阿里云大数据平台

数据安全白皮书

阿里云大数据平台

2016 年 1 月

目　录

第一章 阿里云大数据平台安全性简介

阿里云大数据平台旨在将阿里云在大数据方面的能力和技术分享给数据工作者、数据创业者、大学及科研机构,目的是为了降低数据创业创新的门槛,繁荣大数据生态,实现数据淘宝的梦想。

阿里云大数据平台以阿里云在云计算和大数据方面多年沉淀的安全能力为基础，为数据创业创新提供全链路的安全保障。使用户在充分享受云计算大数据高性能、低成本优势的同时，保证数据不被攻击窃取而泄露。同时，阿里云郑重承诺：客户数据是客户的重要资产资产，云计算平台不得移作它用。并作为云计算中第一家，发出了《数据保护倡议书》。这份公开倡议书明确：运行在云计算平台上的开发者、公司、政府、社会机构的数据，所有权绝对属于客户；云计算平台不得将这些数据移作它用。平台方有责任和义务，帮助客户保障其数据的私密性、完整性和可用性。

大数据平台的安全性由阿里云和阿里巴巴集团的专业安全团队支持。其业界领先的能力在多年反网络攻击、反作弊、反个人数据泄露中反复得到验证，也得到国际国内多个权威机构和标准的认证和奖励。 除此之外，大数据平台也有自己专业数据安全团队，来确保数据提供者在安全可控的云环境中能以“可用不可见”的方式与数据创新创业者合作，并对数据生产、使用和销毁的全生命周期进行审计的使用来保障数据提供者的利益。阿里云大数据平台的数据安全团队积极参与国家大数据安全标准的工作，正在以大数据平台的数据安全能力参与制定国家大数据共享和交换安全技术标准。

第二章 数据安全体系

阿里云大数据平台的数据安全架构图如图 2.1。大数据平台建立在安全性在业界领先的阿里云上，并集成了最新的阿里云大数据产品。这些大数据产品的性能和安全性在阿里巴巴集团内部已经得到多年的锤炼。大数据平台的运维借鉴阿里云、淘宝网的安全运维经验，对人员资质、数据泄露等的风险进行监控。除了大数据平台的安全团队，阿里云和阿里集团的大安全团队也为大数据平台提供了坚强后盾，提供海量日志分析、案件追踪、实时风险监控的能力。

图 2.1 阿里云大数据平台数据安全体系图

阿里云大数据平台拥有业界领先的保护数据合作共享安全的能力。在如图 2.2 所示的多租户的数据合作业务场景下，平台提供了用户自有数据的私有空间和用与敏感数据共享合作的交换空间。 交换空间采用了先进的“可用不可见”的数据合作方式，对数据合作方的自由数据探索和导出进行监控。确保只有合作各方都同意的结果数据可以导出至出口区并服务用户的应用系统。平台通过对数据生产、存储、共享、开发和使用的全链路监控、审计，保障了数据合作共享的安全和相关各方的利益。

