



第十届中国IDC产业年度大典
The 10th Internet Data Center Conference

大数据产业与政策展望

魏凯
中国信息通信研究院
2016年1月7日



1. 产业观察
2. 政策展望
3. 我院工作

2013-2014年：沸腾 buzzing



IDCC

0 大数据将成为信息时代下一个引爆点

1 人人都想要它



2 人们总是夸张得谈论它



3 每个人都以为别人正在



7 除了上帝, 任何人的话都必须经过事实说话, 大数据浪潮, 更是生活... 管理学之父德鲁克有言, 预测未来... 则是当下... 大数据... 何以既便利又危及我们每个人的生活? Good

云计算
移动社交
数据可视化

Big Data

The Industry

将成为信息时代下一个引爆点

Surge Tipping Point

5 但其实只有很少的人在做



6 更少的人知道应该如何正确去做

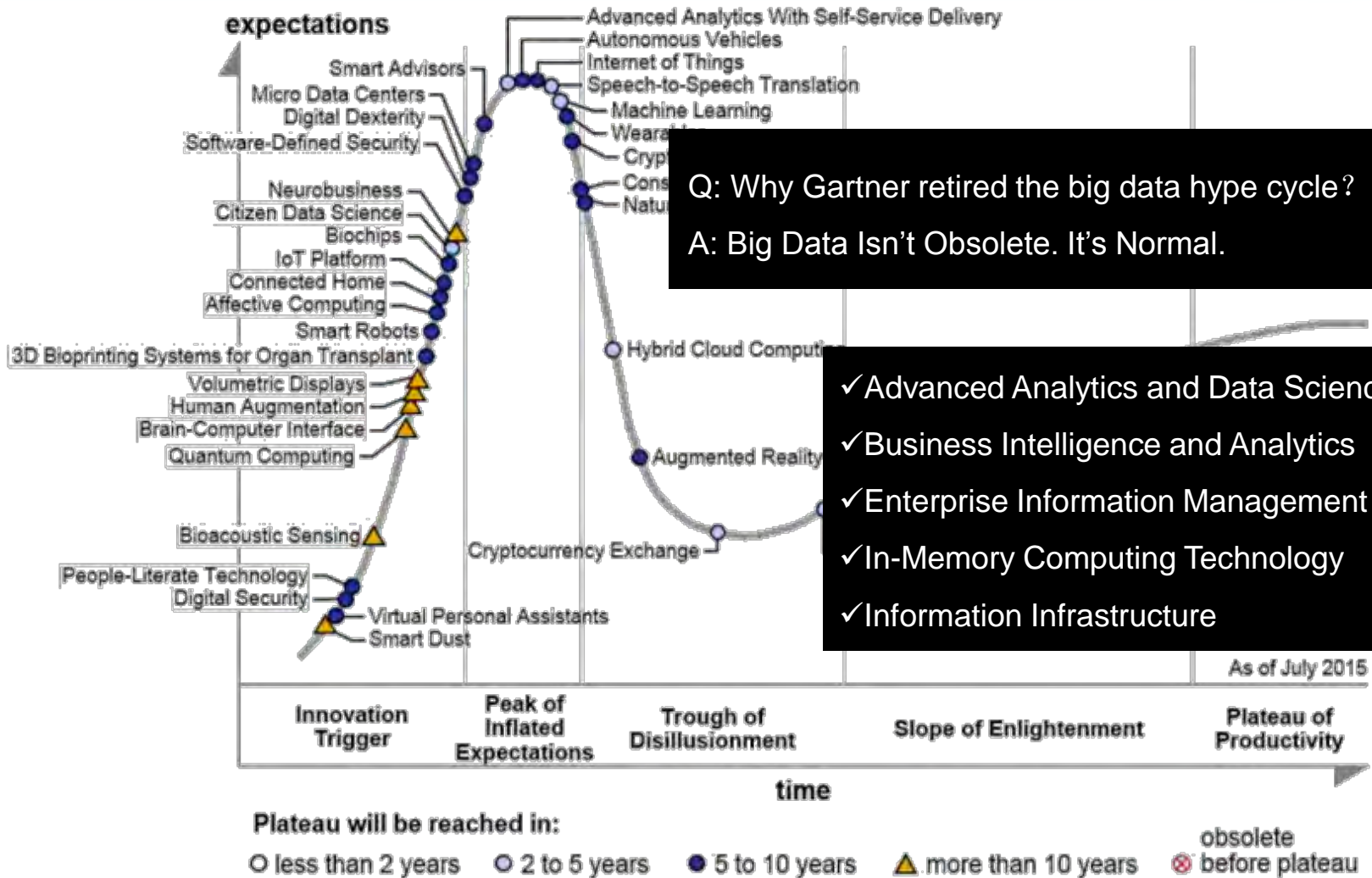
MPP, MapReduce, GFS, data-mining grids, cloud based infrastructure, A/B testing, association rule learning, classification, cluster analysis, crowd-sourcing, data fusion and integration, ensemble learning, genetic algorithms, machine learning, natural language processing, neural networks, pattern recognition, anomaly detection, predictive modelling, regression, sentiment analysis, signal processing, supervised and unsupervised learning

