

网易大数据 应用及实践分享

网易猛犸 朱伟



01

从技术出发，网易大数据的诞生和发展

02

赋能商业成功，大数据实践案例分享

03

如何成为一名合格的大数据工程师

PART 01

从技术出发，网易大数据的诞生和发展

I 定位

互联网本身就是一个建立在数据之上的服务行业，互联网领域的大数据应用。

核心目标

进一步拓展和发掘数据价值的应用途径，以实现数据价值的充分利用。

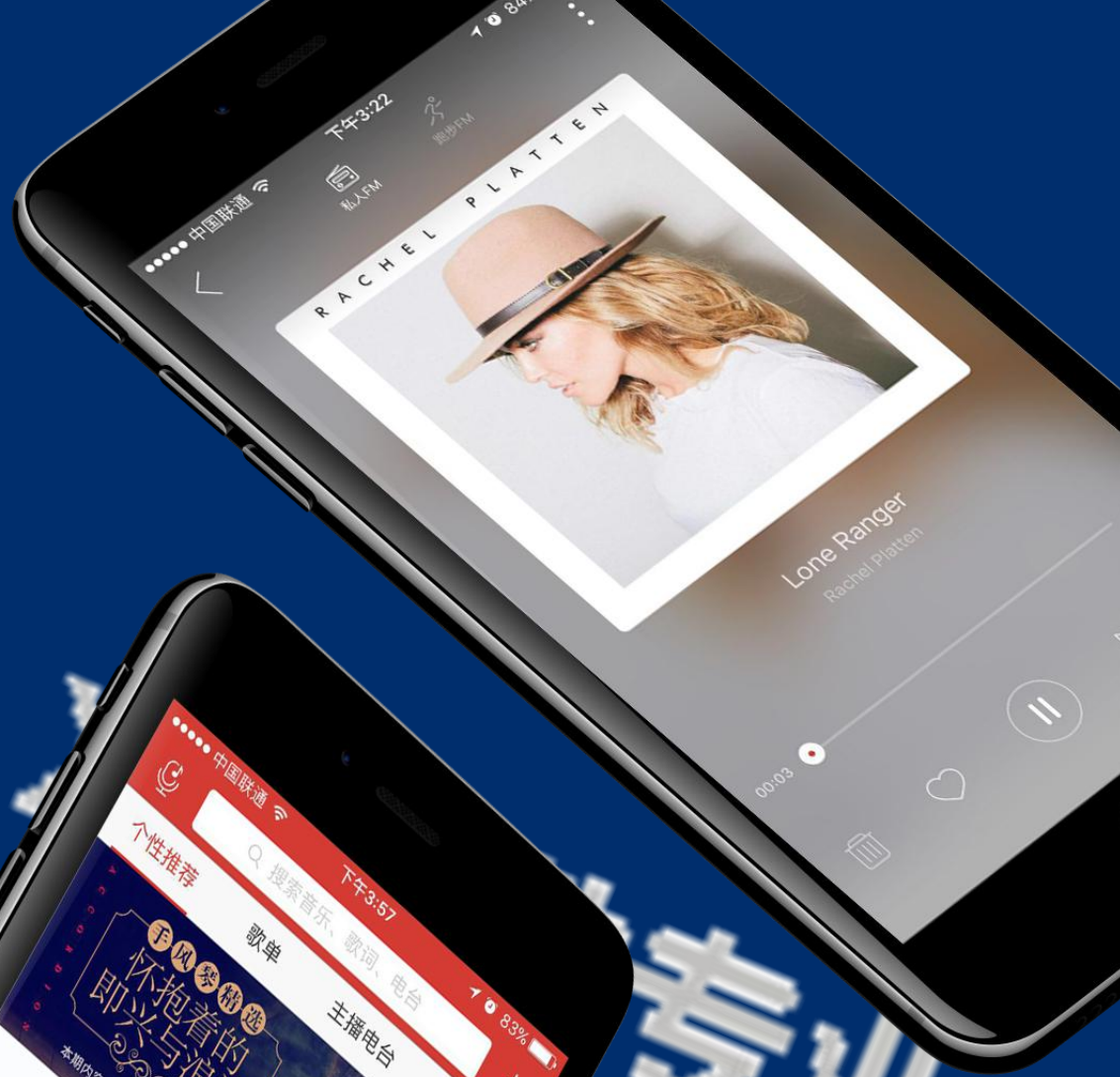
大数据定位

