

主要面向各企事业单位、政府部门等的信息化数字化部门、业务经营部门，从事基本的旅游数据查询、数据整理、数据统计、运营报表制作、数据标注、数据可视化、大数据平台使用与问题答疑等工作，根据已有的数据基础，实现数据的基本分析与应用。旅游大数据思维指导

**1.1 旅游大数据业务指导**

**1.1.1** 能够理清政府官方发布的旅游相关数据的来源、类型、采集方式等。**1.1.2** 能够遵守《旅游法》、《旅行社条例》处理旅游业务。**1.1.3** 能够将旅游大数据应用在一个具体旅游工作或业务场景中。**1.1.4** 能够理解大数据与旅游相结合的落地路径。

**1.2 旅游大数据咨询设计**

**1.2.1** 能够在主管的指导下，对旅游大数据管理进行规划设计。**1.2.2** 能够在有参考样例的基础上，使用至少 1 种思维导图工具（如 Visio、在线协作绘图软件等），编辑简单的数据思维导图、流程图。**1.2.3** 能够看懂旅游监管部门发布统计报告的关键数据，明确旅游统计的调查对象和调查范围。**1.2.4** 掌握旅游行业最新的发展变化，主动判断旅游行业政策、业务态势。

**1.3 法规制度规范指导**

**1.3.1** 能够严格遵守旅游行业相关的重要法律法规。**1.3.2** 能够持续关注旅游行业具体业务最新的政策变化。**1.3.3** 了解并严格遵守大数据管理与应用相关的重要法律法规。**1.3.4** 能够持续关注大数据管理与应用方面最新的政策变化。

**2. 旅游数据收集与采集**

**2.1 旅游数据采集规划**

**2.1.1** 能够按照正确的步骤和流程进行数据采集，不合格率不超过 30%。**2.1.2** 能够分析业务问题，明确所需要的数据以及数据来源、采集方法。**2.1.3** 能够正确使用项目的数据采集标准，确保采集数据的标准化与合理性。**2.1.4** 能够准确定位已有的内外部数据、旅游相关网站渠道等。

**2.2 常见涉旅数据采集**

**2.2.1** 能够借助常用的办公软件进行快速高效的在线调研。**2.2.2** 能够借助常用的办公软件（如 Word、Excel 等），将线下纸质文件中的数据信息转化为电子数据。**2.2.3** 能够了解旅游企业内部业务软件相关的基本功能，筛选出旅游分析所需要的数据项。**2.2.4** 能够了解旅游监管软件相关的基本功能，筛选出旅游分析所需要的数据项。

**2.3 互联网数据智能采集**

**2.3.1** 了解互联网常见的数据分析网站，有效登录和查询这些网站的相关数据。**2.3.2** 了解涉旅数据常见的官方网站，正确登录和查询这些网站的相关数据。**2.3.3** 能够借助智能工具，在主管的指导下，正确爬取与旅游业务相关的有效网络数据。**2.3.4** 了解至少 1 种网络爬虫工具（如 Excel 网页爬虫、后羿采集器、八爪鱼爬虫软件、爬山虎采集器等）。**2.3.5** 能够了解业务系统日志数据采集服务的常规业务信息项。

**3. 旅游大数据管理与治理**

**3.1 大数据编目管理**

**3.1.1** 能够对多渠道和手段采集的数据进行规范合理的数据归集与整理，归集成功率 80%。**3.1.2** 能够对不同类别的旅游数据资源进行简单的归类、编目，80%编目合理。**3.1.3** 能够对不同类别的旅游数据资源进行标签设计，标签创建成功率达到 80%。**3.1.4** 能够按照大数据管理平台的操作手册，基于数据编目与管理进行基本的操作。

**3.2 数据安全与隐私保护**

**3.2.1** 能够合规合法地进行数据采集、分析与管理，符合数据安全、隐私的相关法律法规的要求。应了解数据安全、隐私相关法律法规。**3.2.2** 能够区分不同信息化系统所需的信息安全等级，误差率不超过 20%，应了解信息安全等级保护的等级划分、核心要求。**3.2.3** 能够参与旅游信息系统二级等级保护所需的制度建设工作。**3.2.4** 能够参与配合信息系统二级等级保护制度以外其他建设和改造工作。

**4. 旅游大数据分析与赋能**

**4.1 旅游大数据分析**

**4.1.1** 掌握常见的数据统计指标，能够根据业务分析需求，选择正确的统计指标，并进行计算，计算正确率达到 70%。**4.1.2** 能够从业务目标和实际旅游场景的角度，合理分析所获取与管理的大数据，分析结果合理性达到 70%。**4.1.3** 能够看懂旅游企业经营、旅游政府监管等的数据分析报表，简单分析出一些问题或者结果，应了解旅游相关业务中核心数据之间的关系。**4.1.4** 能够对所获取的旅游数据进行时间、空间的对比分析，应了解业务数据产生的逻辑关系。**4.1.5** 了解至少 1 种大数据分析工具（如 Excel、在线数据分析软件）。

**4.2 旅游大数据应用与赋能**

**4.2.1** 能够以目标和结果为导向，合理进行旅游大数据的应用。**4.2.2** 能够基于旅游大数据业务场景，制作有针对性的数据分析报告。**4.2.3** 了解至少 3 个旅游大数据场景应用（如景点客流监测、酒店入住分析、

行业数据统计、旅游黄金周监测等)。4.2.4 能够通过对企业 / 政府自有数据以及外部数据的分析,深刻洞察旅游市场和旅游消费者。提供全方位的数据赋能,例如制定合理的市场营销策略、政府监管决策等。

**4.3 大数据可视化**

**4.3.1** 能够根据要展示的数据内容,选择合适、合规的可视化图表,合理性达到 70%。**4.3.2** 能够使用至少 1 种非程序式可视化工具(如 Excel、PPT 等),对数据进行可视化解释。**4.3.3** 能够结合数据可视化,编写旅游大数据分析报告,报告数据的准确率达到 80%。**4.3.4** 能够完成数据统计图表化、数据结果展示化。主要面向企业、银行和非银行机构的投融资业务数据采集和投融资数据测算等岗位,能利用大数据技术抓取基本金融数据并进行数据处理分析,业财税融合,从事投资数据测算和融资数据测算等工作。

**1.大数据技术基础应用**

**1.1 Python 基础编程**

**1.1.1** 能正确运用 Python 语言的基本语法规则。**1.1.2** 能正确运用 Python 语言编写选择和循环结构应用程序。**1.1.3** 能正确完成 Python 编程常见错误信息排查与纠错。**1.1.4** 能正确运用 Python 语言编写符合业务需求的简单 Python 程序。

**1.2 大数据抓取**

**1.2.1** 能正确运用 Tushare 金融大数据中心接口包采集股票、国债利率和贷款基准利率等数据。**1.2.2** 能正确运用 Tushare 金融大数据中心接口包采集行业发展状况和市场行情数据。**1.2.3** 能正确运用 Tushare 金融大数据接口包采集投资、融资业务所需的相关数据。

**1.3 大数据清洗与存储**

**1.3.1** 能正确对数据进行一致性检查。**1.3.2** 能正确处理数据的无效值和缺失值。**1.3.3** 能正确检测和消除重复记录。**1.3.4** 能正确将采集的数据存储到 Excel 文件。

**2.投资数据测算**

**2.1 投资现金流量测算**

**2.1.1** 能正确计算企业或项目投资的建设期和经营期。**2.1.2** 能正确计算现金流出量,包括原始投资额、垫支营运资本。**2.1.3** 能正确计算营业现金毛流量,包括税后营业收入和税后付现营业费用。**2.1.4** 能根据企业或项目投资,计算终结期回收额,包括回收垫支的营运资本、回收长期资产的净残值或变现价值。**2.1.5** 能根据企业或项目投资总流入量和总流出量,计算净现金流量。

**2.2 主要投资指标测算**

**2.2.1** 能合理运用税收政策,正确测算股权资本成本和税后债务资本成本。**2.2.2** 能根据企业或项目投资,运用一般模式测算资本成本或运用加权资本成本法测算项目资本成本。**2.2.3** 能根据投资项目现金流量和资本成本,正确计算投资净现值。**2.2.4** 能根据经营期各年净现金流量现值与建设期原始投资现值,正确计算净现值率。**2.2.5** 能根据查表法、试算法和内插法,正确计算投资项目内部收益率。