又是一场炒作?

2015年-国外再热炒?



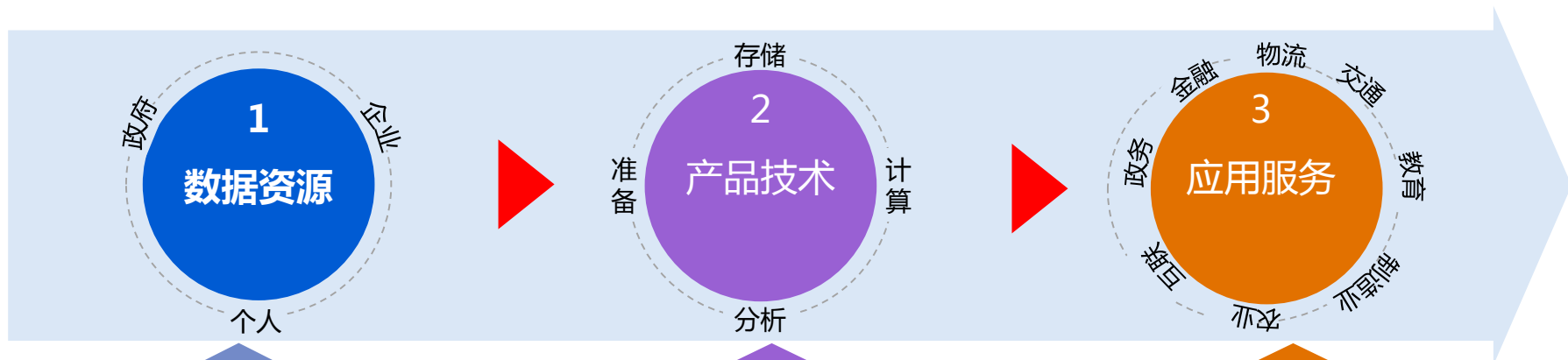
IDCC



产业体系-数据资源+产品技术+应用服务



IDCC



(一) 数据资源提供商

任何企业都可成为数据公司

各种数据共享、开放和交易平台机构

数海、数据堂、九次方大数据、贵阳大数据交易所.....

数据拥有者

新浪微博大数据、微信大数据、四维图新、慈铭体检、大众点评、恒生电子、考拉征信.....

(二) 产品技术提供商

开源厂商和传统闭源厂商

应用软件

西安美林数据、龙信数据、用友、金蝶、东软.....

基础软件

星环、明略、拓尔思、亚信、科大讯飞、爱数.....

硬件

华为、联想、浪潮、曙光

技术服务运维支持

(三) 应用服务提供商

大众创业、万众创新

政府应用

神州数码、易华录.....

行业应用

船讯网、荣之联、随手房、春雨医生、华大基因.....