充分运用数据统计、分析、挖掘、实验等技术，大力推进 服务智能化与决策科学化，全面提升产品规划设计运营水平与服务质量。

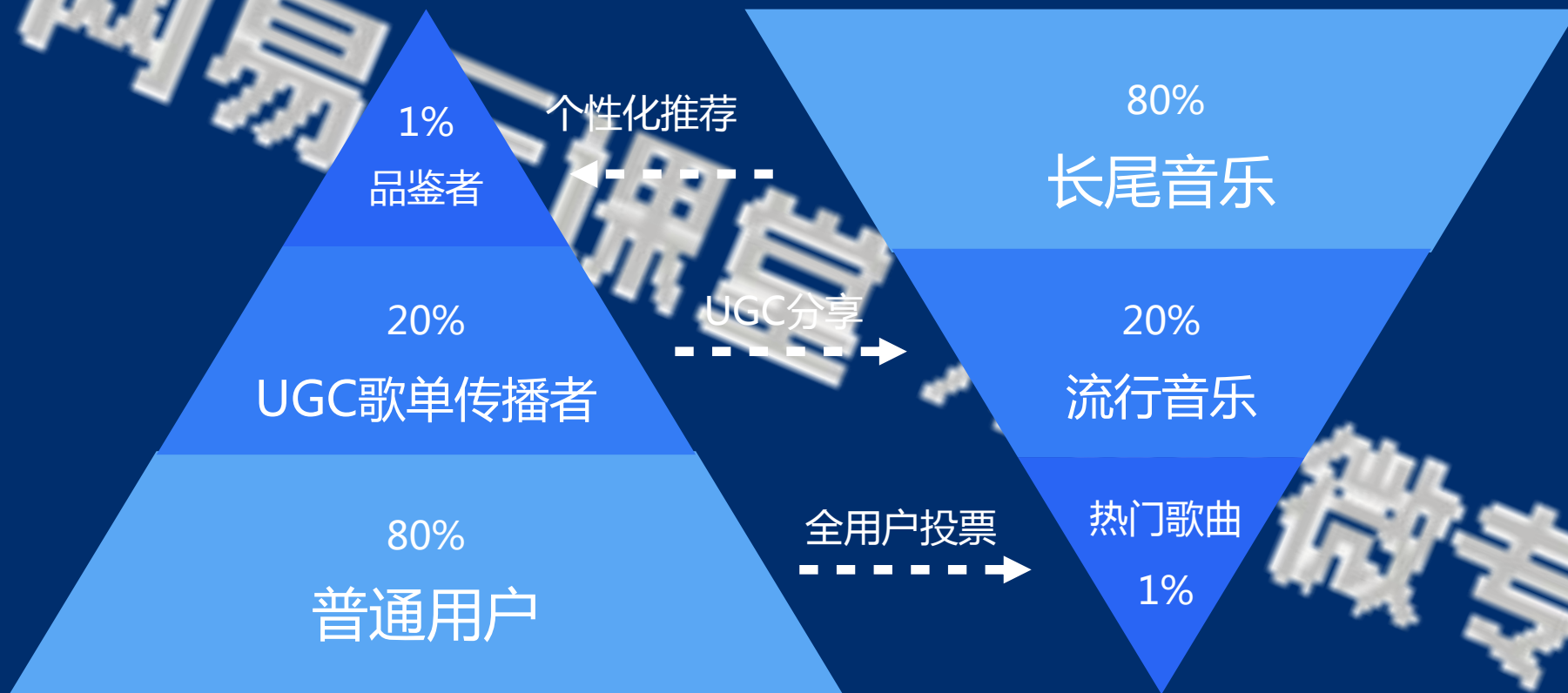
服务智能化：云音乐个性化推荐

云音乐个性化推荐

- 首页个性化推荐
- 私人FM推荐
- 每日歌曲推荐

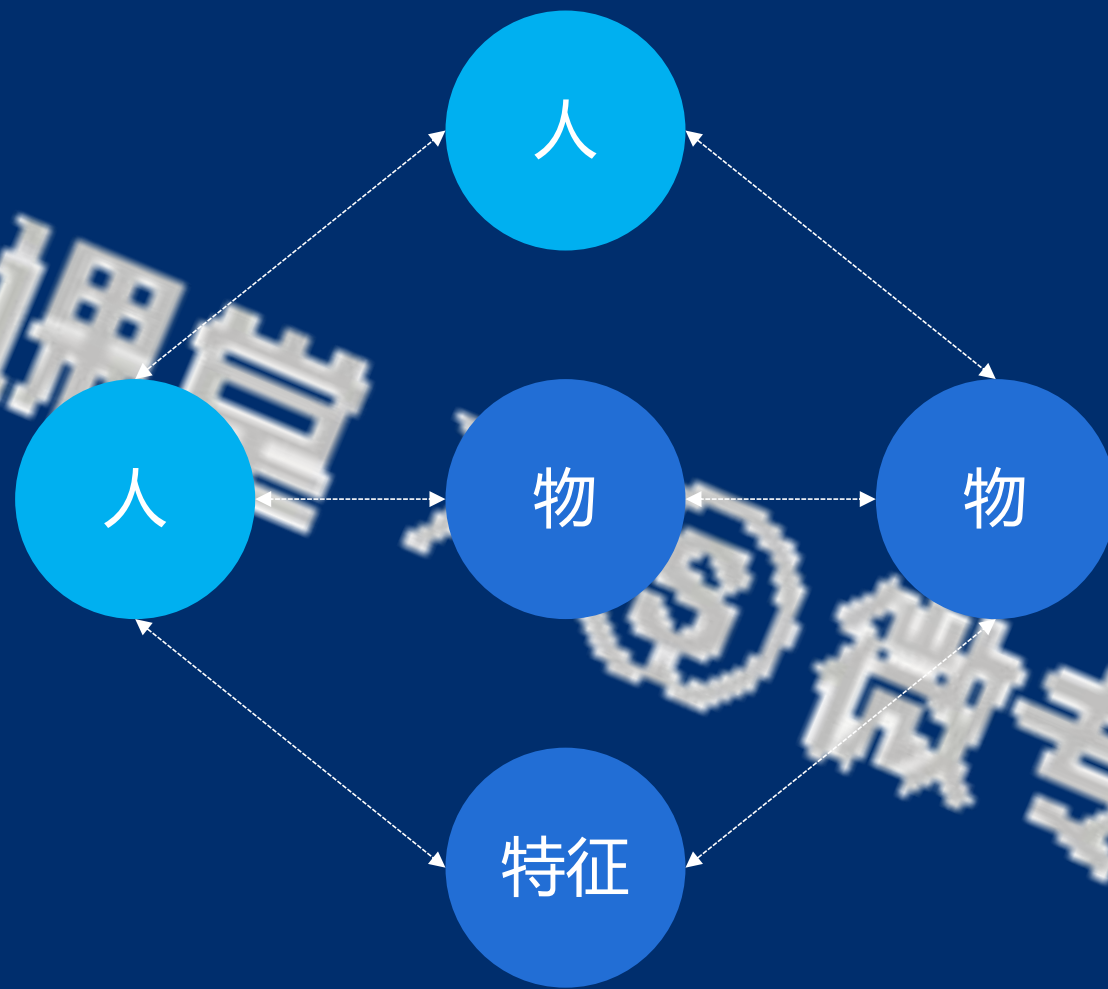


音乐推荐的传播路径：UGC×推荐



服务智能化：个性化推荐

- 召回
- 过滤
- 排序



I 服务智能化：更多案例



内容反垃圾



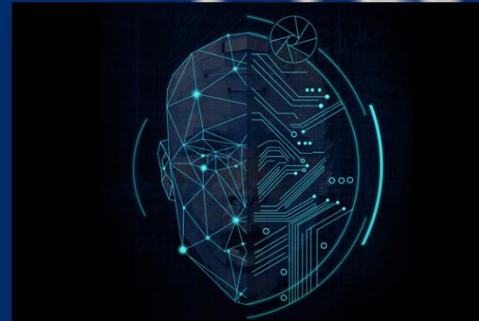
客服机器人



反作弊



金融反诈骗



视频鉴黄

微专业

I 决策科学化：精准营销

案例

针对在1~3个月内买过尿不湿，且消费能力较高的用户，推出一个购买牛栏奶粉满388-50活动
加入购物车后未购买 and 订单额大于200 and 有高毛利商品-> 发送一张-10优惠券给用户

用户标签

性别 年龄 地域 消费能力 ...

行为标签

加入购物车 浏览未购买 收藏
最近购买时间 ...