**2.3 其他投资指标测算**

**2.3.1** 能根据累计净现金流量,正确计算静态投资回收期。

**3.融资数据测算**

**3.1 资金需要量测算**

**3.1.1** 能熟练利用大数据技术采集资金市场相关信息数据。**3.1.2** 能利用大数据技术预测市场占有率,准确测算销售量和销售额,并测算销售增长额。**3.1.3** 能根据预计期销售额和基期销售额,准确测算销售增长额。**3.1.4** 能根据销售量和收入数据,正确运用销售百分比法、高低点法、回归直线法测算总资金需要量。**3.1.5** 能根据需增加的资金和企业内部可筹集资金,准确测算外部资金需要量。

**3.2 资金成本测算**

**3.2.1** 能从金融大数据中心准确采集国债利率、贷款基准利率、市场利率、债券利率和沪深 50 股票收益率等数据资料。**3.2.2** 能根据采集的贷款基准利率,准确计算银行借款资金成本。**3.2.3** 能根据采集的债券利率,准确计算债券成本。**3.2.4** 能根据采集的国债利率、市场利率、债券利率和沪深 50 股票收益率,准确计算企业发行普通股和优先股成本。

**3.3 杠杆系数测算**

**3.3.1** 能根据销售收入和变动成本,正确计算边际贡献和边际贡献率。**3.3.2** 能根据边际贡献和固定经营成本,正确计算息税前利润。**3.3.3** 能利用企业内部数据计算息税前利润变动率和产销量变动率,正确计算经营杠杆系数。**3.3.4** 能利用企业内部数据采集的普通股每股利润变动率和息税前利润变动率,正确计算财务杠杆系数。**3.3.5** 能根据经营杠杆和财务杠杆系数,正确计算总杠杆系数。根据业务需求和工作流程,主要使用 Excel、Power BI 和 Python 等工具,完成数据获取、数据处理、数据存储、数据分析、数据可视化、报表制作等工作。

**1.数据管理**

**1.1 数据获取**

**1.1.1** 能够根据业务需求基于 Excel 或 PowerBI 实现数据录入与整理。**1.1.2** 能够根据业务需求基于 Excel 或 PowerBI 实现 CSV 等文

件的数据获取 1.1.3 能够根据业务需求基于 Excel 或 PowerBI 实现网站数据获取 1.1.4 能够根据业务需求基于 Excel 或 PowerBI 实现 MySQL 等关系型数据库的数据获取 1.2 数据存储与查询 1.2.1 能够根据业务需求基于 MySQL 等关系型数据库实现 Excel、TXT 等文件存储与加载 1.2.2 能够根据业务需求基于 MySQL 等关系型数据库实现将数据存储为 Excel、TXT 等文件 1.2.3 能够根据业务需求基于 SQL 语句实现数据查询获取 1.2.4 能够根据业务需求使用 MySQL 进行数据库的同步、备份 1.3 文件操作 1.3.1 掌握 Python 基本环境安装、Python 的基础语法、程序流程控制语句以及函数的使用 1.3.2 能够根据业务需求基于 Python 实现 Excel、TXT 等文件的数据读取 1.3.3 能够根据业务需求基于 Python 实现 Excel、TXT 等文件的数据写入 1.3.4 能够根据业务需求基于 Python 实现 Excel、TXT 等文件的批量合并 1.3.5 能够根据业务需求基于 Python 实现 Excel、TXT 等文件的批量拆分 2. 数据处理与分析 2.1 数据探索 2.1.1 掌握 Python 的列表、字符串、字典、元组、集合等数据结构的使用及常用操作 2.1.2 掌握 Python 面向对象编程 2.1.3 能够根据业务需求基于 Python 或 Excel 实现数据结构、数据分布的查看 2.1.4 能够根据业务需求基于 Excel 实现数据的频数分析 2.1.5 能够根据业务需求基于 Excel 实现数据的集中趋势分析 2.1.6 能够根据业务需求基于 Excel 实现数据的离散程度分析 2.2 数据处理 2.2.1 能够根据业务需求基于 Excel 或 PowerBI 实现数据缺失值、重复值、异常值的识别与处理 2.2.2 能够根据业务需求基于 Python、Excel 或 Power BI 实现数据类型转换 2.2.3 能够根据业务需求基于 Python、Excel 或 Power BI 实现数据合并 2.2.4 能够根据业务需求基于 Python、Excel 或 Power BI 实现数据提取、筛选等操作 2.3 数据分析 2.3.1 能够根据业务需求基于 Excel 或 PowerBI 实现同比、环比分析 2.3.2 能够根据业务需求基于 Excel 或 PowerBI 实现相关分析 2.3.3 能够根据业务需求基于 Excel 或 PowerBI 实现预测分析 2.3.4 能够根据业务需求基于 Excel 或 PowerBI 实现其他分析 2.4 数据可视化 2.4.1 能够根据业务需求基于 Excel 或 PowerBI 实现柱状图绘制 2.4.2 能够根据业务需求基于 Excel 或 PowerBI 实现散点图绘制 2.4.3 能够根据业务需求基于 Excel 或 PowerBI 实现饼图绘制 2.4.4 能够根据业务需求基于 Excel 或 PowerBI 实现其他个性化图表绘制 3. 报表管理 3.1 报表制作 3.1.1 能够根据业务需求基于 Excel 或 PowerBI 实现数据透视表制作 3.1.2 能够根据业务需求基于 Excel 或 PowerBI 实现数据分析报表制作 3.1.3 能够根据业务需求基于 Excel 或 PowerBI 实现报表的图、表和文字的布局 3.1.4 能够根据业务需求基于 Excel 或 PowerBI 实现报表的美化 3.2 报表部署 3.2.1 能够根据业务需求基于 Power BI 实现报表相关数据、文字描述和图表的发布 3.2.2 能够根据业务需求基于 Power BI 实现仪表板的创建与设置 3.2.3 能够根据业务需求基于 Power BI 实现对发布后的数据共享 3.2.4 能够根据报表部署过程输出规范化的说明文档 3.3 报表更新维护 3.3.1 能够根据业务需求定期更新报表 3.3.2 能够根据业务需求进行报表维护 3.3.3 能够根据业务变化持续迭代优化报表 3.3.4 能够及时发现和定位报表数据问题主要面向大数据行业、大数据处理与应用领域企事业单位的数据采集标注专员、初级大数据 ETL 工程师、大数据开发工程师助理等岗位，利用正在运行的数据系统，根据项目文档及大数据处理流程，能使用图形化或可视化工具完成工程化的数据采集、数据抽取、数据转换、数据加载、数据系统应用等。

1. 大数据工程化采集 1.1 网络数据采集 1.1.1 根据网页标题采集需求，能使用可视化爬虫工具，获取目标标题数据。1.1.2 根据网页文本采集需求，能使用可视化爬虫工具，智能解析文本数据内容。1.1.3 根据数据标签采集需求，能使用可视化爬虫工具，智能解析标签数据。1.1.4 根据循环遍历采集需求，能使用可视化爬虫工具，实现智能翻页采集。 1.2 离线数据采集 1.2.1 根据离线数据采集要求，能使用图形化工具，查询数据源的存储路径信息，确定要抽取的离线数据数量，获取数据源准确的位置信息。1.2.2 根据数据格式，能使用图形化工具，创建转换对象并配置相应结构化数据输入模块。1.2.3 根据目标数据系统类型及原始数据集格式，能使用图形化工具，配置相应数据输出模块。1.2.4 根据配置完成后的模块，能使用图形化工具，获得离线采集

数据。1.2.5 根据数据内容，能使用图形化工具创建存储采集数据的数据表结构。

1.3 作业调度 1.3.1 根据任务调度要求，能使用调度工具，创建作业的图形视图，调度所需要的模块，获取起始到完结所需的任务模块。1.3.2 根据所有转换模块，能使用调度工具，

配置各模块的关联信息，获得符合数据传输流程的模块调度链。1.3.3 根据各模块的关联关系，能使用调度工具，对所有模块进行参数配置，获得数据连通的作业图形视图。1.3.4 根据配置完成的作业属性及任务调度需求，能使用调度工具执行作业，检查目标数据系统数据的