企业应用

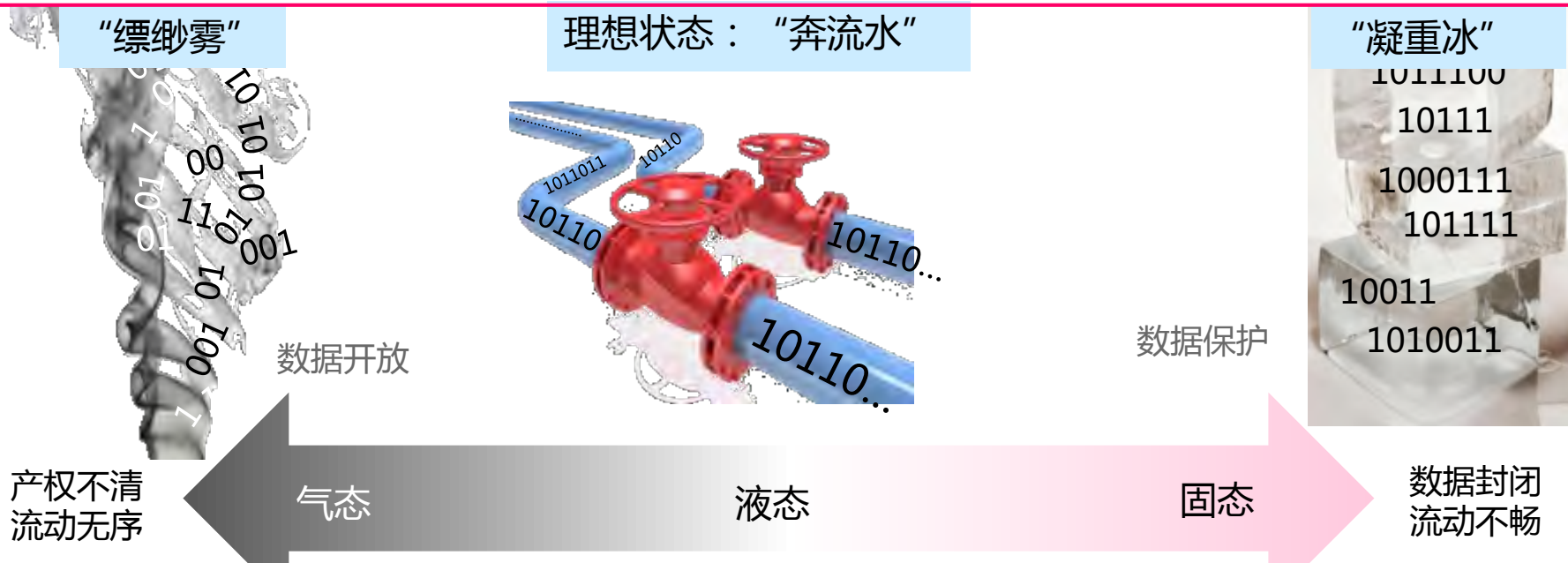
亿赞普、百分点、秒针.....

数据“结构性短缺”普遍存在



IDCC

我国数据资源两大问题：企业线上数据产权不清，流动无序。政府及公共机构数据封闭使用，流动不畅。



- 数据资源权益不清
- 个人数据外泄严重
- 跨境数据流动问题凸显

- 政府及公共机构数据开放程度低
- 企业稳定获取公共数据困难
- 公共信息资源的价值难以释放

纠结-数据的责权利如何界定？



IDCC



政府

事业单位

国企

其他企业和个人

关系国家安全的数据

关系商业秘密的数据

关系个人隐私的数据

不含以上内容的数据

我国确认数据权利归属面临的困难

1、——数据归属单位不清。和国外不同，我国大量公共信息资源实际掌握在公共部门手中，公共部门为数据的采集、软硬件系统、运维付出了一定的成本，甚至为政府部门维护数据库。

按照网信办的内部指导意见，我国公共信息资源责任的主体包括4类：1) 负有公共事务管理职能和公共服务职能的政府部门、2) 市政公用企事业单位、3) 公共服务事业单位、4) 经政府部门授权实施公共服务的企业单位。

