01

第1章 大数据技术概述

授课教师：[李阳源]

15138617705@163.com

主页：https://sites.google.com/view/dalguri





Spark平台搭建与应用

大数据技术概述

Spark的设计与运行原理

实验环境搭建和使用方法

RDD编程

Spark SQL

Spark Streaming

Structured Streaming

Spark MLlib



第1章 大数据技术概述


BIG
DATA
ANALYTICS
COMPLEX
SOCIAL
INTERNET
TECHNOLOGIES
STORAGE
SYSTEMS
PETABYTES
INCLUDE
SOFTWARE
RESEARCH
INFORMATION
SIZE
PARALLEL
SEARCH
HUNDREDS
DATABASES
NETWORKS
RECORDS
MANAGEMENT
SETS
SHARED
DISK
EXAMPLES
EVERY
ALSO
TOOLS
TARGET
SENSOR
DEFINITION
MOVING
MANAGE
ARCHIVES
CAPTURE
PRACTITIONERS
BUSINESS
NEEDED
TIME
LARGE
ONE
PERFORMANCE
STORE
LOGS
COST
SAN
MASSIVELY
GROW
DIFFICULTY
GARTNER
WORKING
ELAPSED
TERABYTES
TOLERABLE
ABILITY
CASE
DISTRIBUTED
GENOMICS
CURRENT
MAY
WITHIN
USING
PROCESS
RELATIONAL
CONTINUES
LARGER
TENS
CITATION
BIOLOGICAL
PROCESSING
CAPACITY
SET
USE

目录

CONTENT

大数据时代

大数据概念

大数据的影响

大数据关键技术

大数据计算模式

代表性大数据技术



1.1 大数据时代

1.1 大数据时代



核心思维方式的3大转变：

1.全量数据分析 (More)：从抽样转向分析所有可获得的数据

2.重量不重精 (Messy)：接受略有不准确但海量的数据，而非完美准确的小数据

3.从因果关系到相关关系 (Correlation)：关注正在发生什么的模式和相关性，而非探究为什么发生

- **Big Data: A Revolution That Will Transform How We Live, Work, and Think**
- **作者**：维克托·迈尔-舍恩伯格（Viktor Mayer-Schönberger），肯尼思·库克耶（Kenneth Cukier）
- **出版年**份：2013年

