阿里云大数据安全实践

演讲人：苏建东

IT时代→DT 时代

零售、制造、金融、医疗…

- **数据成为核心要素**
- **数据驱动商业变革**
- **数据共享创造价值**

典型的大数据存、通、用场景及问题

■数据无处不在（系统、人）

■系统、组织间数据边界模糊

■数据是生产资料，需要流动

■数据要关联、聚合

■海量数据加密性能、成本问题

■数据流动、处理要实时

■数据转移，owner和权利问题

- 不同部门数据存在同一个大数据平台，是否有一个超级管理员可以看到所有数据？
- A部门数据要共享给B部门做计算，能否实现数据不搬家，可用不可见？
- 某省人口库一张大表，如何能授权给某个地市仅仅该地市的数据？
- 不同安全等级的数据存在一张表里，能否根据数据安全级别和访问者安全级别决定访问权限？

大数据安全挑战

大数据技术和平台安全挑战

- 传统安全措施难以适配
- 平台安全机制严重不足
- 应用访问控制愈加困难

数据安全和个人信息保护挑战

- 数据安全保护难度加大
- 个人信息泄露风险加剧
- 数据真实性保障困难

大数据安全法规标准挑战

- 大数据安全法规标准尚需完善
- 缺乏大数据安全最佳实践

阿里云大数据安全实践思路

自主可控的平台安全体系

建设自主可控的大数据技术，原生的安全机制

以数据为中心的安全

从以网络和系统为中心的防护转向以数据为中心的安全

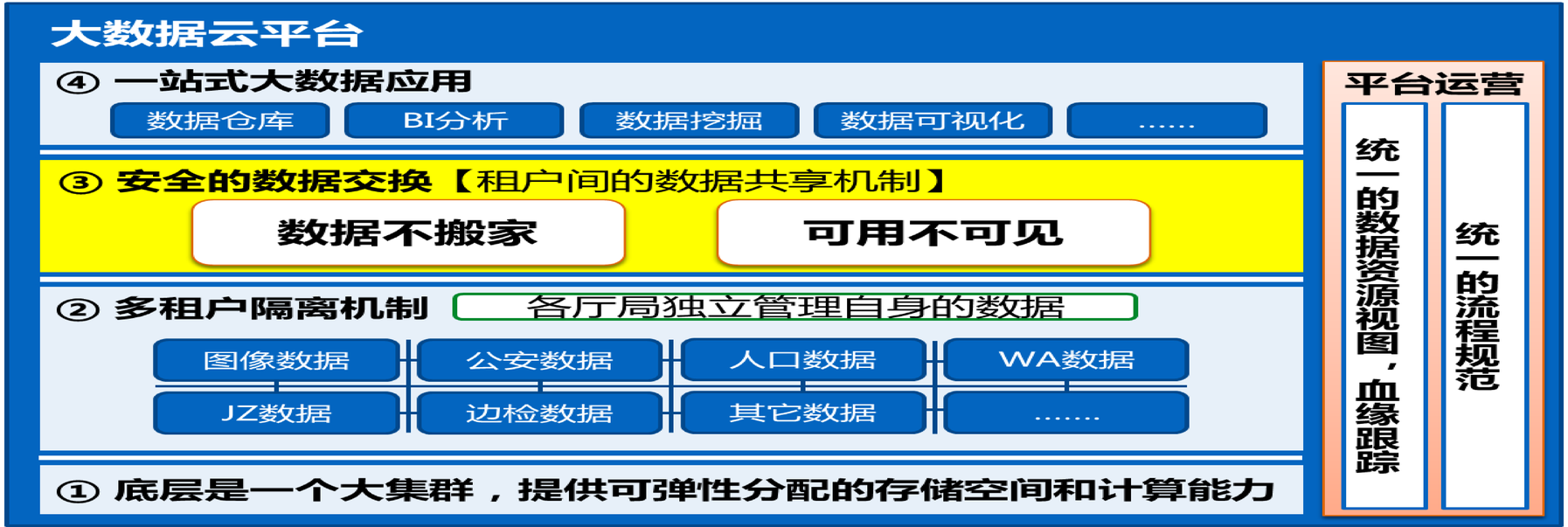
法规标准建设

积极参与大数据法规标准建设；推动大数据安全最佳实践

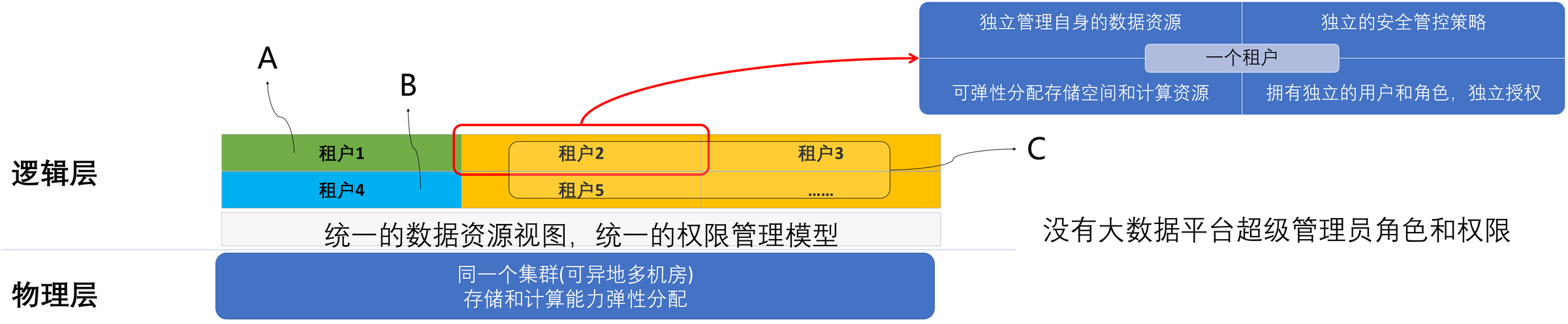
阿里云大数据安全实践


数据安全治理
组织保障
国家法律法规标准建设与实践
数据安全管理
数据安全技术
数据安全生命周期管理
采集
存储
使用
传输
共享
销毁
数据安全技术
数据资产管理
资产发现
数据分类
数据分级
血缘管理
数据权限控制
身份管理
双因素认证
细粒度授权
强制访问控制
数据监控审计
数据库审计
人员操作审计
共享调用审计
数据传输监控
数据安全防护
传输存储加密
数据脱敏
数据水印
数据防泄漏