2、——缺乏完善的个人信息保护法律体系。
个人数据、企业数据、公共数据的边界不清晰

规则缺失制约数据流通交易



IDCC

| 序号 | 单位名称 | 进展情况 | 备注 |
|----|---------------|--|---|
| 1 | 北京数据交易所（筹） | 北京市金融工作局牵头筹建，近期将向北京市政府汇报筹建方案。 | 划拨现有闲置交易所改制。 |
| 2 | 上海数据交易所（筹） | 1、经营范围：商业公司互联网数据产品（一期）。 2、业务模式：使用权交易；多源数据聚合；实时在线交易。 | 2015年内成立上海数据交易中心，2016年筹建上海数据交易所。 |
| 3 | 广州数据交易所（筹） | 广州市大数据管理局牵头筹建。 | 一期成立广州数据交易中心，省政府报批流程中。 |
| 5 | 贵阳大数据交易所 | 1、业务指导单位：贵阳市大数据办、贵阳市经信委； 2、经营范围：政府数据、金融数据、征信数据； 3、业务模式：产权交易；线下交易为主。 | 1、存在问题：政府数据占到总量的60%以上；直接参与数据加工与定价；数据产权争议；个人隐私保护争议。 2、组建形式：贵州阳光产权交易所参股。 |
| 6 | 长江大数据交易所 | 1、业务指导单位：武汉市经信委； 2、经营范围：数据产品（含政府数据产品）、模型算法。 3、业务模式：倡导使用权交易。 | 组建形式：武汉光谷知识产权交易所参股。 |
| 7 | 陕西大数据交易所 | 1、业务指导单位：陕西省工信厅、西安西咸新区； 2、经营范围：政府数据、工业数据、个人数据等。 | |
| - | 其它正在筹建的数据交易机构 | 8、徐州大数据交易所（已成立）；9、江苏省大数据交易中心（即将成立）；10、重庆大数据交易所（重庆市经信委）；11、沈阳大数据交易中心（沈阳市大数据管理局）；12、哈尔滨大数据交易中心；13、青岛航运数据交易所；14、河南数据交易机构；15、其它省市和地区（如济宁、东莞、新疆中亚商品交易中心等） | |

大数据技术还在快速演进中



IDCC

SQL/ACID

NoSQL/CAP

SQL/NewSQL

面向企业应用的商业智能提出，面向数据仓库的数据分析 (OLAP) 技术兴起

开源Apache Hadoop逐渐兴起，大幅推动互联网大数据应用

Spark、Flink等新一代融合大数据平台，SQL on Hadoop兴起

1995

2000

2005

2010

2015

分布式技术提出

谷歌提出分布式文件系统、分布式数据库和分布式计算框架，奠定大数据技术基础

实时计算技术提出

流计算、图计算、交互式分析、内存计算等技术不断演进

前大数据时代

大数据技术创新格局



IDCC

国际格局

阵营一
理论原创地



输出：原创论文

分布式系统+人工智能为代表的系列论文

| 领域 | 论文数 |
|------------|-----|
| 人工智能和机器学习 | 347 |
| 算法与理论 | 319 |
| 人机交互与可视化 | 237 |
| 自然语言处理 | 216 |
| 数据管理与挖掘 | 212 |
| 机器识别 | 202 |
| 信息检索和Web | 158 |
| 软件系统 | 119 |
| 分布式系统与并行计算 | 110 |
| 安全与隐私 | 157 |

影响

阵营二
技术制高点



输出：开源+技术服务

Hadoop、Hive、Storm、Spark、Flink等大数据开源技术

Apache Hadoop
Cassandra (NoSQL)
Apache Thrift(跨语言框架)
Apache Hive(Hadoop分析)
RocksDB(Flash DB)
Presto (SQL on Hadoop)
Open Compute(数据中心)
Apache Storm(实时计算)
Open Academy(培训项目)
Apache Spark
Apache Flink...

影响

阵营三
产业先锋队



输出：商用产品

Hadoop/Spark两大商用技术平台和一体机

传统数据库数据仓库软件

Hadoop/Spark发行版的混搭解决方案

我国

?

?

我国大部分企业处于该阵营的跟随者，但比较分散，尚未形成实力

互联网企业热



- 大互联网企业纷纷把大数据作为战略：中国互联网前十名全部有“大数据战略”：百度大脑、阿里“从IT到DT”、腾讯“大数据连接的未来”、京东大脑、小米“大数据不赚钱就破产”……
- 大批创业公司也以互联网大数据为创投方向：秒针系统、百分点、闪银奇异、宜信、上海拍拍贷、北京云知声……

政府大数据产业



布局拉开序幕

政府热

- 23个省出台了74个与大数据相关的政策和规划。上海、贵州、重庆、天津等省市出台大数据发展规划，
- 全国已建大数据产业园区超过10个。
- 呈现中央高度重视、地方积极部署的良好局面。

行业：大数据应用整体上处于起步阶段，快速渗透，各行业渗透程度不一

应用：以银行业为例



IDCC





1. 产业观察
2. 政策展望
3. 我院工作



习总书记指出“世界进入以信息产业为主导的新经济发展时期”
基于信息网络技术融合而成的新兴产业将以新经济引领新常态。

2015年，国务院围绕信息通信出台9个文件

——数据资源领域

“运用**大数据**加强对市场主体服务和监管”
“促进**大数据**发展行动纲要” 两个文件

——技术基础领域

“云计算创新发展培育信息产业新业态”、
“提速降费” 两个文件

——应用创新领域

“互联网+”、“大力发展电子商务” 两个文件，
“大众创业万众创新” 三个文件

十八届五中全会确定实施“国家大数据战略”



