

作业 2：查阅资料，针对国内阿里云技术解决方案，分析描述其云计算技术平台架构，飞天操作系统，大数据计算框架及数据中台技术。

一、阿里云云计算技术平台架构

阿里云的一级体系架构图如图 1 所示，自下而上依次为阿里云数据中心、Linux 操作系统和大规模分布式计算系统（飞天），飞天操作系统包括分布式文件系统、任务调度等在内的八个内核模块，对外提供弹性计算服务、开放存储服务在内的六个开放服务。在此基础上基于 ACE 云引擎实现应用及其他第三方服务。

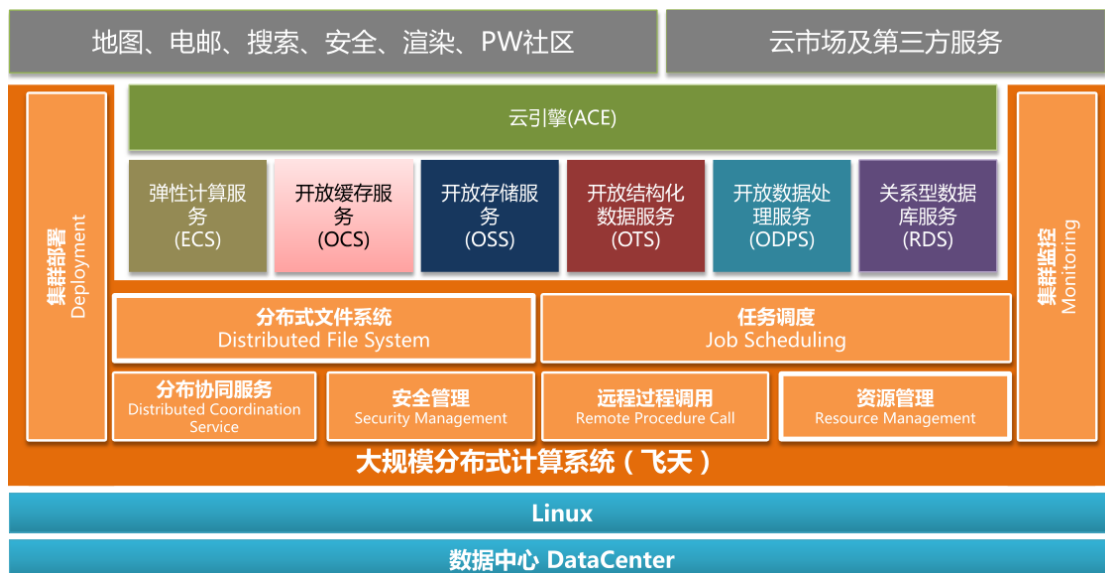


图 1 阿里云一级体系架构图

二、飞天操作系统

飞天是由阿里云开发的一个大规模分布式计算系统,其中包括飞天内核和飞天开放服务。

飞天内核负责管理数据中心 Linux 集群的物理资源,控制分布式程序运行,隐藏下层故障恢复和数据冗余等细节,有效提供弹性计算和负载均衡。

飞天开放服务为用户应用程序提供了计算和存储两方面的接口和服务,包括弹性计算服务 (Elastic ComputeService, 简称 ECS)、开放存储服务 (Open Storage Service, 简称 OSS)、开放结构化数据服务 (Open Table Service, 简称 OTS)、关系型数据库服务 (Relational Database Service, 简称 RDS) 和开放数据处理服务 (Open Data Processing Service, 简称 ODPS),并基于弹性计算服务提供了云服务引擎 (Aliyun Cloud Engine, 简称 ACE) 作为第三方应用开发和 Web 应用运行和托管的平台。

飞天操作系统的体系架构如图 2 所示。橙色部分为飞天内核,蓝色部分为飞天开放服务。



图 2 飞天操作系统体系架构图

飞天内核包括八个部分,其中伏羲包括资源调度和任务管理两个部分:

1. 协调服务 (女娲)。为飞天提供高可用的协调服务 (Coordination Service),是构

建各类分布式应用的核心服务，它的作用是采用类似文件系统的树形命名空间来让分布式进程互相协同工作。

2. 远程过程调用 (夸父)。是飞天平台内核中负责网络通信的模块，它提供了一个 RPC 的接口，简化编写基于网络的分布式应用。
3. 安全管理 (钟馗)。是飞天平台内核中负责安全管理的模块，它提供了以用户为单位的身份认证和授权，以及对集群数据资源和服务进行的访问控制。
4. 分布式文件系统 (盘古)。是一个分布式文件系统，盘古系统的设计目标是将大量通用机器的存储资源聚合在一起，为用户提供大规模、高可靠、高可用、高吞吐量和可扩展的存储服务，是飞天平台内核中的一个重要组成部分。
5. 资源管理和任务调度 (伏羲)。是飞天平台内核中负责资源管理和任务调度的模块，同时也为应用开发提供了一套编程基础框架。伏羲同时支持强调响应速度的在线服务和强调处理数据吞吐量的离线任务。
6. 集群监控 (神农)。是飞天平台内核中负责信息收集、监控和诊断的模块。它通过在每台物理机器上部署轻量级的信息采集模块，获取各个机器的操作系统与应用软件运行状态，监控集群中的故障，并通过分析引擎对整个飞天的运行状态进行评估。
7. 集群部署 (大禹)。是飞天内核中负责提供配置管理和部署的模块，它包括一套为集群的运维人员提供的完整工具集，功能涵盖了集群配置信息的集中管理、集群的自动化部署、集群的在线升级、集群扩容、集群缩容，以及为其他模块提供集群基本信息等。