入库情况，并获得入库信息反馈。2. 大数据工程化处理 2.1 数据清洗 2.1.1 根据采集的原始数据集，能使用图形化工具删除离线数据集中的缺省数据，获得完整的数据集。2.1.2

根据完整数据集，能使用图形化工具，简单标记、删除内容重复的离线数据，获得无干扰的数据集。2.1.3 根据无干扰数据集，能使用图形化工具，归一处理离线数据中不符合标准单位要求的字段，获得标准化数据集。2.1.4 根据标准化数据集，能使用图形化工具，标记、

删除离线数据中不符合数据质量要求的数据，获得可用数据集。2.2 数据计算 2.2.1 根据数据清洗后的数据集，使用图形化工具连接各数据表并根据处理需求进行简单数值计算字符切分等处理，获得正确处理后的多表数据集。2.2.2 根据多表数据集，能使用图形化工具，

连接关联处理多表数据，获得关联整合数据集。2.2.3 根据关联整合数据集，能使用图形化工具，进行数据聚合、分组，获得关联计算数据集。2.3 数据派生 2.3.1 根据给定数据标签规则，能使用图形化工具依据给定数据标签，创建各主题的标签库。2.3.2 根据不同主题的

标签，能使用图形化工具，对关联计算数据集中的数据，进行标签派生及验证，获得含标签的数据集。2.3.3 根据已标签数据集，能使用图形化工具，根据主题划分数据表，创建业务主题数据表。2.3.4 根据主题数据表的格式规范，能使用图形化工具，对同主题数据集

进行全字段的数据组织，获得符合业务主题的宽表。3. 大数据工程化应用 3.1 OLAP 系统应用 3.1.1 根据给定数据表结构，能使用图形化工具，在指定项目中选择数据的加载方式，配置数据系统的数据表，获得符合业务分析需求的模型。3.1.2 根据配置完成的模型，

能使用图形化工具，以查看表格的形式进行检查验证，完成模型的调试。3.1.3 根据调试后的模型，能使用图形化工具，创建并使用数据立方，编写简单查询语句进行数据查询，正确查询目标数据。3.2 搜索系统应用 3.2.1 根据给定数据结构，能使用图形化工具，关联

分布式搜索组件，并执行脚本完成索引结构创建。3.2.2 根据索引结构，能使用图形化工具，将数据完整地导入到分布式搜索组件中，获得可全文检索的数据系统。3.2.3 根据搜索业务需求以及搜索组件中数据，能使用图形化工具，进行数据查询，正确查询目标数据。3.3 报表系统应用 3.3.1 根据报表模板，能使用合适的可视化工具，配置所需展示数据源的连接信息，正确连通数据系统。3.3.2 根据报表数据及离线数据需求，能使用可视化工具，设置数据

透视表、图表等，生成符合需求的静态数据报表。3.3.3 根据数据报表，能使用可视化工具，利用数据报表的筛选、自定义排序、统计等功能，以指定的格式处理并存储数据报表结果。能够根据业务情况，辅助进行大数据采集；能够根据大数据预处理需求，完成基本的大数据预处理工作；能够根据大数据可视化需求，运用数据分析和可视化工具，进行基本的分析，以及数据的静态和动态可视化展示；能够辅助进行数据标准的架构体系设计，能够进行元数据和主数据的创建、查询、维护等基本操作。1.大数据采集 1.1 大数据采集基础知识了解 1.1.1 了解大数据的概念、来源和意义，了解大数据的特征和结构，了解大数据的表现形式和应用场景。1.1.2 熟悉大数据采集的概念，及其在大数据治理过程中的意义。1.1.3 熟悉大数据采集的数据来源、数据规模、数据类型和数据存储方式等。1.1.4 熟悉大数据采集的方式（数据录入、文本文件导入、数据库文件导入、网页爬虫等）及其基本原理。1.1.5 熟悉至少一种大数据采集通用工具（Python、Excel、Flume、Sqoop、八爪鱼等）的使用。1.2 大数据采集方案制定 1.2.1 能够根据业务情况，辅助进行大数据采集业务需求调研。1.2.2 能够根据数据采集需求，辅助进行数据资源的梳理。1.2.3 能够根据梳理好的数据资源，

辅助确定数据采集的方法和工具。1.2.4 能够辅助编写详细的大数据采集方案。

### 1.3 大数据采集实施

1.3.1 能够进行数据的快速录入。1.3.2 能够进行至少一种常用数据库（Access、Mysql、MongoDB、Oracle 等）的快速导入。1.3.3 能够基于录入和导入的数据，进行增、删、改、查等基本操作。1.3.4 能够掌握基本的数据查询技巧。1.3.5 能够按照数据标准进行数据内容的规范与处理。

## 2.大数据预处理

### 2.1 大数据预处理基础知识了解

2.1.1 了解大数据预处理的概念、环节，及其在大数据处理过程中的意义。2.1.2 熟悉数据清洗的概念和意义，并能够进行去重、空值处理等基本的数据清洗。2.1.3 熟悉数据集成的概念和意义，并能够进行数据合并、数据表关联等简单的数据集成。2.1.4 熟悉至少一种数据预处理工具（Excel、Python、Sql、存储过程、Kettle 等）的使用。

### 2.2 大数据预处理方案制定

2.2.1 能够理解大数据预处理需求。2.2.2 能够根据大数据预处理需求，梳理不同类型和结构的数据。2.2.3 能够针对至少一种类型和结构的数据，选定合适的大数据预处理工具和方法。2.2.4 能够辅助编写大数据预处理方案。

### 2.3 大数据预处理实施

2.3.1 能够掌握至少一种类型数据文件（txt、csv、html、xml、xls 等）的加载。2.3.2 能够进行数据的排序、筛选和分类汇总。2.3.3 能够处理数据中存在的格式问题。2.3.4 能够处理数据中存在的內容问题。2.3.5 能够进行数据列的合并与拆分。

## 3.大数据可视化

### 3.1 大数据可视化基础知识了解

3.1.1 了解大数据可视化的概念。3.1.2 了解可视化在大数据治理过程中的意义。3.1.3 了解大数据可视化的形式和应用。3.1.4 熟悉至少一种大数据可视化工具（Excel、Tableau、Python、R 等）的使用。

### 3.2 大数据可视化方案制定

3.2.1 能够理解大数据可视化需求。3.2.2 能够进行基本的描述性和探索性数据分析。3.2.3 能够理解数据可视化的目标。3.2.4 能够辅助编写大数据可视化方案。

### 3.3 大数据可视化实施

3.3.1 能够进行数据可视化的创建、编辑和布局等。3.3.2 能够使用大数据可视化工具，实现数据探索过程的图像化静态和动态展现。3.3.3 能够使用大数据可视化工具，实现数据分析结果的图像化静态和动态展现。3.3.4 能够辅助编写大数据可视化报告。

## 4.大数据治理

### 4.1 大数据治理基础知识了解

4.1.1 了解大数据治理的概念和关键要素。4.1.2 了解大数据治理的对象和内容。4.1.3 了解大数据治理的框架和核心准则。4.1.4 了解大数据治理的实施保障。

### 4.2 数据标准体系建设

4.2.1 了解数据标准的定义和内容。4.2.2 了解数据标准管理的组织架构和设计流程。4.2.3 了解数据标准的制定和执行流程。4.2.4 能够根据数据标准设计的要求，进行数据标准化。4.2.5 能够根据业务需求，辅助进行数据标准的架构体系设计。4.2.6 能够在企业、社会、政府等的大数据治理过程中，参与数据标准体系的建设。

### 4.3 元数据管理

4.3.1 了解元数据的概念、类型和应用。4.3.2 了解元数据管理的概念和意义。4.3.3 了解元数据的生成、获取和维护。4.3.4 了解元数据的存储、整合和共享。4.3.5 能够进行元数据的创建、查询和维护等基本操作。

### 4.4 主数据管理

4.4.1 了解主数据的概念、类型和应用。4.4.2 了解主数据管理的概念和意义。4.4.3 了解主数据的获取和维护。4.4.4 了解主数据的存储、集成和共享。4.4.5 能够根据业务实施方案，进行主数据的创建、查询和维护等基本操作。