1.1 大数据时代



信息技术发展的三次浪潮

- 第一次浪潮是个人电脑（PC）的普及
- 第二次浪潮是互联网的兴起
- 第三次浪潮，就是以“大数据”、“云计算”、“物联网”为代表的新时代



第三次信息化浪潮

1.1 大数据时代



1.1 大数据时代



第1次信息化浪潮（1980年代）——个人计算机（PC）的普及

1.1 大数据时代



手 纸和笔记录和处理相关的信息

1.1 大数据时代



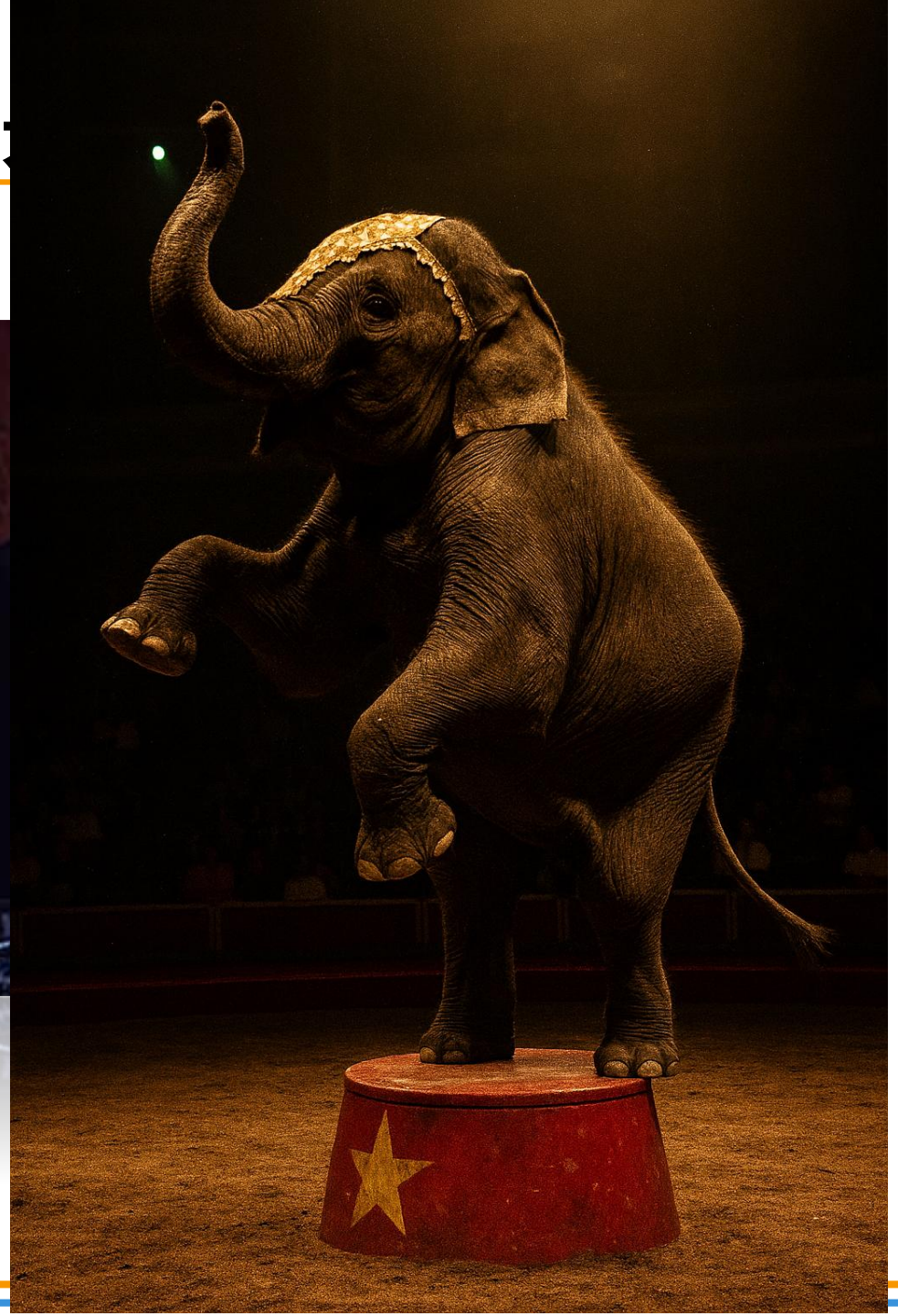
自动化处理的时代

1.1 大数据时代



社会生产力的发展







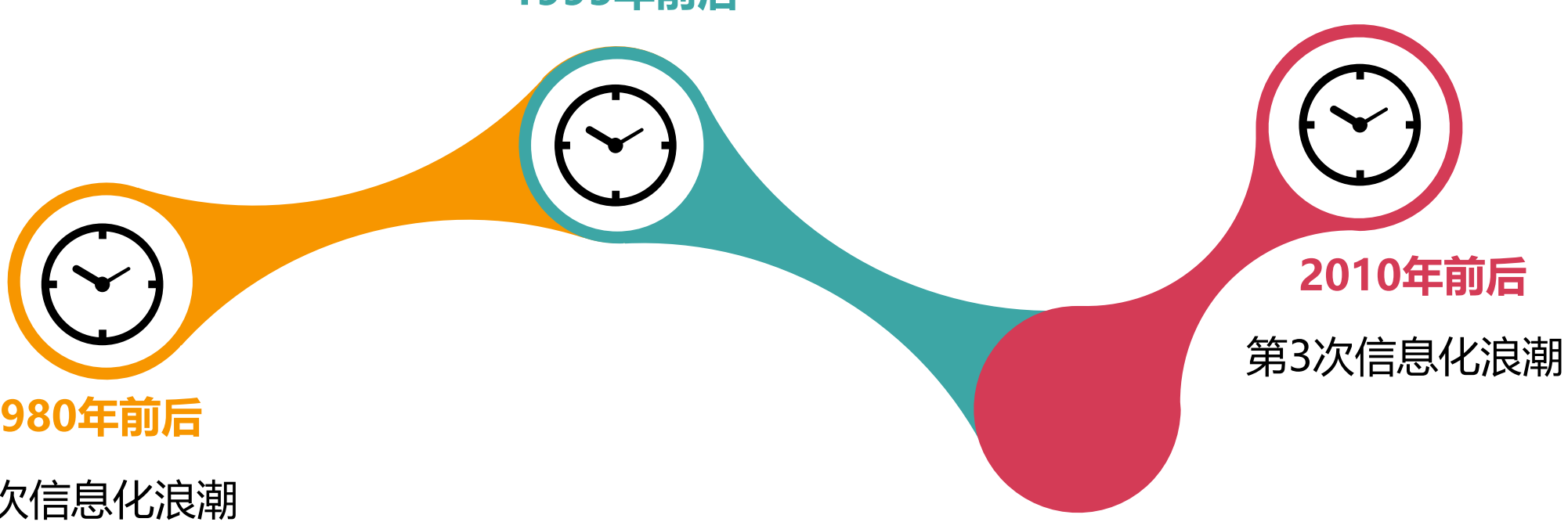
路易斯·郭士纳

IBM前首席执行官

大象能跳舞

"The elephant can dance."