用户画像



营销方案

效果监测

A/B测试

实时监测

I 决策科学化：更多案例



电商定价



促销设计



功能上线



留存分析

PART 02

赋能商业成功，大数据实践案例分享

I 大数据业务需要什么？



I 迭代产生价值



I 数据驱动文化

技术门槛越来越低，而精神门槛从未降低

网易云课堂 微专业

I 一个中心两个平台三个团队

数据资产中心

- 集成所有有价值的数据，进行数据融合与深加工
- 目的：打破产品数据分界
- 价值：提高效率

核心数据团队

- 分析团队
- 算法团队
- 产品团队

网易 大数据

开发计算平台（猛犸）

- 存储和管理全公司数据资产
- 提供高效、易用的大数据分析挖掘应用开发环境
- 提供大规模查询和计算能力

可视化分析平台（有数）

- 支撑产品分析与决策、产品营销推广

网易猛犸 大数据开发计算平台

I 开发计算平台提升数据加工效率



海量数据
数据孤岛



数据资产
缺乏管理



数据开发
门槛高

I 数据集成：数据接入

Mammut

数据开发任务运维自助分析数据管理实时流计算

项目中心intern-滨江

+ 新建

未上线 订单量预测模型

开发模式 线上模式

基本属性

任务 数据

全选

设置并运行

运行结果

查看资源信息

保存提交上线

收起

数据库传输-ods_rrs_pool_noinfor

完成取消

数据流转

来源

类型MySQL

库大数据导数源

模式

表单

SQL

表mytest

列

整表导入（新增列也自动导入）

选择列

列操作符条件

+ 添加

多个条件为"AND"关系

切分键请输入或选择切分键

并行度1

更多参数请输入参数

失败自动重试

目的

类型HDFS

目标路径\${azkaban.flow.start.day}

删除数据

保存格式