IDCC



国家大数据战略

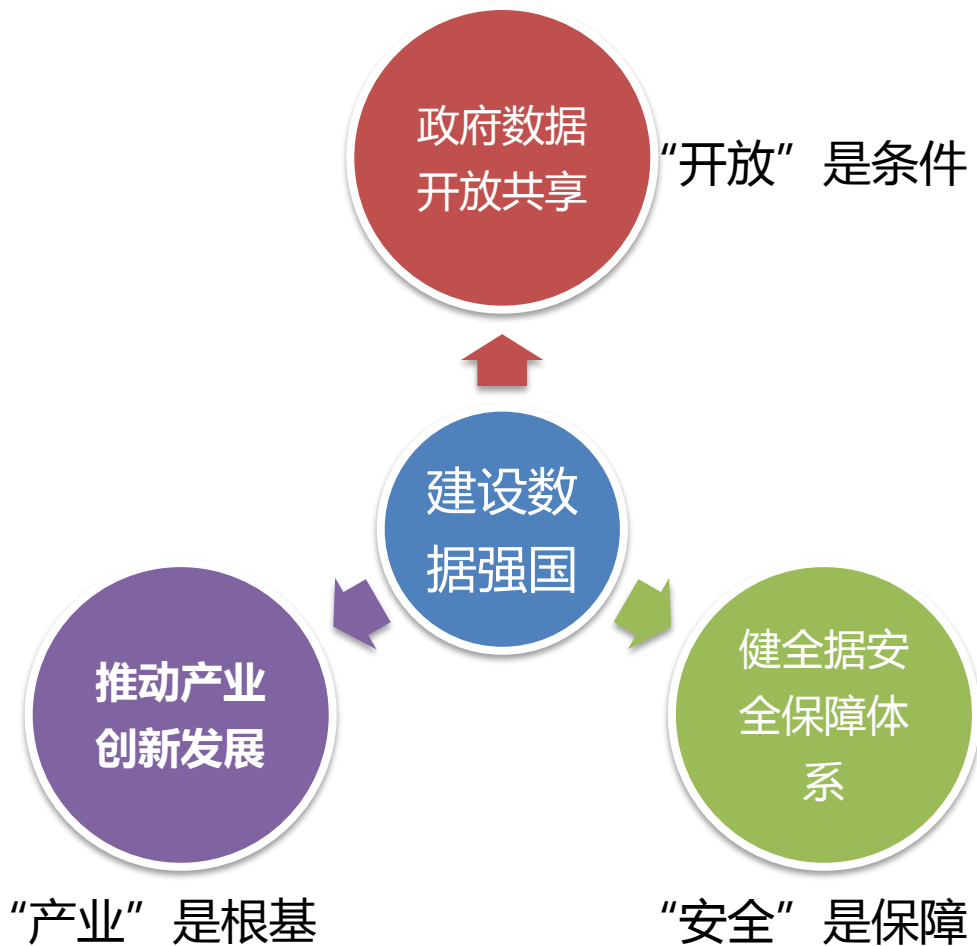
中国制造2025

互联网+行动计划

网络强国战略

十八届五中全会提出要实施网络强国和大数据战略，实施互联网+行动计划，实施《中国制造2025》等举措
“国家大数据战略”是一个核心

三位一体建设数据强国



产业发展是《行动纲要》重要内容



IDCC

《促进大数据发展行动纲要》

全面推动我国大数据发展和应用，加快建设数据强国

加快政府数据开放
推动资源整合
提升治理能力

推动产业创新发展
培育新业态
助力经济转型

健全安全保障体系
强化安全支撑
提高管理水平

政府数据资源共享开放工程

国家大数据资源统筹发展工程

政府治理大数据工程

公共服务大数据工程

工业和新兴产业大数据工程

现代农业大数据工程

万众创新大数据工程

大数据关键技术及产品研发与产业化工程