1.1 大数据时代



1.1 大数据时代



1995年前后，以互联网的普及为标志

迎来第二次信息化浪潮



1.1 大数据时代



互联网在美国的军方

1.1 大数据时代



军用



民用



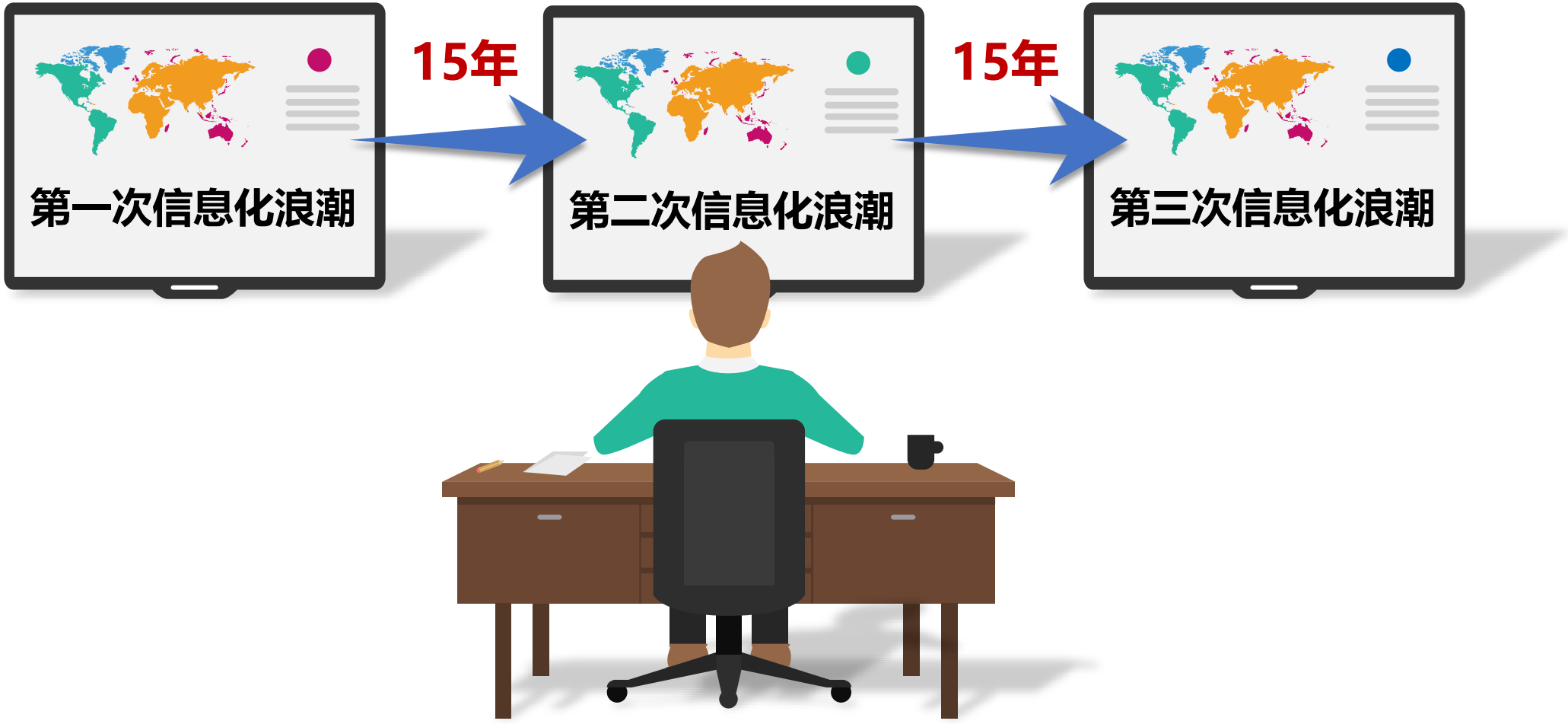
1.1 大数据时代



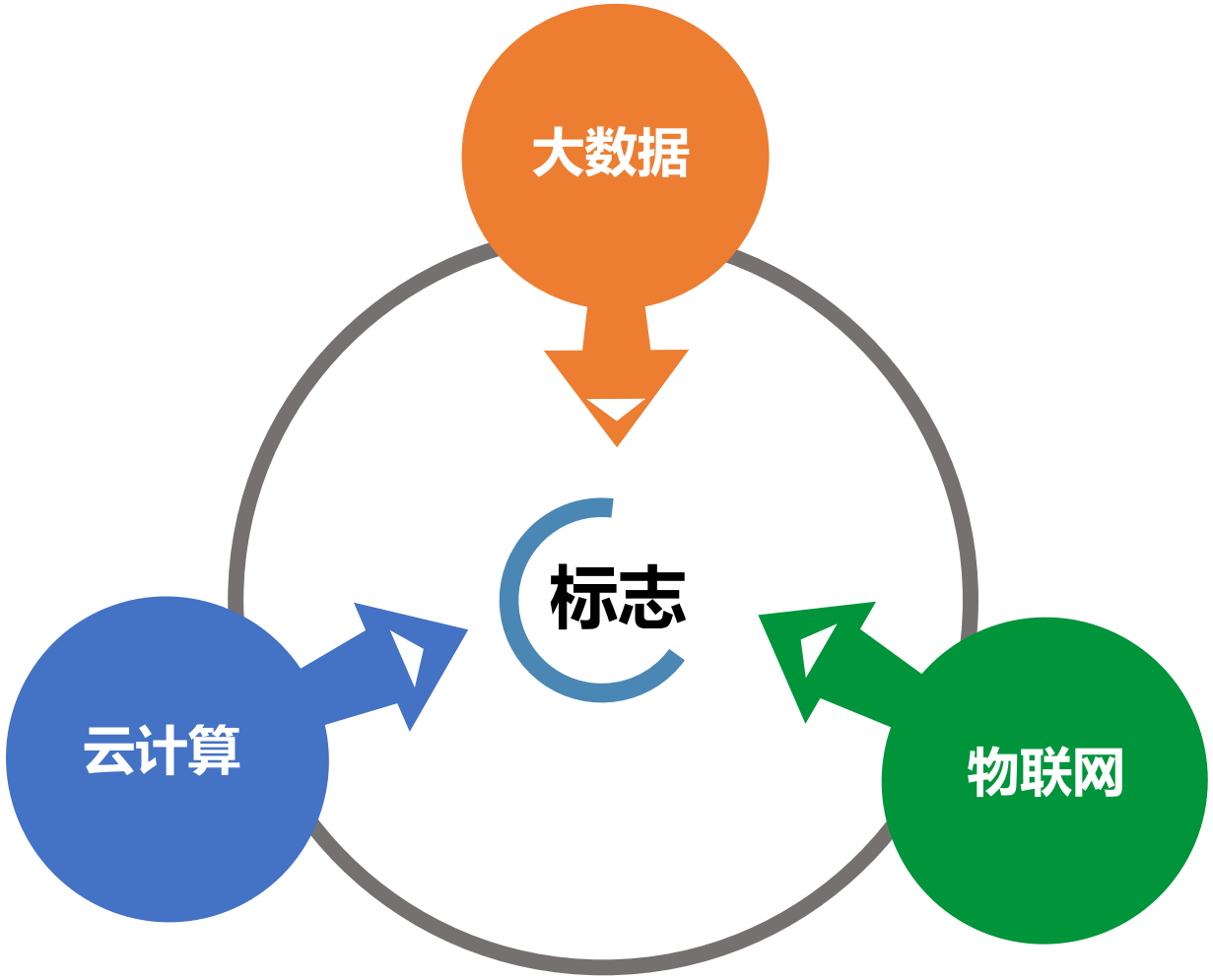
互联网解决什么问题呢?