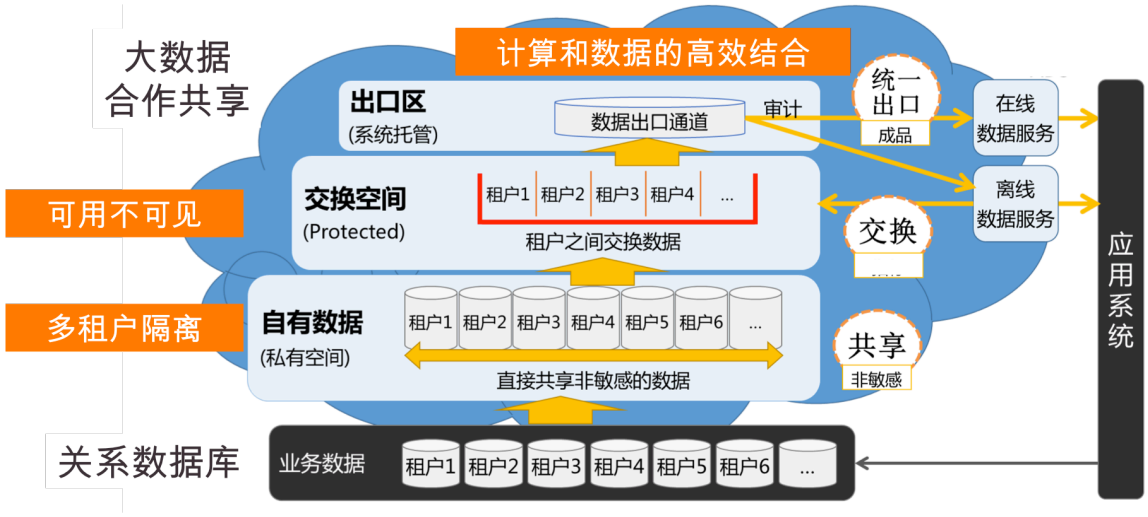

图 2.2 阿里云大数据平台数据合作共享安全

第三章 安全管理与认证

4.1 阿里云安全性

阿里云遵循“生产数据不出生产集群”的安全策略,覆盖从数据存储、数据访问、数据传输到数据销毁等多个环节的数据安全控制要求。阿里云基于阿里巴巴集团十多年信息安 全风险管控经验,以保护数据的保密性、完整性、可用性为目标,制定防范数 据泄露、篡改、丢失等安全威胁的控制要求,根据不同类别数据的安全级别 (例如:生产数据是指安全级别最高的数据类型,其类别主要包括用户数据、业务数据、系统数据等),设计、执行、复查、改进各项云计算环境下的安全 管理和技术控制措施。

阿里云已获得国际国内权威机构的安全认证,包括:

- ISO27001 国际认证。ISO 27001 是一项被广泛采用的全球安全标准,采用以风险管理为核心的方法来管理公司和客户信息,并通过定期评估风险和控制措施的有效性来保证体系的持续运行。
- 英国标准协会全球首张云安全国际认证金牌(CSA-STAR)。CSA-STAR 是一项全新而有针对性的国际专业认证项目,以 ISO/IEC 27001 认证为基础,为提供和使用云计算的任何组织,从沟通和利益相关者的参与;策略、计划、流程和系统性方法;技术和能力;所有权、领导力和管理;监督和测量等 5 个维度,综合评估组织云端安全管理和技术能力。
- 国家信息安全等级保护测评。信息安全等级保护是指对国家安全、法人和其他组织及公民的专有信息以及公开信息和存储、传输、处理这些信息的信息系统分等级实行安全保护。

阿里云还为有更高数据安全要求的行业,如金融、医疗行业,提供了符合专业需求的安全增值服务。这些安全方案包括专属物理集群、专线接入、异地灾备、行业资质认证等。设计、执行、复查、改进各项云计算环境下的安全 管理和技术控制措施。此外,阿里云正在进行保护医疗健康数据安全隐私的美国 HIPPA 认证。

更多阿里云安全信息见阿里云官网链接。

4.2 阿里巴巴数据安全管理

阿里巴巴作为中国最大的电子商务公司,拥有庞大的数据资产,在数据安全领域有着丰富的经验。为了保护客户数据、业务数据和公司数据,制定了完善严格的数据安全政策《阿里巴巴集团数据安全

规范（总纲）》（以下简称“总纲”），从数据流动中的各个环节进行安全控制，每个环节有相应的规则制度，从而规避不当行为带来的风险或损失。

大数据平台的数据安全管理工作围绕着数据的生命周期，如数据的生成、使用、存储、传输和销毁各阶段，完全遵循阿里巴巴集团的数据安全管理策略。

阿里巴巴集团现行的数据安全策略包括但不限于如下内容：

（一）机制保障

建立集团、事业部和部门多层级的数据安全组织保障。指定事业部、业务部门负责人为各自业务范围的数据安全第一责任人，建立事业部和部门的数据安全接口人机制，实现集团数据安全管理。

