[1 引言 1](#_Toc492577631)

[1.1 编写目的 1](#_Toc492577632)

[1.2 项目背景 1](#_Toc492577633)

[1.3 定义 1](#_Toc492577634)

[1.4 参考资料 2](#_Toc492577635)

[2 可行性研究的前提 2](#_Toc492577636)

[2.1 要求 2](#_Toc492577637)

[2.2 目标 4](#_Toc492577638)

[2.3 条件、假定和限制 4](#_Toc492577639)

[2.4 可行性研究方法 5](#_Toc492577640)

[2.5 决定可行性的主要因素 5](#_Toc492577641)

[3 对现有系统的分析 5](#_Toc492577642)

[3.1 处理流程和数据流程 5](#_Toc492577643)

[3.2 工作负荷 5](#_Toc492577644)

[3.3 费用支出 5](#_Toc492577645)

[3.4 人员 5](#_Toc492577646)

[3.5 设备 5](#_Toc492577647)

[3.6 局限性 5](#_Toc492577648)

[4 所建议技术可行性分析 6](#_Toc492577649)

[4.1 对系统的简要描述 6](#_Toc492577650)

[4.2 处理流程和数据流程 6](#_Toc492577651)

[4.3 与现有系统比较的优越性 6](#_Toc492577652)

[4.4 采用建议系统可能带来的影响 6](#_Toc492577653)

[4.5 技术可行性评价 6](#_Toc492577654)

[5 所建议系统经济可行性分析 7](#_Toc492577655)

[5.1 支出 7](#_Toc492577656)

[5.2 5.2效益 7](#_Toc492577657)

[5.3 收益/投资比 7](#_Toc492577658)

[5.4 投资回收周期 7](#_Toc492577659)

[5.5 敏感性分析 7](#_Toc492577660)

[社会因素可行性分析 8](#_Toc492577661)

[5.6 法律因素 8](#_Toc492577662)

[5.7 用户使用可行性 8](#_Toc492577663)

[6 其他可供选择的方案 8](#_Toc492577664)

# 引言

## 编写目的

本文档意在通过系统化的方法，对意图建设的软件产品（下称产品），从技术、资金和人员的角度，进行详尽可靠的分析和阐述。本文档应该作为产品立项的可靠性依据、作为潜在开发人员（包括产品经理、设计和实现者以及质量保证人员）进行产品开发和产品确认的重要参考资料。

## 项目背景

### 建议的软件名称

产品定位为通用数据处理平台，建议的软件项目名称为：Dawn。

其中，数据预处理部分的名称代号为：PancakeofMountHuang（PMH）。

### 项目提请信息

|  |  |
| --- | --- |
| 提出者 | 邱依强（Arvinsc@foxmail.com） |
| 开发者 | Dawn-Team（https://github.com/orgs/Dawn-Team/people） |
| 所面向的用户 | 对于数据处理、分析和呈现，有较为经济的方式的需要的个人、教育机构和企业 |
| 软件实现单位 | Dawn-Team |

### 与其他软件的关系

作为一个全新的数据处理平台，产品主要与数据库系统软件有较为密切的关系；在产品开发的后期，当提供对外API时，产品作为被依赖对象与其他软件有联系。

## 定义

|  |  |
| --- | --- |
| Dawn | 产品名称与项目代号 |
| Dawn-Team | Dawn-Team，Dawn开发组成员，详细清单位于：<https://github.com/orgs/Dawn-Team/people> |
| PancakeofMountHuang（PMH） | 数据预处理子系统部分的开发代号与名称 |
| BI | Business Intelligence |

## 参考资料

【列出有关资料的作者、标题、编号、发表日期、出版单位或资料来源，可包括：

1. 项目经核准的计划任务书、合同或上级机关的批文；
2. 与项目有关的已发表的资料；
3. 文档中所引用的资料，所采用的软件标准或规范。】

# 可行性研究的前提

## 要求

### 产品功能

产品功能主要包括：数据（文档型）预处理（下称数据预处理）、数据分析、数据呈现、非文档型数据处理、开放性接口设计。以下是详细说明：

数据预处理。作为数据分析的前提，数据预处理需要完成对原始数据的数据清洗以及对于数据的在分析前所需要预先的运算。数据清洗即对文档型数据中的缺失值和异常值，通过适当的算法进行处理，以满足后续数据分析的需要的过程；数据所需要的预先的运算是指，为了满足特定的数据分析需要（如：原始数据已知功耗与对应的时间，可以预先运算得出对应的功耗），对已知数据进行一定的加工运算以满足后续数据分析需要的过程。

