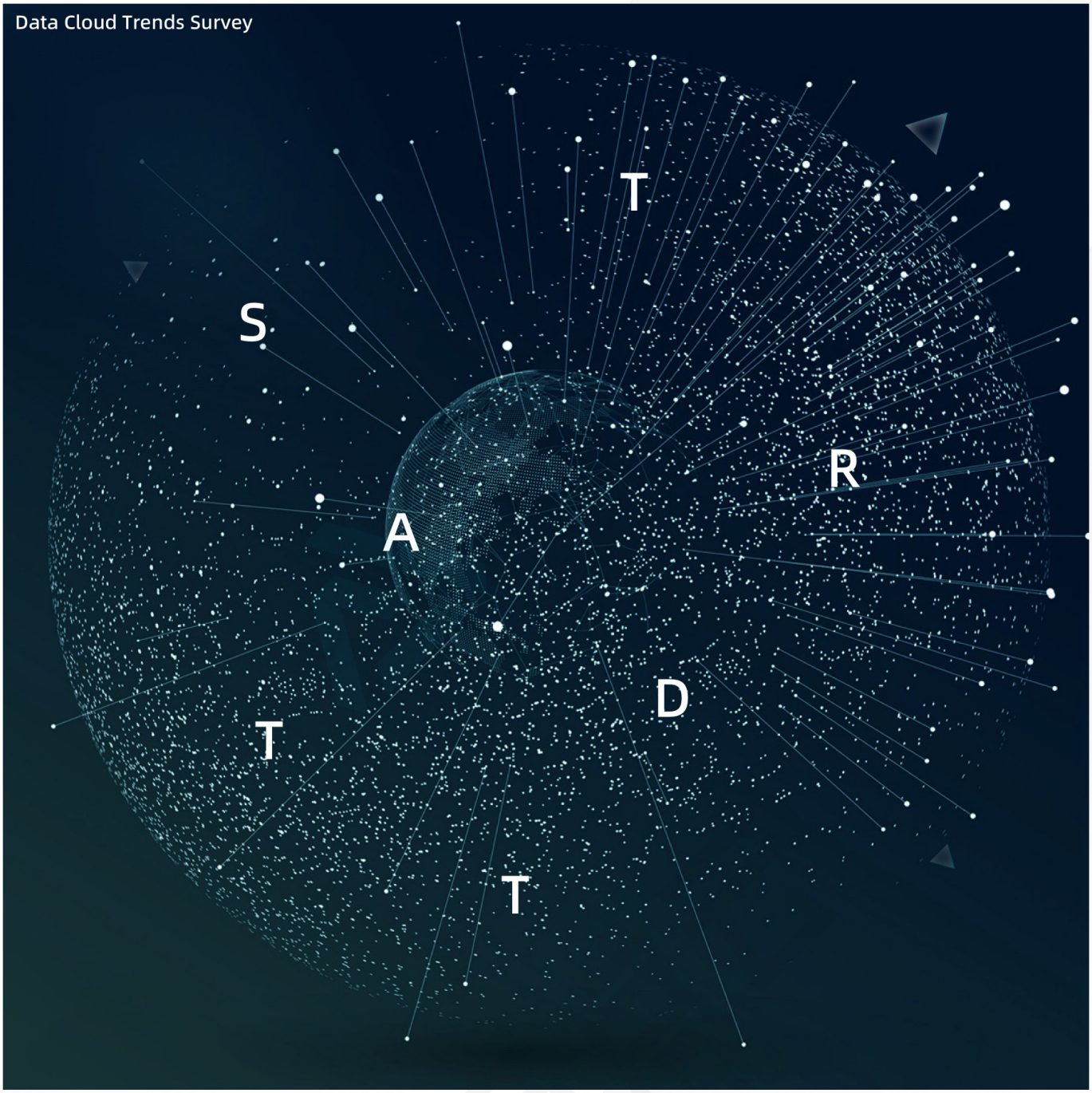


数据云趋势报告（2022）

更安全统一的数据未来

奇点研究院
StartDT Research

Contents

目录

1 什么是数据的未来 1

2 未来属于数据交换 4

3 迄今为止，唯独数据只增不减 6

4 海量数据带动云计算增长 8

5 数据从“上云”走向“云上” 10

6 数据时代走向 3.0 12

7 拥抱数据云，构建统一基础设施 14

8 定义：一站式企业级数据云 16

结语：数据未来将无死角地降临 20

01.
什么是数据的未来

2012年，Steve Lohr在《纽约时报》上撰文描写“Data-ism”（数据主义）：“大数据这个词现在时常被人们随意使用，然而其语义十分模糊。简单地说，这个包罗万象的词条一般有三层含义：首先，它指代一揽子的技术；其次，它有可能引发一场度量数据规模的革命；最后，它为人们未来将会、甚或是应该如何制定政策提供了一个新视角、一种新理念。”

差不多同时间，2010年，维克托·迈尔·舍恩伯格（Viktor Mayer-Schönberger）在《经济学人》上发布了长达 14页对大数据应用的前瞻性研究，并在两年后出版了《大数据时代：生活、工作与思维的大变革》一书。在书中，他指出大数据的核心就是预测：“大数据将为人类的生活创造前所未有的可量化的维度。”他进一步指出大数据时代最大的转变是“要相关关系，不要因果关系”。这颠覆了人类过去的视角，进一步从数据延伸出了全新的人类与世界交流的方式。

时间进入 21世纪 20年代，Steve Lohr的描述已然成为现实。世界各国纷纷转变视角，把数据发展作为一项战略规划。对于数据的占有和使用，影响着国家竞争力的走向，以及企业在市场竞争中的地位。2020年，中共中央、国务院《关于构建更加完善的要素市场化配置体制机制的意见》第一次把数据提到生产要素的高度上来，并进一步明确要加强数据资源整合和安全保护。一年之间，从基础设施到数据市场，数据要素的可能性热议中国。（图 1）