（二）人员保障

阿里巴巴集团员工，都需要签署保密协议，并受集团的安全政策和商业行为准则等制度的约束，公司内部开展各类培训对员工进行全方位的宣传教育。如员工未按要求违规操作，则有对应的违规处分原则配套落地，最高处分包括立即辞退员工，并有可能被追究民事直至刑事责任。

同时，明确具有管理职责的员工的监管责任：各级审批人对数据安全负有监管责任，如有疏忽将视情节給予处分。

(三) 环境保障

所有数据的使用都必须在隔离的网络安全域中使用，不同等级数据对使用的网络安全域有对应要求，不允许安全级别高的数据在低级别的网络安全域中使用。

（四）数据使用安全措施保障

- 实施数据等级保护策略，按照数据价值和敏感度对数据进行等级划分，根据数据安全分级，有对应的保护策略和要求；
- 保证数据完整性，建立数据的灾难恢复和备份机制；
- 所有数据使用都必须授权，不允许未经授权使用数据；
- 对安全网络域进行分级，有监控机制和访问/下载控制；
- 定期对数据可行性进行评估，对不可用或不再用数据及时销毁。

（五）对外数据合作中的数据安全措施保障

- 对外合作、交流中设计的对外开放数据都必须报备集团数据安全工作小组报备，评估确认后方可以对外开放；
- 合作中涉及用户的个人数据，应遵循集团的隐私保护策略；
- 严格遵循数据最少够用原则，严禁将超过业务约定用途的数据对外开放；
- 对外开放的数据中若涉及敏感类型的数据，应对数据执行脱敏处理；

- 数据开放过程中必须明确第三方使用数据的法律责任，且第三方使用数据的环境必须满足数据的安全性保护要求。

第四章 阿里云大数据平台数据产品安全性

4.1 阿里云数据产品安全性

大数据平台中使用了阿里云的应用引擎、数据平台等产品。它们的安全性介绍如下：

（一）云引擎 ACE：ACE 提供 PHP 和 JAVA 安全沙箱，为防止恶意应用而对应用权限进行一些限制。另外 ACE 禁止应用本地文件读写，读写文件可以使用阿里云的 OSS（开放存储服务），OSS 存储空间不受容量限制。ACE 由经验丰富的阿里运维和安全团队支持，协助解决网络攻击，网站挂马，漏洞扫描，代码行为分析等，并对服务异常进行告警。更多信息见阿里云官网链接。

（二）云数据库 RDS：主从双机热备架构，具有多重安全防护措施。主节点发生故障，秒级切换至备节点，服务可用性高达 99.95%。自定义访问 IP 白名单。可防 DDoS 攻击，SQL 注入告警。可自动多重备份，数据可靠性高达 99.9999%。

（三）开放数据处理服务 ODPS：采用多层次数据存储和访问安全控制机制，各个层次（操作系统，分布式计算系统和 ODPS）的用户系统解耦合，层次之间授权采取对当前任务最小够用原则。减少单个层次被攻击导致整体数据泄露的可能。数据存储多个备份，所有计算在沙箱中进行，保护数据不丢失，不泄露。ODPS 提供多种灵活、功能强大的访问控制，包括强制访问控制（Mandatory Access Control（MAC））和基于数据分级的访问控制。使用户更好地控制自有数据的分享。

4.2 阿里云大数据平台产品集成安全性

阿里云大数据平台的数据流转主要涉及三个子系统，分别为数据市场、数据加工及应用服务。根据用户的需求场景，数据会涉及到几个子系统之间的流程。阿里云大数据平台对这几块之间的数据流转，建立相应的机制以保障数据在各系统对安全管控上的连惯性。

