

在大数据应用爆炸的年代

从现状预见未来

分享人：Simon Yang

下一代
软件研发
SOFTWARE
DEVELOPMENT

一个典型的大数据应用诞生过程



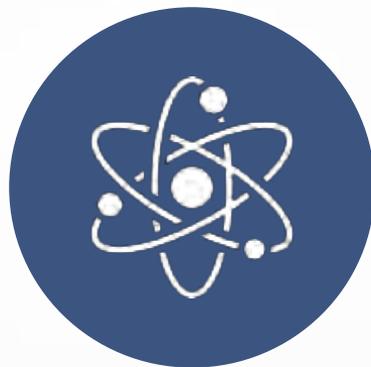


大数据产业
发展与繁荣

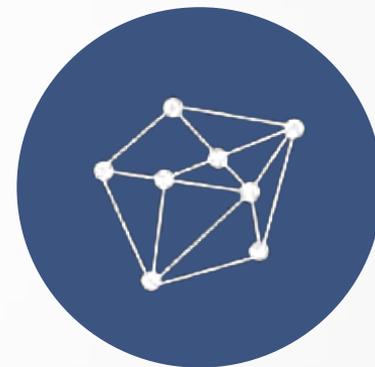




数据人才



技术应用



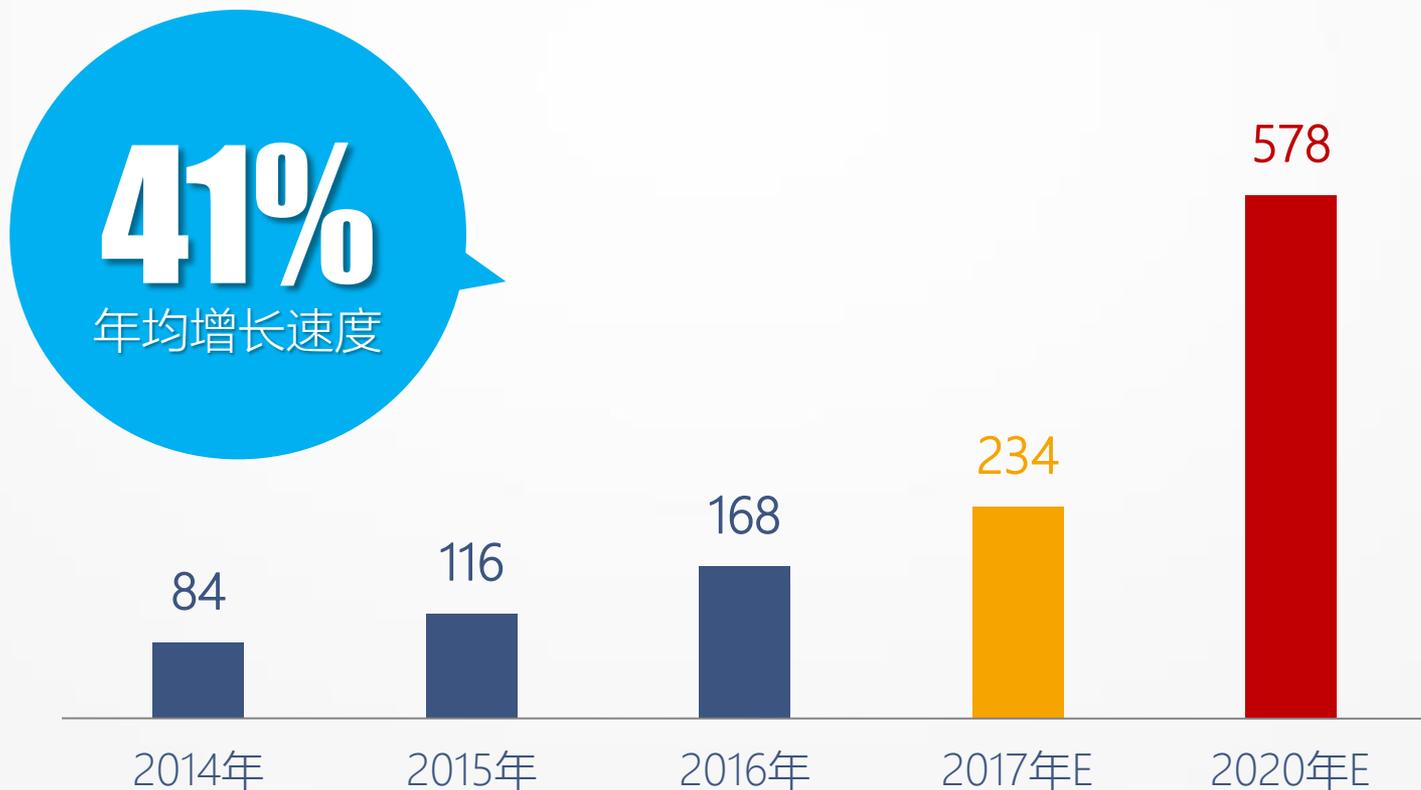
生态体系



数据人才

我国大数据市场规模前景广阔

我国大数据市场规模（亿元）



大数据产业发展配套政策密集出台

- 截止2016年底，已有21个省区市明确出台大数据产业规划，4个省出台相关规划和发展举措。

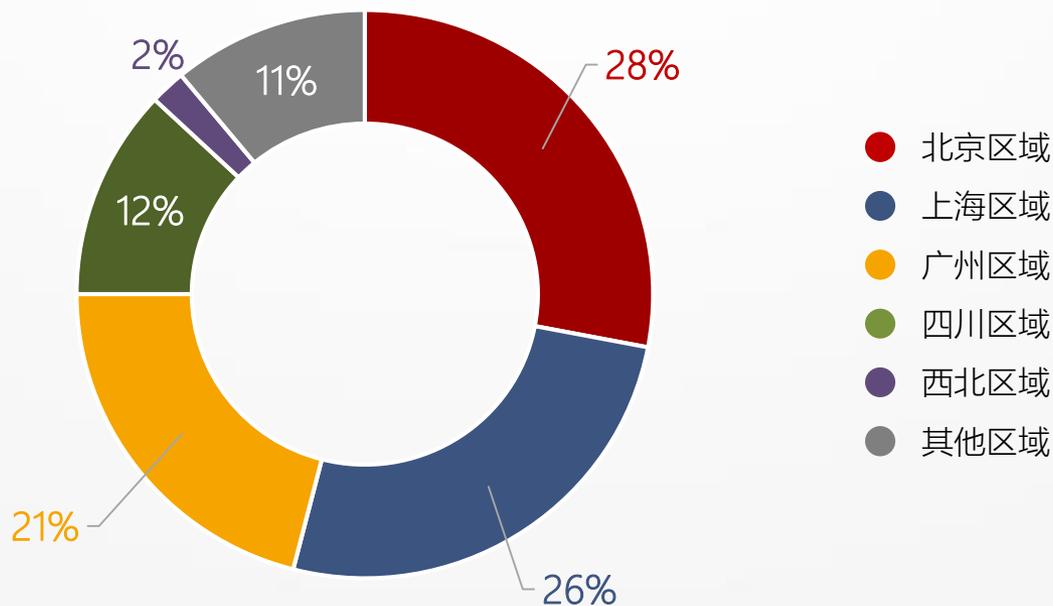


北京	上海	重庆	广东	浙江
江苏	吉林	辽宁	河北	陕西