中共中央、国务院《关于构建更加完善的要素市场化配置体制机制的意见》	首提数据要素 →	推进政府数据开放共享； 提升社会数据资源价值； 加强数据资源整合和安全保护。
国家首次官宣明确 “新基建”3大方面 7大领域	信息基础设施 →	以工业互联网为代表的通信网络基础设施；以人工智能、云计算为代表的新技术基础设施；以数据中心为代表的算力基础设施。
总理发布 政府工作报告	首提培育数据市场 →	推进要素市场化配置改革； 促进人才流动，培育技术和数据市场，激活各类要素潜能。
港交所掌门人李小加数据相关提案引起广泛关注	加速数据要素市场培育 →	聚焦以下 5大问题：数据确权；数据的安全和隐私保护；助力数据共享，以产生更大价值；统一数据市场的监管标准；培育数据交易生态体系
工信部发布《关于工业大数据发展的指导意见》	工业大数据 →	重点任务涉及数据汇聚、数据共享、数据应用、数据治理、数据安全、产业发展、组织保障 7个方面

△ 图 1：2020年“数据要素”相关发布

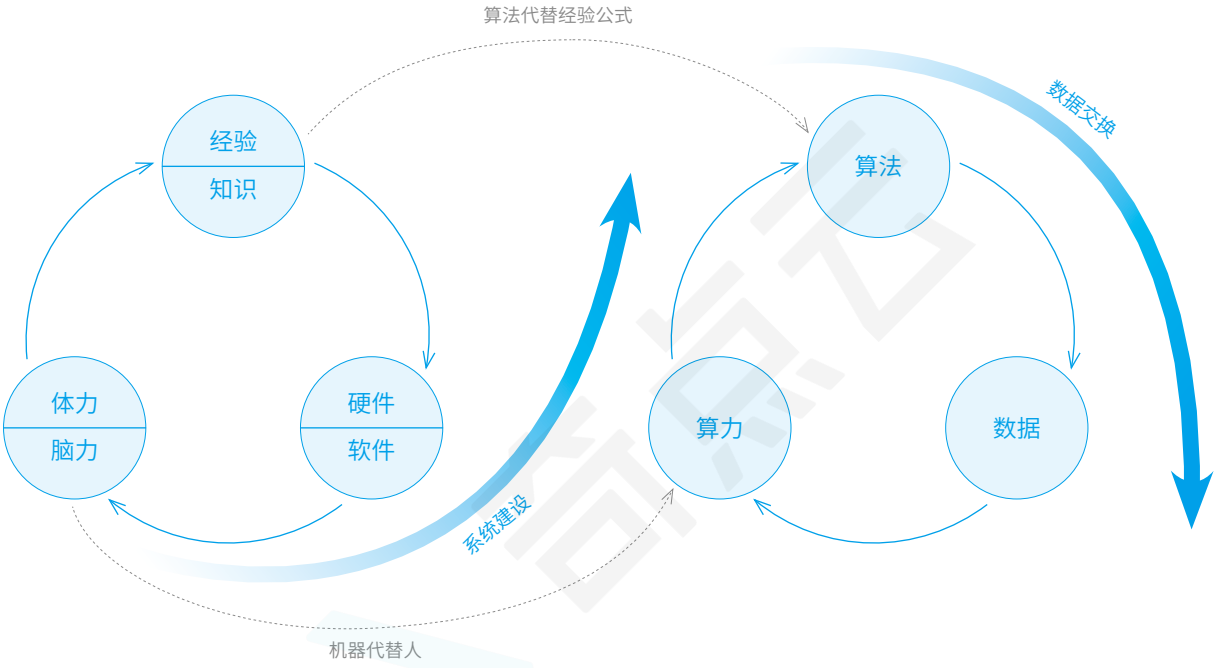
“数据要素化”是人类第一次对数据的未来拥有了相似的集体想象。这种想象可能来自于对当下的迷茫和不满。

2019年底，中国社会科学院发布的《社会蓝皮书》指出，中国城市发展从“增量”时代走向了“存量”时代，从社会经济到企业经营，过去的经验似乎不再有效。“你可能什么都没做错，你只是太老了。”马化腾的这句话真切地说明了时代变迁给企业、给个人带来的不安和恐惧。

为了消解变化背后的不确定性，企业越来越依赖于对机器智慧的想象，既要解决当下企业面临的问题，也要为更长远的发展提供空间。如果说计算机帮助企业收集信息，从经验试错走向了分析择优，那么，面对越来越复杂的计算机系统和企业组织，今天企业就必须要去思考和解决如何让数据自动化流动起来，消解复杂系统本身的不确定性，优化整体资源配置效率。

“让正确的数据，在正确的时间，以正确的方式传递给正确的人和机器”，安筱鹏在《重构：数字化转型的逻辑》里认为，数据自动流动是企业竞争的关键，而算法则是将不确定性转化为确定性的最佳路径。

所以，从企业数字化转型的终局来看，企业技术能力的中心将由“系统建设”走向“数据交换”。用机器代替人去完成更高效率、更低成本的存储和计算，用算法代替经验公式去更准确高效地复用经验和知识，这就是未来数字化转型的定义。（图 2）



△ 图 2：从“系统建设”走向“数据交换”

02.
未来属于数据交换

过去主流的数据交换方式，来自于云厂商通过其上的平台（如淘宝）、通道（如微信）和工具（如手机），统一收集、处理、使用数据，数据只允许在平台内交换，构成了一个个封闭的中心型数据交换体系，每个节点和中心之间一对一地开展数据交换，即“以平台为中心”的交换方式。