1.1 大数据时代



1.1 大数据时代



1.1 大数据时代



信息化浪潮	发生时间	标志	解决问题	代表企业
第一次浪潮	1980年前后	个人计算机	信息处理	Intel、AMD、IBM、苹果、微软、联想 戴尔、惠普等
第二次浪潮	1995年前后	互联网	信息传输	雅虎、谷歌、阿里巴巴、百度、腾讯等
第三次浪潮	2010年前后	物联网、云计算和大数据	信息爆炸	将涌现出一批新的市场标杆企业

1.1 大数据时代

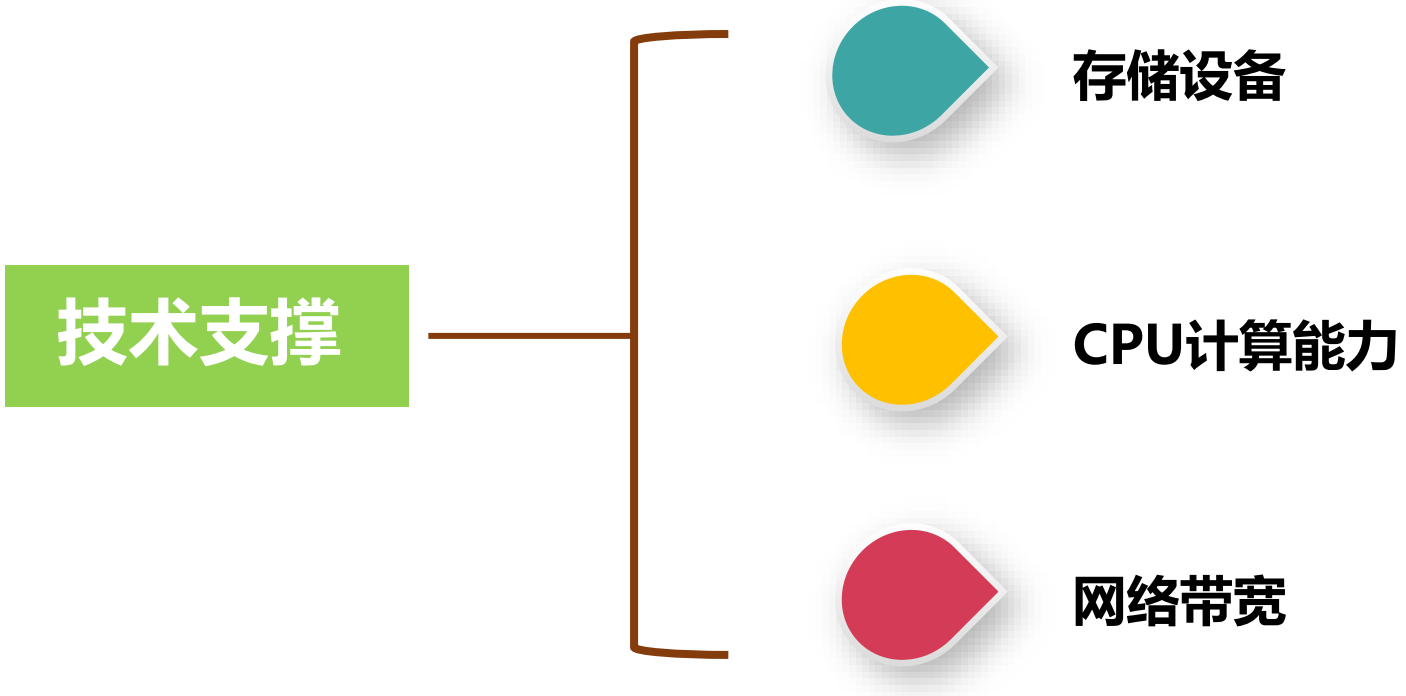


大数据时代为什么会到来?