Text文件

ParquetJSON

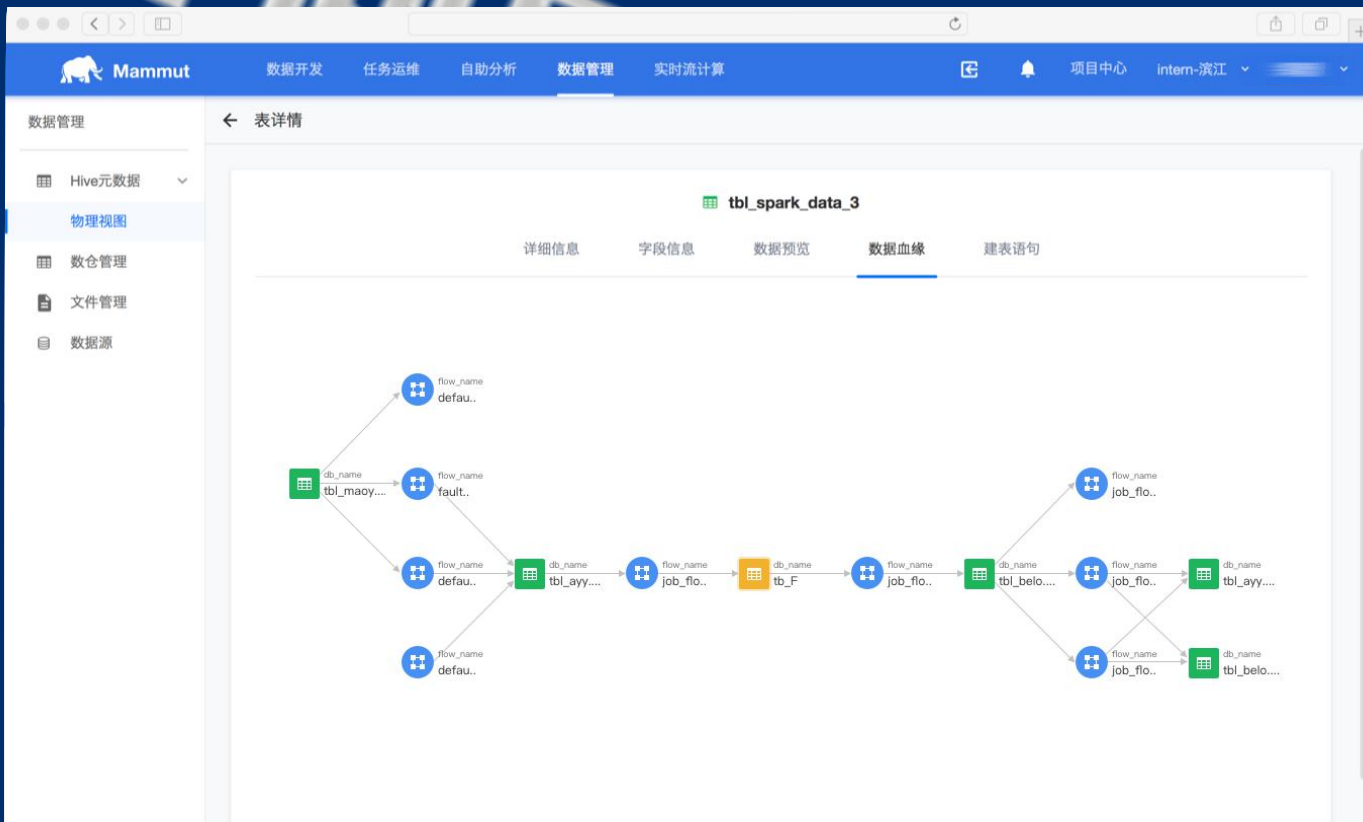
字段分隔符,

使用Izo压缩并生成index

I 开发效率：可视化开发



数据资产管理：数据血缘/主题管理/角色权限



The screenshot shows the Mammut project management system. The left sidebar has a '角色管理' (Role Management) menu. The main content area is titled '风控' (Risk Control) and '风控团队' (Risk Control Team). It includes a '成员' (Members) list, a '数据权限' (Data Permissions) section, and a '功能管理' (Function Management) section with various task and data management options.

The screenshot shows the '数据源' (Data Sources) page in the Mammut application. The left sidebar contains a tree view with the following structure:

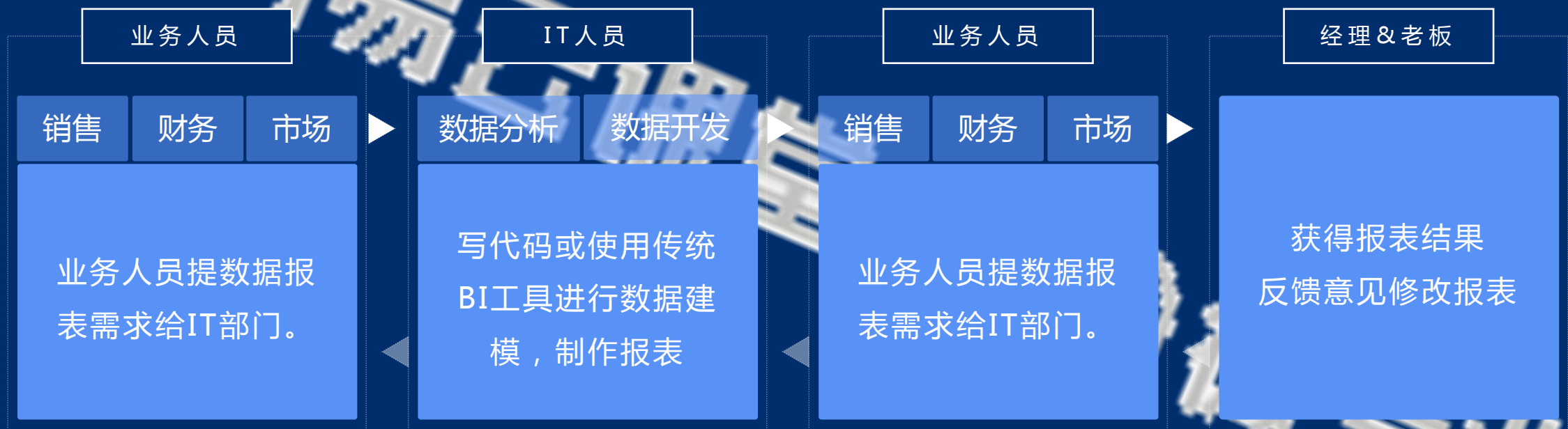
- 数据源
 - 请输入名称
 - 用户基本信息
 - test_second
 - 用户行为日志
 - 用户操作记录
 - 用户登录登录
 - 物流
 - 背景音乐系统
 - 测试方案第3版CMMO
 - 消息数据
 - 数据基础库
 - 数据工具基
 - 数据中间层
 - github
 - oa