三、大数据计算框架

阿里云 MaxCompute 大数据计算服务是一种快速、完全托管的 TB/PB 级数据仓库解决方案。随着数据收集手段不断丰富，行业数据大量积累，数据规模已增长到了传统软件行业无法承载的海量数据（TB、PB、EB）级别。MaxCompute 致力于批量结构化数据的存储和计算，提供海量数据仓库的解决方案及分析建模服务。

由于单台服务器的处理能力有限，海量数据的分析需要分布式的计算模型。分布式的计算模型对数据分析人员要求较高且不易维护。数据分析人员不仅需要了解业务需求，同时还需要熟悉底层分布式计算模型。MaxCompute 为客户提供完善的数据导入方案以及多种经典的分布式计算模型，客户可以不必关心分布式计算和维护细节，便可轻松完成大数据分析。

MaxCompute 的系统架构如图 3 所示，以数据为中心，内建多种计算模型和服务接口，满足广泛的数据分析需求。一切服务“开通”即用，更好地赋能数据业务。

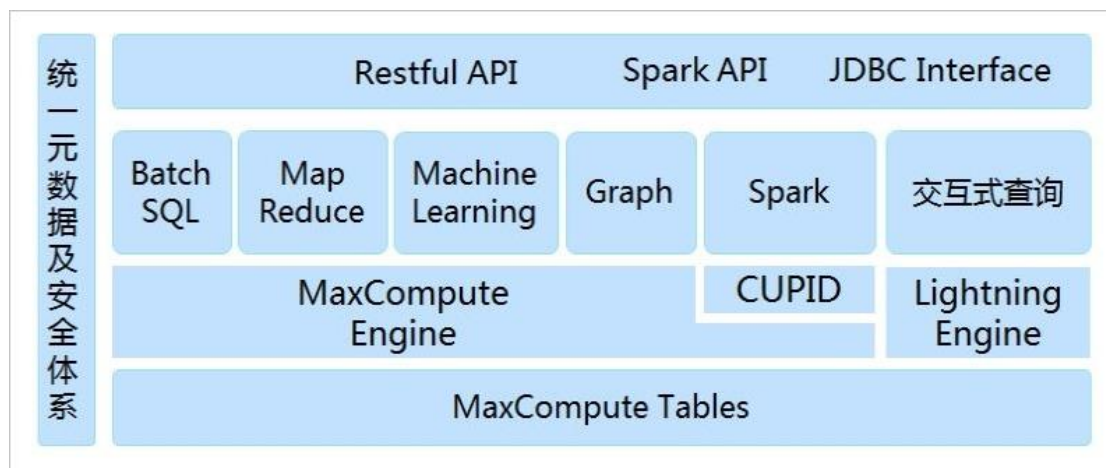


图 3 MaxCompute 系统架构

MaxCompute 系统自下而上依次为 MaxCompute 数据表，MaxCompute 引擎和 CUPID 或轻量级引擎，数据库查询系统，和数据库接口系统。整个 MaxCompute 系统由统一元数据及安全体系所支撑保障。

MaxCompute 可以与其它阿里云服务集成使用，发挥强大效用：

1. 数据分析平台：DataWorks 是基于 MaxCompute 计算和存储，提供工作流可视化开发、调度运维托管的一站式海量数据离线加工分析平台。MaxCompute 和 DataWorks 一起向用户提供完善的 ETL 和数仓管理能力，以及 SQL、MR、Graph 等多种经典的分布式计算模型，能够更快速地解决用户海量数据计算问题，有效降低企业成本，保障数据安全。
2. 数据集成：MaxCompute 可以通过数据集成加载不同数据源（例如：MySQL 数据库等）数据，同样也可以通过数据集成把 MaxCompute 的数据导出到各种业务数据库。
3. 机器学习：机器学习 PAI 是基于 MaxCompute 的一款机器学习算法平台。它实现了数据无需搬迁，便可进行从数据处理、模型训练、服务部署到预测的一站式机器学习。创建 MaxCompute 项目，开通机器学习，即可通过机器学习平台的算法组件对 MaxCompute 数据进行模型训练等操作。
4. 数据可视化：数据在 MaxCompute 进行加工处理后，将 Project 添加为 QuickBI 数据源，即可在 QuickBI 页面对 MaxCompute 表数据进行报表制作，实现数据可视化分析。
5. 数据挖掘：AnalyticDB for MySQL 是海量数据实时高并发在线分析（Realtime OLAP）的云计算服务，与 MaxCompute 结合实现大数据驱动业务系统的场景。通过 MaxCompute 离线计算挖掘，产出高质量数据后，导入分析型数据库，供业务系统调用分析。
6. 云存储数据：对象存储 OSS 是海量、安全、低成本、高可靠的云存储服务，MaxCompute2.0 支持直接通过外部表方式访问对象存储中的表数据并进行处理。
7. 搜索引擎：阿里云开放搜索 OpenSearch 是一款自主研发的大规模分布式搜索引擎