完全自主可控的大数据平台

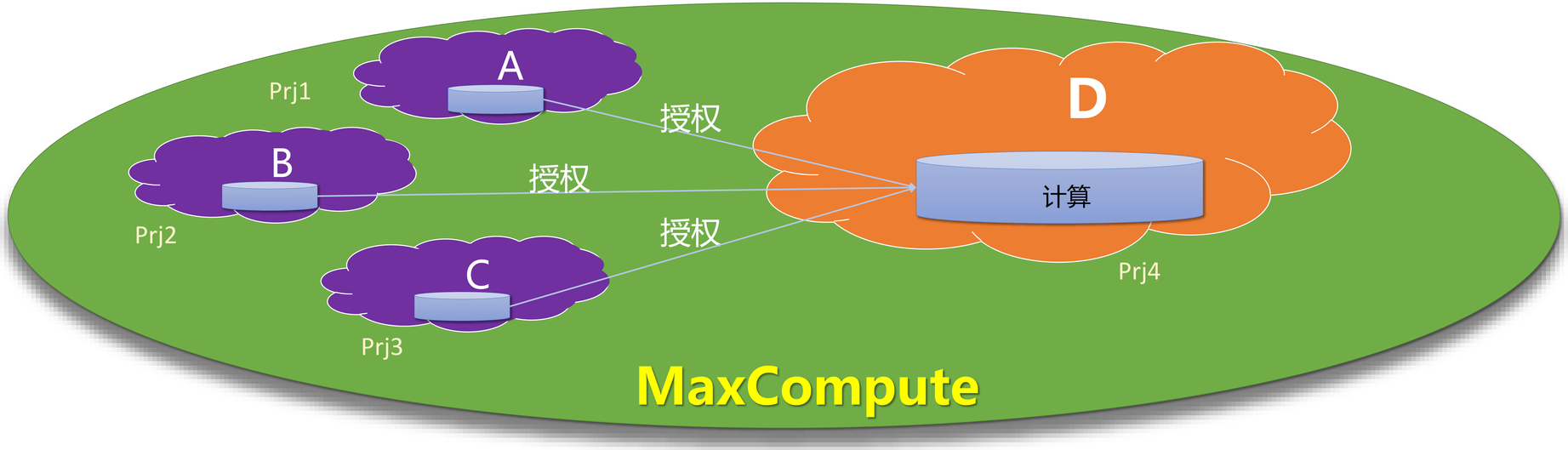


原生的安全机制：多租户隔离

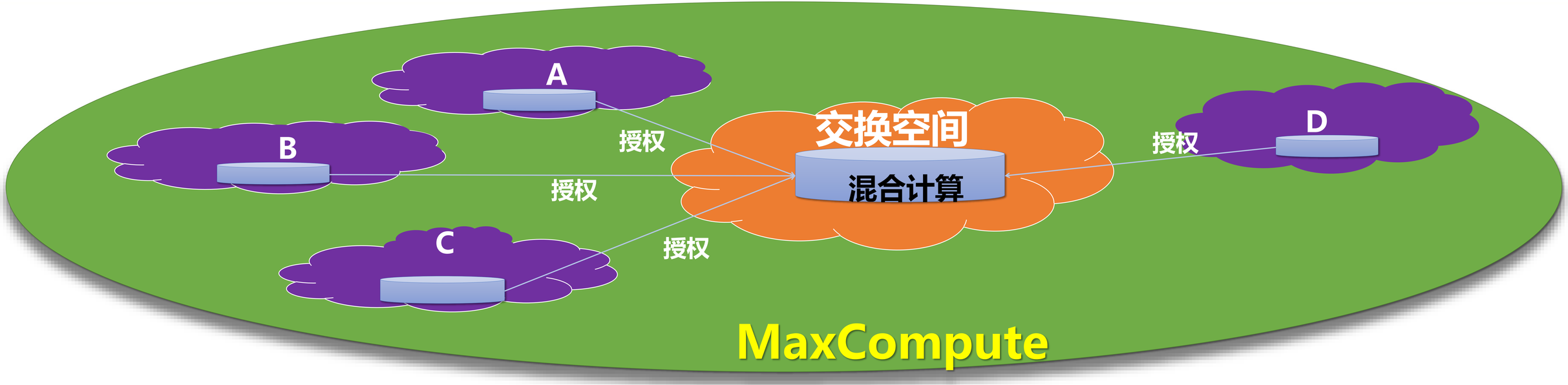


没有大数据平台超级管理员角色和权限

原生的安全机制：租户间数据直接授权



原生的安全机制：租户间数据通过交换空间共享



原生的安全机制：打包授权机制

- Project之间，通过package将资源打包后授权。
- 对方安装package之后，可以将package里面的资源二次授权给内部用户/角色。
- Package是虚拟对象，并非实体存储。

原生的安全机制：多种授权机制

原生的安全机制：多种授权机制

ACL

- 基于实体对象的授权
- 支持类似于SQL92定义的GRANT/REVOKE语法，它通过简单的授权语句来完成对已存在的租户对象的授权或撤销授权

Policy

- 对一组对象进行授权
- 带限制条件的授权
- 使用访问策略语言(Access Policy)来描述授权。策略语言目前支持20种访问条件（即从20个维度来限制对一张表的访问，例如访问来源IP地址）

Label

- 强制访问控制
- 对数据和人分别设置安全等级标记之后，不允许用户读取敏感等级高于用户等级的数据，除非有显式授权

MaxCompute原生安全机制与Hadoop安全机制的对比

	hadoop	MaxCompute(odps)
权限	基于hdfs的rwx权限	完善的权限控制模型（acl、policy、role、label）
隔离	无	基于project的安全隔离模型
容错	文件回收站，依赖用户配置	多层回收站机制，保留多次数据操作快照版本，任意恢复数据到某个快照
多租户	无	MaxCompute支持完善的多租户机制，通过存储和计算配额的方法可以让多个用户分享一个集群的资源。 所有的计算任务都运行在安全沙箱中，通过进程和JAVA沙箱，配合运行时的签权方法，保证数据安全。 MaxCompute提供丰富的授权管理手段，包括ACL，角色授权， Policy授权、跨Project授权以及Label机制，可以提供精确到列级别的安全方案，满足一个组织或者跨组织间的授权需求。 用户访问需要认证，用户操作需要鉴权，提供完整的审计功能。 安全要求较高的项目，可以提供项目保护机制，防止数据泄露。 系统级别、Project级别和表级别的IP访问控制白名单设置。