1.1 大数据时代

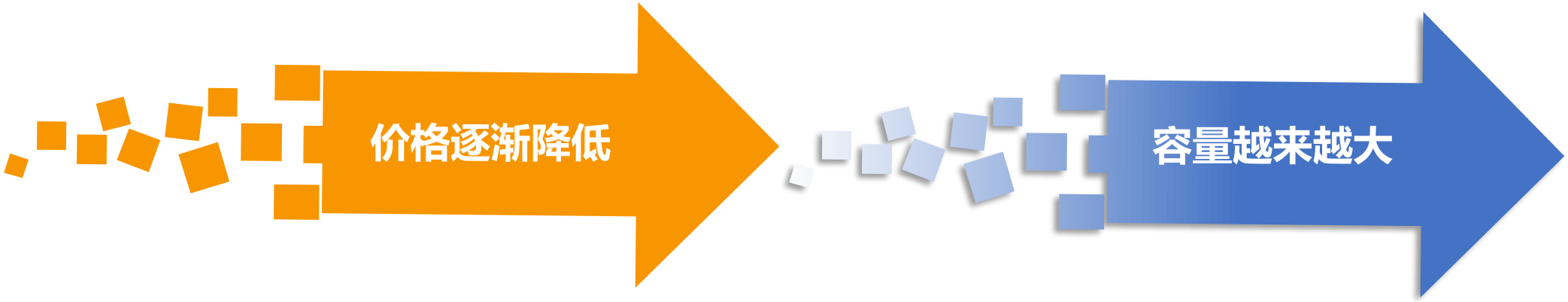




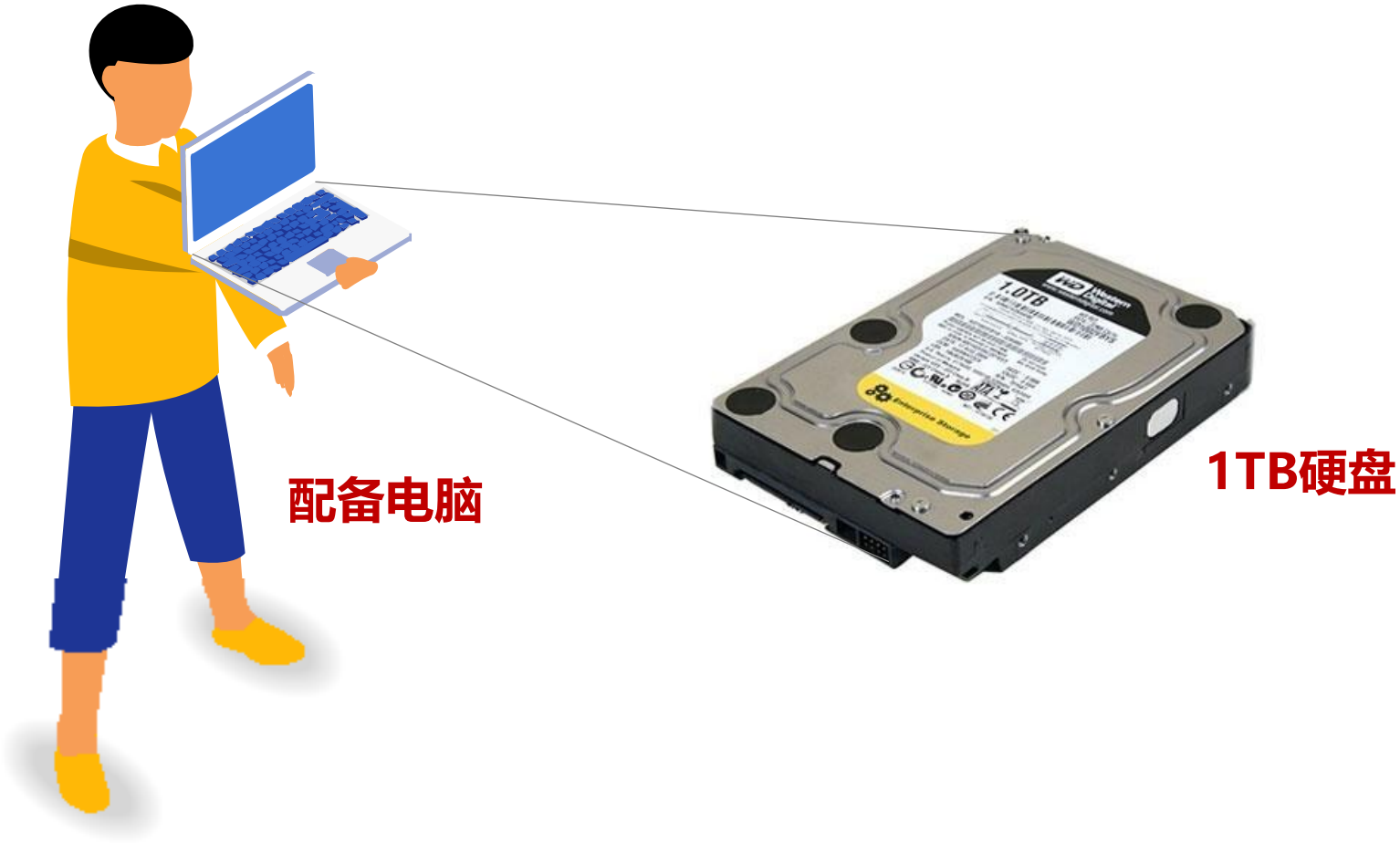
1.1 大数据时代



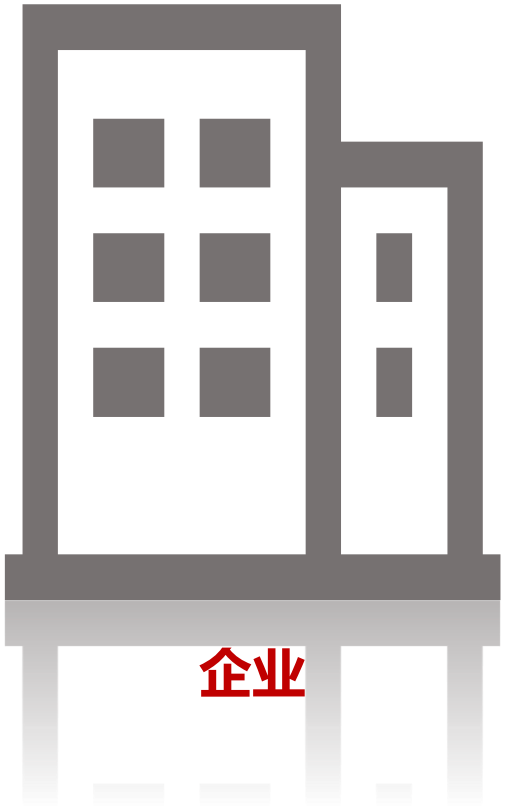
1.从存储设备来讲



1.1 大数据时代