根据项目需求，独立完成简单的大数据开发、组件运维、平台监控管理等工作。

## 1.大数据平台集群硬件管理

### 1.1 大数据平台服务器基础管理

1.1.1 按照大数据平台管理指导书，能够独立登录 Linux 服务器；1.1.2 按照大数据平台管理指导书，能够独立完成 Linux 文件目录的创建、查看、删除等操作；1.1.3 按照大数据平台管理指导书，能够独立完成 Linux 文件的创建、查看、移动、复制、删除等操作；1.1.4 按照大数据平台管理指导书，能够独立掌握 Linux vi/vim 文本编辑操作。

### 1.2 大数据平台集群管理

1.2.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动集群；1.2.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止集群；1.2.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看集群健康状态；1.2.4 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看集群历史警报情况。

## 1.3 大

数据平台节点管理 1.3.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序添加新的节点；1.3.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序删除节点；1.3.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动节点所有角色；1.3.4 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止节点所有角色；1.3.5 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看节点运行状态。1.4 大数据平台资源监控管理 1.4.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看集群资源使用情况；1.4.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看节点资源使用情况。2. 大数据平台组件可视化监控管理 2.1 HDFS 组件运行可视化监控管理 2.1.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 HDFS 组件；2.1.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 HDFS 组件；2.1.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看 HDFS 组件的运行健康状态。2.2 Yarn 组件运行可视化监控管理 2.2.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 Yarn 组件；2.2.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 Yarn 组件；2.2.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看 Yarn 组件的运行健康状态。2.3 Zookeeper 组件运行可视化监控管理 2.3.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 Zookeeper 组件；2.3.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 Zookeeper 组件；2.3.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看 Zookeeper 组件的运行健康状态。2.4 Hive 组件运行可视化监控管理 2.4.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 Hive 组件；2.4.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 Hive 组件；2.4.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看 Hive 组件的运行健康状态。2.5 HBase 组件运行可视化监控管理 2.5.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 HBase 组件；2.5.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 HBase 组件；2.5.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看 HBase 组件的运行健康状态。3. 大数据组件使用 3.1 SQL 基本操作命令使用 3.1.1 按照大数据平台开发指导书，能够独立完成数据库的创建、查看、删除等操作；3.1.2 按照大数据平台开发指导书，能够独立完成表的创建、查看、删除等操作；3.1.3 按照大数据平台开发指导书，能够独立完成数据的写入、更新、删除等操作；3.1.4 按照大数据平台开发指导书，能够独立完成表数据的简单查询；3.1.5 按照大数据平台开发指导书，能够独立完成表数据的条件过滤查询；3.1.6 按照大数据平台开发指导书，能够独立完成表数据的子查询。3.2 HDFS 组件使用 3.2.1 理解 HDFS 存储机制；3.2.2 按照大数据平台开发指导书，能够独立通过 Shell 命令的方式完成 HDFS 目录的创建、删除、修改等；3.2.3 按照大数据平台开发指导书，能够独立通过 Shell 命令的方式完成 HDFS 文件的上传、查看、下载等；3.2.4 按照大数据平台开发指导书，能够独立通过 Shell 命令的方式完成 HDFS 文件及目录权限的查看、授予、删除等。3.3 Yarn 组件使用 3.3.1 理解 Yarn 的基础原理；3.3.2 了解 Yarn 的三种任务调度机制；3.3.3 按照大数据平台开发指导书，能够独立通过 Shell 命令的方式完成 Yarn 中任务的提交、查看、终止等。3.4 Hive 组件使用 3.4.1 理解 Hive 基础原理；3.4.2 按照大数据平台开发指导书，能够独立完成 Hive 数据库的创建、查看、删除等操作；3.4.3 按照大数据平台开发指导书，能够独立完成 Hive 内部表的创建、查看、删除等操作；3.4.4 按照大数据平台开发指导书，能够独立完成 Hive 外部表的创建、查看、删除等操作；3.4.5 按照大数据平台开发指导书，能够独立完成 Hive 表数据的写入操作；3.4.6 按照大数据平

台开发指导书，能够独立完成 Hive 表的简单查询、过滤查询等操作。 3.5 HBase 组件使用 3.5.1 理解 HBase 存储机制；3.5.2 按照大数据平台开发指导书，能够独立通过 Shell 命令完成 HBase 命名空间的创建、查看、删除等操作； 3.5.3 按照大数据平台开发指导书，能够独立通过 Shell 命令完成 HBase 表的创建、查看、禁用、删除等操作； 3.5.4 按照大数据平台开发指导书，能够独立通过 Shell 命令完成 HBase 表数据的写入操作；3.5.5 按照大数据平台开发指导书，能够独立通过 Shell 命令完成 HBase 表数据的全表查询、过滤查询等操作。

4.大数据 ETL 工具使用 4.1 Sqoop 工具使用 4.1.1 熟悉 Sqoop 工具使用规范；4.1.2 按照大数据平台开发指导书，能够独立编写 Sqoop 脚本完成查看数据库、查看数据表等操作；4.1.3 按照大数据平台开发指导书，能够独立编写 Sqoop 脚本完成全量数据导入操作；4.1.4 按照大数据平台开发指导书，能够独立编写 Sqoop 脚本完成全量数据导出操作。

4.2 Flume 工具使用 4.2.1 熟悉 Flume 工具使用规范；4.2.2 按照大数据平台开发指导书，能够基本掌握 Flume 中 Exec Source、Avro Source 等的使用方法；4.2.3 按照大数据平台开发指导书，能够基本掌握 Flume 中 Memory Channe 的基础使用方法；4.2.4 按照大数据平台开发指导书，能够基本掌握 FLume 中 Logger Sink、Avro Sink 等的使用方法。

4.3 Kafka 工具使用 4.3.1 熟悉 Kafka 工作原理； 4.3.2 按照大数据平台开发指导书，能够基本掌握 Kafka Shell 命令用法。根据操作文档的要求，独立完成云计算平台搭建、操作系统安装、Hadoop 平台配置与部署，能够对常见故障进行识别、判断和处理，满足日常运维、故障管理和日常巡检要求，以及承担数据采集、预处理、存储和处理等基础工作任务。

1. 平台搭建 1. 1 云计算平台搭建 1. 1. 1 能根据操作文档搭建 Openstack 私有云平台，安装 Nova、Swift、Glance、Dashboard 等组件。1. 1. 2 能在专业人员指导下，协助完成私有云平台搭建过程中的排错工作。1. 1. 3 能根据操作文档，正确使用公有云资源，如云服务器、对象存储服务、云网络服务、云安全服务、云数据库服务等。 1. 2 操作系统安装 1. 2. 1 1. 2. 2 1. 2. 3 1. 2. 4 能独立配置操作系统所需虚拟环境。能独立完成操作系统的安装与设置。能独立进行操作系统补丁安装。能同时使用图形化和命令行界面进行操作。 1. 3 Hadoop 系统部署 1. 3. 1 1. 3. 2 1. 3. 3 1. 3. 4 排查。 能根据操作手册进行 Hadoop 单节点部署。能根据操作手册进行 Hadoop 伪分布式部署。能根据操作手册进行 Hadoop 集群部署。能在搭建好的集群上运行自带测试用例，进行错误 2. 统调优 2. 1 运维管理 2. 1. 1 能运用供应商提供的系统运维管理工具或者可视化界面，独立完成大数据架构下各个模块（如：HDFS、spark、配置信息）的日常运维管理操作，如：系统状态监测、系统日志收集、日常巡检等。2. 1. 2 能运用 Shell 命令完成对各个节点的日常运维操作，如：运维脚本查询、存储使用情况、机器运行状态、网络防火墙配置。2. 1. 3 能根据业务需求，运用数据库管理工具(Pl sql , navicat)，完成数据库日常监控和运维。2. 1. 4 能协助高级技术支持人员梳理优化运维管理流程，补充系统运维操作手册等相关文档。 2. 2 故障管理 2. 2. 1 能运用日常运维实例和网络检索实例，定位日常简单故障，评估可能造成的影响范围，并及时上报。2. 2. 2 能运用系统故障处理的常用方法和工具，独立分析常见故障的原因，提出改进建议和方法措施。2. 2. 3 能根据软件提供的系统故障处理手册，运用故障诊断工具或系统自带的故障诊断命令，进行应急处理，保障系统的稳定运行。2. 2. 4 能在故障闭环后，根据故障现象、解决过程、核心原因、处理方案等多方面因素，编写相关故障解决文档。 2. 3 日常巡检 2. 3. 1 能使用基础运维工具完成硬件资源信息的收集汇总并协助硬件管理人员完成周期性更新勘误。2. 3. 2 能使用基础运维工具完成软件部署信息的收集汇总并独立完成软件服务部署信息的周期性更新勘误。2. 3. 3 能使用基础运维工具完成网络资源分配信息的收集汇总，并协助网络资源管理人员完成对网络拓扑结构和资源 IP 的周期性更新勘误。2. 3. 4 能根据既有常用组织架构体系，并通过供应商所提供的配置管理工具，完成对具体岗位职务的日常运维工作。 3. 应用部署 3. 1 数据采集