这种交换方式成为主流的原因包括：

1

数据处于非确权状态，需要依赖平台进行交换；

2

数据技术发展刚刚起步，企业缺少交换的基础设施；

3

企业对数据的重要性认知不足，更多地把数据作为工具，而没有视为资源。

对于互联网平台以外的大部分企业而言，当前数据交换的主要矛盾不是数据“可用”和“易用”的问题，而是数据“有”和“没有”的问题：

从技术角度来看，企业通常面临数据生产（没有数据）和数据采集（有数据但没采集）能力的缺失；

从业务角度来看，面临的是缺乏数据想象力，以及数据应用需求的单一化（可视化报表）的问题；

从组织角度来看，面临的则是数据战略和专业人才的缺失问题。

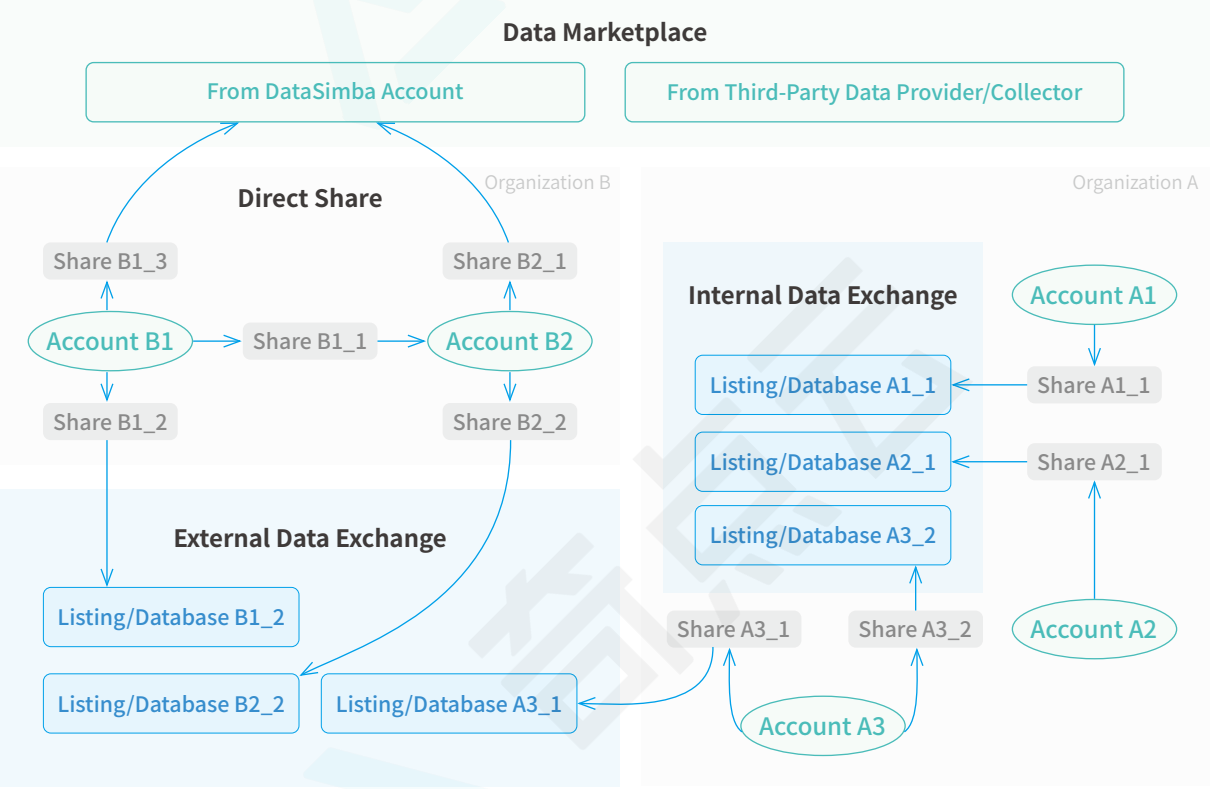
但这种现状正在发生改变。

首先，大型企业集团同样希望在集团内部实现“以平台为中心”的数据交换，尽管面临数据能力不足、数据量少、数据需求少的现状，平台化仍然是大势所趋；

其次，随着云计算市场的发展，越来越多的企业希望摆脱云依赖，走向多云，不依附于某一家云厂商；

最后，数据确权和反垄断也给互联网平台的数据使用带来了更大的挑战，存量竞争让平台数据价值降低。

不仅如此，大型企业集团希望数据也能实现与产业上下游的互联互通，以及和第三方公司的数据交换，最终，实现数据要素的市场化交易。即统一构建云上数仓，统一构建云上账号体系，最终实现 Direct Share—Internal Exchange—External Exchange—Data Marketplace（数据共享—内部交换—外部交换—数据生态市场）的数据交换体系。（图 3）



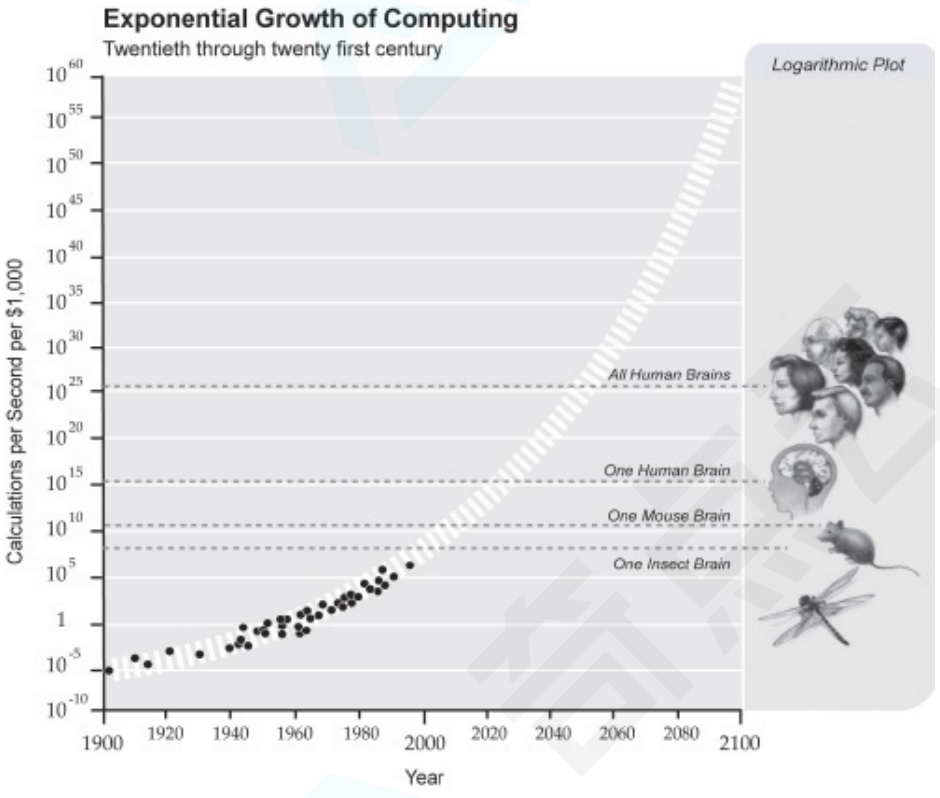
△ 图 3：数据交换体系

数据交换，意味着新的工作方式、新的商业增长模型、新的商业模式、新的商业角色…机会就孕育在风暴之前。

03.
迄今为止，唯独数据只增不减

1980年，雷·库兹韦尔（Ray Kurzweil）想要找到一种方法，用于测量和追踪技术发展的速度。
他用 1000美元能购买到的计算能力随时间的变化情况来绘制这条曲线。（图 4）