以数据为中心的安全

采集	存储	使用	传输	销毁
●资产打标 ●血缘管理 ●安全等级打标	●存储加密 ●密钥管理中心	●网络准入 ●身份管理和访问控制 ●操作监控和日志审计 ●虚拟桌面、数据防泄漏 ●数据脱敏	●VPN加密 ●HTTPS加密	●硬盘消磁机 ●硬盘粉碎机 ●硬盘折弯机

中台化	国产加密	传输加密	存储加密	数据脱敏
• 密钥管理中心 • 加密SDK	• 国产密码机云服务化	• HTTPS统一接入 • 证书集中管理	• 服务端加密 • 客户端加密	• 静态脱敏 • 动态脱敏

安全、合规、智能的大数据安全管家

数据安全痛点	面对不同的挑战	阿里云实践	达到安全目标
监管合规要求	个人隐私数据保护	隐私数据识别、脱敏	满足数据资产保护的合规性
	数据出境	数据导出监控	
	数据交换	数据发布脱敏、同态加密等	
数据管理	数据如何分类分级	数据智能分类分级	对数据生命周期的安全管理
	无法掌握敏感数据分布	敏感数据分布热力图	
数据风险	敏感数据明文展示	脱敏SDK、IDE脱敏集成	通过智能数据风险识别，提升数据安全运营效率
	担心数据泄露	数据导出监控与风险识别	
	敏感数据违规操作难以发现	智能数据操作风险识别	

核心能力

发现&评估

- 定义敏感数据
- 发现和定位敏感数据
- 数据分级分类
- 安全评估

监控&识别

- 实时操作监控
- 可视化展示
- 数据导出风险识别
- 数据操作行为风险识别

审计

- 风险事件实时推送
- 提供SIEM对接的API
- 合规要求审计
- 集成审计流程

加固&优化

- 合规要求
- 安全策略
- 脱敏
- 加密

发现&评估

定义敏感数据

- 内置默认敏感数据定义
- 提供不同行业敏感数据定义模板
- 支持用户自定义敏感数据定义

敏感数据定位

- 处理PB级敏感数据识别
- 数据识别准确率98%
- 小时级识别时效性
- 识别计算效率高

数据分级分类

- 基于机器学习和用户自定义结合的分级分类方案
- 符合网络安全法等个人隐私数据分级保护要求
- 分级分类准确率为99%

安全评估

- 环境安全基线评估
- 账号安全评估
- 角色&授权安全评估

操作监控&风险识别

- 秒级操作监控
- 覆盖所有DDL,DML任务
- 字段级精准识别

实时操作监控

可视化展示

- 易上手的高可视化监控报表
- 支持自定义报表
- 支持数据导出用于数据离线分析

- 覆盖所有数据出口监控
- 对敏感数据进行重点监控
- 基于机器学习的数据泄露风险监测模型

数据泄露风险识别

数据操作风险识别

- 覆盖所有数据操作行为
- 基于机器学习和自定义场景的数据操作风险模型
- 自定义规则配置以满足特定的场景

审计

风险事件实时预警

- 实时推送风险模型识别的事件
- 提供基于钉钉的风险预警服务窗
- 提供电话、短信、邮件、钉钉等不同事件提醒方式

提供SIEM对接的API

- 通过API推送到SIEM平台
- 数据格式自定义配置

合规要求审计

- 提供基于合规要求的审计模板
- 提供监管要求的审计模板
- 提供行业安全管理审计模板

审计流程

- 内置审计任务流程
- 对接钉钉服务窗，提供审计流转流程

加固&优化

合规要求

- 适合不同行业的合规要求，例如PCI、央行、证监会、保监会等
- 按照合规要求提供数据安全保障功能

安全策略

- 网络访问策略
- 权限安全等级策略
- 数据泄露应急安全策略

脱敏

- 提供显示脱敏接口
- 为阿里云数加、蚂蚁采云间等数据IDE提供结果脱敏
- 为开发和测试环境提供物理数据脱敏

细粒度化访问控制

- 表级别权限管理
- 字段级别权限管理
- 行级权限管理

THANKS