山西	河南	甘肃	青海	四川
湖北	安徽	云南	贵州	广西
福建	内蒙古	黑龙江	湖南	海南

大数据人才分布不均

- 全国大数据人才主要分布在经济发达的一二线城市及大数据行业发展优先的城市，各地人才资源抓紧培养，保持和推进产业稳步向前发展。

我国大数据人才区域分布比例



大数据人才储备缺口严重

46万

当前我国
大数据人才缺口

数联寻英，大数据人才报告

150万

未来3-5年
大数据人才缺口

数联寻英，大数据人才报告

1/3

目前大数据岗位
合适人员招聘比率

Gartner Group

14-19万

大数据科学家
人才缺口

麦肯锡

150万

大数据分析师和经理
等决策岗位缺口

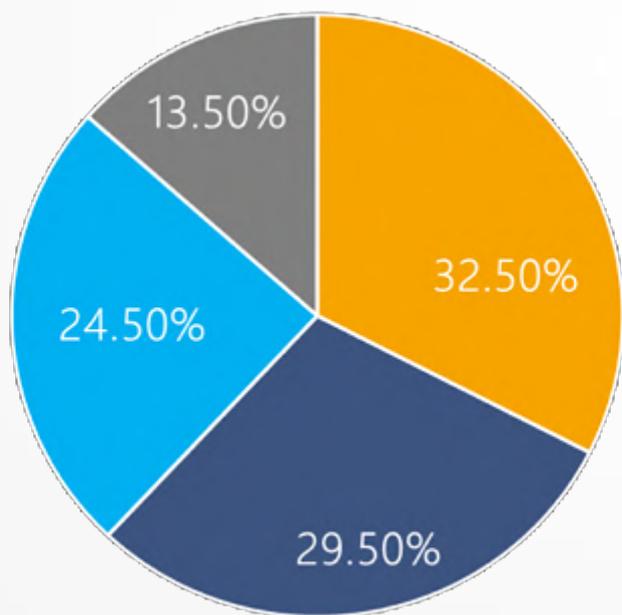
麦肯锡

90%

数学/经济学相关专业
硕士及以上学历

Burtch Research

现有3000+公司的大数据使用情况