根据计算结果，库兹韦尔惊讶地发现，技术并不是线性发展，而是呈指数级上升的加速度发展。



△ 图 4：Exponential growth of computing. 20th to 21st centuries

图源： Ray Kurzweil and Kurzweil Technologies, Inc.

库兹韦尔把这一规律命名为“加速回报定律”，把技术发展延伸至未来的某个巨变时刻称为“奇点”，人类社会在奇点来临之后将呈现完全不一样的状态。

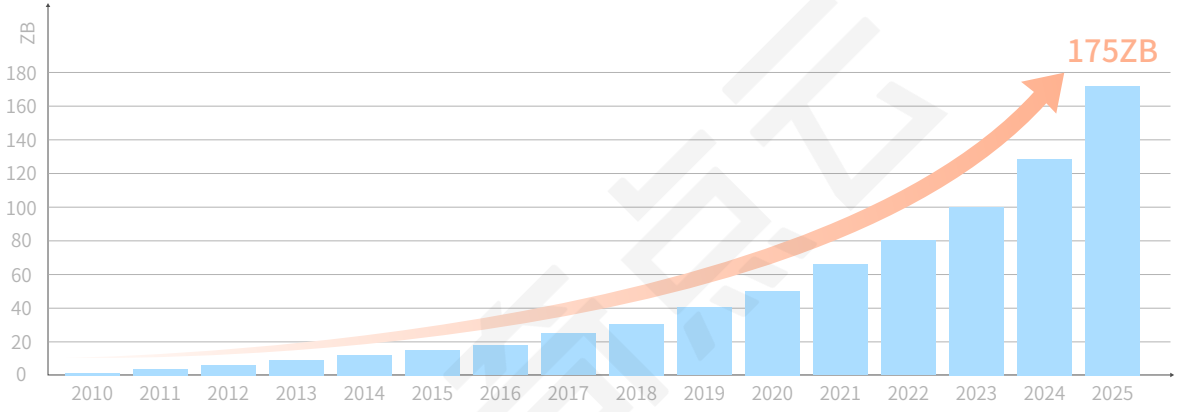
不管是芯片领域的摩尔定律，亦或是生物科技领域，一系列的技术进步曲线看起来都呈现出了指数级增长的特征。2015年之后，全球范围内，数据同样开启了指数级增长曲线。

有两个数据可以佐证：

1. 2015年，这一年人类产生的数据，近似等于 2015年之前所有数据的总和；

2. 2015年之后，人类产生数据总和以每年 30%-40%的速度在增长。根据 IDC的预测，2025年，人类数据总和将突破 175ZB。（图 5）

可以说，2015年就是数据奇点开启的元年。



△ 图 5：全球数据圈的每年规模

图源：数据时代 2025，IDC全球数据圈

04.
海量数据带动云计算增长

互联网行业巨头最早感知到数据指数级增长带来的云计算机会。

早在 2006年，亚马逊就成立了 AWS（Amazon Web Services，亚马逊云服务），开始尝试提供S3存储服务、SQS消息队列、EC2虚拟机，从而证明“云是可行业务”。

2008年，微软、谷歌两大巨头躬身入局，微软发布了 Windows Azure技术社区预览版，开始提供服务托管和线上化服务；谷歌发布了 Google App Engine预览版，用于开发者基于专有 Web框架开发 App并部署，这可以被视为 PaaS的起源。

在国内，2008年，阿里巴巴开始筹办和起步“云计算”；2009年，阿里云成立。（表 1）

进入 2020年，企业的云计算使用走到了新的拐点。

1. **2020年，我国公有云市场规模首次超过私有云**。同期，全球云计算市场保持稳定增长，但增速逐渐放缓。公有云市场规模达 689亿元，增长 57.6%，私有云市场规模达 645亿元，增长 22.8%。

2. **企业对云的投入增加**。2020年 IDG的云计算调研报告指出，云计算相关支出占比达到企业总 IT预算的 32%，且比例还在增加。

3. **云原生趋势**。2020年在云中运行的应用程序中，有 54%是从本地基础结构迁移到云上的现有应用程序，而 46%是“专门为云计算而构建”的应用程序。

4. **企业采用“多云”趋势**。2020年 IDG的云计算调研报告显示，当组织利用多个公共云时，49%的受访者提到的主要目标是利用“最好的平台和服务选项”。其次“节省成本 /优化成本”占 41%，“提高灾难恢复 /业务连续性”则占 40%。如果仅放大企业目标，“避免供应商锁定”是第二目标，占比达 40%。