3.1.1 能运用大数据技术(SparkStreaming、Kafka 等)完成数据采集系统的搭建和基础配置。

3.1.2 能运用 Python 编写基础的爬虫脚本,完成对指定网页的数据爬取。

3.1.3 能运用 Flume 技术,完成大数据架构体系的系统日志收集。

3.1.4 能运用 ETL 工具( Kettle ),完 Excel/json/csv 等通用格式的数据采集任务。

3.1.5 能运用 Kettle 构建定时调度作业,完成对已有采集任务的自动优化。

3.2 数据预处理

3.2.1 能运用 ETL 工具( Kettle ),完成数据转换,合并与拆分动作。

3.2.2 能运用 ETL 工具( Kettle ),对关键字段完成数据脱敏。

3.2.3 能运用 Python 语言,实现对已采集数据的清洗。

3.2.4 能通过 ETL 工具( sqoop ),完成从关系数据库到 NOSQL 数据库的数据迁移。

3.3 大数据存储

3.3.1 能运用大数据技术(如分布式文件系统 HDFS)维护、管理数据存储系统,优化存储资源利用率和计算效率。

3.3.2 能独立完成 HBase 分布式集群环境部署。

3.3.3 能运用 HBase Shell 完成日常节点维护和数据查询导出等维护动作。

3.4 大数据处理

3.4.1 能运用 spark 技术,完成多中间件日志格式的解析和系统各种基础指标的日常统计。

3.4.2 能运用 Neo4j 技术,完成对图形化数据的可视化输出。

3.4.3 能运用日常数据库技术,优化 ETL 工具流程,监控并维护例行数据 ETL 任务。

3.4.4 能运用标注工具 Labelme,完成对日常巡检图像数据的标注和预警。根据工业企业的业务需求,完成工业大数据软件的安装部署,能够监控集群的运行状态,进行简单的运维处理;能够进行简单的数据库管理、数据查询及处理操作;能够使用工具完成简单的数据采集与数据清洗;能够进行 Hadoop 分布式文件系统 HDFS 的存储;能够使用工具完成简单的数据可视化。

工业大数据软件安装与维护

1.1 工业大数据软件安装部署

1.1.1 能够根据工业大数据业务场景的需求,选择合适的工业大数据软件,并根据运行要求,选择合适的软硬件环境;

1.1.2 能够独立安装 Linux 操作系统;

1.1.3 能够根据给定的系统软硬件环境,独立完成工业大数据软件伪分布式的安装与部署;

1.1.4 能够根据给定的系统软硬件环境,独立完成工业大数据软件分布式集群的安装与部署。

1.2 工业大数据软件安装测试

1.2.1 能够完成 linux 软件安装后的测试工作,验证软件安装部署的正确性;

1.2.2 能够完成大数据 hadoop 软件伪分布式安装后的测试,验证软件安装部署的正确性;

1.2.3 能够完成大数据 hadoop 软件分布式集群安装后的测试,验证软件安装部署的正确性。

1.3 工业大数据软件维护

1.2.1 能够使用工具监控工业大数据软件的运行状况;

1.2.2 能够根据工业大数据平台软件的运行状况,完成软硬件资源的调整;

1.2.3 能够根据业务的需求,调整节点的部署。

2. 关系型数据库应用开发

2.1 关系型数据库部署与管理

2.1.1 能够根据工业大数据业务场景的需求,选择合适的操作系统,独立安装 mysql 数据库;

2.1.2 能够使用适合的客户端软件远程访问数据库服务器;

2.1.3 能够根据业务需求,选择合适的存储引擎并创建数据库;

2.1.4 能够根据业务需求,创建用户并分配用户适当的权限。

2.2 关系型数据库的数据查询

2.2.1 能够根据业务需求,完成单表数据的条件查询和排序操作;

2.2.2 能够根据业务需求,完成多表连接查询;

2.2.3 能够根据业务需求,使用分组函数完成分组统计功能;

2.2.4 能够根据业务需求,使用子查询实现高级查询功能;

2.2.5 能够根据业务需求,使用常用函数完成复杂的查询功能。

2.3 数据操作与事务控制

2.3.1 能够根据业务需求,进行业务数据的增加操作;

2.3.2 能够根据业务需求,进行业务数据的修改操作;

2.3.3 能够根据业务需求,进行业务数据的删除操作;

2.3.4 能够根据业务需求,完成事务的控制,保证数据的一致性。

2.4 数据对象管理

2.4.1 能够根据业务需求,完成普通表的维护,并设置有效的数据约束;

2.4.2 能够根据业务需求,选择合适的分区表,实现大数据的分区管理;

2.4.3 能够根据业务需求,实现视图、索引对象的管理。

3. 工业大数据存储与运维

3.1 分布式文件系统 HDFS 数据操作

3.1.1 能够完成分布式文件系统的数据上传;

3.1.2 能够完成分布式文件系统的数据下载;

3.1.3 能够完成分布式文件系统的数据的查看。

3.2 分布式文件系统 HDFS 文件操作

3.2.1 能够完成 Hadoop 分布式文件系统文件所属权限的修改;

3.2.2 能够完成



Hadoop 分布式文件系统文件的删除、重命名、复制等操作； 3.2.3 能够完成分布式文件系统目录的创建。

3.3 分布式文件系统管理操作 3.3.1 能够监控分布式文件系统集群状态； 3.3.2 能够将分布式文件系统集群置于安全模式； 3.3.3 能够对分布式文件系统集群健康情况进行检测； 3.3.4 能够对分布式文件系统中的文件进行健康状况检测。

4. 工业数据采集与清洗 4.1 工业数据采集 4.1.1 能够根据不同的业务场景选择数据采集策略或工具； 4.1.2 能够使用工具采集工业网关数据； 4.1.3 能够独立使用工具实现 MES 系统数据库与大数据系统之间的数据迁移； 4.1.4 能够独立实现 MES 系统日志数据的采集。

4.2 工业数据清洗 4.2.1 能够根据业务需求，完成数据质量的评估； 4.2.2 能够根据业务需求，制定数据清洗的策略； 4.2.3 能够根据业务需求，运用数据清洗工具实现数据清洗。

5. 工业大数据相关工具应用 5.1 分布式数据库 HBase 应用 5.1.1 能够基于已有的 HBase 数据库实现分布式数据库的创建； 5.1.2 能够独立对 HBase 数据库进行数据查询操作； 5.1.3 能够独立对 HBase 数据库进行数据增加操作； 5.1.4 能够独立对 HBase 数据库进行数据删除操作。

5.2 分布式数据仓库 Hive 应用 5.2.1 能够基于已有的 hive 数据仓库实现分布式数据仓库的创建； 5.2.2 能够根据业务需求，实现基本外部表和内部表的创建； 5.2.3 能够根据业务需求，完成数据的加载。 5.2.4 能够根据业务需求，完成数据的查询。

5.3 内存数据库 Redis 应用 5.3.1 能够使用内存数据库实时操作工业数据，进行数据的插入操作； 5.3.2 能够使用内存数据库实时操作工业数据，进行数据的查询操作； 5.3.3 能够使用内存数据库实时操作工业数据，进行数据的删除操作； 5.3.4 能够选择数据库、监视实时转储收到的请求等操作。