3.2.1 数据市场

当出现有数据交换的需求时，会出现两种形态的交易商品：数据、服务。

数据供应商可以将自己的数据打包后发布到数据市场中，在数据市场中合作使用的数据，可以指定数据的使用范围，如数据可以被用户下载、数据只允许在安全的管控环境中加工不允许用户看到真实的数据。数加会根据市场中的数据属性，对数据做不同等级的管控，如下文提到的合作共享业务就是针对这类场景的解决方案。

数据供应商也可以根据业务特点，将不同的数据开发成不同的服务，以数据服务的形式供数据需求方使用。数据服务的使用同样也可以限定用户的范围，如服务可以在用户自有系统中使用、必须在平台管控的环境中使用等。

4.2.2 数据加工

数据加工分用户自有数据加工和共享数据加工两种场景。

用户自有数据加工在整个过程中用户对所有数据拥有绝对的管控权，可以进行管理、下载、授权等相关的操作。而共享数据加工，主要是针对两方及以上的数据合作模式，当数据供应商希望数据的隐私性得到保护，不被其他人复制时，阿里云大数据平台提供一个管控的共享区域，在该区域中数据供应商可以将数据放进来，数据使用者可以进入该共享区进行相应的业务数据开发，当出现数据需要与数据系统对接时，由数据供应商对相应的加工、结果进行审核确认。这种合作共享的模式在保障数据价值最大化的同时，也保障了数据的安全性。

4.2.3 应用服务

在以服务形式进行数据交换的场景，虽然数据通过服务的方式开放给第三方使用，但仍有一些涉及到敏感性的数据会透出。针对这类场景，平台会为提供一套在线的应用托管环境，用户的在线加工、处理数据逻辑只能在这个环境中进行，数据拥有方有权对加工的逻辑进行审查以确认数据安全是否得到保障。阿里云大数据平台提供相关的日志监控、审计能力，用于发现一些违反初期规则的行为泄露数据。

第五章 数据业务安全性

5.1 数据自用业务

用户拥有全部的控制权限，包括对数据、成员管理等。用户可以根据需要，将其他用户加入到自己的空间中对数据进行加工生产，将结果数据的导出。数据的存储、用户及权限管理，由阿里云开放数据处理服务（ODPS）提供，认证采用消息签名机制，具有更高的安全性。

5.2 数据合作共享业务

担保方式交换使用数据，由平台作为担保方，创建可供用户使用的共享项目空间，该项目空间的特点为：

1）用户无法看到真实的全量数据。

2）用户无法将该项目中的真实数据移出。

双方基于该机制将自己的数据授权到交换区中，使用方根据开发环境的样本数据做相应的开发，当任务发布后，由担保方来执行相应的任务，数据对于使用方可用不可见。

项目空间的数据导出，系统会进行数据的血缘分析，由各数据拥有方确认业务场景是否符合对数据的管控

第六章 运维管理安全性

6.1 平台管理

阿里云大数据平台管理权限仅限于经授权的平台运营人员；安全工程师审计并回溯运营人员权限审批、操作行为记录、数据访问是否合规；为了保护阿里云客户和自身的数据资产安全，阿里云采用一系列控制措施，以防止未经授权的访问。

6.1.1 基本安全保障

阿里云大数据平台数据的安全密级遵从重点保障客户隐私数据、公司机密与商业秘密的原则，根据数据类型、数据保密性要求、数据访问授权的对象不同进行了规范。

任何团队和个人不得从事以下危害数据安全的行为：

(一) 未经授权查阅他人邮件、旺旺等通讯记录；

(二) 未经授权使用技术手段获取敏感数据（如：爬虫）

(三) 未经授权使用他人在信息系统网络中未公开的信息；

(四) 盗用他人名义发送电子邮件或执行未授权操作；

(五) 未经授权对公司系统中存储的信息（包括系统文件和应用程序）进行增加、修改和删除等，破坏数据完整性的行为；

(六) 将任何具有密级要求的信息，在未授权的情况下，发送或存储在外部公司的服务器或服务上。(如：Apple iCloud、百度云盘、印象笔记、外部开源代码库等)。

