**INFO 103 – Project description**

Group member names:

Group member IDs:

Assigned: 10/23/19; Due: 11/10/19

Instructions: Groups may have a maximum of **four** and a minimum of **two** people. Turn in one document for your whole group. （presentation）

1. **(5 pts)** Your project should fit into one of three categories. Please select one, below.
   1. **Data product or project development.** Study the development of an existing or hypothetical data product or project, focusing on value and marketability.

**数据产品开发：**“数据产品”在数据科学中具有特殊的含义——基于数据开发的产品的统称。数据产品开发是数据科学的主要研究使命之一，也是数据科学区别于其他科学的重要区别。与传统产品开发不同的是，数据产品开发具有以数据为中心、多样性、层次性和增值性等特征。数据产品开发能力也是数据科学家的主要竞争力之源。因此，数据科学的学习目的之一是提升自己的数据产品开发能力。（<https://www.sohu.com/a/224047326_774914>）

农业-交通-配送一体化

* 1. **Data collection and curation.** Study an existing or hypothetical data source, focusing on its challenges in curation, storage, and collection.
  2. **Data technology and processing.** Study an existing or hypothetical processing methodology or big data technology, focusing on problems and applications.

1. **(15 pts) Who’s on this project’s team?** Whether existing or hypothetical, what team member specialties did or would this specific project require? Identify three or more **roles** from the data science spectrum, and discuss how the skills of the individuals filling these roles support the needs of this data science project. **Note: it may help to complete task 3 before completing this task!**

**Use more Week 1 matrices**

**2.（15分）谁是该项目团队的成员？ 无论是现有的还是假设的，该团队成员需要做什么专业知识？ 从数据科学领域确定三个或更多角色，并讨论填补这些角色的个人技能如何满足该数据科学项目的需求。 注意：这可能有助于完成任务3，然后再完成此任务！**

**使用更多的第1周矩阵**

数据收集、数据处理、数据储存对应的科学家

1. **(25 pts) Identify your specific topic of study, and provide a summary, below.** Now that you have selected your project category, you must select and describe the specific topic on which you will focus. Using complete sentences, this should include a clear description of any **product**, **data**, **market**, **technology**, or **processing method**. In addition, this section should include a discussion of the topic’s **impact** and **use**. **This summary should be 1–2 pages in length, single spaced at font size 12.**

**3.（25分）确定您的特定学习主题，并在下面提供摘要。 现在您已经选择了项目类别，您必须选择并描述您 将关注的特定主题。 使用完整的句子，应包括任何产品，数据，市场，技术或加工方法的清晰描述。 此外， 本节还应讨论该主题的影响和使用。 此摘要的长度应为1-2页，单行间距为12号字体。**

[[1]](#endnote-1)Topic of study:

创造一个数据综合管理系统:管理农场和物流， 在保证食品是健康、安全的前提下，使农场主和农产品加工商能保证农产品（蔬菜、猪肉、牛奶等）快速地从农场送到餐桌上。

Description of:

Product:

Data: 传感器收集到的农产品的生长数据、物流运输的实时数据（交通情况、最优路线、API---交通管制系统实时检测的数据）

Market: 单个用户、餐厅、超级生鲜市场（沃尔玛、盒马···）

Technology: LoRa（远距离无线电技术） & IoT（物联网技术）& Sensor（传感器）

1

市场：单个客户、餐厅、超级市场（沃尔玛、合马）

算法估计周围的人口密度等来计算需求量

数据：根据之前的各个商品的购买量来预测当天的需求量，然后配送。

在一开始，我们的农场就要确定目标产量（根据当地城市的具体情况 人口 习惯啥的，这些信息的收集就需要大数据，调查啥的） 因为只有这样才能定量比较垂直和传统哪个方案更优秀。（这个是展开选址前的第一步）

>>>

新鲜果蔬是居民的必需品，包括水果、蔬菜等，对居民的生活有着重要的影响。传统生鲜行业采用“产地—各级批发市场—菜市场”的模式，中间环节多，运输效率低，终端价格高。

我们的这个项目设计的data product 的市场主要有：个人，餐厅，超级市场，生鲜市场。

首先需要根据收集和统计的数据进行市场调查，市场需求调查：包括市场的商品需求种类、市场商品需求量、及需求时间。即了解消费者在何时何地需要的商品种类和数量。市场环境调查：了解市场购物习惯。那样就可以控制农场的目标产量，起到成本控制的作用。

通过计算各个地区的市场需求，我们将农场选在最适宜的地方，目的是为了使物流更加便捷，降低仓储和物流的成本。

Market

Fresh fruits and vegetables are a necessity for residents and have an important impact on the lives of residents. The traditional fresh-keeping industry adopts the model of “place of origin – wholesale market at all levels – vegetable market”, with many intermediate links, low transportation efficiency, and high terminal prices.

The market for the data product designed by our project mainly includes individuals, restaurants, supermarkets, and fresh markets.

First, we need to do market surveys based on collected and statistical data. Market demand surveys include the types of commodity demand in the market, the demand for market commodities, and the demand time. That is to understand the types and quantities of goods that consumers need when and where. That way we can control the farm's target production and the cost of the farm.

Then, by calculating the market demand in each region, we select the farm in the most suitable place, in order to make logistics transport more convenient and reduce the cost of warehousing and logistics transport.

