

# 百分点大数据理念和实践

2016.02

# 一些大数据定义

大数据指的是所涉及的数据量规模巨大到无法通过传统方式，在合理时间内达到截取、管理、处理、并整理成为人类所能解读的信息。

《维基百科》

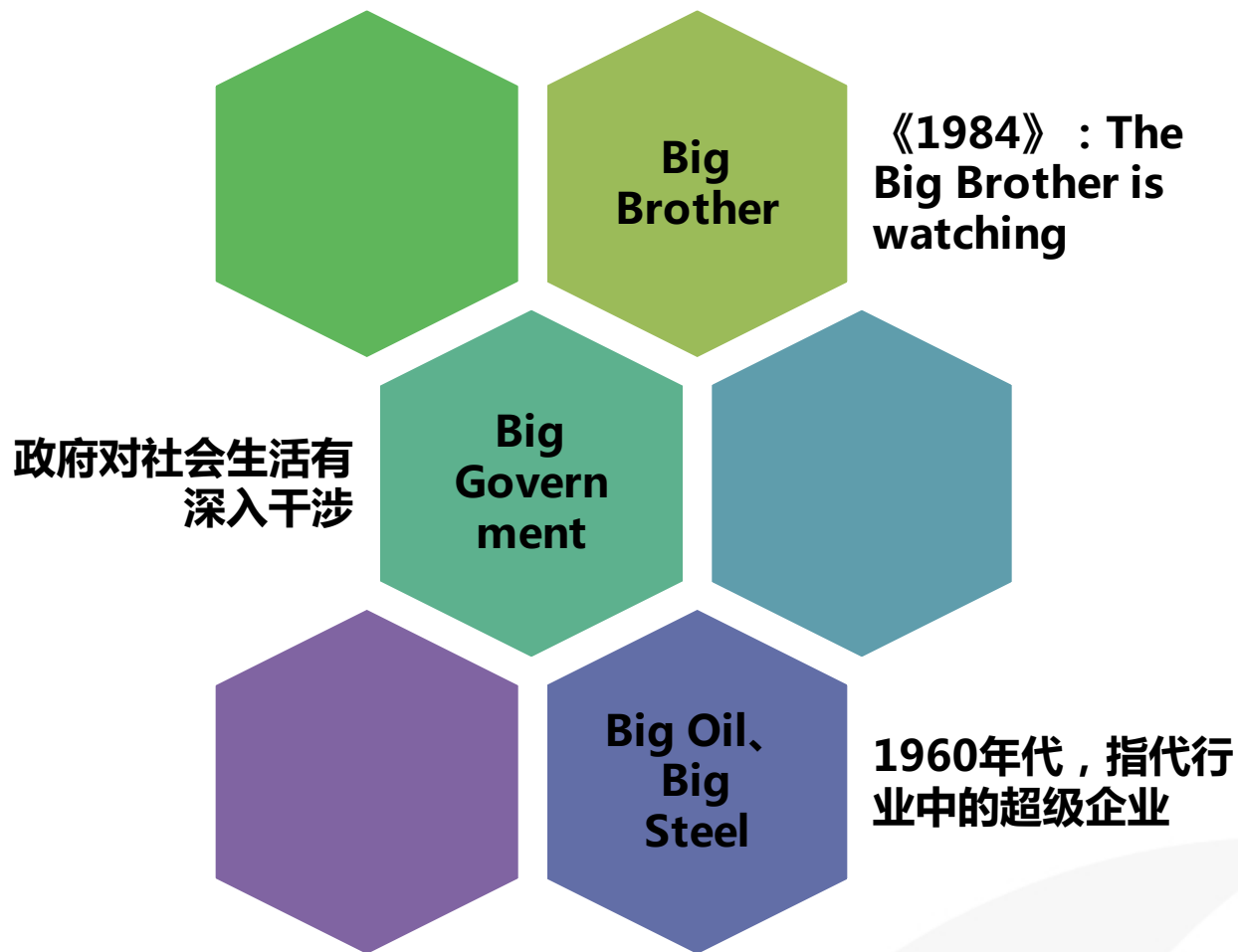
大数据指人们在大规模的数据基础上可以做到的事情，而这些事情在小规模数据的基础上是无法完成的。

《大数据时代：生活、工作与思维的大变革》，2012

大数据集是“庞大的、多样化的、复杂的、纵深的和/或分布式的,它由各类仪器设备、传感器、网上交易、电子邮件、视频、点击流，以及现在与未来所有可以利用的其他数字化信号源产生”。

