



2015 杭州·云栖大会
COMPUTING CONFERENCE

阿里巴巴大数据实践之路

陈鹏宇（不老）
阿里云数据事业部·高级专家



目录
content

存：数据上云

通：数据打通

用：数据化运营

生态：运营数据





马云：阿里巴巴是一家数据公司





阿里大数据发展历程

● 时间线

- 09年：阿里定位为一家数据公司
- 12年：成立数据事业部
- 14年：大数据产品正式对外服务

● 认知线

- Data 1.0：DW和BI为主，“看”
- Data 2.0：数据化运营，“用”
- Data 3.0：运营数据，“生态”



阿里大数据体系结构

前台
业务



中台



公共数据产品和服务

公共数据服务
One Data



统一的数据仓库

后台



统一的大数据技术平台

公共技术平台
One Platform

ODPS, ADS

数据共享

算法共享

知识共享

生态
建设

运营
管理

数据安全

数据质量

数据标准

成本管理

阿里大数据体系结构



2015 杭州·云栖大会
COMPUTING CONFERENCE

前台
业务

③用

中台

公共数据服务
One Data

②通

后台

公共技术平台
One Platform

①存

④生态

①存：数据上云



数据平台技术演进



Hadoop, GP, Oracle, ...

如何做到一个平台？

1

统一平台，数据大集中

统一的存储，统一的计算引擎，统一的数据开发平台

2

资源共享，弹性分配

基于ODPS多集群技术，由数以万计的服务器提供超级计算能力，按需弹性分配给各数据开发团队

3

数据隔离，分权管理

基于ODPS多租户机制，各部门可独立管理自身的数据，独立做数据授权

按需弹性分配的計算資源大池

ODPS：一個計算引擎

飛天：一台計算機

集群1



集群2



集群n



②通：数据打通



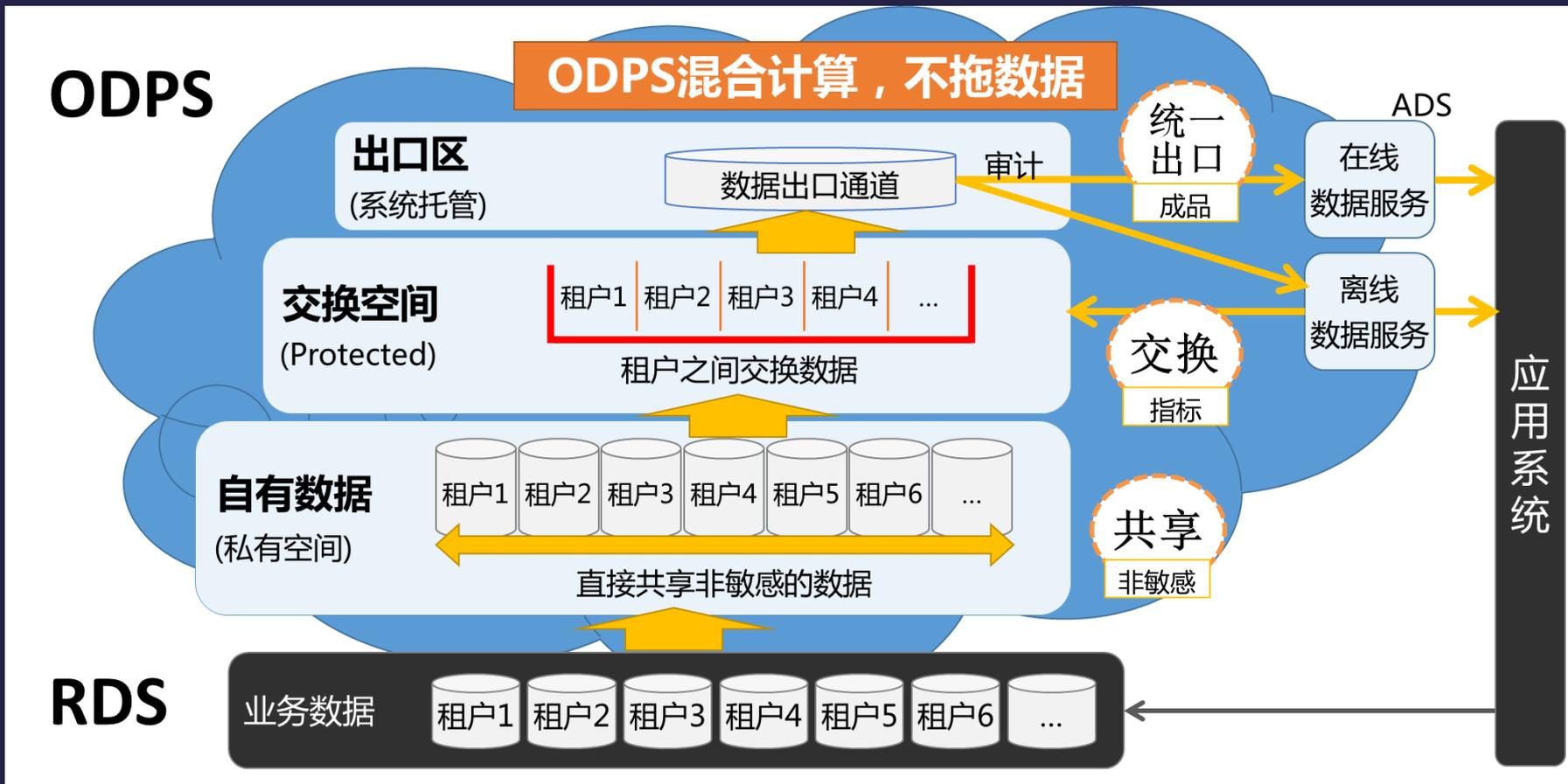
如何打通多个业务数据？



数据隔离与共享：普通模式



数据隔离与共享：担保模式



数据整合与服务

One Service : 公共数据服务

人群画像

特征工程

行为库

关系库

.....