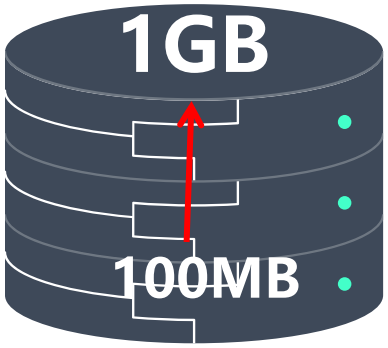
1.1 大数据时代



企业



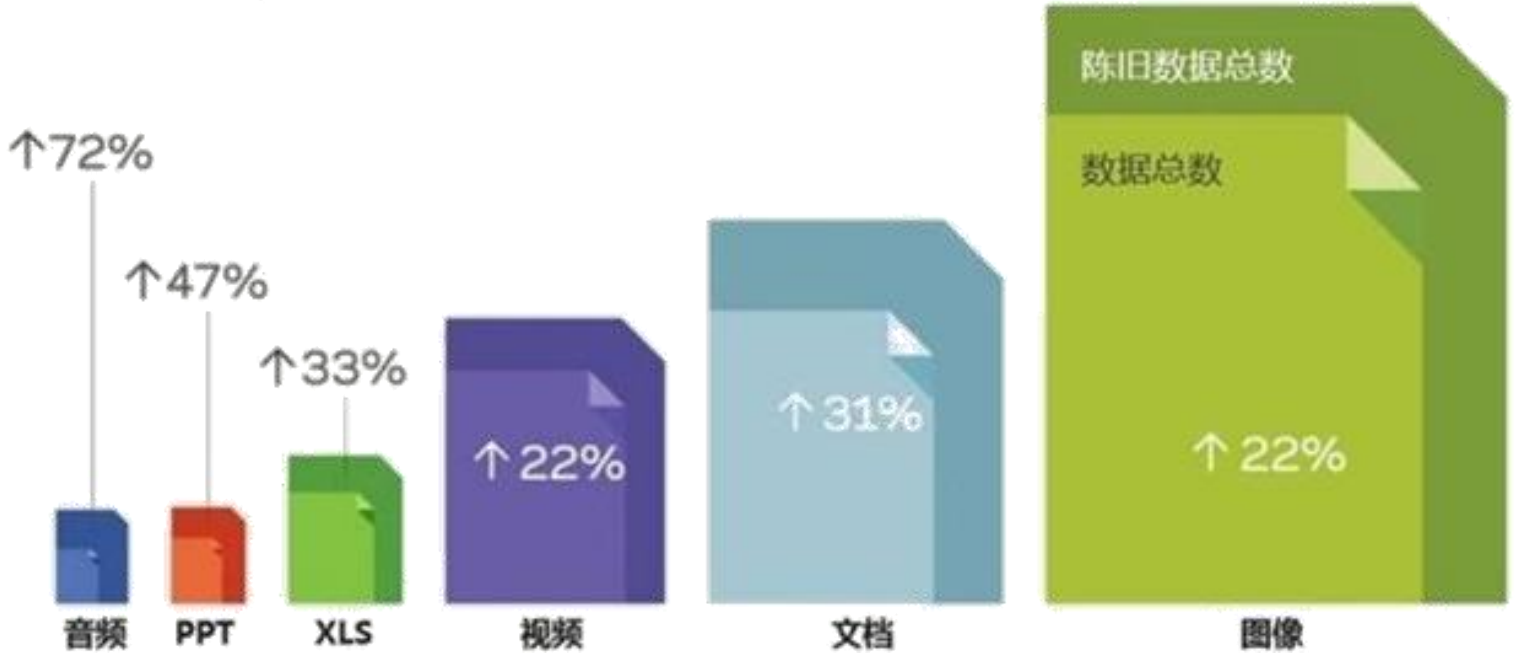
更多资金



存储设备

1.1 大数据时代

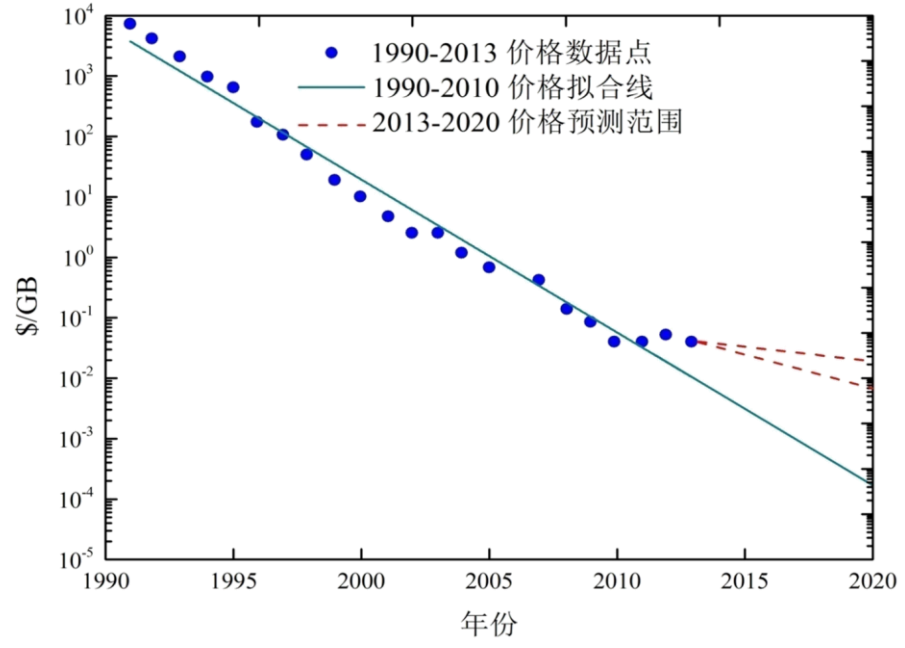
企业数据越来越多