	云时代 1.0 上云，从无到有	云时代 2.0 大数据，技术升级	云时代 3.0 数据平台，效率升级
数据量	截至 2010，全球数据量达 2ZB	截至 2015，全球数据量达 15ZB，增速达 25%	截至 2020，全球数据量达 64ZB，增速超 50%
产业格局	· 规模导向，巨头垄断 · “云计算”成型，但未解耦“云计算”、“大数据”等核心概念	· “云”从面向个人的网盘服务，转为面向企业的企业级云服务 · 国外云厂商正式加入中国市场竞争 · 更多互联网头部企业进入市场，并开始出现有特色产品服务和自主研发能力的独立云计算创业企业	· 企业趋向采用“多云”/“混合云”策略 · 大量第三方独立服务商出现，与云厂商互利共生
技术趋势	· 虚拟机层面实现存储与计算分离，逐步形成适用于云上的应用架构	· IaaS、PaaS、SaaS技术层级分流出现，各领域深化发展，相互对接 · PaaS层整体技术方向转变，尽可能解除原有平台、语言、框架的限制，保证云端用户的技术灵活性和可移植性	· IaaS层技术深化，容器化与微服务形成趋势，Kubernetes胜出成为事实标准，“云原生”出现 · PaaS层摒弃大一统的应用程序架构，逐步形成标准的可复用中间件，支撑与 IaaS层融合联动 · PaaS层 DBaaS逐步发展，云数仓成型
产品结构	· 云厂商将“存储服务”开辟为单独产品类别（例如 Amazon S3、阿里云 OSS） · 可选容量和性能量级的“托管磁盘”服务（例如 Amazon EBS、阿里云盘） · 云上大数据分析服务，即开源大数据技术在云上实现及适配（例如 Amazon EMR、阿里云 EMR）	· IaaS层：CPU升级带来云上虚拟机升级，产品结构更加丰富（例如通用型、计算型等） · IaaS层：存储类服务功能细化（例如 Amazon EFS等） · PaaS层：发展自动扩容、负载监控等“通用型”服务，解除平台绑定（例如 AWS Elastic Beanstalk，Azure App Service） · PaaS层—DBaaS：云上数据库服务快速发展（例如 Amazon RDS，阿里云 RDS）	· IaaS层：云厂商提供基于 IaaS的容器运行环境和编排引擎（例如 Amazon EKS、Azure Kubernetes Service） · IaaS层：容器可作为独立计算单元直接运行，形成无服务计算（例如 AWS Fargate、阿里云 ECI） · PaaS层—DBaaS：云厂商完全自行实现数据内核和引擎（例如 Amazon Aurora、Azure Cosmos DB、阿里云 PolarDB） · PaaS层—DBaaS：第三方数据服务商出现，完善升级开源数据库 /引擎，同时与云厂商合作（例如 Databricks、Snowflake、Kyligence）
计费方式	· 虚拟机按不同 CPU/内存比例计费	· 灵活计费（例如包年包月、竞价实例、预留实例等）	· 灵活计费，同时出现“垂直云”、“领域云”解决方案（例如游戏云、视频云、工业云、金融云等）
市场玩家	· 巨头入场，包括以 Amazon、Microsoft、Google为代表的国外玩家，以及以阿里云为代表的国内玩家	· 国外巨头玩家进入中国市场 · BAT及移动、电信、联通三大运营商纷纷重金入局 · 独立云计算企业创立，例如 UCloud、QingCloud等	· IaaS、PaaS玩家分流 · 出现大量第三方独立服务商

△ 表 1：云时代发展历程

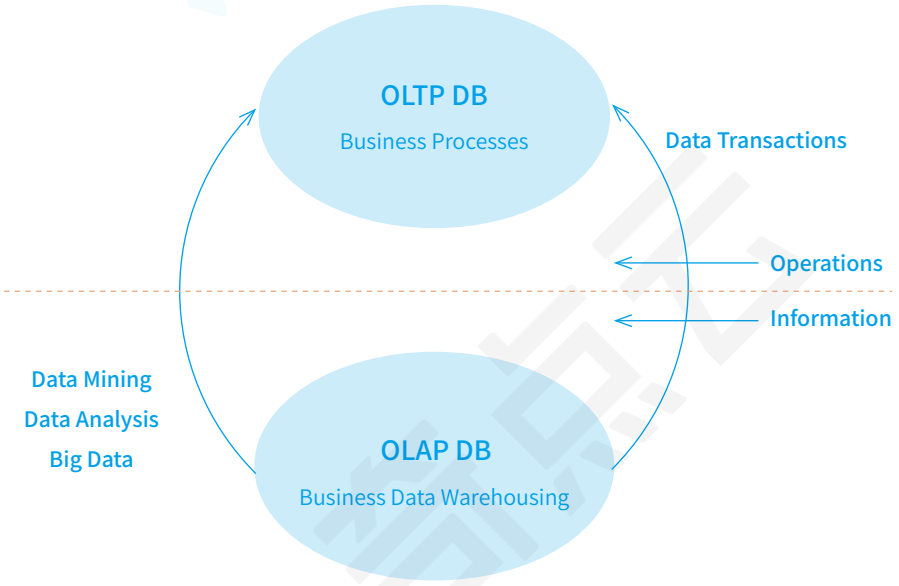
05.
数据从“上云”走向“云上”

随着计算机技术的诞生和成熟，计算机开始运用于数据管理，与此同时，数据管理技术也迅速发展。传统的文件系统难以应对数据增长的挑战，也无法满足多用户共享数据和快速检索数据的需求。

20世纪 60年代，数据库（DataBase）和数据库管理系统（DBMS，Database Management Systems）应运而生，旨在解决日常工作的 OLTP（On-Line Transaction Processing，联机事务处理）问题。

数据库是“按照数据结构来组织、存储和管理数据的仓库”，是一个长期存储在计算机内的、有组织的、可共享的、统一管理的大量数据的集合。

伴随数据库的大规模应用，数据越存越多。为了研究数据之间的关系，挖掘数据隐藏的价值，人们越来越多地需要使用 OLAP（On-Line Analytical Processing，联机分析处理）来进行数据分析。



△ 图 6：OLTP与 OLAP

图源：Difference between OLAP and OLTP in DBMS, GeeksforGeeks

1988年，为解决企业数据集成问题，IBM公司的研究员 Barry Devlin和 Paul Murphy创造性地提出了一个新的术语——数据仓库（Data Warehouse）。

	数据库	数据仓库
面向对象	面向顾客的，用户操作员，客户和信息技术人员的事务和查询处理	面向市场的，用于知识工人的数据分析
数据内容	管理当前数据，一般这种数据比较琐碎，很难用于决策	管理大量历史数据，提供汇总和聚集机制，而且在不同的粒度层上存储和管理信息
数据设计	采用实体联系数据模型和面向应用的数据库设计	系统采用星形或雪花模式和面向主题的数据库设计
数据视图	关注一个企业或部门内部的当前数据，不涉及历史数据或不同单位的数据	经常需要跨越不同版本的数据库模式
访问模式	主要由短的原子事务组成，一般需要并发控制和恢复机制	大部分是只读操作