6.1.2 认证控制

阿里云每位员工拥有唯一的用户账号和证书，这个账号通过有线和无线网络接入用来识别每个人在阿里云网络内的活动情况并作为阻断非法外部连接的依据，而证书则是作为抗抵赖工具用于每位员工接入所有阿里云内部系统的证明。员工入职后，人力资源部会给予一个用户账号，并按照其岗位类别和职级进行授权，离职后，人力资源部将通过系统将禁止该账号访问阿里云网络。阿里云密码系统强制策略用于员工的密码或密钥（例如登陆工作站）。包括密码定期修改频率、密码长度、密码复杂度、密码过期时间等。阿里云针对生产数据及其附属设施的访问控制除去采用单点登录外，均强制采用双因素认证机制，例如像证书和一次性口令生成器

6.1.3 授权控制

访问权限及等级是基于员工工作的功能和角色，最小权限和职责分离是所有系统授权设计基本原则，阿里云员工访问公司的资源只授予有限的默认权限。例如访问邮件和阿里云内部办公系统。如根据特殊的工作职能，员工需要被授予权限访问某些额外的资源，则依据阿里云安全政策规定进行申请和审 批，并得到数据或系统所有者、安全管理员或其他部门批准。所有批准的审计记录均记录于工作流平台，平台内的控制权限设置的修改和审批过程的审批政策确保一致。

6.1.4 审计

阿里云所有信息系统的日志和权限审批记录均采用碎片化分布式离散存储技术进行长期保存，以供审计人员根据需求进行审计

6.1.5 传输

数据在传输过程中，须遵循如下原则：(1)　必须根据网络域安全策略对网络域进行分级；(2)在不同等级的网络域边界必须由监控机制和访问控制。

6.2 人员管理

在入职前，阿里云在国家法律法规允许的情况下，通过一系列背景调查手段来确保入职的员工符合公司的行为准则、保密规定、商业道德和信息安全政策，背景调查手段涉及刑事、职业履历和信息安全等方面，背景调查的程度取决于岗位需求。

在入职后，所有的员工必须签署保密协议，确认收到并遵守阿里云的安全政策和保密要求，而在这些安全政策和保密要求中关于客户信息和数据的机密性要求将在每一位新员工入职培训过程中被重点强调。除去针对新员工信息安全课程的培训，阿里云依据员工工作的不同角色进行额外信息安全培训，确保不同角色员工管理的用户数据必须按照安全策略执行。最后，阿里云通过对员工进行企业价值观考核的方式检验每位员工是否以诚信、敬业的态度来管理每位客户的云端数据，保证其对客户、合作伙伴和竞争对手的尊重；

员工在离职、转岗等工作交接过程中需要做好保密和安全加固工作，主管应提供协助和指导，例如：

(一)　　对工作岗位、涉及到的涉密数据提取流程，应该对交接员工提供培训和安全宣导。

(二) 对岗位中涉及到的特权账户进行账户密码的交接时，需要进行密码更新与相关备案工作。

(三) 对涉密岗位需要根据岗位要求和人力资源部规定签署保密协议。

(四) 经过授权取得涉密数据后，在使用时应当注意不超过授权范围使用，并采用适当手段防止意外泄露，（如：纸质文件封签，数字文件加密等）使用完成后应当及时销毁，纸质文件需要碎纸。

集团员工不允许擅自通过各种媒介披露或与公司以外的成员（包括但不限于家人、亲友、客户及第三方合作方）披露与工作相关的任何数据；

阿里云提供机密报告机制以确保员工可以匿名报告任何违反安全政策、商业道德的事件。 与数据加工、生产等特殊岗位相关的员工需签署《阿里巴巴集团数据安全管理规范》