大数据产业支撑能力提升工程

网络和大数据安全保障工程

完善组织实施机制

加快法规制度建设

健全市场发展机制

建立标准规范体系

加大财政金融支持

加强专业人才培养

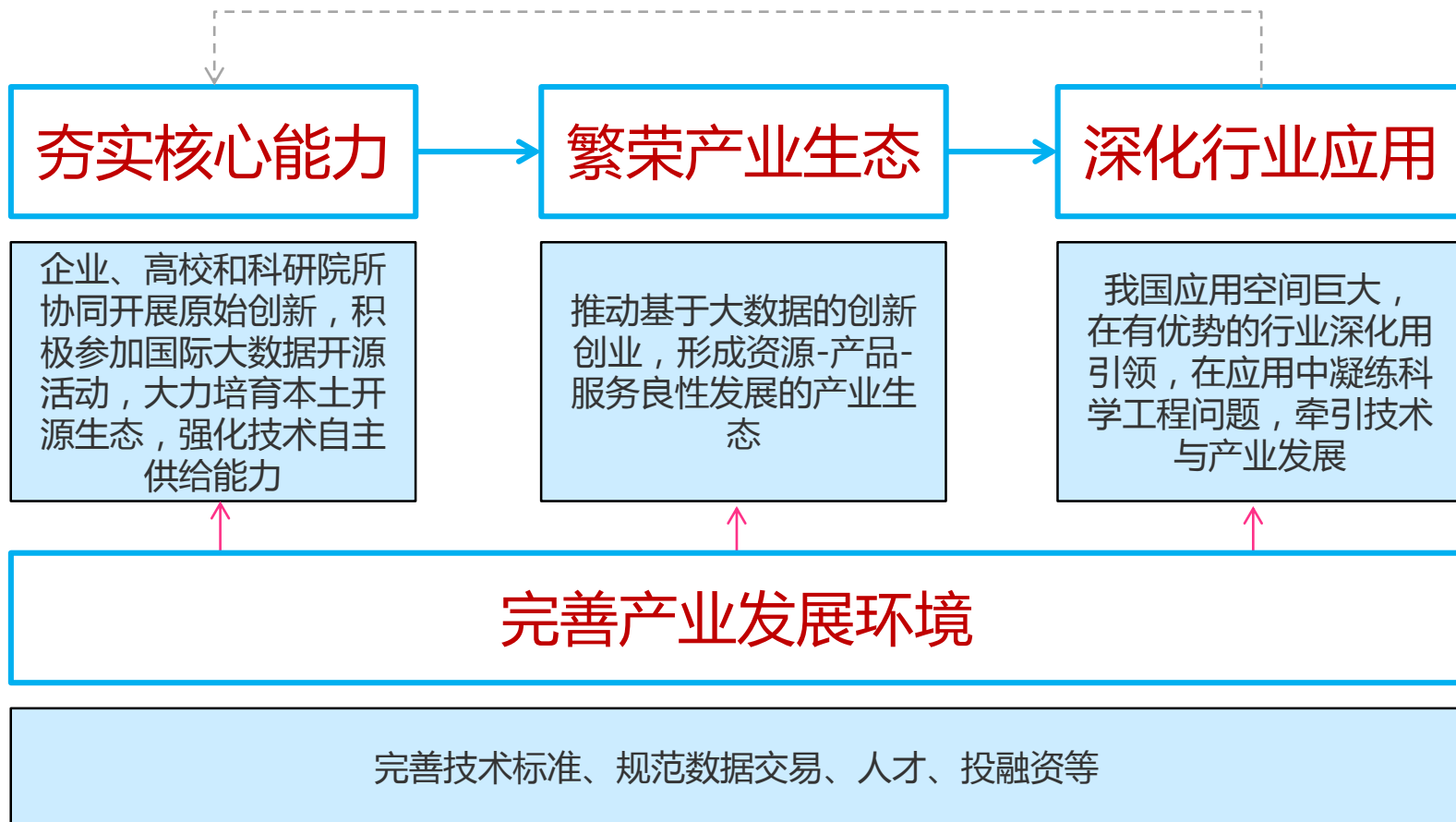
促进国际交流合作

五个主要目标：

- 一是打造精准治理、多方协作的社会治理新模式
- 二是建立运行平稳、安全高效的经济运行新机制
- 三是构建以人为本、惠及全民的民生服务新体系
- 四是开启大众创业、万众创新的创新驱动新格局
- 五是培育高端智能、新兴繁荣的产业发展新生态



应用需求牵引推动技术发展





实施8个推动工程

1. 大数据核心产品超越工程
2. 大数据资源共享流通工程
3. 大数据产业支撑平台工程
4. 大数据支撑创新创业工程
5. 电信互联网大数据引领工程
6. 工业大数据创新示范工程
7. 大数据安全体系加固工程
8. 大数据标准体系建设工程