△ 表 2：数据库与数据仓库

表源：《关于数据中台、数据平台、数据仓库、数据湖等数据概念的对比解析》，CSDN

1990年，Bill Inmon出版了名为《构建数据仓库》（Building the Data Warehouse）的里程碑著作，数据仓库自此真正开始应用，用来处理面向主题的（Subject Oriented）、集成的 （Integrated）、相对稳定的（Nonvolatile）、反映历史变化的（Time Variant）数据集合，支持管理决策。随着数据仓库在 20世纪 90年代得到广泛运用，主数据（Master Data）和元数据（MetaData）也逐渐为人所知。

根据舍恩伯格在《大数据时代》中对于大数据的分析定义，当我们讨论大数据，更多地都是在讨论数据仓库对海量数据分析的影响。而数据仓库和云计算的结合，才是 2000年大数据时代真正的基石技术。

20世纪 90年代以来，每十年左右，跟随着云计算发展的脚步，数据技术也在快速迭代。

2003-2006年，谷歌先后发表了分布式文件系统（Distributed File System）、大数据分布式计算框架（MapReduce）、NoSQL的数据库系统（BigTable）三篇重要论文，奠定了今天大数据计算的理论和实践基础。2002年，基于谷歌的论文，在 Apache Nutch 项目上发展而来的 Hadoop吹响了开启大数据时代的号角。

2010年前后，伴随着云计算的快速发展，数据仓库领域持续出现了技术进化，数据湖（DataLake）、大数据平台（Big Data Platform）、云数仓（Cloud DataWarehouse）、湖仓一体（LakeHouse）等新名词层出不穷。

其中，尤为值得注意的是“云数仓”：从根本上，云数仓和传统数据仓库并没有很大不同，但随着数据处理复杂度和数据新类型的增加，Gartner认为，云数仓充分利用云的特性，让企业不仅仅能免除采购、维护昂贵基础设施的过程，更具有以下特性：

- **按需采购，按使用付费**
- **可支持数据分享的安全能力**
- **作为服务交付**

简单解释，云数仓利用云计算，突破了原有的存算边界，实现了近乎无限的可扩容性和升级空间，但同时又降低了存算成本和运维成本，降低了数据管理和数据使用的难度，让非技术用户可以直接访问数据。

云数仓的广泛应用，代表了数据正在全面拥抱公有云，从上云走向云上，标志着数据领域全面进入了新时代。

06.
数据时代走向 3.0

综上所述，和云时代类似，我们可以把整个数据仓库之后的大数据时代划分成三个阶段。（图 7）

	数据时代 1.0 数据仓库	数据时代 2.0 数据平台	数据时代 3.0 数据云
核心趋势	企业开始积累和分析数据	数据成为核心竞争力 企业间展开数智化“军备竞赛”	企业开始期待数字化转型带来直接的效率提升，开源节流成为主要诉求
产业格局	规模导向，巨头垄断	单点技术，蓬勃开源	统一平台，新巨头出现
海外典型	ORACLE teradata. 数千亿美元估算市值	cloudera 发展受阻，$50亿被收购	snowflake databricks 千亿美元市值
国内典型	无典型公司	TRANSWARP 星环科技 阿里云	?
经济性变迁 *	~90万 / 年	相比 1.0时代 **存算成本降低 60%**	相比 2.0时代，**存算成本每年下降 10-20%**；用户应用数据通过账单模式量化且可控

*该数字体现的是每 10TB数据一年存算所需的成本

△ 图 7：数据时代三阶段

数据时代 1.0：数据仓库阶段

核心技术：**软硬件一体化数仓（1980、1990年开始）**

代表产品：Oracle、IBM DB2、SyBase、Teradata、SQL Server

特点：

· 非多云支持；
· 高部署成本；
· 共享存储架构；
· 规模导向，巨头垄断。

局限性：

· 扩容成本高昂；
· 可扩展性较差；
· 仅能满足管理层对传统报表分析的需求，无法应用于更大规模、更多用户的数据分析。

数据时代 2.0：数据平台阶段

核心技术：**开源化数仓构建大数据平台（2000年开始）**

代表产品：GreenPlum、Cloudera、Splunk、DataWorks、DataPhin

特点：

- 基于 Hadoop、Hive、Spark 等开源框架和 Presto 等开源 MPP（Massive Parallel Processing）架构，提升了负载均衡能力，更加适应于大数据分析场景，足够实现海量数据的存储和计算；
- 采取软硬件解绑、适配通用性硬件的分布式架构，并成功为企业提供了更为灵活、成本更低的扩容能力。

局限性：

- 数据仓库规模有限；
- 查询效率低（分区数据库、表分区、多维集群是拉动查询性能的“三驾马车”）；
- 上云困难；
- 数据孤岛；
- 受大型机、专有硬件和软件架构的牵制，其扩展性仍难达到数千节点级别；
- 由于开源化数仓仍需依托于线下机房，在日后性能需扩充时仍要购买硬件，后期成本相对较高。

数据时代 3.0：数据云阶段

核心技术：**云上化数仓为核心构建数据云（2010年开始）**

代表产品：Snowflake Data Cloud、DataSimba

特点：

- 实现了近乎无限的可扩容性和升级空间，彻底摒弃了笨重的硬件采购和线下部署过程；
- 性能与规模上，每秒可支持数千个并发用户和查询，并具有根据业务需求扩大或减小规模的能力；
- 由于云计算自身拥有“基础设施即代码”的特性，云上化数据仓库资源的分配基本可以自动化完成，可极大降低前两代数据仓库在后期会产生的运维成本。具体的自动化功能包括自动化的数据摄取、查询调优、数据处理和数据集成；
- 直接与数据湖和对象储存集成；
- 自助服务，更庞大、更复杂的仓库访问与管理得到简化；
- 简化解决方案，让非技术用户能直接访问数据；
- 具备并行处理、压缩、分区、索引、查询优化和动态资源配置等方面的高级功能。

07.

