
阿里云行业认证：大数据分析师 (ACP 级) 认证考试大纲

阿里云大数据行业认证-大数据分析师认证介绍：

阿里云大数据分析师认证 (Alibaba Cloud Certified Professional-Data Analyst, ACP) 是阿里云大数据行业认证体系中的职业认证, 同时也是一个跨平台、通用型专业技术认证。主要包括数据分析相关的知识体系, 如数据库知识、数据质量控制、数据编程、机器学习、数据分析工具、机器学习、数据可视化, 主流大数据技术等; 介绍了数据分析在行业中的实际应用与项目管理方法, 及相关的技术技能和技能, 包括 8 个知识与技能模块: 大数据基础知识、大数据存储技术、数据分析工具、数据可视化、数据编程、数据项目质量控制、数据项目设计与执行、机器学习。通过该行业技术认证可以有效证明持证者具备以下能力:

- 具备大数据相关的基础知识
- 了解大数据分析职业的特点及行业对大数据分析人员的职业要求
- 了解大数据存储技术的特点, 能够熟练使用传统关系型数据库, 了解数据仓库的基本知识, 能够使用开源大数据技术、阿里云数加等管理和使用数据
- 掌握 SQL 语言编程技能, 能够根据项目需要进行数据库管理和数据编程
- 熟练掌握数据可视化相关工具, 如 Quick BI、DataV 等, 并且能设计与开发可视化大屏和商业报表
- 掌握数据质量管理的特点和要求, 能够在数据分析中判断数据质量对项目的影晌并提供相应解决方案

- 掌握数据分析的质量控制流程，利用数据预处理技术合理处理脏数据
- 基于对数据分析项目的编程方法，保证程序的运行效率和数据分析结果的质量
- 能够运用 5W2H 分析法梳理数据项目的目标、范围，根据对业务要求的理解设计合理的数据分析方案
- 掌握机器学习技术的使用和应用场景，如聚类分析、决策树、关联分析等
- 能够独立撰写数据分析项目报告

阿里云大数据分析师专业认证的报名方式：

通过阿里云官网提供的入口进行在线报名：

<https://edu.aliyun.com/certification/acp08>

阿里云大数据分析师认证所需具备的前置知识：

通用 IT 的知识：

- 具备基础的 IT 知识，熟练使用 Windows、Mac OS、Linux 等操作系统中的至少一种
- 了解大数据相关的基础知识，如定义、特征、实际应用案例等
- 了解关系型数据库的基本概念：数据库，表，索引，视图，存储过程，函数等
- 了解云计算、开源大数据 Hadoop 生态圈的主要产品、阿里云大数据相关的主要产品和服务，包括 MaxCompute、RDS、DataWorks、QuickBI，PAI、DataV 等
- 了解软件工程的基本流程

阿里云大数据分析师认证相关的学习方法、学习资料及培训课程：

建议的学习方法：

- 强烈推荐开通阿里云大数据相关的产品并进行实际的操作
- 结合阿里云官网的文档同步学习

<https://help.aliyun.com>

更多学习途径：

- 参加阿里云官方提供的面授培训课程，可以帮助您快速的学习阿里云弹性计算相关产品的概念、使用、操作及解决方案。相关的课程链接：

<https://www.aliyun.com/support/courses>

- 通过阿里云的自助实验平台开放实验室进行在线的动手实验练习

<https://edu.aliyun.com/lab>

- 通过阿里云官网的论坛进行学习、提问和交流 <https://bbs.aliyun.com>

认证考试形式和试卷结构：

一、试卷满分及考试时间

试卷满分为 100 分，70 分为通过，考试时间为 120 分钟

二、考试形式

线下上机考试

三、答题方式

闭卷

四、试卷内容

产品名称	试题比例
大数据基础知识	10%
大数据存储技术	10%
数据分析工具	10%
数据可视化	10%
数据编程	25%
数据项目质量控制	15%
数据项目设计与执行	10%
机器学习	10%

五、试卷题型

题目类型	题目数量	分值
单选题	50 题	每题 1 分
多选题	30 题	每题 1 分
判断题	20 题	每题 1 分

六、试卷内容范围

- 大数据基础知识考试内容：
 - ✓ 了解大数据的定义、特点等
 - ✓ 了解数据的类型和不同的分析处理方法
 - ✓ 了解大数据相关的概念、实际的应用案例、适用的场景等
 - ✓ 了解云计算的特点、云计算与大数据的关系

