云和恩墨
ENMOTECH

ACOUG All China Oracle User Group 中国 Oracle 用户组 | IT大咖说 知识分享平台

恩墨学院 大数据

YUNHE ENMO (BEIJING) TECHNOLOGY CO.,LTD

云和恩墨 成就所托

姓名
联系方式

1 如何理解大数据

2 如何与客户沟通大数据项目

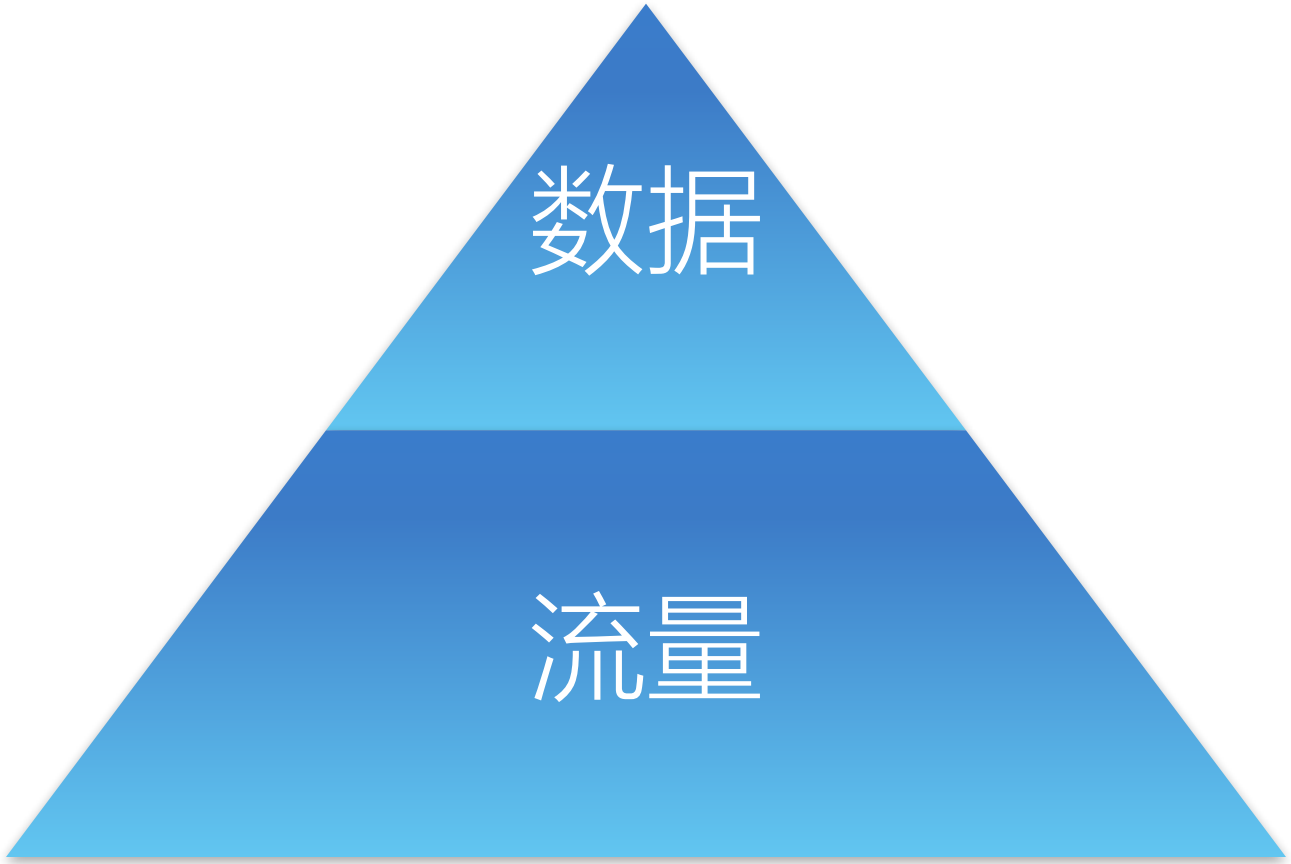
3 为什么来恩墨学习大数据

4 关于公司大数据业务的思考

理解大数据-1：数据的力量

公司	Alphabet	Facebook	腾讯	阿里巴巴	百度
总收入(亿美元)	817.62	179.28	158.41	122.93	102.23
广告营收(亿美元)	732.23	170.83	26.90+87.14	77.04	100.78
广告占比	89.6%	95.3%	17.0+55.0%	62.7%	98.6%