拥抱数据云，构建统一基础设施

2020年，随着公有云日益占据市场主导地位，各大数据公司都推出了云数据仓库（Cloud DataWarehouse）和云数据平台（Cloud Data Platform）产品，并进而用数据云（Data Cloud）来统一称呼其平台数据产品集。

我们收集了各大厂商的定义如下：

企业名称	该企业对 Data Cloud的定义 *	产品名称	产品特色 **
Snowflake	· 数据云消除了数据孤岛，支持企业无缝地统一、分析、共享数据，甚至将数据货币化。 · 数据云最终是一个由数千家企业和组织组成的**生态系统**，企业不仅连接自己的数据，还通过共享和消费，将数据、数据服务连接起来。	Snowflake Cloud Data Platform	· Seamless Data Experience, Multiple Clouds · Secure Data Sharing · Snowflake Data Marketplace · Near-Unlimited Scale, Concurrency, and Performance
Google	· 数万客户利用 Google Cloud 构建数据云，这是一种**统一、开放**的数据驱动型转型方法，内建 AI 能力，在速度、规模和安全性方面拥有无与伦比的优势。 · Google 为希望构建数据云的各种规模的企业**提供数据和软件**。	Google Cloud	· 业界唯一统一的数据平台 · 开放式数据平台 · 无论企业处于“云之旅”的哪个阶段，Google Cloud 的数据库解决方案都能满足需求
Cloudera	· **企业数据云**：企业级安全和治理，多功能数据分析，弹性的云计算体验，无孤岛，无锁定。 · **混合数据云**：在混合云和多云环境中提供服务。	Cloudera Data Platform	· The world' s first Enterprise Data Cloud · We deliver a Hybrid Data Cloud for any data, anywhere, from the Edge to AI. · 在任何云环境和数据中心中管理和保护数据生命周期
Oracle	· **云数据平台**的主要目的是为组织提供一种更容易使用其数据的方法，同时保护数据、管理数据，并提供数据的集成视图。 · 云数据平台是一个**集成解决方案**，支持机器学习、第三方分析和 ISV 应用程序。 · 云数据平台是**广泛深厚的跨应用程序、平台和基础设施的组合**，可为企业提供工具和能力，构建其成功的云之路。	Oracle Autonomous Database	· 通过多模型融合数据库和基于机器学习的自动化全生命周期管理，**降低运营成本达 90%** · Oracle 自治数据库在 Oracle 云基础设施上原生运行，为事务处理和数据仓库提供负载优化的云服务 · 自治：自治驱动、自治安全、自治修复
IBM	· **Data Fabric** 是一个新兴的架构，旨在解决**混合数据环境**中出现的数据挑战。其基础理念是充当数据端点之间的虚拟连接组织，在去中心化和全球化之间取得平衡。 · Data Fabric 架构能够在分布式环境中实现动态和智能数据统筹，创建一个即时可用的信息网络，为企业提供动力。	IBM Cloud Park for Data	· Data Fabric（数据编织）与 Data Virtualization（数据虚拟化）：Data Virtualization 是支持 Data Fabric 方法的技术之一，支持访问并统一分布在分布式存储和云中的数据，而无需对其进行移动或复制 · 集成的现代云数据仓库：自动发现和整理数据，向用户提供可操作的知识资产，同时自动执行策略以保护使用，从而简化对数据的访问 · 容器化

△ 表 3：各大厂商对 Data Cloud的相关表述

*译自各企业官方表达。
**摘录或译自各企业官方表达

纵览其定义，尽管在数据产品定义、功能上有各自不同的描述，但我们可以发现有如下共性：

统一性：

就像一家公司只能有一个 ERP，一家公司最终也只能有一个数据云。统一的数据平台是未来企业建设数据云的必备，也是消灭孤岛所必需。

开放性：

强调数据云要提供对开源标准和 API的集成能力，确保数据可以被简单地广泛使用，不管何种需求，都可以使用数据云来应对。

中立性：

云原生架构要求平台使用开放式计算和开放式存储，同时确保可移植性和可扩展性，支持多云，避免受制于特定供应商。

安全性：

在所有环境中均保证数据的安全、治理和控制，确保数据的共享和使用得到安全保障，确保客户对数据的充分可控。

08.
定义：一站式企业级数据云

从上述共性出发，我们尝试定义数据云如下：

数据云是企业自有的、无孤岛、无绑定、安全可控的数据基础设施，为企业在其内部统一连接、共享数据资产，构建企业内外数据生态，助力企业全域数据管理与应用。

在技术上，数据云平台需要具备跨平台、云原生、自主可控、数据安全的特性。

跨平台

依托数据云，企业可以使用统一的账号权限体系，对多个 IaaS云基础设施、多个 Workspace（域）进行管理，实现分级多域、跨云跨平台部署，提升协作与管控效率。

云原生

数据云具备微服务、容器化编排、存算分离、CI/CD等云原生技术特性，优化架构，为企业降低数据应用在 IaaS云上的存算成本，提升研发效率。

自主可控

内置自主可控的数据存算引擎，能帮助企业快速构建其自有可控、智能化、轻量级的大数据基础平台；同时，支持数据基础设施的平滑迁移，避免受制于特定供应商，实现技术可控、成本可控。

数据安全

内置数据安全引擎，架构设计以数据为中心，支持企业智能、高效地完成数据全生命周期安全管控，同时，保障数据云交换全场景安全可靠，为企业数据资产护航。

奇点云数据云生态

奇点云 StartDT Data Cloud，提供安全可控的数据技术与可信易用的数智产品，以助力企业成功搭建算法驱动、统一开放的数据云，释放企业自有数据生产力，充分激活数据价值。

