

阿里云大数据专业认证（ACP 级）考试大纲

阿里云大数据专业认证介绍：

阿里云大数据专业认证（ACP 级 - Alibaba Cloud Certification Professional）是面向使用阿里云大数据产品的架构、开发、运维类人员的专业技术认证，主要涉及阿里云大数据类的几款核心产品，包括大数据计算服务（MaxCompute，原 ODPS）、Data IDE（原 BASE）、数据集成（原 CDP 云道）等。通过该技术认证可以有效证明该认证人员具备以下能力：

- 具备大数据的存储、处理以及分析的基础知识
- 能够根据企业的业务需求，基于阿里云的产品制定有效的技术解决方案和企业最佳实践
- 能够熟练的使用阿里云的大数据计算服务、Data IDE 以及数据集成等产品
- 能够诊断基于阿里云大数据产品构建的业务系统在运行中出现的常见问题并找到相应的解决方案

阿里云大数据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://market.aliyun.com/products/57252001/cmgi012944.html?spm=5176.7930894.374513.7.J5Ui9T#sku=yuncode694400000>)

阿里云大数据专业认证所需具备的知识：

阿里云相关的知识：

- 熟悉阿里云大数据相关产品的基本概念，包括大数据计算服务（MaxCompute，原 ODPS）、Data IDE（原 BASE）、数据集成（原 CDP 云道）等（下同）
- 了解阿里云大数据相关产品的特点、应用场景以及与其他云产品配合使用，包括表格存储（Table Store，原 OTS）、云数据库 RDS、分布式

关系型数据库 DRDS、分析型数据库（Analytic DB，原 ADS）等，能根据实际业务场景进行基础的架构设计

- 掌握阿里云大数据的相关产品的基本操作，包括通过管理控制台、客户端等进行交互式操作，通过产品提供的方式（如 SQL、MapReduce、UDF、API 或 SDK 等）进行数据处理
- 了解各产品的最佳使用实践，能识别使用过程中常见的问题，且能给出合适的解决方案

通用 IT 的知识：

- 了解大数据相关的基础概念以及相关知识，如分布式存储、分布式计算、数据仓库、数据分析等
- 了解关系型数据库的基本概念和常用技术，包括事务、索引、OLTP、建模方法、SQL（DDL、DML）等
- 具备使用关系型数据库或者其他产品进行数据处理的经验
- 了解数据仓库的基本概念和常用技术，包括维度、事实表、即席查询、数据挖掘、OLAP、星型模型、ETL 等
- 具备一定的编程经验，能使用至少一种常见的编程语言如 Java、Python 等进行简单的程序开发
- 了解软件开发生命周期，了解项目的建设或者项目管理

阿里云大数据专业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。相关的课程如下：

- ✓ ACP61003 全面理解阿里云上的数据处理和分析
(<http://www.aliyunedu.net/pingnan/front/courseinfo.php?spm=5176.7930894.374513.11.J5Ui9T&id=5>)
- ✓ ACP61005 全面理解阿里云安全防护产品-云盾
(<http://www.aliyunedu.net/pingnan/front/courseinfo.php?spm=5176.7930894.374513.12.J5Ui9T&id=8>)
- 通过阿里云的自助实验平台云中沙箱进行在线的动手实验练习
(<https://www.aliyunedu.net/>)
- 通过阿里云的在线培训学习平台进行阿里云大数据相关的线上课程的学习
(<http://www.chinamoocs.com/>)
- 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，考试时间为 120 分钟

二、考试形式

在线考试

三、答题方式

闭卷

四、试卷内容

| 产品名称 | 试题占比 |
|----------------|------|
| 大数据计算服务 | 60% |
| Data IDE | 25% |
| 数据集成 | 10% |
| 和其他云产品配合的应用及架构 | 5% |

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

- 大数据计算服务考试内容：
 - ✓ 熟悉大数据计算服务基本概念，包括项目空间、表、分区、资源、任务等
 - ✓ 了解大数据计算服务的组成架构和各组成部分功能
 - ✓ 掌握大数据计算服务的特点、优势以及适用场景
 - ✓ 掌握大数据计算服务的连接和使用方式，包括使用客户端 odpscmd、管理控制台、Java SDK 等
 - ✓ 掌握大数据计算服务的数据上传和下载，可以熟练的使用 tunnel 命令行工具，了解 Tunnel SDK
 - ✓ 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
 - ✓ 熟悉大数据计算服务的用户自定义函数，包括 UDF、UDAF 以及 UDTF，可以编写简单的自定义函数
 - ✓ 熟悉大数据计算服务的 MapReduce 编程框架，可以配置 eclipse 的环境，编写简单的 MR 程序
 - ✓ 了解大数据计算服务的 Graph 编程框架，包括基本概念、处理流程等，可以编写简单的 Graph 程序
 - ✓ 了解大数据计算服务 DataHub 的相关概念和使用方法

- ✓ 熟悉大数据计算服务的安全和权限管理的概念和实际操作，包括用户、角色、授权（ACL&Policy）、项目空间保护、例外以及安全等级等
- Data IDE 考试内容：
 - ✓ 熟悉 Data IDE 的基本功能模块，包括数据开发、数据管理、运维中心、组织管理以及项目管理等
 - ✓ 了解 Data IDE 的基本特点，包括角色隔离、环境隔离等
 - ✓ 可以使用项目管理和组织管理模块搭建环境
 - ✓ 熟练使用 Data IDE 的数据开发模块进行设计开发，包括建表、任务开发、资源上传、数据上传、新增函数等
 - ✓ 熟练使用 Data IDE 的数据开发模块进行工作流任务和节点任务的开发设计，并且可以配置合适的依赖和周期性调度
 - ✓ 熟练使用数据管理模块进行数据管理，包括血缘分析、表的使用权限申请和授权等
 - ✓ 对于使用过程中出现的问题，能够识别、定位，对其中基本的问题能进行修复
- 数据集成考试内容：
 - ✓ 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 DataX 组件，包括概念、功能以及实际使用
 - ✓ 了解并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，比如切分键配置等
 - ✓ 能够对使用过程中出现的常见问题进行排查，并能解决其中的基本问题

- ✓ 对于数据集成中的脏数据能够进行正确的处理
- 应用和架构考试内容：
 - ✓ 了解其他相关云产品的特点和应用场景，包括云数据库 RDS、分布式关系型数据库 DRDS、表格存储（Table Store，原 OTS）、分析型数据库（Analytic DB，原 ADS）等
 - ✓ 了解大数据计算服务、Data IDE 以及数据集成等如何与其他相关产品配合使用
 - ✓ 能根据实际的应用场景，结合阿里云产品的特点设计合理的基础架构