- 正在搭建大数据平台，处于测试阶段。
- 已经在生产环境实践大数据，并有成功用例。
- 已经做了足够的了解，开发准备就绪。
- 基本不了解。

中国企业大数据研发团队规模



几种大数据人才候选人特征

业务能力突出者



技术潜力者



无经验潜力者



已经是综合领先者



立即着手培养企业大数据人才



- 从理解企业的业务开始要比从对于机器学习的理解开始来得更为重要。找一个真正熟悉企业业务的人员，并教给这些人员相关的分析能力。

业务



- 找到那些现在还不是，但具备成长为超级巨星潜力的人才，他们的教育和培训背景至关重要，考察对象集中在统计学，计算机科学等相关专业，有时包括物理专业。



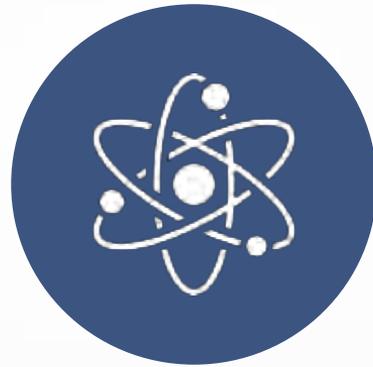
- Excel是学习大数据分析基本功能的一款入门级的学习工具，能够快速学习并掌握Excel的人一定是大数据分析的潜在人才，他们对于其他分析工具也能够快速上手。