出台5项政策措施

1. 建立统筹协调机制
2. 健全相关政策法规
3. 加大资金扶持力度
4. 完善投融资政策
5. 建设复合型人才队伍



1. 产业观察
2. 政策展望
3. 我院工作



工信部

- 大数据产业十三五规划

网信办

- 政府数据开放标准
- 数据资源管理体系研究

发改委

- 国务院大数据行动纲要
- 行动纲要落实工作

科技部

- 大数据重大工程实施方案

面向大数据国家重大决策需求

面向大数据产业发展关键问题

促进技术与产品成熟

大数据平台基准测试标准
基准测试活动与测试服务

推动数据资源流动

国内第一个大数据交易规则
推动大数据资源合作

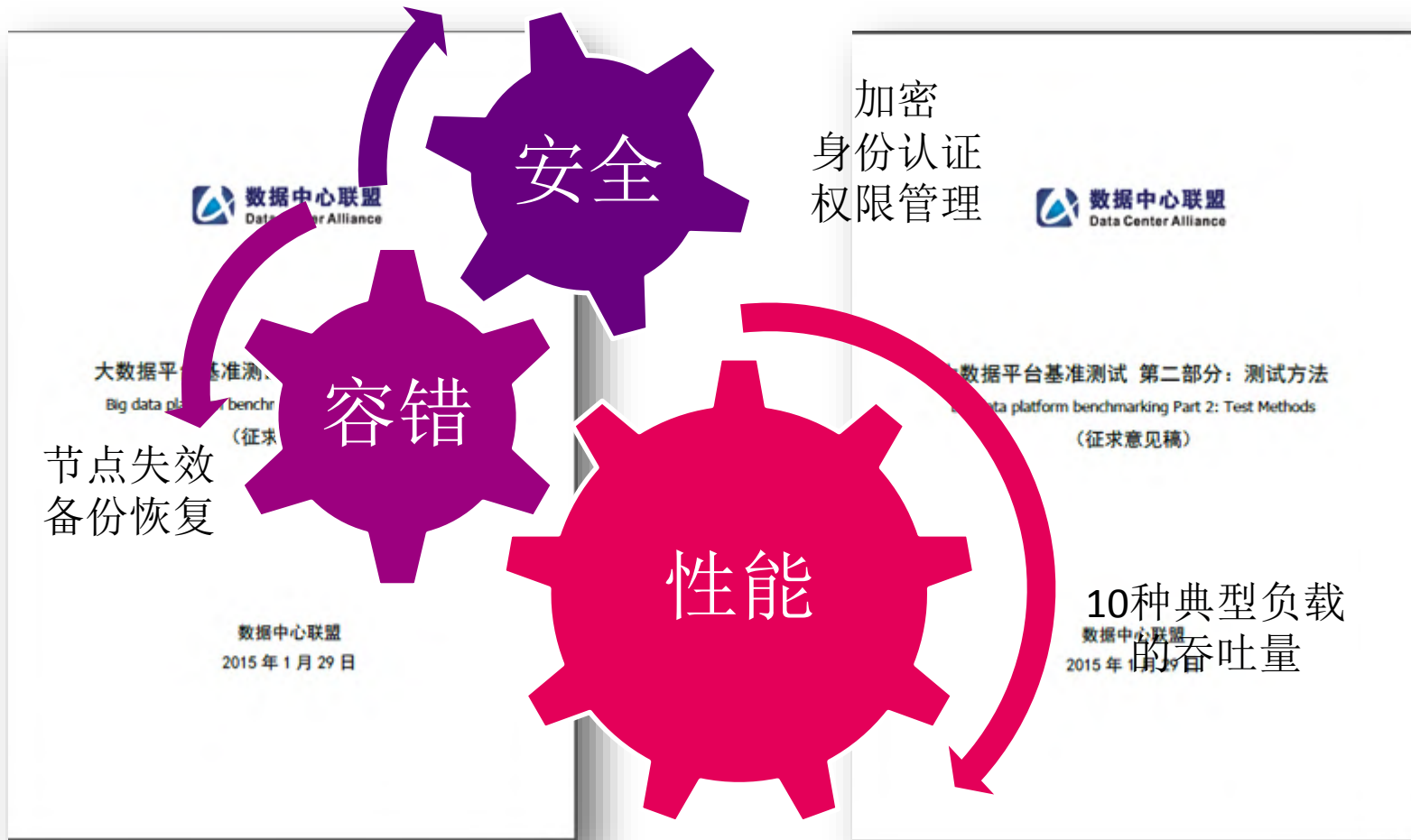
大数据人才培养

1+N课程体系（筹备）
人才认证



数据中心联盟
Data Center Alliance

数据中心联盟标准 → 行业标准 → 国家标准 → 国际标准



大数据测试活动



IDCC



数据中心联盟

Data Center Alliance

第一轮（2015年5-7月）



中国移动通信
CHINA MOBILE



HUAWEI



TRANSWARP

U CLOUD

全球首次在**统一的测试平台**上（16台节点）进行测试

国内**第一个工业级大数据**基准测试活动，4家国内顶级大数据商业发行版厂商