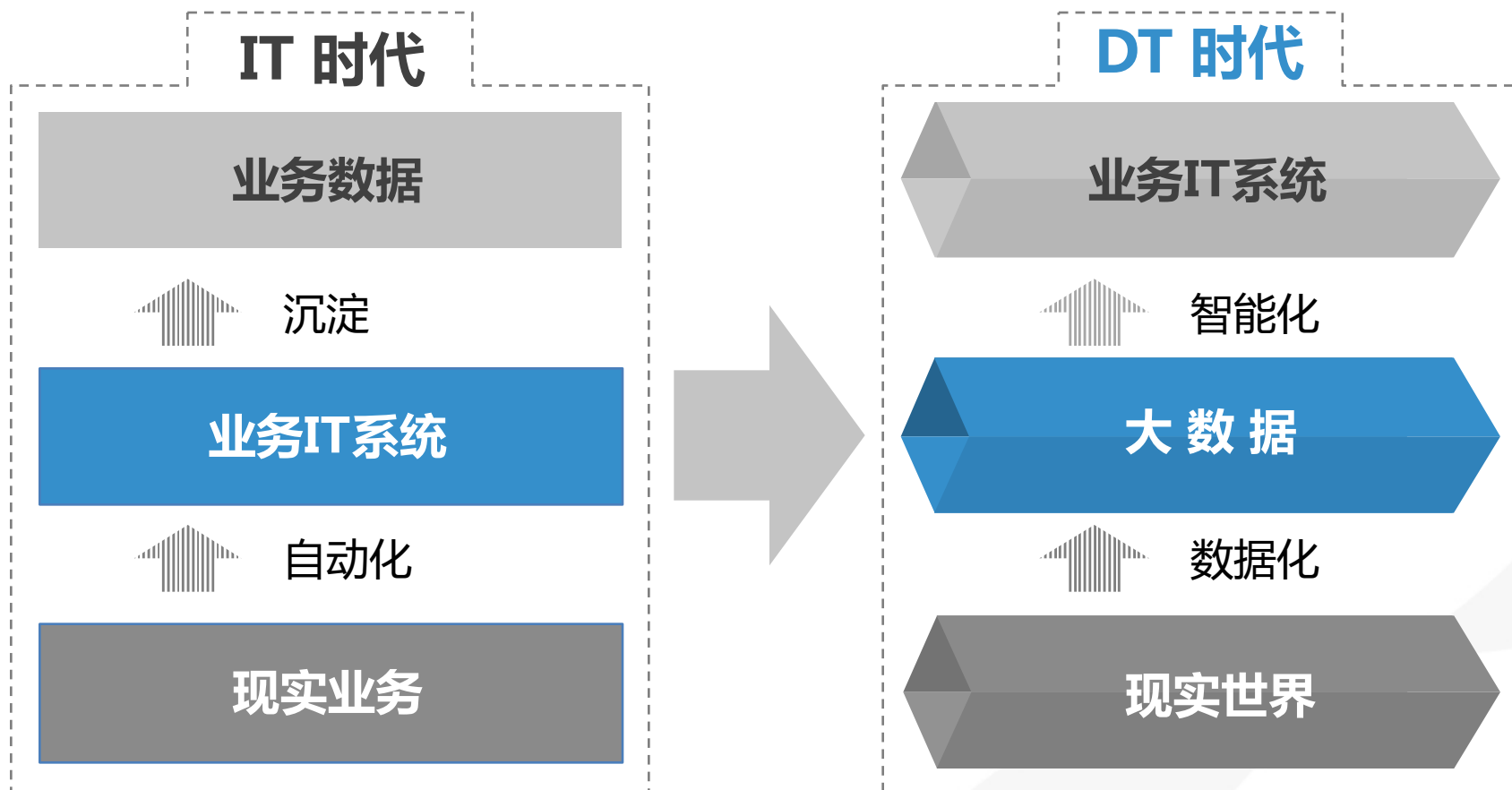
白宫《大数据白皮书》，2014

# “Big Data” 之 “BIG”



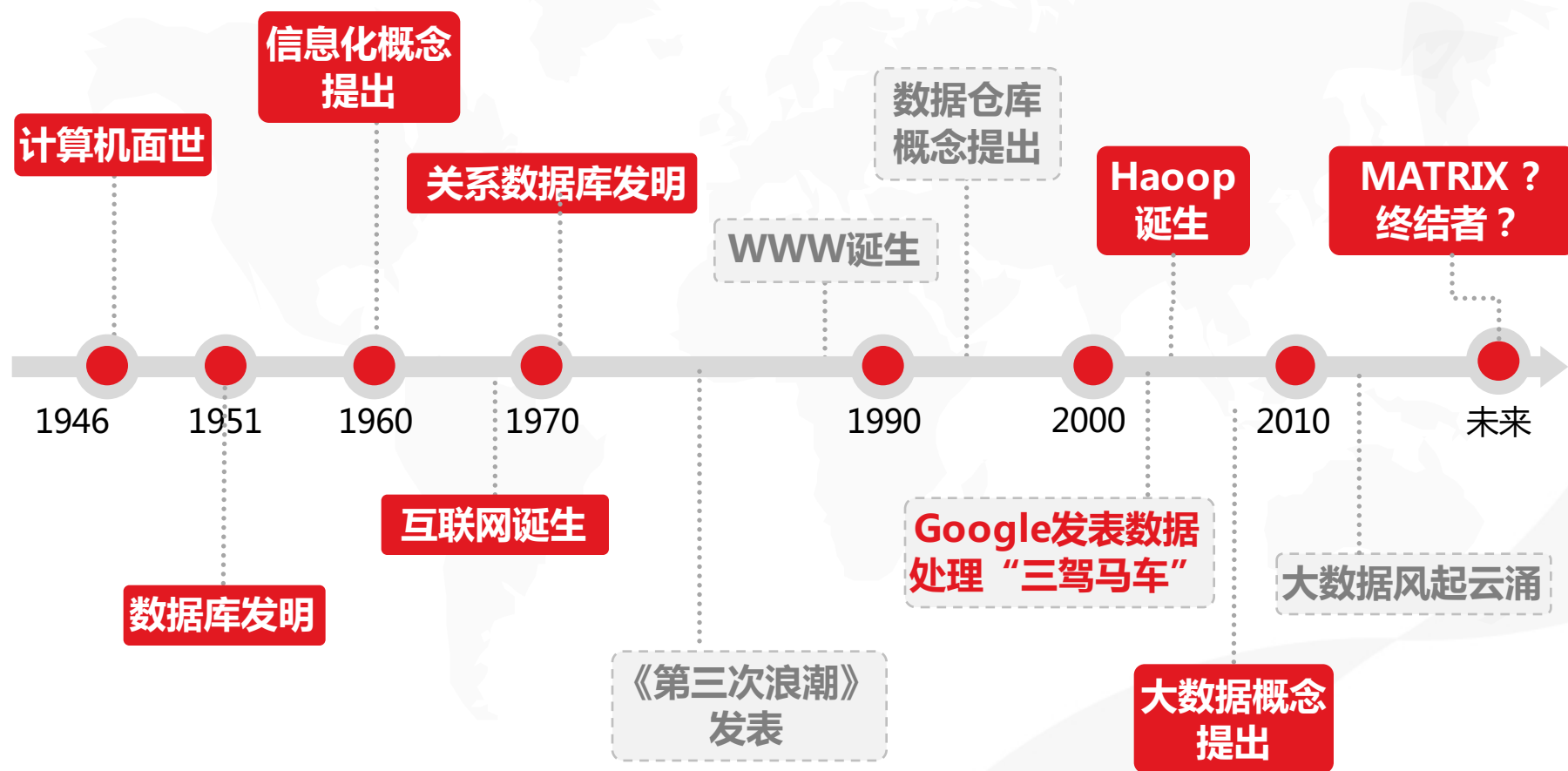
# “Big Data” 之 “BIG”

大数据是信息化技术的自然延伸，意思是无处不在的数据



# 大数据时代大事记

大数据是一个强技术的概念，与Hadoop技术密不可分



# 什么是大数据：我的答案

- 对事物发展全面细致的**数据化**的结果
  - 量化一切，记录一切
  - 多侧面，有深度
  - 不为特定目标而记录
- 海量数据的管理应用的**技术和方法**
  - 采集、存储、追溯、查询、分析、挖掘

# 大数据的特点

## Volume 体量

- 事无巨细的数据记录
- 过程数据远多于结果数据（传统数据）

## Velocity 速度

- 数据稍纵即逝，必须实时采集
- 商机时不再来，必须实时应用

## Variety 多样

- 多种数据来源，内容五花八门
- 多种数据结构，标准无法统一

## Value 价值

- 分析造成“结果”的“原因”
- 发现潜在的“结果”