Excel



- 通过建立一个导师计划，让有经验的专家来培训年轻人才，考察一下那些可能只有本科学历的员工，看看他们是否对于数据分析方面感兴趣。





技术应用

大数据技术的主要应用领域

金融

电信

医疗

政务

教育

工业

电商

农业

环保

生物

零售

媒体

公安

能源

交通

对应的典型行业应用主题

风险
管控

流量
经营

临床
决策

智慧
社保

就业
预测

排产
计划

潜客
获取

产品
溯源

排污
管控

制药
研发

交叉
销售

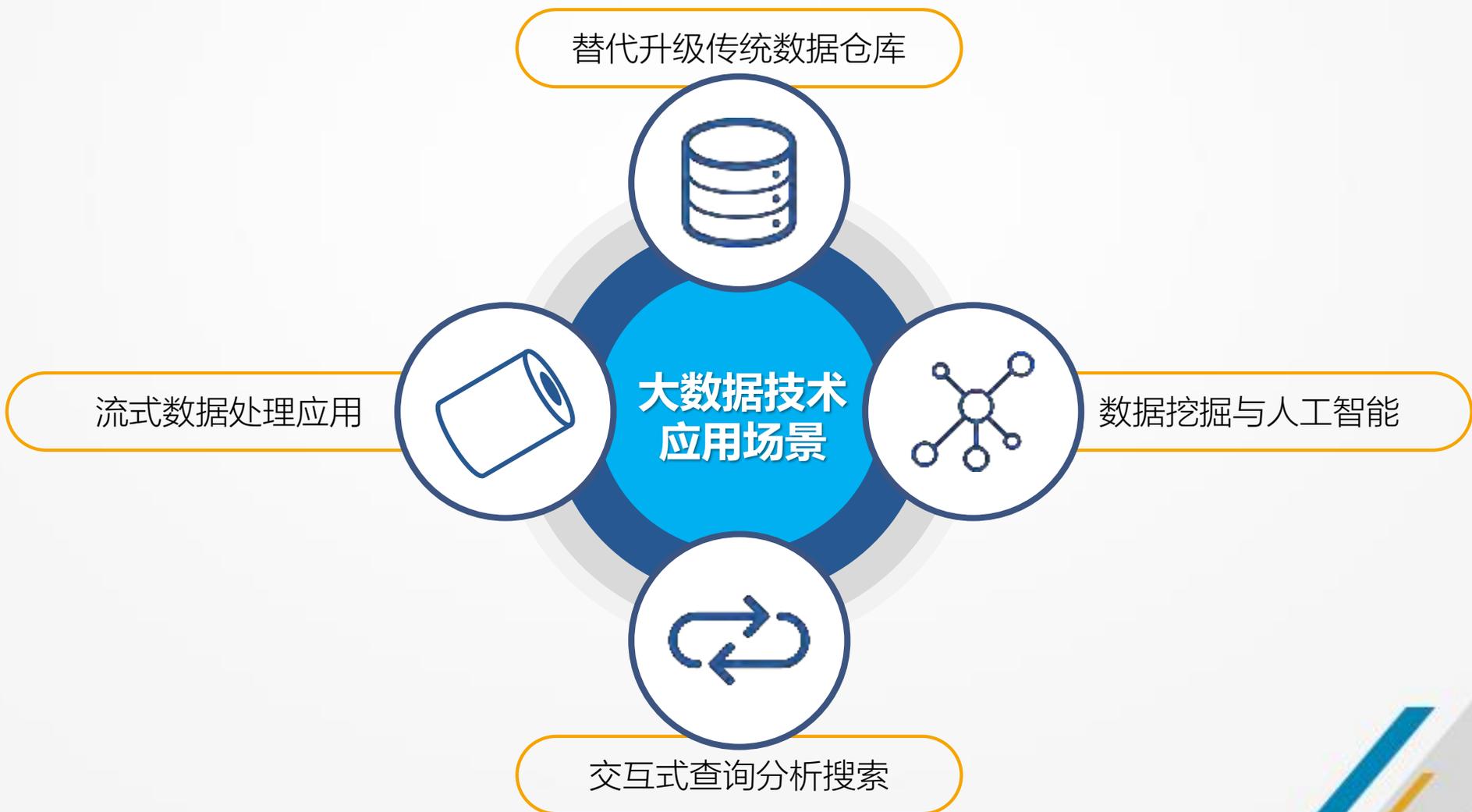
节目
推荐

刑侦
反恐

设备
维护

智能
调度

共同的大数据技术应用场景



大数据技术发展趋势

混合数据存储

基础

挖掘深度标签

核心

融合数据库架构

方向

大数据
技术

行业知识库

要素

异构数据关联

关键



大数据技术发展趋势



混合数据存储

- 大数据存储一方面需要提供超大容量的存储空间，另一方面需要支持对海量数据的智能检索和分析。为了兼容各种类型的大数据应用，大数据存储需要提供混合的数据存储模型，支持文件、对象、键值、块等多种访问接口作为大数据技术的基础。