国内**第一个多个任务负载**的大数据基准测试活动，覆盖性能、容错性、安全与运维

国内外**第一次对大数据公有云**服务进行测试（优刻得）

人民银行、信息产业部《关于商业银行与电信企业共享企业和个人信用信息有关问题的指导意见》（银发〔2006〕112号）

第一条遵循“由易到难，逐步推进”的工作原则。可从**共享企业和个人欠缴电信费用信息起步，逐步扩大信息共享范围**。.....

《电信和互联网用户个人信息保护规定》
（工信部2013年第24号令）

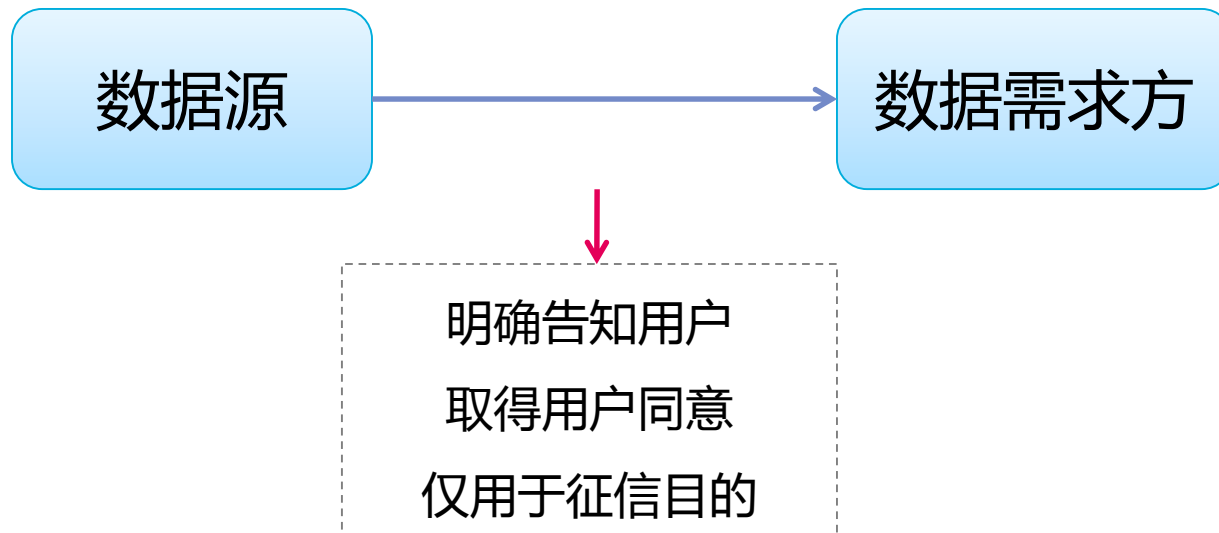
第十条 电信业务经营者、互联网信息服务提供者及其工作人员对在提供服务过程中收集、使用的用户个人信息应当严格保密，**不得泄露、篡改或者毁损，不得出售或者非法向他人提供**。

《征信业管理条例》
（国务院2013年第631号令）

第十三条 **采集个人信息应当经信息主体本人同意**，未经本人同意不得采集。但是，依照法律、行政法规规定公开的信息除外。

需要进一步更紧密衔接

- 征信条例**授权**征信机构：**依法采集个人信息**
- 人大决定和工信部24号令：**不得出售或者非法向他人提供**
- 刑法修正案禁止：**违反国家规定，将个人信息，出售或非法提供给他人**



提出了一套技术方案（规范）和推进机制去执行合规性要求



第十届中国IDC产业年度大典
The 10th Internet Data Center Conference

谢谢！