1. **(25 pts) Support your topic with reference materials and briefly summarize what each piece of reference material tells you about your topic.** These materials should relate to key components of your topic, and might include **academic papers**, describing related methods or data sets; **news articles**, indicating market demand, changing use, or ethical concerns; **blog posts**, describing products, updates, and opinions; **patents**, describing system details; **APIs**, describing data and software access; or **any documentation** detailing structure, licensing, and use. Note that all projects should have substantial reference material, **regardless of topics being existing or hypothetical.** For example, a hypothetical project might focus on references for market potential, competition, and predecessor methods, while an existing project might focus on its history, system descriptions, product updates, and ethical impacts.

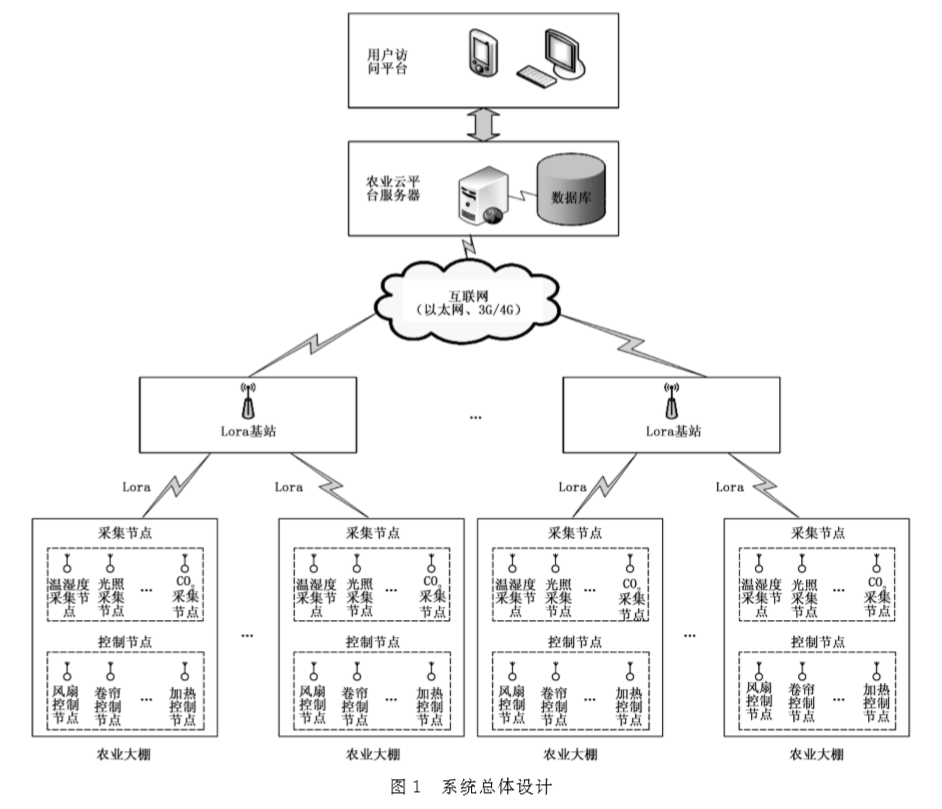
4**.（25分）用参考资料支持您的主题，并简要总结每篇参考资料告诉您的主题。 这些材料应与您主题的关键 组成部分相关，并且可能包括学术论文，描述相关方法或数据集； 指出市场需求，用途变化或道德问题的新 闻文章； 博客文章，描述产品，更新和意见； 专利，描述系统细节； API，描述数据和软件访问； 或任何详 细说明结构，许可和使用的文档。 请注意，无论存在的主题还是假设的主题，所有项目均应具有大量参考资 料。 例如，一个假设的项目可能专注于市场潜力，竞争和先前方法的参考，而现有的项目可能专注于其历 史，系统描述，产品更新和道德影响。**

1. **(30 pts) Identify how your topic intersects with the different areas of the data science life cycle.** Each stage of the data science life cycle is listed below. Using complete sentences, fill in how your topic crosses over into each area. If your topic generally does not cross over into an area, please indicate why this is the case in that stage’s area. Note that even if your topic does not specifically engage in a particular stage of the life cycle, **it may still have significant impacts to other to other data science projects.** For example, even though a **data collection and curation** may have a greater focus on **preparation**, there may be important implications in **acquisition** for others who want to use this data. Be sure to include these impacts.

**5.（30分）确定您的主题如何与数据科学生命周期的不同领域相交。 下面列出了数据科学生命周期的每个阶段。 使用完整的句子，填写主题如何跨越每个区域。 如果您的主题通常不涉及某个区域，请说明为什么在该阶段的区域是这种情况。 请注意，即使您的主题没有专门参与生命周期的特定阶段，也可能会对其他数据科学项目产生重大影响。 例如，即使数据收集和管理可能更侧重于准备工作，对于希望使用此数据的其他人来说，数据采集也可能具有重要意义。 确保包括这些影响。**

**Data acquisition:**

**数据采集：**



终端节点包括采集节点和控制节点，图一中，温湿度采集节点、光照采集节点、二氧化碳采集节点等称为采集节点。

**Data preparation:**

**数据准备：**

**Hypothesis and modeling:**

**假设和建模：**

**Evaluation & interpretation:**

**评价与解释：**

**Deployment:**

**部署：**

**Operations:**

**操作方式：**

**Optimization:**

**操作方式：**

1. [↑](#endnote-ref-1)