什么是大数据-2：数据变现



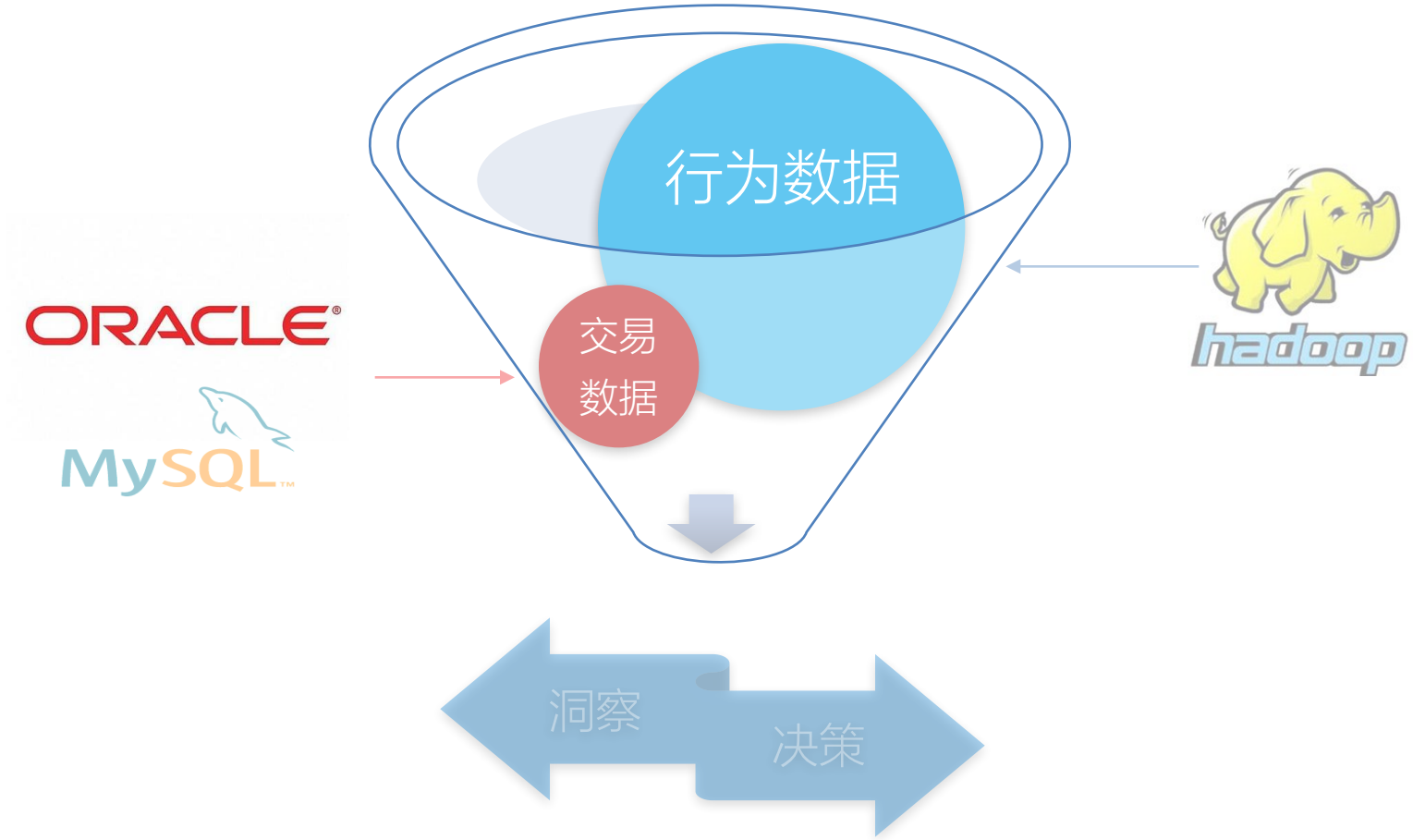
数据的力量-高效变现

100w 点击 -> 1w 广告费

50%男用户广告：1w广告费

%50女用户广告：1w广告费

如何与客户沟通大数据项目



有价值的行为数据

有价值的行为数据
决策行为：阿里
转化（Conversion）：下单购买/购物车
对应着非常明确的用户兴趣，价值最高
阿里数据量小，但是价值高，市值高
主动行为：百度
搜索、广告主动点击、搜索点击
在意图支持下主动产生的行为
百度、谷歌数据量大，价值稍弱，市值没有阿里高
半主动行为：今日头条
分享、网页浏览
量大、用户意图比较弱
被动行为
强加给用户
广告的被动浏览
用户ID*
重要：没有用户ID，数据没有价值
拥有强账号体系利用行为数据的效率高
标识用户的方式
Web/Wap: Cookie
存续性差7-14天：被杀毒软件，手动清除
跨域名时需要映射
解决：Cookie Mapping
IOS：IDFA
存续性好于Cookie
ID For Ad
Android：Android ID
IMEI 设备号，中国可以用
FingerPrint（IP+User Agent）
弱用户标识，不得已为之
社交关系
用户兴趣，用户画像的平滑