One Data : 一个数据仓库

ODPS : 一个计算引擎

飞天 : 一台计算机



全局元数据管理 全链路血缘跟踪

账号 权限 计量

③用：数据化运营



阿里数据化业务版图

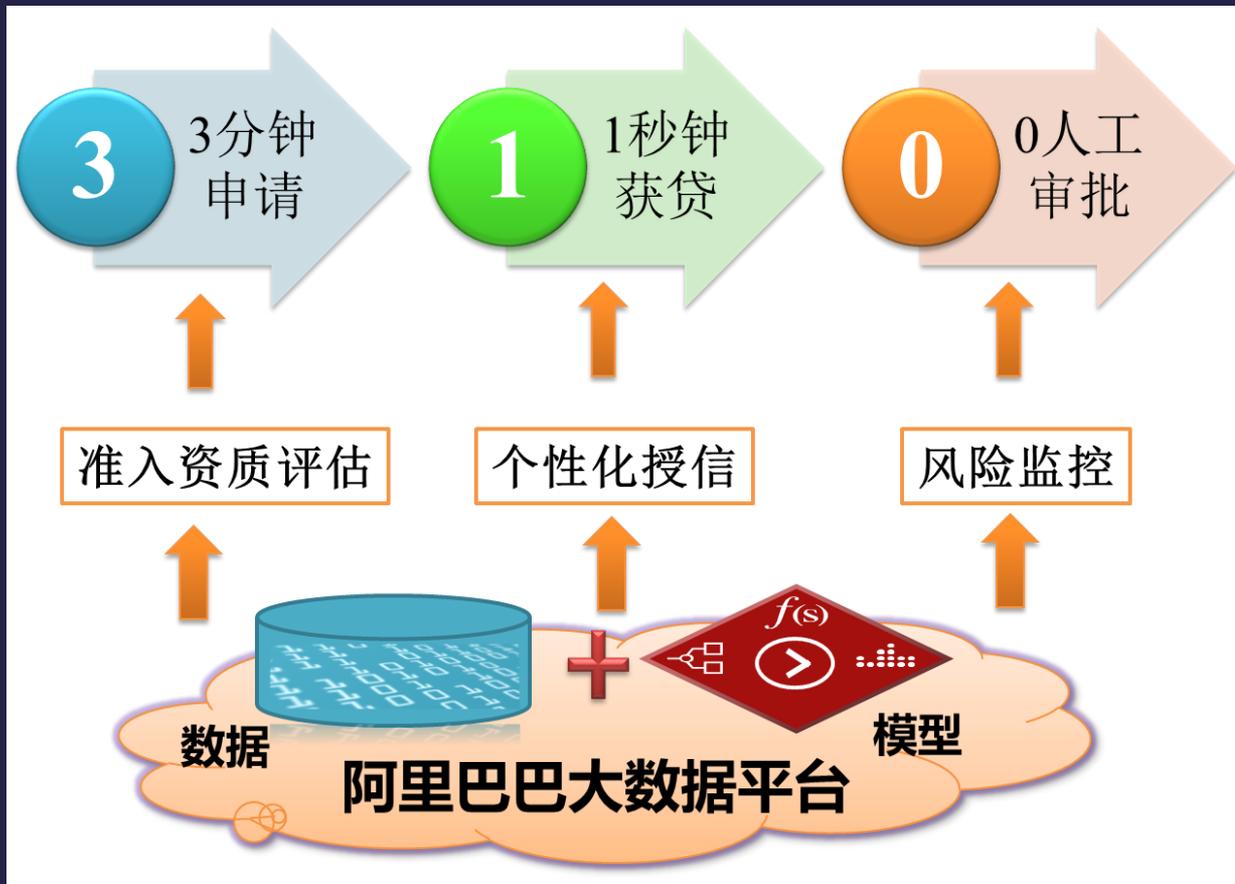


Facts :

几十个业务部门

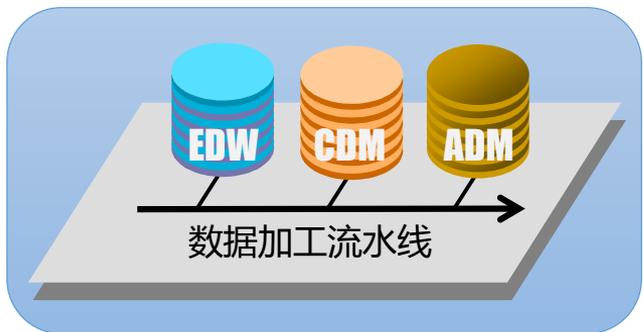
数百个应用