数据分析。数据分析是指，对已经通过了数据预处理的数据，使用诸如聚类、回归分析、关联分析等相关算法，对数据进行深入地发掘而获取其中更多的信息和知识的过程。

数据呈现。数据呈现即对数据分析的结果，根据其使用的算法，选用合适的方法（例如采用散点图、条形图、和饼图），进行有效的呈现（例如以可视化的方式呈现）。

非文档型数据处理。面向机器学习和BI的数据常常不仅限于文档型数据，而常常包括如图像、音频、视频甚至复杂的声光电传感器数据，这些数据不能通过简单的数据预处理进行有效地清洗和转换，媒体数据处理旨在解决这样的相对复杂的问题（例如处理一个人脸图片数据集中的图片尺寸、图片内容以更加符合对其进行数据分析的需要）。

开放性接口设计。在系统自身功能和稳定性相对完善的情况下，为丰富本系统的运作和扩展易用性，对已有的功能进行一定的包装，对外设计一套可以直接使用系统核心组件的过程，即为产品的开放性接口设计。

### 产品性能

除传统的关于产品在一定限制条件下对一定负载能够进行有效的响应外，产品需要对进入系统的大量数据处理、数据分析和接口使用有良好的调度处理能力，需要能够在将要出现高负载的情况下及时告知用户并保证数据的安全性和产品的稳定性。

### 输入与输出

从全局视角观察，系统的输入是用户给出的一份或者多份原始数据，输出是对于数据进行发掘而得出来的信息和知识。从各个子系统的角度观察，有如下的输入输出关系：

从数据预处理系统以及非文档型数据处理系统看，以用户提供的数据源（如文档型的Excel、CSV文件和DB连接，非文档型的图像和声音文件）作为输入，以处理后的规范化的结构化数据集（文档型数据）或特征值、非文档型数据集（非文档型数据）作为输出。

从数据分析的角度看，预处理后的结构化数据作为输入，通过一定算法得出的结果、结论作为输出。

从数据呈现看，数据分析的结果作为数据呈现的输入，多样的数据图表、文字表达作为输出。

从开放性接口看，使用接口的系统所传入的命令和参数作为输入，根据被调用的接口而返回的数据（如JSON、URL、状态码等）作为输出。

### 基本数据流程与处理流程



1. 基本数据流程和处理流程

由图1所示，整个系统的工作流程根据请求来源的不同，可以分为由系统内部请求所导向的基础性功能处理路径和由外部系统调用开放API进而相对独立使用基础性功能的处理路径。需要进一步说明的是，在系统视角（全局视角）看,基础性功能应当按照路径一步一步进行（由于数据的输入输出存在着一定的依赖关系，见[2.1.3输入与输出](#_输入与输出)），才能够作为一次完整的数据处理，而相对比之下，调用开放性API的过程可以独立使用基础性功能中的任何一个模块，在使用后可以立即结束整个数据处理流程。而从每一个独立的子系统视角看，子系统可以独立于其他系统完成其自身的功能，而不受制于其他子系统，在整个系统的实现中，可以选择性对系统内开放单个子系统的独立性功能。简而言之，要处理一整套数据处理任务，是需要各个子系统协同工作，缺一不可的，但是，用户可以灵活地根据需要单独使用其中的一个部分；与此同时，相对与系统内部，为外部调用本产品的系统提供服务的开放性API则是直接工作在单个的独立的子系统中的，子系统之间的联系是需要外部调用者负责的。

### 安全与保密要求

安全性与保密要求在产品中主要表现为两个方面，一是对用户自身使用系统（下称：用户导向的安全）的用户名、密码、邮箱、使用记录和Cookie信息的安全保障和系统数据的故障转储处理，另一方面是对于用户投入到产品中的数据（下称：数据导向的安全）的安全性保证。

用户导向的安全。用户在使用产品期间（包括非在场情况，即处于对系统并非实时操作的状态）的数据，需要设计严格的安全与保密机制，包括用户信息和使用记录的保存与读取、使用系统的安全验证、中止使用系统（停止服务与完全注销<非退出登录>用户）的功能、在客户端使用cookie和本地存储的告知服务以及数据备份的功能。