大数据技术发展趋势



融合数据库架构

- 除了面向关系型数据库之外，面向各类应用，接口灵活，功能丰富且高效的NoSQL数据库也得到了蓬勃发展。在应用类型多样、数据种类繁多的大数据平台中，融合关系型数据库、列数据库、内存数据库、图数据库等多种数据库的混合数据库架构，方能满足多种场景下的数据处理需求。

大数据技术发展趋势



异构数据关联

- 各企业和机构的系统平台均累积了海量的数据，这些数据结构各异且相对独立，在没有建立起关联关系的情况下难以进一步提升数据价值。将这些多源异构数据进行关联和融合，挖掘数据之间的相关性，能够为数据分析奠定坚实的基础，最大限度地发挥数据价值。

大数据技术发展趋势



行业知识库

- 随着“互联网+”战略的实施,各产业尤其是传统产业,纷纷进行互联网化转型。在“互联网+”的浪潮下,面向多个行业,深挖行业知识详情,构建行业知识库,形成完整的行业知识体系,能有效推动数据应用与价值落地,是产业互联网发展的关键。

大数据技术发展趋势



挖掘深度标签

- 数据挖掘越来越多地应用到各个行业，使用数据挖掘技术而打造用户深度标签，已经逐渐成为大数据技术应用的热点。通过针对具体场景的数据挖掘，深入分析用户行为，打造多层次、多角度的用户深度标签，使得大数据应用更加精准，业务能够更加贴近用户，更好地满足用户的需求。

跟着数据走的五个关键步骤

▶ 数据采集



▶ 数据存储



▶ 数据可视化



▶ 数据预处理



▶ 数据分析挖掘

一款优秀大数据平台产品的必备素质

1

共存

迁移并存

RDBMS迁移
混搭架构
传统架构平滑过渡
.....

2

强大

功能性能

数据挖掘
机器学习
混合负载
.....

3

灵活

技术组件

开源组件调用
知识库引用
UDF运用
.....

4

快速

零编码开发

大数据ETL
调度管理
平台运维
.....