6. 工业数据可视化 6.1 工业数据可视化方案设计 6.1.1 能够根据业务需求，针对不同的业务场景选择不同的图表展现形式； 6.1.2 能够根据数据可视化方案，选择适合的数据可视化工具； 6.1.3 能够根据业务需求，进行数据可视化整体方案的设计。

6.2 工业数据可视化实现 6.2.1 能够对选择的数据可视化工具进行基本的配置； 6.2.2 能够基于数据可视化方案设计，使用数据可视化工具制作单变量的柱状图、饼图、折线图等基本图形； 6.2.3 能够基于数据可视化方案设计，使用数据可视化工具制作多变量的散点图、气泡图、盒须图等图形。

6.3 工业数据可视化数据管理 6.3.1 能够根据业务需求，连接文件数据源； 6.3.2 能够根据业务需求，连接 Excel 数据源； 6.3.3 能够根据业务需求，连接 mysql 数据源； 6.3.4 能够根据业务需求，实现多表数据连接。根据业务需求，在业务主管的指导下，完成基本的宏观、行业、上市企业以及金融产品相关数据统计汇总、数据采编、数据清洗、数据结构化、数据存储、初级数据咨询服务等工作。

金融数据采集 1.1 数据合规认知 1.1.1 掌握宏观、行业、上市公司、金融产品常见数据指标，并能说出其代表的含义； 1.1.2 熟悉并严格遵守数据采集相关法律法规； 1.1.3 能够有效分析数据合规需求； 1.1.4 能够在合法的情况下进行数据库采集。

1.2 数据源获取 1.2.1 能够找到宏观、行业、上市企业以及金融产品数据源； 1.2.2 熟悉各数据源的数据披露规律； 1.2.3 能够快速定位到宏观行业、金融产品以及上市企业数据出处及数据内容在公告中的具体位置； 1.2.4 能够对来自多个数据源的数据进行有效性分析并做出选择。

1.3 金融数据手动采集 1.3.1 能够使用 PDF 编辑器、WPS 等工具将数据源获取到的资料转化为可编辑的数据； 1.3.2 通过质检工具对采集数据进行质检； 1.3.3 对质检结果发现的问题，能够进行数据修复； 1.3.4 对所采集数据做简单的准确性比对检测。

1.4 python 数据采集 1.4.1 熟悉互联网上常见数据终端种类； 1.4.2 掌握网站型互联网应用的常用开发语言、技术框架； 1.4.3 能够运用网页开发者工具来分析网页内容； 1.4.4 能够使用编译好的 python 采集模型，通过调参获取不同网站等互联网应用的静态数据并进行抽取。

2. 金融数据清洗及存储 2.1 数据清洗 2.1.1 具备使用 Excel 或 Python 工具进行脏数据的初步清洗，处理常见的异常数据； 2.1.2 具备使用 Excel 或 Python 工具进行数据规范； 2.1.3 能够完成数据分解；能够对网站等互联网应用的静态数据进行准确拆解； 2.1.4 能够保证拆分后

的数据属性都有明确的意义和目的。

**2.2 数据存储**

**2.2.1** 能够根据互联网应用数据的特点合理选择适当的数据结构；

**2.2.2** 熟悉常见数据库的特点；

**2.2.3** 能将数据保存为不同类型文件；

**2.2.4** 具备将清理后的数据存储到目标库或数据表的能力。

**2.3 数据查询**

**2.3.1** 能够正确安装 mysql 数据库；

**2.3.2** 掌握 SQL 基本数据查询命令的语法；

**2.3.3** 能够使用 SQL 数据库语句对所采集数据进行查询；

**2.3.4** 熟练使用 SQL 语句进行数据库错误排查。

**3. 金融数据应用**

**3.1 数据咨询**

**3.1.1** 能够根据咨询需求，规划要调研的数据指标，整理需求文档；

**3.1.2** 能说出金融数据的在不同业务场景中下的应用价值；

**3.1.3** 能够解答数据基础分析理论问题；

**3.1.4** 能够解答数据分析技术相关问题。

**3.2 数据可视化呈现**

**3.2.1** 能够对所需呈现的数据，选择合适的图表模型进行呈现；

**3.2.2** 能够说出常见的几类图表的优缺点和使用限制；

**3.2.3** 能够使用常见的可视化工具，如 excel、Tableau、云图等

进行快速绘图；

**3.2.4** 能够排版设计做出可视化图表报告。

**3.3 数据调研报告撰写**

**3.3.1** 能够对调研数据进行简单业务分析和规律描述；

**3.3.2** 知道常见个股调研报告所需的数据指标，能够编写个股调研图文报告；

**3.3.3** 知道常见行业调研报告所需的数据指标，能够编写行业调研图文报告；

**3.3.4** 知道常见企业经营调研报告所需的数据指标，能够编写企业经营调研图文报告。

主要面向物流行业相关企业或各大中型企事业单位的物流部门、运营或服务部门，从事仓储数据采集、仓储运营及设备基础分析、日常数据监控与报表制作等工作，具有从数据库进行基础数据采集、加工处理、监控与汇报等的能力。