The main content area displays a table of data sources:

数据源名称	操作
<input type="checkbox"/> mysql_online_userid	db_mysql_online_userid 取消关联
<input type="checkbox"/> mysql_factory_room_info	mymysql_factory_room_info 取消关联
<input type="checkbox"/> mysql_supervisor_info	mymysql_supervisor_info 取消关联
<input type="checkbox"/> mysql_line_info	mymysql_line_info 取消关联
<input type="checkbox"/> mysql_device_info	mymysql_device_info 取消关联
<input type="checkbox"/> mysql_device_maintain	mymysql_device_maintain 取消关联

网易有数 企业级大数据可视化分析平台

I 大数据分析平台提升分析决策效率

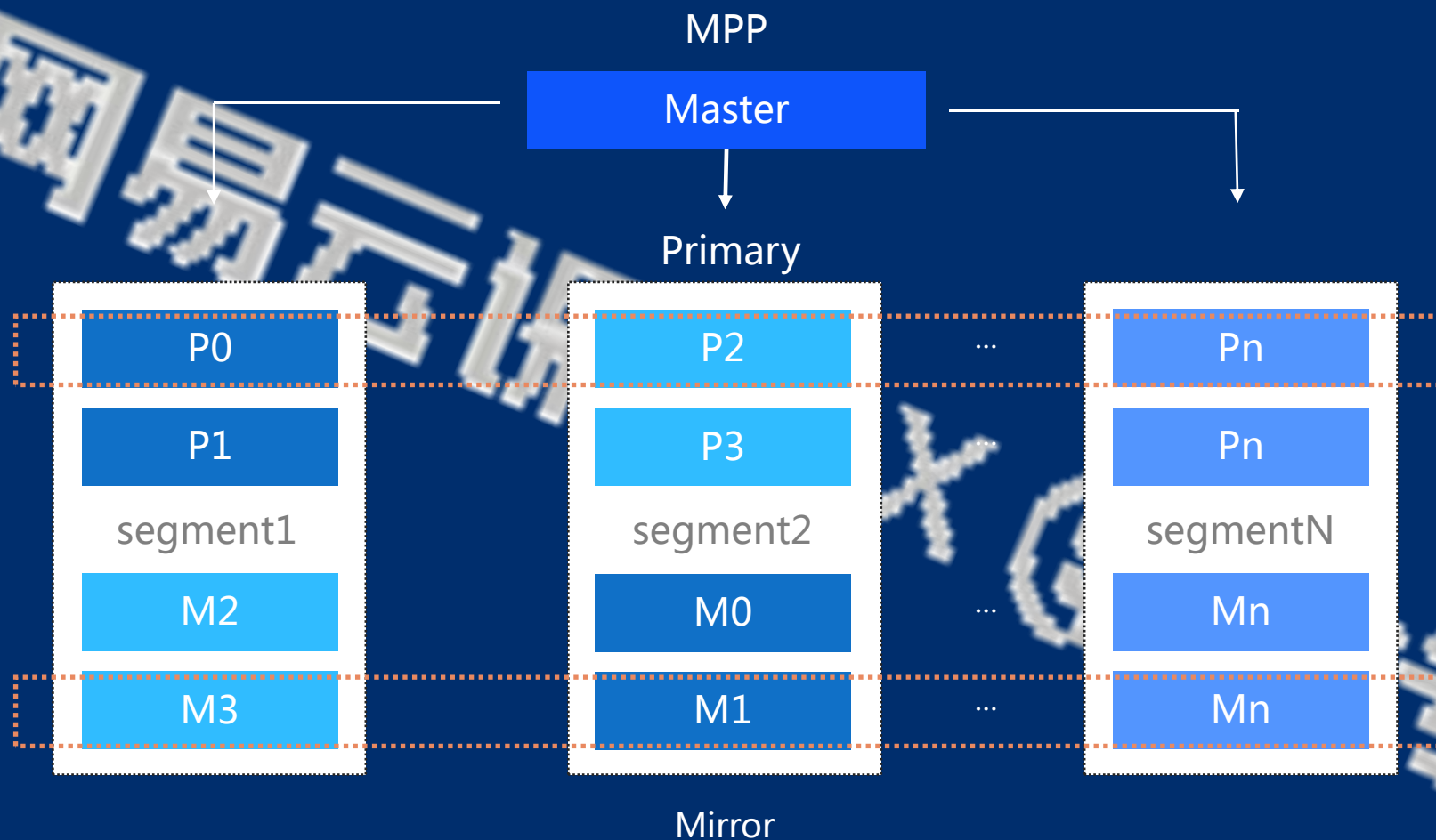


分析流程长

涉及人员多

决策周期长

I 并行计算MPP



- 数据源隔离
- 异构数据源集成
- 强大并行计算

PPT模式的报告制作，更易学，更易用



I 多端展示

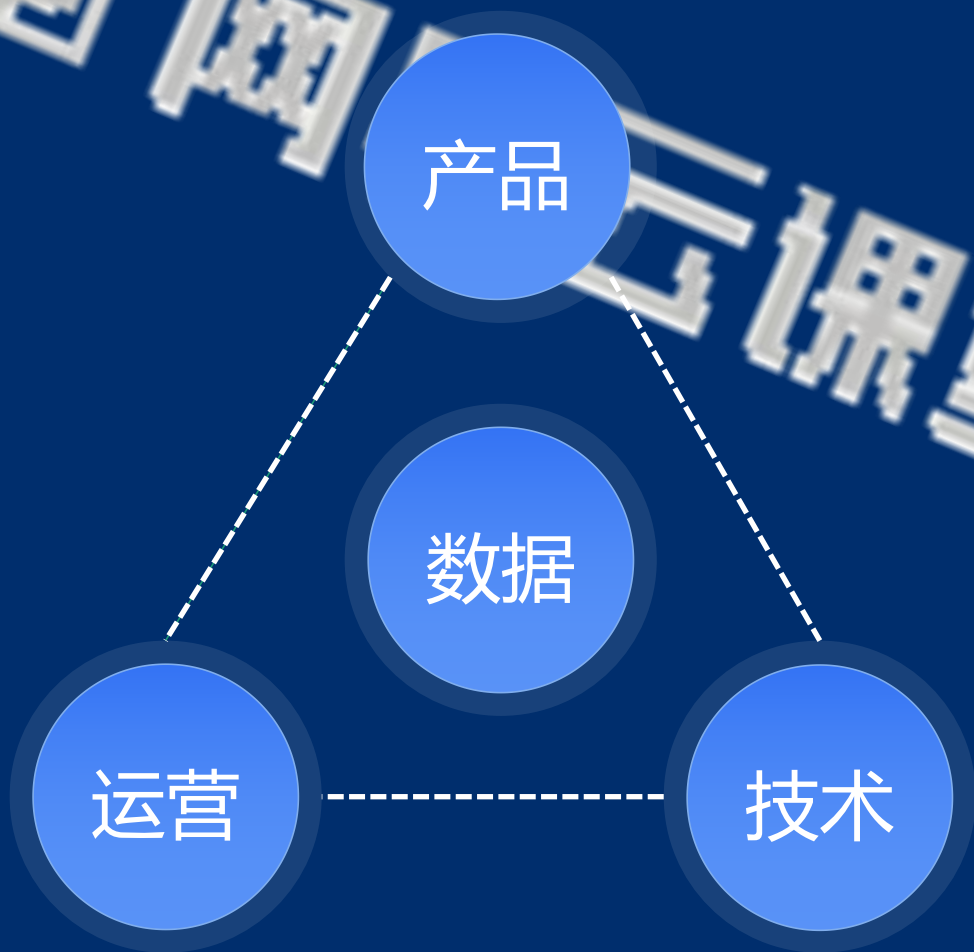


随时随地阅览报表，自适应各种尺寸

PART 03

如何成为一名合格的大数据工程师

I 互联网业务架构

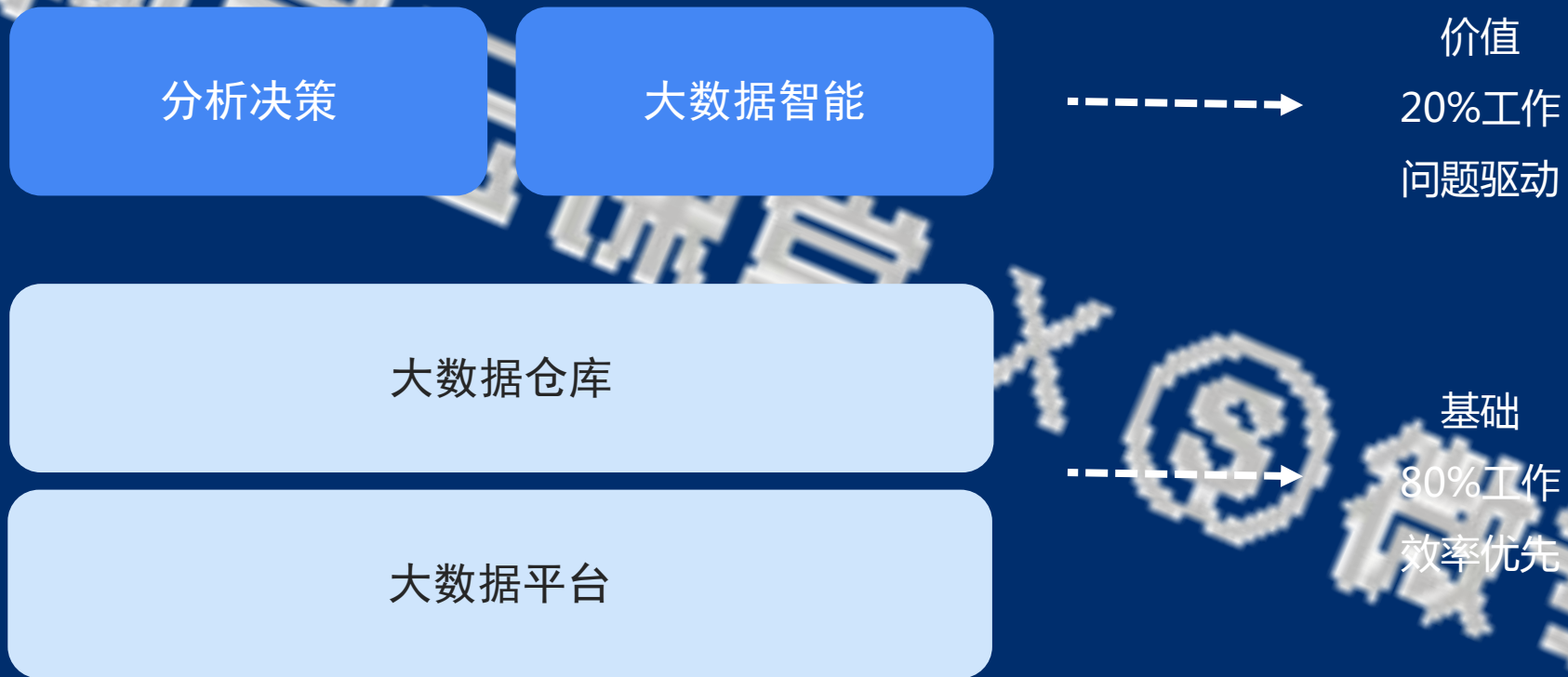


数据团队起到至关重要的作用，并将不断壮大

互联网业务架构

微专业

I 大数据业务架构



I 大数据团队角色及技能栈

数据分析师

- 适当的编程能力
- 数据到业务解读
- 数据分析工具：BI工具（有数、MSTR等）、用户行为分析类等

平台开发工程师

- 分布式存储/分布式计算理论
- 多租户、高可用、数据一致性等

大数据团队

数据开发工程师

- 数据仓库建设方法论
- 编程开发/调优：Spark、Hive、Flink等
- 熟悉大数据平台各组件应用场景

算法工程师

- 数据建模、数据挖掘
- 熟悉机器学习算法应用场景及使用

⑤ 微专业 / 大数据开发工程师

挑战年薪50万，成为大数据工程师

- 完整Hadoop生态组件+真实网易案例+资深工程师经验+网易云私有集群在线实操+全SSD硬盘支持 -



¥2698 原价：¥3800

¥200 07月16日前有效

立即购买

200元优惠券 享受2498元购课

爱学习的你，带上小伙伴一起结伴学习！

有疑问可以加QQ群或者加微信咨询课程（右上角）

I 学员问题

- 1、我想了解数据分析师与数据开发工程师的不同。
- 2、大数据开发工程师的薪水怎么样？平均工资多少？
- 3、想学习大数据的话需要什么基础，分别掌握到什么程度才可以？
- 4、一般面试这方面的工作对学历会有限制吗？可以讲讲会面试哪些东西吗？
- 5、学完课程以后怎么进入大数据行业工作呢？



THANKS

<https://bigdata.163.com/>
bigdata-bd@hz.netease.com