5

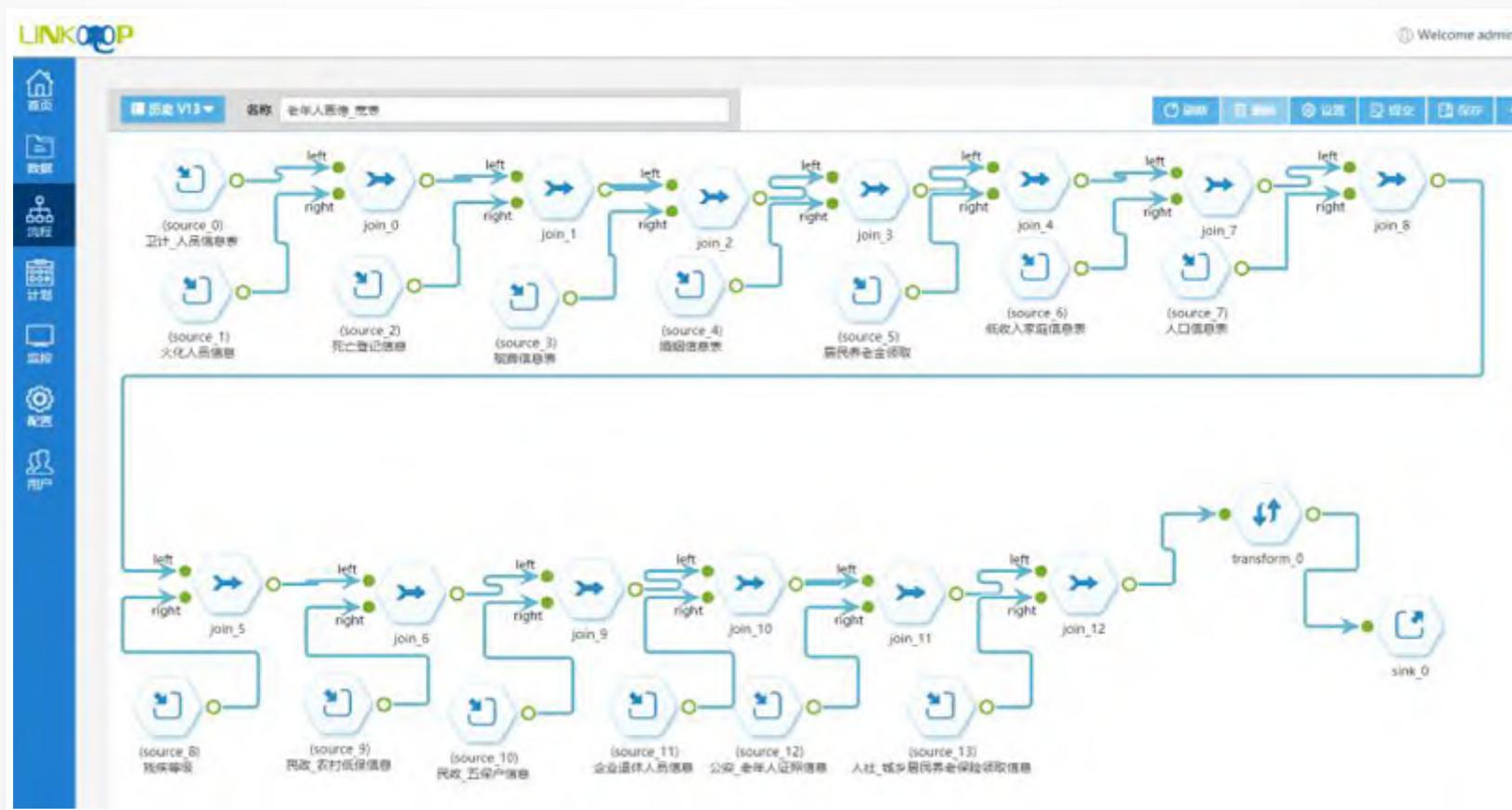
兼容

传统技术

传统SQL
现有分析工具
已有接口
.....