**1. 仓储数据采集**

**1.1 仓储数据采集指标确定**

**1.1.1** 熟练掌握仓储数据分析流程，了解仓储数据分析需求，明确数据采集的目标。

**1.1.2** 熟练掌握仓储各类数据指标及其含义。

**1.1.3** 能够确定数据分析的采集指标项、采集范围以及采集注意事项。

**1.1.4** 能判断采集的仓储数据的内容和质量是否符合要求。

**1.2 仓储数据采集渠道和工具使用**

**1.2.1** 熟悉仓储数据主要来源渠道，确定仓储数据采集渠道，保证数据渠道的合法、有效。

**1.2.2** 熟悉数据采集及数据分析常用工具及其主要功能。

**1.2.3** 掌握 SQL 常用语句，能够灵活运用查询、聚合、关联等常用数据查询语句进行数据提取。

**1.2.4** 熟练使用 Excel，调用常用函数进行基本的数据分析。

**1.3 仓储数据采集**

**1.3.1** 能够根据数据采集方案的要求，快速准确采集入库、储存、拣货、分货等作业数据。

**1.3.2** 能够根据数据采集方案的要求，快速准确采集设备利用率、故障和维修记录等设备数据。

**1.3.3** 能够运用 Excel 对所采集的数据进行加工整理。

**2. 仓储运营及设备数据分析**

**2.1 进货和入库作业分析**

**2.1.1** 了解进货和入库作业的流程，熟悉进货和入库作业分析中各项数据指标的基本含义。

**2.1.2** 掌握计算进货和入库作业分析中各项指标的方法，能够根据任务的明细数据，计算各项指标值。

**2.1.3** 能够对关键指标进行描述统计，计算指标的平均值、最小值、最大值、方差、标准差等，并理解描述统计量的实际含义。

**2.1.4** 能够结合业务需求及统计分析的目的，选择适合的统计图描绘数据变化趋势、比例情况等。

**2.2 储存作业分析**

**2.2.1** 了解储存作业的流程，熟悉储存作业分析中各项数据指标的含义。

**2.2.2** 能够独立根据储存作业要求，对库存明细数据进行清洗、转换、排序等数据处理工作。

**2.2.3** 能够独立计算库存总量、库容率、周转速度等反映库存状态的基础指标。

**2.2.4** 能对储存作业数据指标进行描述性统计、趋势分析、同比环比分析以及频数频率分析。

**2.3 盘点作业分析**

**2.3.1** 了解盘点作业的流程，熟悉盘点作业分析中各项数据指标的含义。

**2.3.2** 能够独立根据盘点作业的要求，对盘点明细数据进行清洗、转换、排序等数据处理工作。

**2.3.3** 能够独立计算盘点工作量、盘点差异总量、盘点准确率等数据指标，使用饼图呈现盘点差异分布情况。

**2.3.4** 能够对盘点作业数据指标进行描述性统计、趋势分析、同比环比分析以及频数频率分析。

**2.4 拣货作业分析**

**2.4.1** 了解拣货作业的流程，熟悉拣货作业分析中各项数据指标的含义。

**2.4.2** 能够独立根据拣货作业的要求，对拣货明细数据进行清洗、转换、排序等数据处理工作。

**2.4.3** 能够独立计算拣货整体分析指标和效率指标，如平均每日拣货总单量、订单履约率、每小时出库容器数、每小时每工作站拣选件数等。

**2.4.4** 能

能够对拣货作业数据指标进行描述性统计、趋势分析、同比环比分析以及频数频率分析。 2.5 分货作业分析 2.5.1 了解分货作业的流程,熟悉分货作业分析中各项数据指标的含义。2.5.2 能够独立根据分货作业的要求,对日常采集的分货数据,按照作业流程、时间等字段进行数据分类和汇总。2.5.3 能够对分货明细数据进行清洗、转换、排序等数据处理工作并完成分货总量、分货能力、分货资源投入量以及细部作业时长等数据计算。2.5.4 能够对分货运营数据进行描述性统计、趋势分析、同比环比分析以及频数频率分析。 2.6 智能装备运行状态分析 2.6.1 了解仓库中常见的各种自动化设备及相应的作业场景,能够根据场景判断设备运营收集的数据类型和数据类目。2.6.2 能够根据设备真实性能分析方案,独立完成设备运行状态数据的统计及整理,计算设备运行状态数据的平均值、最大值、最小值和分位值等描述性统计值。2.6.3 能够绘制频率分布直方图并对分布进行初步估计。2.6.4 了解设备异常的定义、分类方法以及等级判定标准,能对异常数据进行描述性分析。2.6.5 能够按照不同的类别或不同的等级统计数据,进行描述性分析,并通过柱状图、饼图等进行可视化。 2.7 智能装备可靠性分析 2.7.1 能够独立根据设备可靠性的要求,对日常采集的物流设备数据按照设备类型、时间等字段进行数据分类和汇总。 2.7.2 能够独立对物流设备的任务数据进行清洗、转换、切分、拼接等数据处理工作。2.7.3 能够独立完成物流设备的寿命、失效率、MTBF、MTTR 以及可靠度等指标计算。2.7.4 能够熟练使用数据分析软件,对物流设备的可靠性数据进行描述性统计、趋势分析、同比环比分析以及频数频率分析等。