数千名工程师



蚂蚁小贷：
百万级客户
日均流水过万
极低坏账率

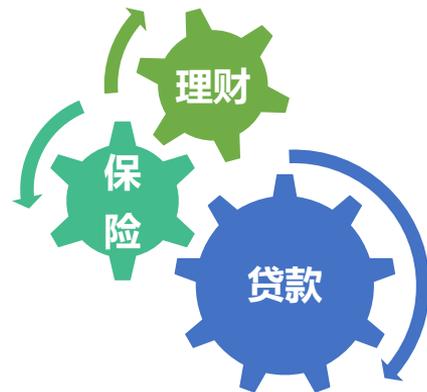
数据仓库



决策引擎



金融业务



数据加工

模型运算

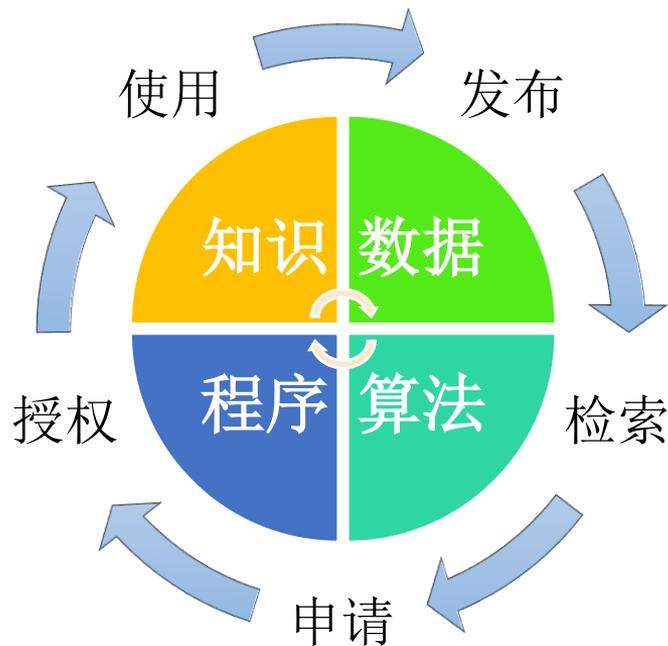
在线系统

④生态：运营数据



如何运营数据？

共建共享的数据社区



数据资产运营管理