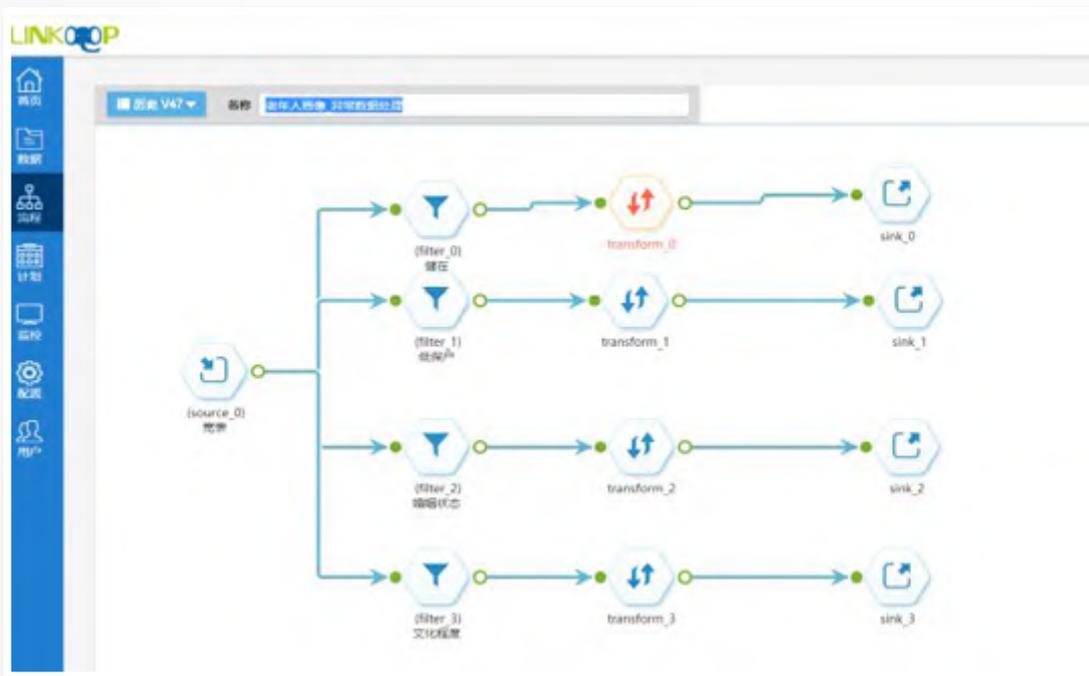
一个界面完成所有复杂设计与操作

- 场景示例：从多个政务系统数据源抽取数据，融合形成宽表，制作市民画像。



新兴技术依旧从容上手

- 场景示例：对客户数据按不同规则进行分群，按贡献度进行筛选分类，导入不同分析组件作为后续的数据输入。



各种人性化细节设计

- 场景示例：点选标签直接锁定相关数据处理作业，快速定位问题或查看相关作业运行情况。

The screenshot displays the LINKOOP web interface. At the top, there's a navigation bar with the LINKOOP logo and user information. Below it, a search bar is visible with the text 'tag_join_dataflow'. A horizontal bar contains various colored tags representing different data jobs. Below this, a table lists the jobs with columns for '名称' (Name), '类型' (Type), '版本' (Version), '创建时间' (Creation Time), and '修改时间' (Modification Time). The table contains 10 rows of data jobs.

名称	类型	版本	创建时间	修改时间
UC11301_数据分发	dataflow	10	2017-02-28 16:42:33	2017-03-02 18:50:07
UC19205_字符串操作_表达式_函数	dataflow	10	2017-02-24 10:59:19	2017-03-02 18:49:24
UC19207_类型转换函数_数字函数	dataflow	6	2017-02-26 09:17:39	2017-03-02 18:47:57
UC19102_离线数据分区_数据汇总函数	dataflow	26	2017-02-23 16:18:26	2017-03-02 18:47:03
UC10103_HIVE可视化数据采样	dataflow	8	2017-02-23 17:16:02	2017-03-02 18:45:37
UC19101_实时数据分区	streamflow	12	2017-02-23 19:43:07	2017-03-02 18:45:09
UC10303_半结构化数据采样	dataflow	6	2017-02-25 18:02:42	2017-03-02 18:44:44
UC10302_文本数据采样	dataflow	5	2017-02-25 16:52:49	2017-03-02 18:42:24
UC10101_IDBC全量数据采样	dataflow	6	2017-03-01 13:28:38	2017-03-02 18:41:49
UC10102_HDFS数据采样函数	dataflow	6	2017-02-23 15:17:50	2017-03-02 18:40:28

