQL等存储**
4. **会使用sparksteaming 在第10章。**

**总结这章主要是 说明了下 具体的机器学习生产环境的实施过程，然后最后给出一个总体的框架图：**



**Summary**  
In this chapter, you learned about the components inherent in a data-driven,  
automated machine learning system. We also outlined how a possible high-level  
architecture for such a system might look in a real-world situation.  
In the next chapter, we will discuss how to obtain publicly-available datasets for  
common machine learning tasks. We will also explore general concepts related  
to processing, cleaning, and transforming data so that they can be used to train a  
machine learning model.

|  |
| --- |
| **总结**  在本章中，您以数据驱动的了解了固有的组成部分，  自动化的机器学习系统。我们还概述了如何可能的高层次  架构这样的系统看起来在真实世界的情况。  在下一章中，我们将讨论如何获得公开可用的数据集  常见的机器学习任务。我们也将探讨相关的基本概念  到处理，清洗，和转换数据，以便它们可以被用于训练  机器学习模型。 |