-
- ✓ 了解大数据相关的技术，如存储、计算、分析等
 - ✓ 了解大数据职业的特点与对人才的要求
 - 大数据存储技术考试内容
 - ✓ 了解分布式存储技术的概念与特点
 - ✓ 了解数据存储技术适用的不同场景，包括数据类型（如结构化、半结构化、非结构化数据）、数据容量、使用场景等
 - ✓ 了解数据库的基本概念与特点，包括可靠性、约束、三范式、适用场景等
 - ✓ 了解数据仓库的基本概念与特点，包括与数据库的区别、ETL 等
 - ✓ 了解 HDFS 与 MaxCompute 的构成与特点
 - ✓ 了解文件存储、数据库存储、分布式存储之间的优缺点
 - ✓ 掌握大数据计算服务的数据上传和下载，可以熟练使用 MySQL、HDFS、MaxCompute 等进行数据存储
 - ✓ 了解 Hadoop、MaxCompute 等产品的基本概念与特点，包括应用场景和局限性
 - 数据分析工具考试内容
 - ✓ 掌握大数据计算服务的 SQL 命令，包括 DDL、DML 以及常见内置函数
 - ✓ 了解 MapReduce 的基本概念与特点
 - ✓ 能够使用 DataIDE 的数据开发模块进行设计开发，包括建表、任务开发、数据上传等
 - ✓ 能够使用 MySQL、MaxCompute、Hive 平台进行数据分析

- 数据可视化考试内容
 - ✓ 了解数据可视化的基本知识，如定义、特点、实现方式等
 - ✓ 了解 Quick BI、DataV 的产品特点和使用场景
 - ✓ 了解常见图表类型的特点和适用场景
 - ✓ 能够使用 Quick BI 设计开发报表和门户
 - ✓ 了解可视化产品的分类和基本设计原则

- 数据编程考试内容
 - ✓ 掌握数据预处理的基本方法
 - ✓ 了解描述性统计分析的概念和特点，包括常见统计量、概率分布、拟合与检验
 - ✓ 了解假设检验的概念和特点，能够根据应用场合正确使用正态分布单样本和双样本和二项分布假设检验
 - ✓ 能够基于项目的目标与范围规划数据分析方案，设计合理的指标
 - ✓ 了解数据分析编程的特点，包括编程手法、编程效率、编程规范和质量控制
 - ✓ 了解指标体系的概念，包括总量指标、结构指标和平均指标，并能够合理应用达到数据分析目的
 - ✓ 掌握数据分析报告撰写的规范

- 数据项目质量控制考试内容
 - ✓ 了解数据质量的 5 个维度的概念和特点
 - ✓ 了解在数据质量的 5 个维度基础上，脏数据的种类、来源、造成的影响

-
- ✓ 掌握处理脏数据的方法，包括对脏数据的检查、修复、清洗、转换等
 - ✓ 了解数据质量问题在数据编程过程中发生的原因，并能够利用质量检验的技术手段保证项目的顺利执行
 - ✓ 了解项目的目标是数据分析项目中衡量数据质量的主要标准，并能够判断数据中的质量问题是否对数据分析项目产生影响
- 数据项目设计与执行考试内容
 - ✓ 了解项目工程管理方法论的定义、特点和实际应用场景
 - ✓ 能够理解数据分析项目的实施流程、重点环节、数据项目执行流程的重要性
 - ✓ 掌握项目设计的特点，包括业务问题数据化、明确项目的目的、范围、和分析维度等
 - ✓ 能够利用项目绩效分析实现项目后数据分析
 - ✓ 能够利用现状、原因、预测的分析方法实现项目前数据分析
 - ✓ 了解临时性项目与经常性项目之间的区别
 - 机器学习考试内容
 - ✓ 了解机器学习常见的算法，如聚类、决策树、关联分析等
 - ✓ 了解机器学习的常见使用流程，包括算法调优和效果评估
 - ✓ 能够使用聚类分析，包括 K-means 算法对相似的顾客分类
 - ✓ 能够使用决策树算法生成商业规则
 - ✓ 能够使用关联分析实现购物篮分析