别人没有为你做到的，我们为你做到了。

ZEROCODE DATA HUB



场景融合操作



超强计算能力



插件扩展功能



全图形化界面



任意底层兼容



生态体系

大数据产业链的主体结构



大数据产业分类和商业模式

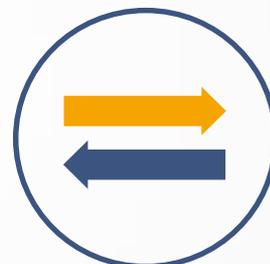
大数据产业的三大主要分类



数据服务



基础支撑



融合应用



多元化的商业模式

BIG DATA LANDSCAPE 2017



BIG DATA LANDSCAPE 2017



我们眼中大数据生态系统的各种角色

数据驱动的企业和跨领域创新者

系统开发与集成服务提供者

行业大数据方案提供者

分析软件提供者

应用软件提供者

数据服务提供者

基础平台软件提供者

大数据平台环境提供者

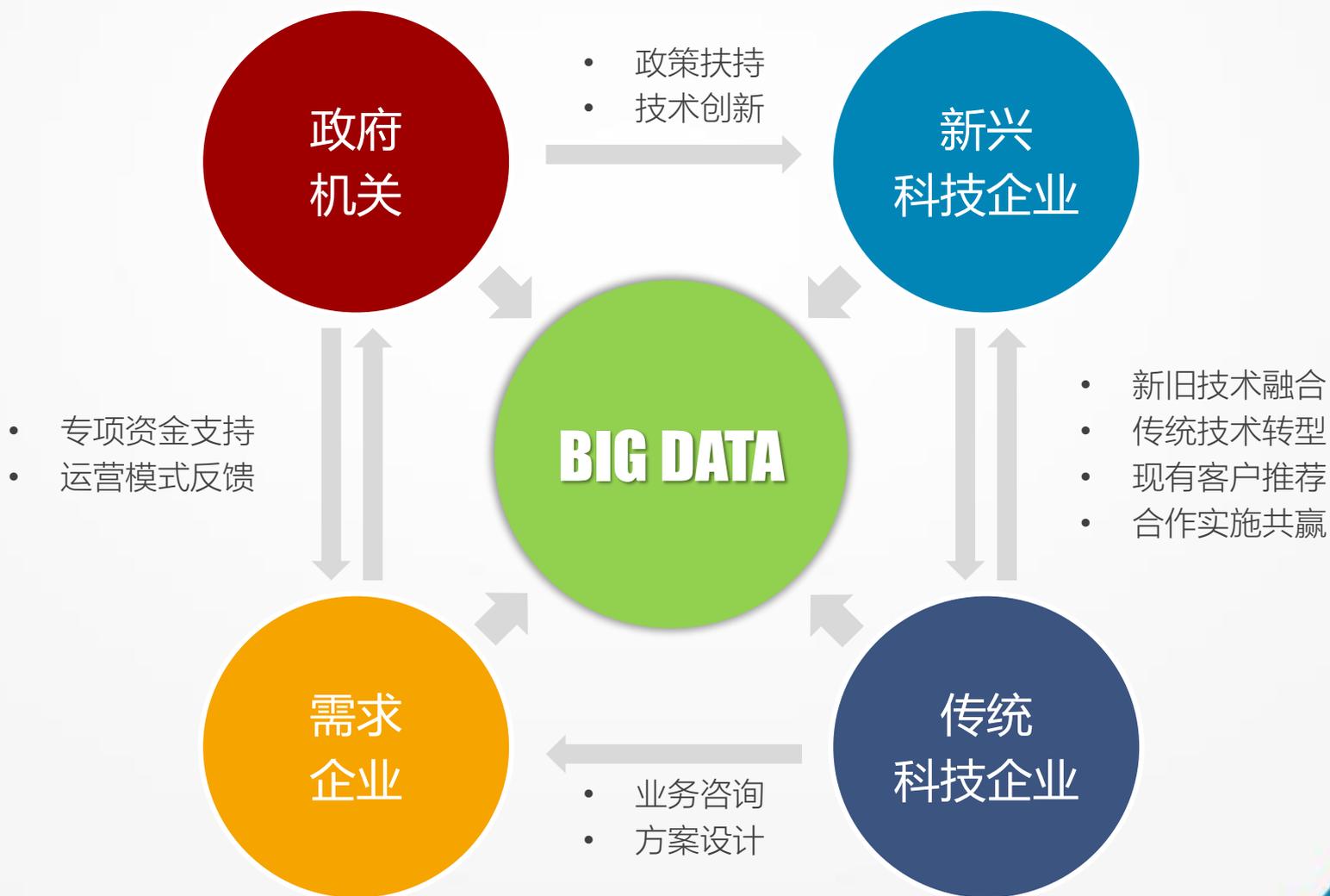
网络服务提供者

IT基础设施提供者

数据生产者

数据提供者

为了大数据生态的共生共荣



大数据生态体系不断壮大的六字要义



OPEN



开放



SHARE



专注



CREATE



创新

通过不断地探索和实践前沿数据科技
提升行业效率和人类生活品质