# “云大物移社”



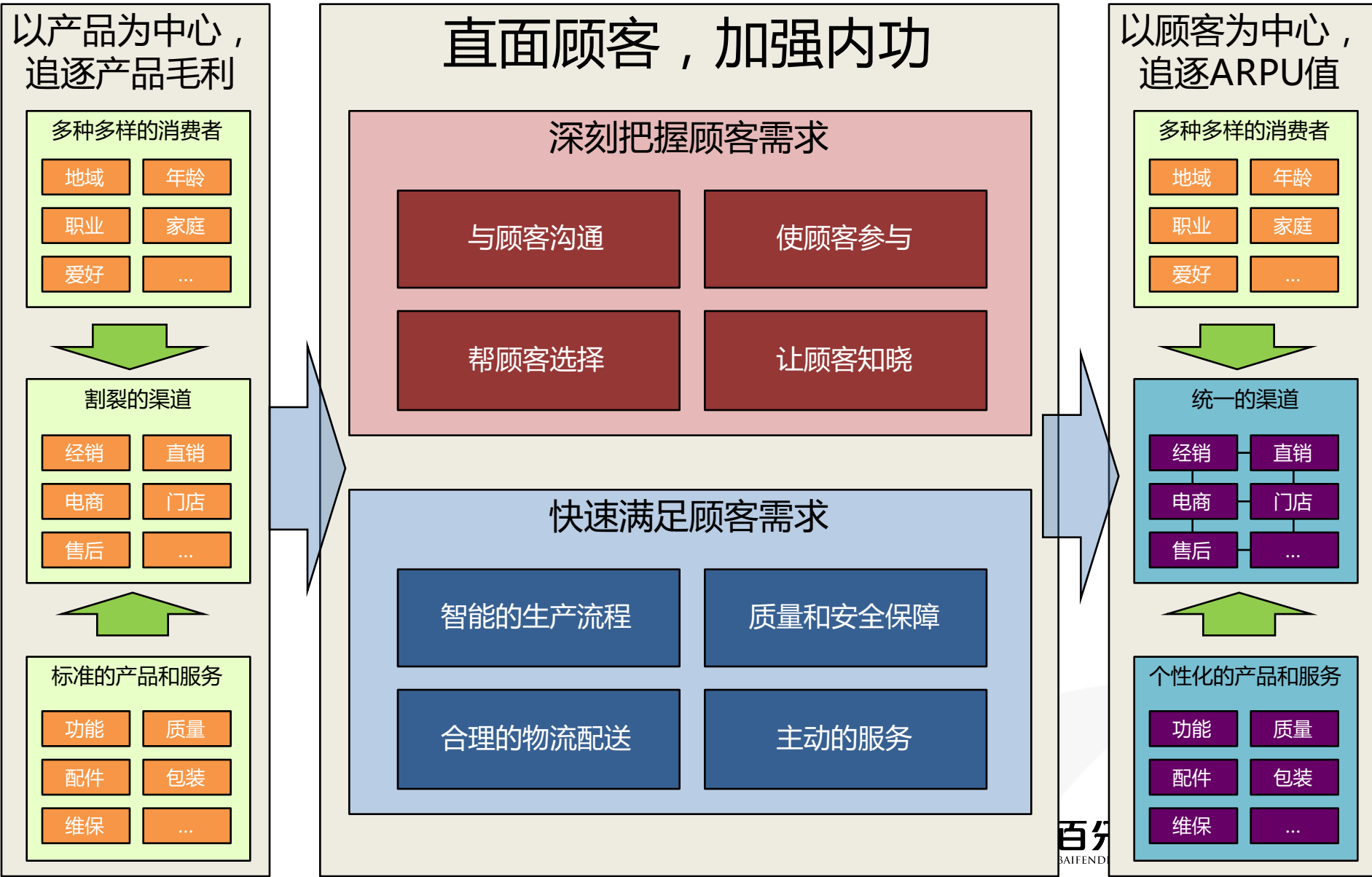


- 
- 云计算 是 骨骼
  - 物联网 是 感官
  - 大数据 是 大脑
  - 移动化 是 策略
  - 社会化 是 灵魂

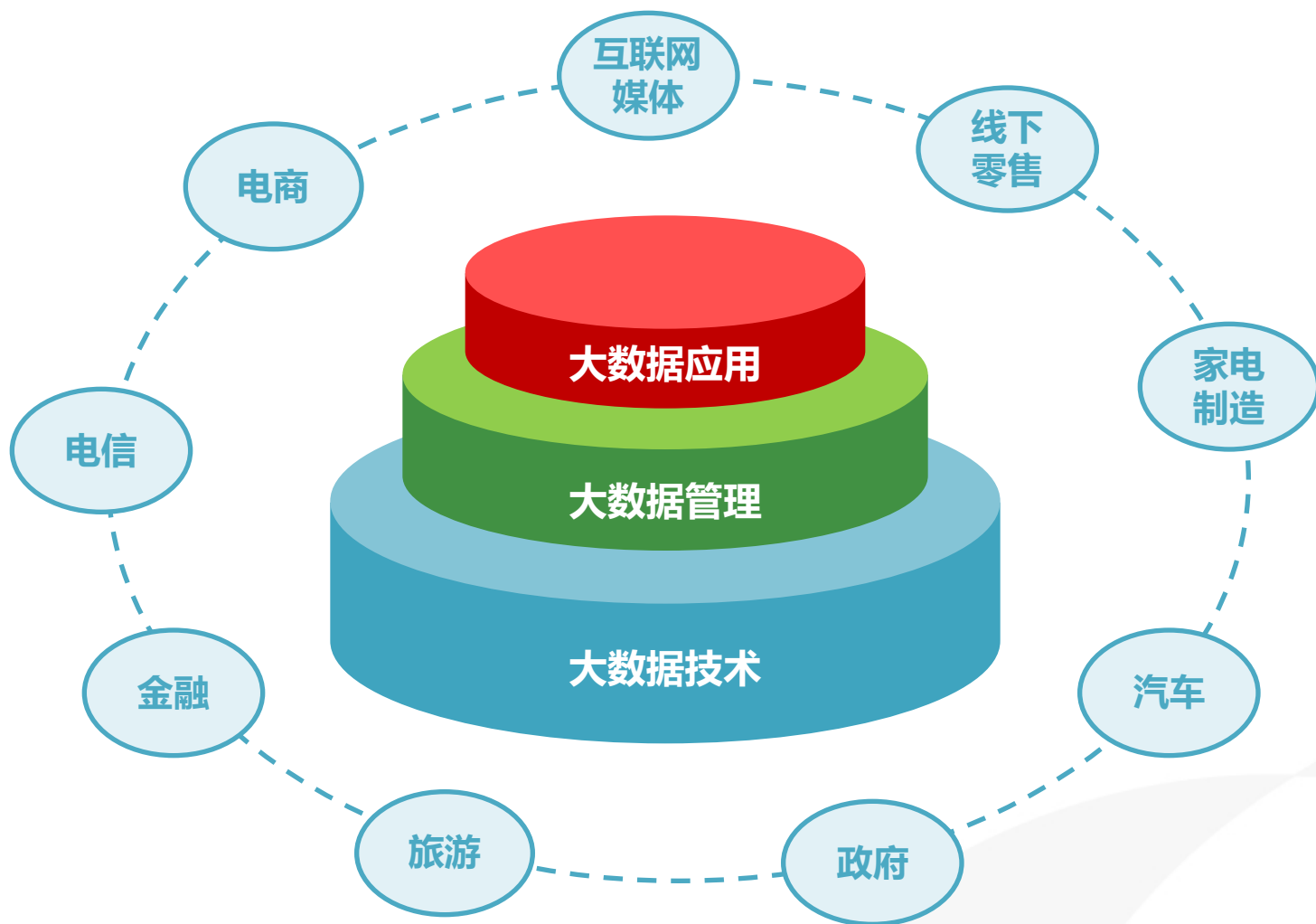
# 全球大数据生态格局

基础	分析	应用	数据
<b>Hadoop:</b> Cloudera MapR	<b>平台:</b> databricks dataspora	<b>广告:</b> eXelate DataXu	<b>数据市场:</b> bluekai DataMarket factual Azure Marketplace
<b>NoSQL:</b> Cloudbant Couchbase	<b>非结构化:</b> Palantir Quid	<b>营销:</b> Lattice Gainsight	<b>数据源:</b> quandl premise xignite plaid
<b>数据管理:</b> Oceansync Datadog	<b>可视化:</b> visual.ly actuate	<b>金融:</b> Zestfinance Lendup	
<b>数据安全性与存储:</b> Stormpath Nimblestorage	<b>日志:</b> splunk loggly	<b>教育:</b> Panorama Knewton	
开源框架	开源数据库	开源计算	开源其他
Hadoop, Yarn Storm, Spark	HBase, MongoDB CouchBase, Neo4J	Mahout, Wabbit GraphLab, R	Zookeeper, Pentaho Talend