擎平台。数据通过 MaxCompute 进行计算处理后，可以在 OpenSearch 平台上通过添加数据源的方式将 MaxCompute 数据接入。

8. 移动数据分析：移动数据分析（Mobile Analytics）是阿里云推出的一款移动 APP 数据统计分析产品，为开发者提供一站式数据化运营服务。当移动数据分析自带的基础分析报表不能满足 APP 开发者的个性化需求时，APP 开发者可以将数据一键同步至 MaxCompute，结合自己的业务需求来进一步加工、分析自己的数据。
9. 日志服务：日志服务能快速完成数据采集、消费、投递以及查询分析等操作。日志数据采集后，需要更多的个性化分析、挖掘，您可以通过 DataWorks 的数据集成将日志服务数据同步到 MaxCompute，通过 MaxCompute 对日志数据进行个性化、深层次的数据分析、挖掘。
10. 身份校验：RAM 是阿里云为客户提供的用户身份管理与资源访问控制服务。可以实现通过 DataWorks 使用 MaxCompute 时，子账户的身份管理。以及 MaxCompute 处理非结构化数据时，通过 RAM 对非结构化数据进行授权……

四、数据中台技术

阿里云数据中台是以 OneData 为内核，满足“快”、“准”、“全”、“统”、“通”的智能大数据体系。目前（2020/12/22）拥有六大工具：

1. Quick Audience，全域消费者运营平台。定位智能用户增长，实现全方位洞察，多渠道触达的增长闭环，实现以人为中心的消费者全生命周期洞察运营。
2. Quick Stock，智能货品运营平台。以货品为中心视角洞察市场，拥有强大的智能算法能力，让企业清楚了解货品全貌，进行更好的商渠匹配。
3. Quick Decision，智能决策平台。面向金融风控、财富管理、个性化保险、智能客服等场景，为企业提供决策全链路的构建、管理及服务。
4. Quick BI，数据可视化分析平台。进入 Gartner 魔力象限的 BI 产品，内置超过 40 种丰富图表，无缝对接云上数据库和自建数据库，0 代码实现数据分析。
5. Quick A+，全域行为洞察分析。跨多端全域行为分析的一站式数据产品，实现自动可视化的数据埋点，万亿级别的日志获取、海量数据实时和离线处理能力。
6. Dataphin，智能数据构建与管理。面向各行各业大数据建设、管理及应用诉求，一站式提供从数据接入到数据消费全链路的智能数据构建与管理的大数据能力，包括产品、技术和方法论等，助力打造标准统一、融会贯通、资产化、服务化、闭环自优化的智能数据体系，以驱动创新。

阿里云数据中台开发秉承 OneData 方法论，包括三个方面内容：

1. OneID，以业务和自然对象为基础，实现全域连接，标签萃取，立体画像。
2. OneService，面向应用提供主题式数据服务，简化大部分数据查询。
3. OneModel，统一模型构建及管理，100%消除指标二义性。OneModel 统一建模过程如图 4-1 所示。OneModel 指导指标的生产、管理、消费，将最终用于分析与

应用中的指标定义为“派生指标”，派生指标又拆解成原子指标、统计周期、业务限定、统计粒度四个子元素，如图 4-2 所示。



图 4-1 OneModel 统一建模过程



图 4-2 OneModel 指标过程

数据中台定位于计算后台和业务前台之间，其关键职能与核心价值是大数据以业务视角而非纯技术视角出发，智能化构建数据、管理数据资产与提供数据调用、数据监控、数据分析与数据展现等多种服务。承技术启业务，是建设智能数据和催生数据智能的引擎；而以数据中台内核价值为中段的数据中台业务模式不是纯数据、不是纯技术、也不是纯业务，它同时关注着与大数据能力相关的上下游，以大数据为中轴线，基于技术而又深入业务，它以数据产品+数据技术+方法论+场景实现的综合性输出，同时为智能化数据、技术极致提升和数据智能化业务负责。

一方面专注于从业务视角，建设标准统一、融会贯通、资产化、服务化、闭环自优化的数据中台智能数据体系，同时极致化追求技术上的降本提效。另一方面，致力于智能数据与业务场景深度融合的业务数据化与数据业务化中的各类智能化价值创新。