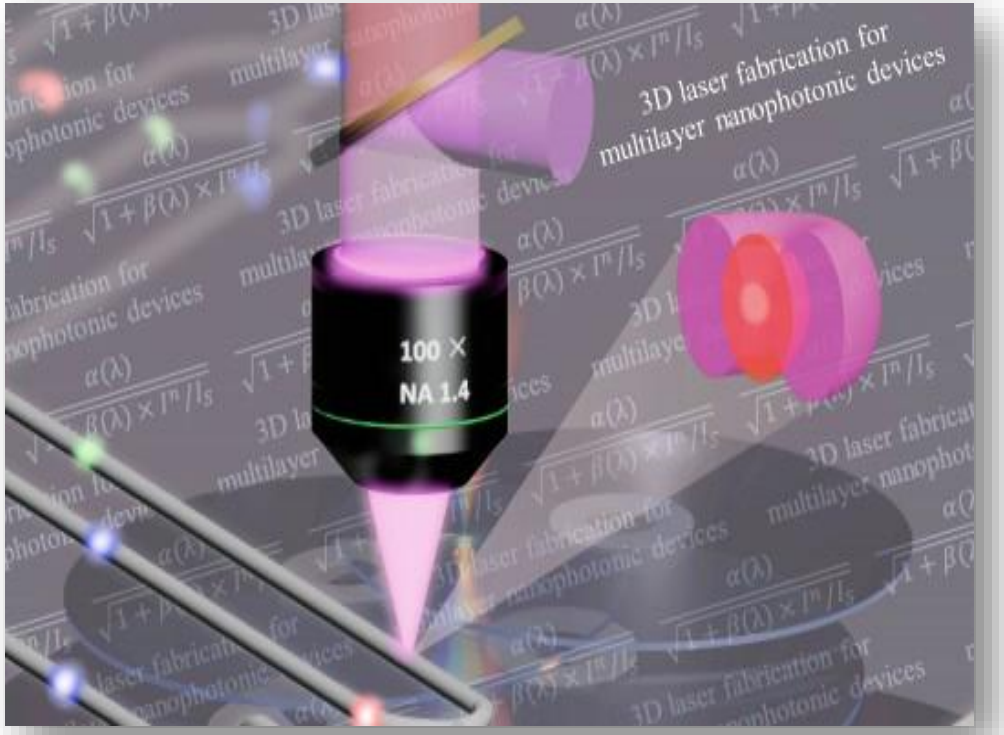


1.1 大数据时代

1.存储设备容量不断增加

存储价格随时间变化情况

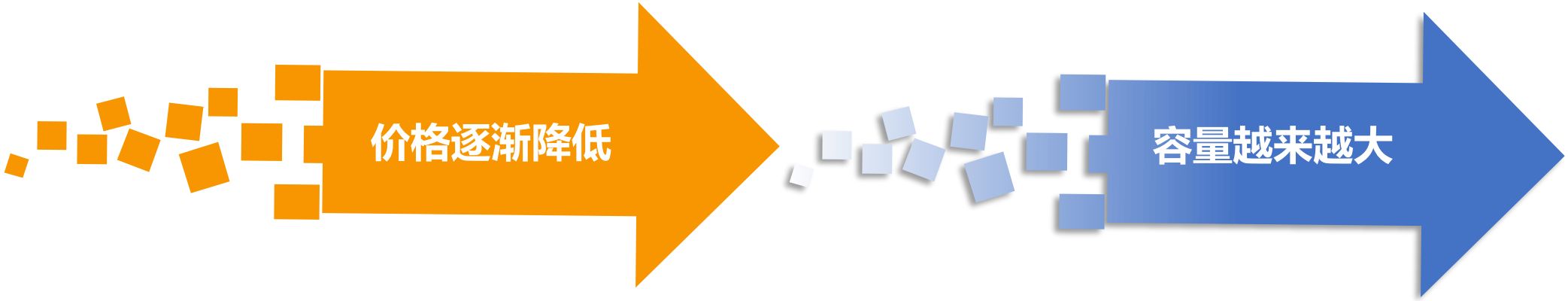




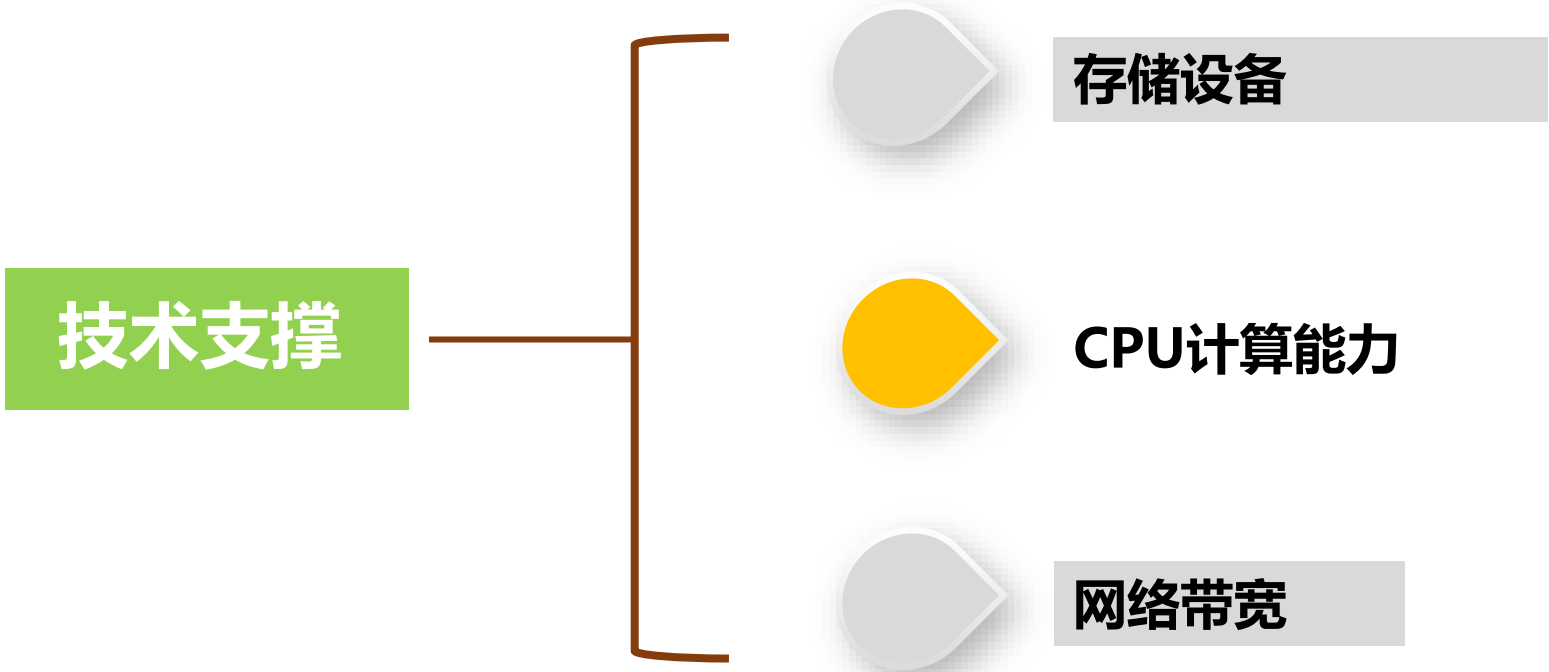