# 百分点的核心：将消费者数据化，驱动智能经营



# 大数据解决方案

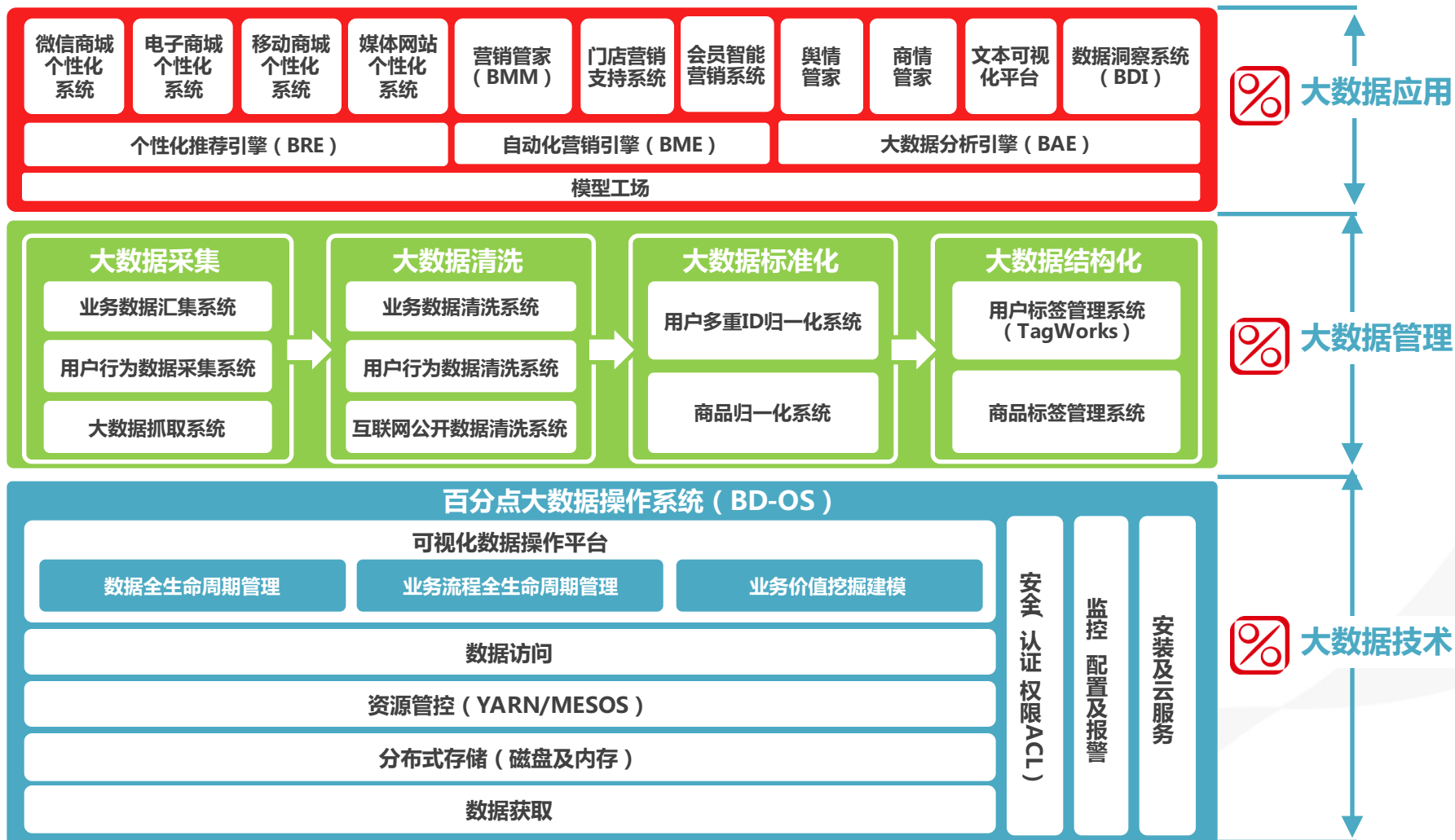


# 用户画像体系：银行版

支持应用场景



# 大数据产品全景图



# 践行于大数据最前沿



Big Data Practitioner