### 3. 仓储运营日常监控与报表编制

#### 3.1 仓储日常数据监控

##### 3.1.1 了解日常仓储运营中应重点监控的数据指标,对数据指标进行日常监控和记录。

##### 3.1.2 熟悉仓储运营和设备指标间关联关系与异常数据特点,并能够及时发现并反馈异常数据和数据变化幅度较大的指标。

##### 3.1.3 熟悉日常仓储运营数据指标异常的常见原因,根据数据监控情况对异常数据出现的原因进行初步分析,及时发现并处理运营、系统或硬件层面的问题。

##### 3.1.4 具备异常数据敏感性和较好的沟通能力,以及一定的问题分析、处理能力。

#### 3.2 仓储日常数据图表制作

##### 3.2.1 熟悉条形图、柱形图、折线图、饼图、散点图等各类图表及其内涵。

##### 3.2.2 能够结合图表的选择原则,依据物流仓储日常运营和设备数据情况,选择合适的图表类型。

##### 3.2.3 能运用数据图表制作方法,熟练使用 Excel 表格处理软件中图表制作功能,高质高效地制作物流仓储数据图表。

##### 3.2.4 能运用数据图表美化设计技巧,优化仓储数据图表的展示效果。

#### 3.3 仓储日常数据报表编制

##### 3.3.1 熟悉仓储日常数据报表及其内涵。

##### 3.3.2 能够依据物流日常运营数据汇报的需求,熟练运用数据报表的制作方法,搭建科学严谨的数据报表框架。

##### 3.3.3 能够依据物流日常运营数据汇报的需求,制作满足需求的日常运营数据报表,如日、周、月度和年度的数据报表。

##### 3.3.4 根据数据报表结构逻辑分析,能够手动制作数据报表。根据业务需求,在业务主管的指导下,完成基本的 数据咨询、数据分析、数据可视化等工作。

### 1. 大数据咨询管理

#### 1.1 大数据技术咨询处理

##### 1.1.1 能够在业务主管的指导下解答客户关于大数据基础理论相关的简单问题。

##### 1.1.2 能够在业务主管的指导下解答客户关于 大数据分析技术相关的简单问题。

##### 1.1.3 能够在业务主管的指导下解答客户关于 大数据可视化技术相关的简单问题。

##### 1.1.4 能够在业务主管的指导下根据实际业务 情况,输出大数据技术说明文档。

#### 1.2 大数据平台咨询处理

##### 1.2.1 能够在业务主管的指导下根据客户需求 推荐合适的大数据平台及方案。

##### 1.2.2 能够在业务主管的指导下根据客户需求 推荐合适的大数据存储计算产品。

##### 1.2.3 能够在业务主管的指导下根据客户需求 推荐合适的大数据可视化产品。

##### 1.2.4 能够在业务主管的指导下解答客户关于 大数据平台使用相关的基础问题。

#### 1.3 大数据平台管理

##### 1.3.1 能够在业务主管的指导下,根据大数据平台使用手册开通和配置大数据计算平台。

##### 1.3.2 能够在业务主管的指导下,根据大数据平台使用手册开通和配置大数据可视化服务。

##### 1.3.3 能够在业务主管的指导下,根据大数据平台使用手册对大数据计算平台进行简单的操作。

##### 1.3.4 能够在业务主管的指导下,根据大数据平台使用手册对大

数据可视化服务进行简单的操作。

## 2、数据分析管理

### 2.1 数据质量管理

#### 2.1.1 能够在业务主管的指导下，根据业务需求及数据质量标准要求，对企业数据进行完整性评估。

#### 2.1.2 能够在业务主管的指导下，根据业务需求及数据质量标准要求，对企业数据进行规范性评估。

#### 2.1.3 能够在业务主管的指导下，根据业务需求及数据质量标准要求，对企业数据进行一致性评估。

#### 2.1.4 能够在业务主管的指导下，根据业务需求及数据质量标准要求，对企业数据进行准确性评估。

### 2.2 数据离线分析

#### 2.2.1 能够在业务主管的指导下，根据业务需求，使用 SQL 进行基本的数据清洗。

#### 2.2.2 能够在业务主管的指导下，根据业务需求，使用 SQL 进行基本的数据离线分析处理。

#### 2.2.3 能够在业务主管的指导下，使用大数据平台实现数据库与大数据平台之间的数据迁移。

#### 2.2.4 能够在业务主管的指导下，根据业务需求，使用大数据平台实现基本的数据离线分析处理。

### 10 2.3 数据仓库搭建

#### 2.3.1 能够在业务主管的指导下进行数据仓库方案设计。

#### 2.3.2 能够在业务主管的指导下，基于数据仓库方案设计，选择合适的大数据产品进行数据仓库搭建。

#### 2.3.3 能够在业务主管的指导下使用数据仓库进行数据的 ETL（抽取、转化、加载）操作。

#### 2.3.4 能够在业务主管的指导下使用数据仓库进行数据的简单处理。

## 3. 数据可视化

### 3.1 数据可视化方案设计

#### 3.1.1 能够在业务主管的指导下，对数据进行简单的预处理操作。

#### 3.1.2 能够在业务主管的指导下，选择不同的图表来展现数据。

#### 3.1.3 能够在业务主管的指导下，根据业务数据展示要求，选择合适的数据可视化工具。

#### 3.1.4 能够在业务主管的指导下，根据业务需求，进行简单的数据可视化方案设计。

### 3.2 数据可视化处理

#### 3.2.1 能够在业务主管的指导下，根据业务需求使用数据可视化工具对数据进行简单的配置和操作。

#### 3.2.2 能够在业务主管的指导下，使用数据可视化工具将数据以图表的形式进行展示。

#### 3.2.3 能够在业务主管的指导下，调整可视化图表的布局和显示内容。

#### 3.2.4 能够在业务主管的指导下，根据数据可视化方案和分析指标，完成 BI 报表中的交互体验优化。

### 3.3 分析报告撰写

#### 3.3.1 能够在业务主管的指导下，导出数据可视化结果，并进行简单的归纳整理。

#### 3.3.2 能够在业务主管的指导下，将数据可视化结果整理成初步的数据分析报告。

#### 3.3.3 能够在业务主管的指导下，对数据分析报告进行排版整理。

#### 3.3.4 能够在业务主管的指导下，根据数据分析可视化结果，发现业务中存在的问题，并体现在数据分析报告中。初步掌握大数据思维，具备运用基础数据技

## 78 能在工作中从事数据初始准备、数据收集整理、数据工具选择、财务与业务应用场景辨识的能力，并具备大数据基本职业操守。

## 1. 大数据初始与基础工作

### 1.1 组织内外大数据类型与量级辨识

#### 1.1.1 理解什么是大数据

#### 1.1.2 了解大数据的不同量级

### 1.2 财务工作中的大数据问题识别

#### 1.2.1 了解财务职能与大数据的关系

#### 1.2.2 识别财务领域的大数据问题

### 1.3 大数据分析分类

#### 1.3.1 了解大数据相关术语

#### 1.3.2 辨识大数据分析类型

#### 1.3.3 理解大数据的价值

## 2. 大数据收集与辨识

### 2.1 大数据在财务及中小企业的应用

#### 2.1.1 从财务人员的视角理解大数据

#### 2.1.2 从财务人员的视角理解数据发展史

#### 2.1.3 从中小企业的视角理解大数据

### 2.2 大数据收集与辨识工作

#### 2.2.1 按来源收集大数据

#### 2.2.2 根据特征辨识大数据

#### 2.2.3 大数据的应用举例：零售调查

### 3. 大数据前瞻与趋势调研

#### 3.1 大数据发展趋势追踪

#### 3.1.1 了解大数据的最新发展趋势

#### 3.1.2 了解大数据领域的专业研究机构

### 3.2 大数据调研报告查询

#### 3.2.1 了解高德纳公司（Gartner）

#### 3.2.2 了解埃森哲公司（Accenture）

#### 3.2.3 了解 NewVantage 公司

## 4. 大数据的业务应用：平台与工具选择

### 4.1 大数据业务场景化应用

#### 4.1.1 理解大数据的应用目标

#### 4.1.2 了解大数据的业务应用场景

### 4.2 大数据应用风险识别

#### 4.2.1 识别大数据的业务应用挑战

#### 4.2.2 了解错误数据带来的风险

#### 4.2.3 知晓 IT 领域的大数据常见错误

### 4.3 大数据应用能力评价

#### 4.3.1 评价组织的大数据能力

#### 4.3.2 了解大数据的应用平台

## 4.4 大数据分析平台与工具了解

#### 4.4.1 知晓主流业务数据分析工具，如 Tableau

#### 4.4.2 了解大数据开源框架 Hadoop、地图还原工具与 R 语言

## 5. 大数据的财务应用：场景与工具选择

### 5.1 大数据现实应用举例

#### 5.1.1 提供大数据的

现实应用举例 5.1.2 从电子商城案例中接触实际企业数据 5.1.3 了解各行业的大数据应用现状 5.1.4 识别大数据的现实应用场景 5.2 利用大数据进行定价优化 5.2.1 使用大数据进行优化 5.2.2 利用大数据优化邮轮线路定价的案例 5.3 各类大数据应用工具解析 5.3.1 了解预测型数据分析工具 5.3.2 了解其他大数据读取工具 5.3.3 了解数据库工具 5.3.4 了解大数据可视化工具 5.4 大数据财务应用场景辨析 5.4.1 从财务人员的视角了解大数据 5.4.2 了解大数据的各类财务应用场景 5.5 财务工作中的大数据应用 5.5.1 在财务工作中应用预测型数据分析 5.5.2 了解证券监管机构的数据分析应用 6. 大数据操守与安全问题考量 6.1 大数据应用合规 6.1.1 辨析大数据应用中的伦理问题 6.1.2 知晓大数据滥用的影响与后果 6.1.3 了解大数据失范行为的实例 6.2 大数据安全管理 6.2.1 解读大数据安全与隐私问题 6.2.2 了解位置追踪与全域视频监控主要面向大数据平台安装配置、大数据组件安装配置、大数据平台基础实施、大数据平台简单维护及监控工作岗位。从事虚拟化软件安装与使用、基于 Linux 系统的常用服务安装配置、安装配置及运行 Hadoop 集群、安装配置及运行核心组件、执行客户大数据平台实施方案、监控大数据平台运行状态、查看管理大数据平台日志信息、监控大数据平台服务和资源状态等工作，掌握大数据平台安装和配置方法，理解 Hadoop 核心组件的功能及工作原理，掌握关键组件安装配置方法，理解大数据平台实施流程，熟悉常用集群监控工具的使用方法。

1. 大数据平台安装 1.1 能安装和使用虚拟化软件 1.1.1 能安装虚拟化软件 1.1.2 能使用虚拟化软件 1.2 能使用虚拟化软件安装 Linux 操作系统 1.2.1 能使用虚拟化软件安装 Linux 操作系统 1.3 能安装 Linux SSH 服务 1.3.1 能下载 Linux SSH 服务 1.3.2 能安装 Linux SSH 服务 1.4 能安装 Linux JDK 1.4.1 能熟练安装 Linux JDK 1.4.2 能熟练配置 JDK 环境变量 1.5 能在 Linux 系统中安装 Hadoop 1.5.1 能熟练下载 Hadoop 安装包 1.5.2 能熟练安装完全分布式模式 Hadoop

2. 大数据平台配置 2.1 能配置主机的网络属性 2.1.1 能熟练配置主机网络 IP 2.1.2 能熟练配置主机网络 DNS 2.1.3 能熟练配置主机名 2.2 能配置集群的网络连通 2.2.1 能配置集群局域网络连通 2.3 能配置集群主机之间 SSH 免密登录 2.3.1 能正确生成 SSH 密钥对 2.3.2 能正确配置 SSH 认证文件 2.4 能配置 Hadoop 文件参数 2.4.1 能正确配置 hdfs.site.xml 文件参数 2.4.2 能正确配置 core.site.xml 文件参数 2.4.3 能正确配置 mapred.site.xml 文件参数 2.4.4 能正确配置 yarn.site.xml 文件参数 2.5 能启停 Hadoop 集群 2.5.1 能正确启动和停止 Hadoop 集群 2.5.2 能查看 Hadoop 集群启动进程

3. 大数据平台组件安装配置 3.1 能安装配置 HBase 组件 3.1.1 能熟练安装 HBase 组件 3.1.2 能熟练配置 HBase 组件 3.2 能安装配置 Hive 组件 3.2.1 能熟练安装 Hive 组件 3.2.2 能熟练配置 Hive 组件 3.3 能安装配置 Zookeeper 组件 3.3.1 能熟练安装 Zookeeper 组件 3.3.2 能熟练配置 Zookeeper 组件 3.4 能安装配置 Sqoop 组件 3.4.1 能熟练安装 Sqoop 组件 3.4.2 能熟练配置 Sqoop 组件 3.5 能安装配置 Flume 组件 3.5.1 能熟练安装 Flume 组件 3.5.2 能熟练配置 Flume 组件

4. 大数据平台实施 4.1 能理解客户的大大数据平台实施方案 4.1.1 能正确理解客户需求 4.1.2 能理解客户大数据平台实施方案 4.2 能执行客户大数据平台实施方案 4.2.1 能按要求正确执行客户大数据平台实施方案 4.3 能制定客户培训方案 4.3.1 能使用文档制作工具 4.3.2 能制定客户培训方案 4.4 能培训客户使用大数据平台 4.4.1 能操作客户大数据平台 4.4.2 能培训客户使用大数据平台 4.5 能解决培训中大数据平台出现的问题 4.5.1 能解决培训中大数据平台出现的问题

5. 大数据平台监控 5.1 能监控大数据平台的运行状态 5.1.1 能熟练使用集群监控工具监控大数据平台的运行状态 5.2 能监控大数据平台的资源状态 5.2.1 能熟练使用集群监控工具监控大数据平台的资源状态 5.3 能查看大数据平台的告警信息 5.3.1 能熟练使用集群监控工具查看大数据平台的告警信息 5.4 能查看大数据平台的服务状态 5.4.1 能熟练使用集群监控工具查看大数据平台的服务状态 5.5 能查看大数据平台的日志信息 5.5.1 能

熟练使用集群监控工具查看大数据平台的日志信息"