数据导向的安全。在用户使用产品时，对于用户向系统输入的数据（用户用于数据预处理或非文档型数据处理的数据），在用户输入的过程中应当提供风险告知的服务，对于这样的数据，系统应该尽量保证其不被其他未经过用户授权的外部系统访问，关心其在系统中的生存情况（是否被意外清除等）。

### 与其他软件系统之间联系产生的需要

产品在使用过程中会产生大量的数据，根据其产生对象，主要分为用户导向的数据（因为使用系统而产生<而与所分析的数据本身无关>的数据）和数据导向的数据（因为需要对数据进行分析而由用户引入的数据或与用户引入数据相关的数据）。对于用户导向的数据，需要使用关系型数据库系统进行数据的保存与修改，对于数据导向的数据，根据需要可能使用到关系型数据库系统与非关系型数据库系统。对于这两类数据库系统的选择，需要根据实际需要进行适当的选择；在数据库系统之间的联系，应当尽可能地提高效率。

### 产品完成交付的期限

产品时间表计划如下（均为完整的开发生命周期）：

|  |  |  |
| --- | --- | --- |
| 系统功能划分 | 计划交付日期 | 备注 |
| 数据预处理 | 2017-11 |  |
| 数据分析 | 2018-01 |  |
| 数据呈现 | 2018-02 |  |
| 非文档型数据处理 | 2018-04 |  |
| 开放型API | 2018-05 |  |

## 目标

### 人力资源的节约

产品将数据从原始数据到数据呈现集成，将复杂的数据预处理过程做到可视化自动化的处理，相比于当前独立使用各种软件进行处理甚至手工处理，能够减少使用者在各种软件之间学习和相互联接的时间，能够重复、批量地处理数据以尽量地节约人力资源在这样的数据处理中的投入。

### 生产效率的提高

产品需要达到能够直观给予用户在操作上的便捷、在数据处理上的批量化、自动化，减少用户对数据处理结果不满意而返工的概率，提高生产效率。

### 管理信息服务的提高

在数据处理的批量化、自动化过程中，产品需要能够有效处理高耗时任务，能够给用户明确的任务执行进度，提高用户对于自己提交的数据处理任务的管理能力。

## 条件、假定和限制

【可包括：

1. 建议开发软件运行的最短寿命；
2. 进行系统方案选择比较的期限；
3. 经费来源和使用限制；
4. 法律和政策方面的限制；
5. 硬件、软件、运行环境和开发环境的条件和限制；
6. 可利用的信息和资源；
7. 建议开发软件投入使用的最迟时间。】

## 可行性研究方法

## 决定可行性的主要因素

# 对现有系统的分析

## 处理流程和数据流程

## 工作负荷

## 费用支出

【如人力、设备、空间、支持性服务、材料等项开支。】

## 人员

【列出所需人员的专业技术类别和数量。】

## 设备

## 局限性

【说明现有系统存在的问题以及为什么需要开发新的系统。】

# 所建议技术可行性分析

## 对系统的简要描述

## 处理流程和数据流程

## 与现有系统比较的优越性

## 采用建议系统可能带来的影响

### 对设备的影响

### 对现有软件的影响

### 对用户的影响

### 对系统运行的影响

### 对开发环境的影响

### 对运行环境的影响

### 对经费支出的影响

## 技术可行性评价

【包括：

1. 在限制条件下，功能目标是否能达到；
2. 利用现有技术，功能目标能否达到；
3. 对开发人员数量的和质量的要求，并说明能否满足；
4. 在规定的期限内，开发能否完成。】

# 所建议系统经济可行性分析

## 支出

### 基建投资

### 其他一次性支出

### 经常性支出

## 效益

### 一次性收益

### 经常性收益

### 不可定量收益

## 收益/投资比

## 投资回收周期

## 敏感性分析

【敏感性分析是指一些关键性因素，如：系统生存周期长短、系统工作负荷量、处理速度要求、设备和软件配置变化对支出和效益的影响等的分析。】

# 社会因素可行性分析

## 法律因素

【如，合同责任、侵犯专利权、侵犯版权等问题的分析。】

## 用户使用可行性

【如，用户单位的行政管理、工作制度、人员素质等能否满足要求。】

# 其他可供选择的方案

【逐个阐明其他可供选择的方案，并重点说明未被推荐的理由。】

8.结论意见

【结论意见可能是：

1. 可着手组织开发；
2. 需待若干条件（如资金、人力、设备等）具备后才能开发；
3. 需对开发目标进行某些修改；
4. 不能进行或不必进行（如技术不成熟，经济上不合算等）；
5. 其他。】