恩墨学院大数据课程

课程、课时、认证	课程概述
管理员培养路径	
Apache Hadoop 管理员培训 培训时长：4 天 CCA Hadoop 管理员	从安装及配置、负载均衡及调整，以及 诊断和解决部署问题等各方面了解 Hadoop 系统管理员的概念和实践。
Apache HBase 培训 培训时长：3 天	学习如何运用 HBase 作为分布式数据存 储实现低延迟查询和高度可扩展性的数据吞 吐量，涵盖 schema 设计、应用程序编写、配 置和维护。
数据分析师培养路径	
数据分析师培训： Pig、Hive 和 Impala 培训时长：4 天 CCA 数据分析师	了 解 Apache Pig、Apache Hive 及 Apache Impala（孵化中）如 何 通 过 筛 选、连接以及用户自定义函数进行数据转换和数据分析。
Spark 和 Hadoop 上的数据科学 培训时长：3 天 CCP 数据科学家	了解数据科学家的工作内容、需要解决 的问题以及其处理现实问题时应用的方法，以获得来自不同行业数据的业务价值。本课程中学员将学习并实现一个推荐系统。

课程、课时、认证	课程概述
开发人员培养路径	
Spark 及 Hadoop 开发员培训 培训时长：4 天 CCA Spark 及 Hadoop 开发员	了解 Apache Spark 的基础知识及其与 Hadoop 整体生态系统的集成方式。本课程 将重温 HDFS 的基础内容，学习如何使用 Sqoop/ Flume 摄取数据，利用 Spark 处理分 布式数据，学习在 Impala 和 Hive 上数据建模，以及在数据存储方面的最佳实践。
Apache HBase 培训 培训时长：3 天	学习如何运用 HBase 作为分布式数据存 储实现低延迟查询和高度可扩展性的数据吞 吐量，涵盖 schema 设计、应用程序编写、配 置和维护。
Cloudera 检索培训 培训时长：3 天	索引 Hadoop 数据和强有力的实时查询功能，并且可将 Cloudera 检索与外部应用程序进行集成。了解如何将全文本、交互式检索和可扩展的、灵活的索引导入到 Hadoop 和企 业数据中心（EDH）中去。

认证	概述	推荐学习
CCA Spark 和 Hadoop 开发员	CCA Spark 和 Hadoop 开发员需具备作为核心开发人员的技能水平，并且可以编写和维护 Apache Spark 和 Apache Hadoop 项目。	• Spark 及 Hadoop 开发员培训
CCP 数据工程师	CCP 数据工程师需具备开发可靠、自主、可扩展的数据流水线的技能，可为各种工作负载产生优化的数据集。	• Spark 及 Hadoop 开发员培训 CCP 数据工程师候选人员应该具备大数据工具使用经验、深厚的 Hadoop 知识、以及专家级的解决实际数据工程问题的能力。
CCP 数据科学家	CCP 数据科学家需具备专家精英群体在大数据工作方面的技能。候选人员须证明解决实际问题的能力，能设计并开发可用于产品环境的数据科学解决方案，并且该解决方案经同行评估认定具备准确性、可扩展性和稳健性。	• Spark 及 Hadoop 开发员培训 • Cloudera CCP 数据科学家准备套件 CCP 数据科学家候选人员应该具备大数据工具使用经验、深厚的 数据科学方面的知识、以及专家级的解决实际数据科学问题的能力。
CCA Hadoop 管理员	CCA Hadoop 管理员需具备配置、部 署、维护和保护公司和组织机构欲部署 CDH 所要求的 Apache Hadoop 集群方面的知识。	• Apache Hadoop 管理员培训
CCA 数据分析师	CCA 数据分析师需具备在基于 Cloudera CDH 部署的大数据平台上使用 Hive 和 Impala 进行分析并产生报告的技能。	• 数据分析师培训：Pig、Hive 和 Impala

合作过的企业客户

关于公司大数据业务的思考

云和恩墨
ENMOTECH

数据驱动 成就未来

Make Your Data Dance