奇点云认为，“数据云产品”、“数据云服务”、“数据云交换”是数据云成立的三大支撑。

数据云产品

以自主可控的数据技术为内核，用数据平台、数据引擎、数据生产力工具等独立、前沿的数智产品，为企业构建一站式企业级数据云，帮助企业化繁为简、统一治理。

△ 图 8：StartDT Data Cloud产品架构

数据云服务

以数据云平台为基座，提供包括咨询、运维、运营、实施与培训在内的高标准服务，用前瞻专业的全局服务陪伴客户成功。

咨询

为企业提供大数据相关的战略规划、能力建设、组织设计、业务设计等咨询服务，为构建企业自身的数据资产和数据能力出谋划策。

运维

为客户提供平台架构优化、数据治理、数据血缘、任务监控等多类型服务的策略、配套方案及落地支持，全流程保障企业大数据平台稳定高效运行，为数据发挥价值保驾护航。

运营

以驱动业务增长为核心逻辑，提供端到端的全链路运营服务，从解决方案到关键场景的执行落地，从系统搭建到业务成员使用，支撑企业成功实践数据运营管理，实现业务增长。

实施

以业务价值为先，关注长期数据能力建设，为企业定制化规划数据中台实施方案，充分利用数据能力服务业务场景。同时，具备标准化的交付矩阵，支撑企业完成落地与知识转移，保障交付成功，支持企业以数智驱动增长。

培训

以培养数据架构师、业务架构师等数据人才为核心，集结业界顶尖的数据技术精英，为企业与个人提供大数据领域的体系化培训与专业认证，助其构建数智战略眼光与实战能力，成为 DT时代的骨干力量。

数据云交换

数据作为生产要素，只有流动起来，才能产生价值。奇点云提供全套的数据云交换“咨询 +产品”解决方案，以助力企业实现安全、可靠、可控的内部数据流通，发现全新增长机遇，实现数据增值。

安全可控

立足客户可控的数据技术
保障数据云交换全场景安全可靠

互联互通

构建隐私计算互联互通生态圈
支撑跨平台的数据云交换

价值增长

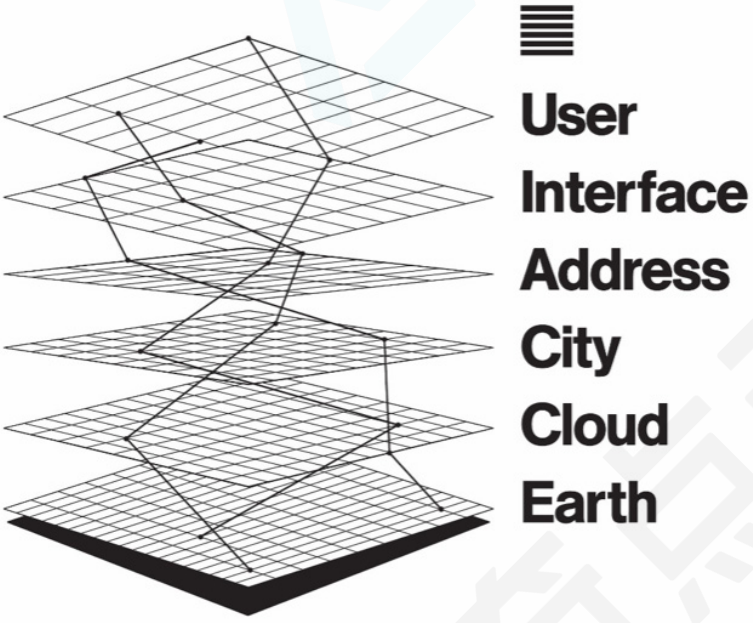
让数据在数据云上流动起来
实现多方共赢与数据增值

技术创新

以实际业务场景为核心
深化算法、加速性能，充分释放数据价值

结语：
数据未来将无死角地降临

数据的巨大影响力并不仅仅浸润至计算机系统和互联网，而是覆盖了人类生活的方方面面——人们同时生活在线下的实体社会和线上的虚拟世界。



△ 图 9：Metahaven of the six layers of The Stack

图源：Benjamin H. Bratton, The Stack: On Software and Sovereignty

美国社会学家本杰明·布拉顿（Benjamin H. Bratton）重新审视了实体世界和虚拟世界的关系，提出了平台和堆栈理论，在他看来，未来城市或许不再由建筑师来塑造，代码塑造了我们新的城市体验，编写代码的程序员才是城市真正的建筑师，“建筑将根据用户手机端的指令来设计，用户随时可以改变楼的内容和功能”。

布拉顿认为，未来世界将由 6个层面构成（图 9），而实体城市仅仅是其中薄薄的一层，在这个世界中，传统的权力、社区、交互等等都将由代码塑造。数据在其中流淌，用于分析和发现新的可能性。

数据的流动构成了堆栈的血脉，塑造着我们未来社区的方方面面，影响着每家企业未来的商业模式和竞争策略，也影响着每个人的生活。

那么，什么是数据的未来？

All for data，data for all.

参考文献

维克托·迈尔·舍恩伯格，《大数据时代：生活、工作与思维的大变革》，2013
安筱鹏，《重构：数字化转型的逻辑》，2019
中国信息通信研究院，《云计算发展白皮书（2020年）》，2020
国泰君安，《产业观察：数据仓库 破茧成蝶》，2021
William H. Inmon, Building the Data Warehouse (Fourth Edition), 2005
Benjamin H. Bratton, The Stack, 2016
IDC, Data Age 2025: The Digitization of the World From Edge to Core, 2018
IDG, 2020 Cloud Computing Survey, 2020
Ray Kurzweil and Kurzweil Technologies, Inc., www.singularity.com
Ray Kurzweil and Kurzweil Technologies, Inc., www.kurzweilai.net/the-law-of-accelerating-returns

著者
何夕

研究团队
星环 石木 北冥 三火

编校
何夕 此方

Data Cloud Trends Survey

免责声明

本文档仅代表奇点云的观察和建议，内容来源于奇点云核心团队的实践经验，所有文案、图表除特别注明来源外均为原创，包括但不限于定义、方法论、发展阶段、建设内容和实施方式。由于实践过程中存在不确定因素，可能导致实际过程和本趋势报告有较大出入。因此，本趋势报告信息仅供参考，不构成任何要约或承诺。奇点云可能不经通知修改上述信息，恕不另行通知。

版权所有 @杭州比智科技有限公司 保留一切权利

非经杭州比智科技有限公司同意，任何单位和个人不得擅自摘抄、复制本趋势报告内容的部分或全部，并不得以任何形式传播。

All for data, data for all.

400-080-0326
www.startdt.com
business@startdt.com
浙江省杭州市余杭区良睦路 1399 号梦想小镇互联网村 15、17 幢