

# 数梦工场 DCP1002考试大纲

## 一、 DCP 1002认证介绍

大数据数据仓库建设是数据价值化不可或缺的一环，大数据仓库建设的好坏最终影响上层大数据的应用，因此数据仓库建设是非常严谨及规范的过程，对数据仓库建设工程师的综合能力要求也是极高的。

DCP1002培训&认证正是为了解决这一问题，让大数据数仓建设变得更标准化、规范化及体系化。该认证围绕数梦工场数据仓库建设理论体系及相应工具展开，涉及到数梦的主要产品有：DataBridge、DataRiver、DThink EMR、DataStudio、DataMall的操作使用。

通过该认证可以有效证明该人员具备以下能力：

- 掌握数梦基于大数据平台的数据仓库解决方案；
- 深刻理解数梦数据仓库体系架构和分层理念；
- 理解从需求调研到数据集成、数据标准、数据建模、数据整合、数据质量、数据集市等建设过程；
- 能够在数仓建设中熟练使用数梦数据仓库建设工具；
- 能依据数梦数仓建设方法及工具完成数据仓库建设工作；

## 二、 DCP 认证报名及要求

数梦工场大数据工程师专业认证的报名方式：数梦工场官网在线报名：[数梦工场DCP 认证官网报名页面](#)。

DCP1002专业认证所需具备的前提基础知识：

- 了解数据仓库基本概念；
- 有数据库基础，并具备一定的 SQL 编程能力；
- 对大数据平台知识（不限于 hive/spark/Maxcompute）有一定了解；

- 有 ETL 开发经验，有一定编程经验（不限于 Python/java）；
- 具备基础 Linux 命令的使用。

### 三、 认证考试说明

- 认证考试形式和试卷结构：
  - 理论闭卷考试（占比40%分值）
  - 上机实践考试（占比60%分值）
- 试卷满分及考试时间  
试卷满分为 100 分；理论考试时间为 80分钟，上机考试时间为240分钟。
- 考试形式  
理论考试采用线上考试（共50道题，单选30道、多选20道）；  
上机考试为集中线下认证考试（综合实验任务）。
- 通过要求  
**机考成绩必须达到70分以上且最后综合总分>=70分以上，认证才算通过。**

### 四、 考试内容知识点

DCP1002考试内容：

- 大数据数仓建设体系方案
  - 了解数据仓库的发展背景；
  - 掌握数据仓库特点及相关基础理论知识；
  - 理解数据仓库分层概念及作用；
  - 数梦大数据产品在数据仓库建设的定位。
- 数据仓库之数据建模
  - DataRiver数据建模名词理解；
  - DataRiver数据建模操作流程；

- DataRiver数据模型映射关系作用。

## ● 数据仓库之数据调研

- 数据调研在数仓项目中的作用、意义以及调研的二大阶段；
- 数据调研的人力保障内容，流程保障；
- 数据调研的在每阶段的定义，实施细节，以及对应的输入输出内容。

## ● 数据仓库之数据集成

- 了解数据集成的基本概念和工作流程；
- 熟练使用数据集成进行多种场景下的数据同步，可以从不同的数据源同步数据到大数据计算服务；
- 了解数据集成中的组件，包括概念、功能以及实际使用；
- 能正确使用数据集成中的字段映射、常量使用、变量使用、导入导出规则等；
- 了解数据集成的最佳实践，掌握含全量集成、增量集成（基于日志及时间戳）、CLOB字段集成、历史数据补录等常见场景的数据集成；
- 掌握相关大数据平台数据简单的数据查看功能（验证数据集成的准确性）。

## ● 数据仓库之数据标准

- 理解数据标准位于数据仓库中的层级，以及数据标准在数据仓库中的作用；
- 理解数据标准中数据元素和数据字典的定义和作用；
- 掌握在 DataRiver 中创建数据元素和数据字典，与数据表形成映射，生成标准化清洗任务；
- 在 DataRiver 中生成可视化清洗任务。

## ● 数据仓库之数据整合

- 理解数据整合在各分层的整合内容及作用；
- 掌握数据整合在数据仓库中的实施策略；
- 掌握数梦DataBridge、DataRiver、DataStudio等产品在数据整合过程中的定位与操作方法。

## ● 数据仓库之数据质量

- 掌握数据质量的概念；
- 理解数据质量控制实施过程与方法；
- 掌握基于River的数据质量实施操作过程步骤；

- 掌握数据质量小组实施成员所具备的角色。

- **数据仓库之数据集市**

- 数据集市的基本概念和类型；
- 数据集市的建设步骤和建设必要性；
- 数据集市的建设注意事项和应用场景；
- 从传统数据集市到数据挖掘。