1.1 大数据时代



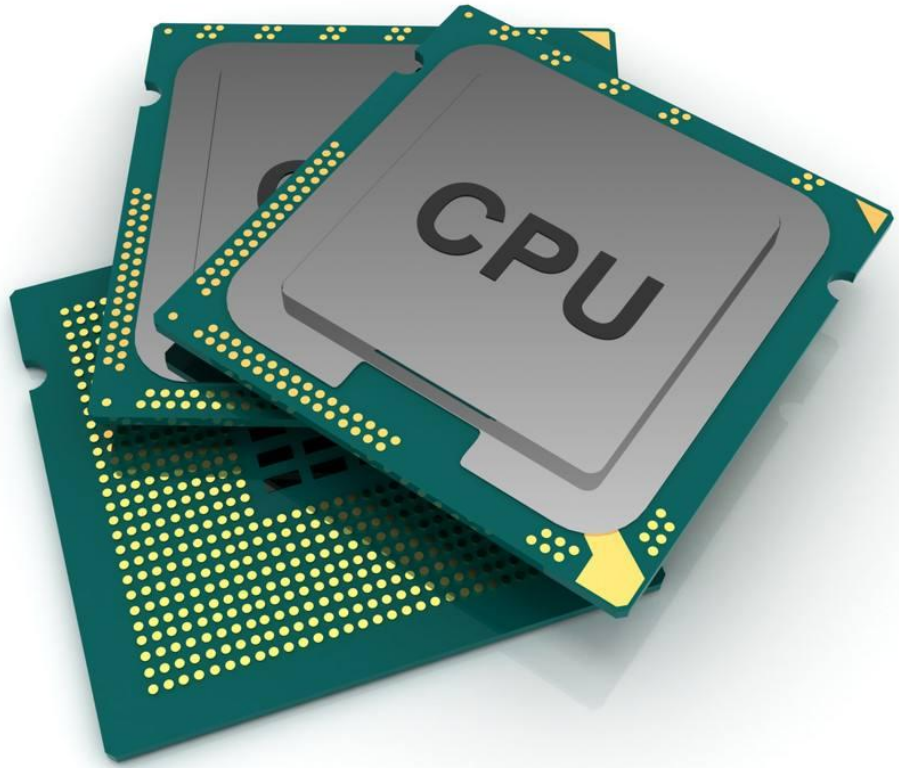
1.存储设备容量不断增加



1.1 大数据时代



2.CPU处理能力大幅提升



1.1 大数据时代

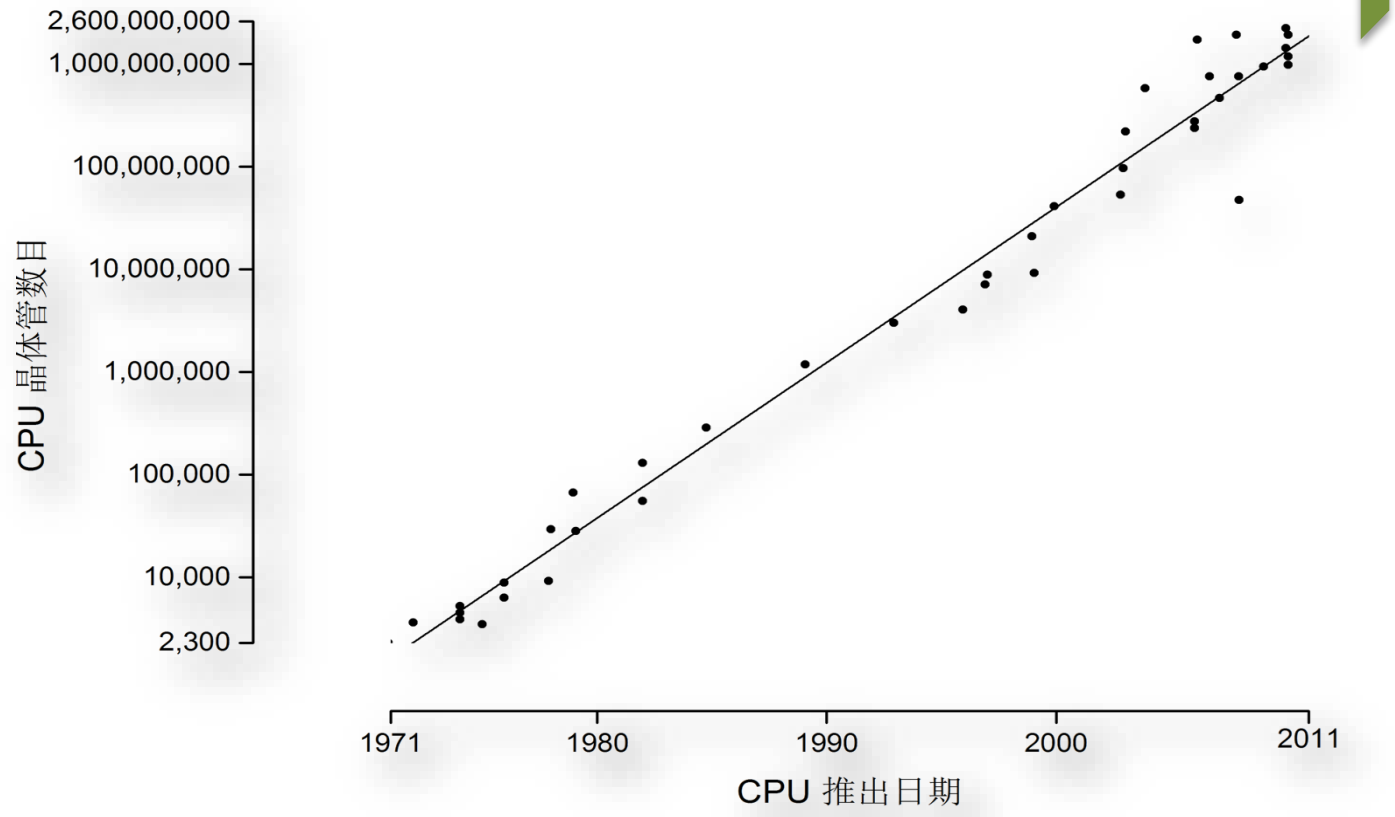


CPU处理能力大概每隔18-24个月会翻一番

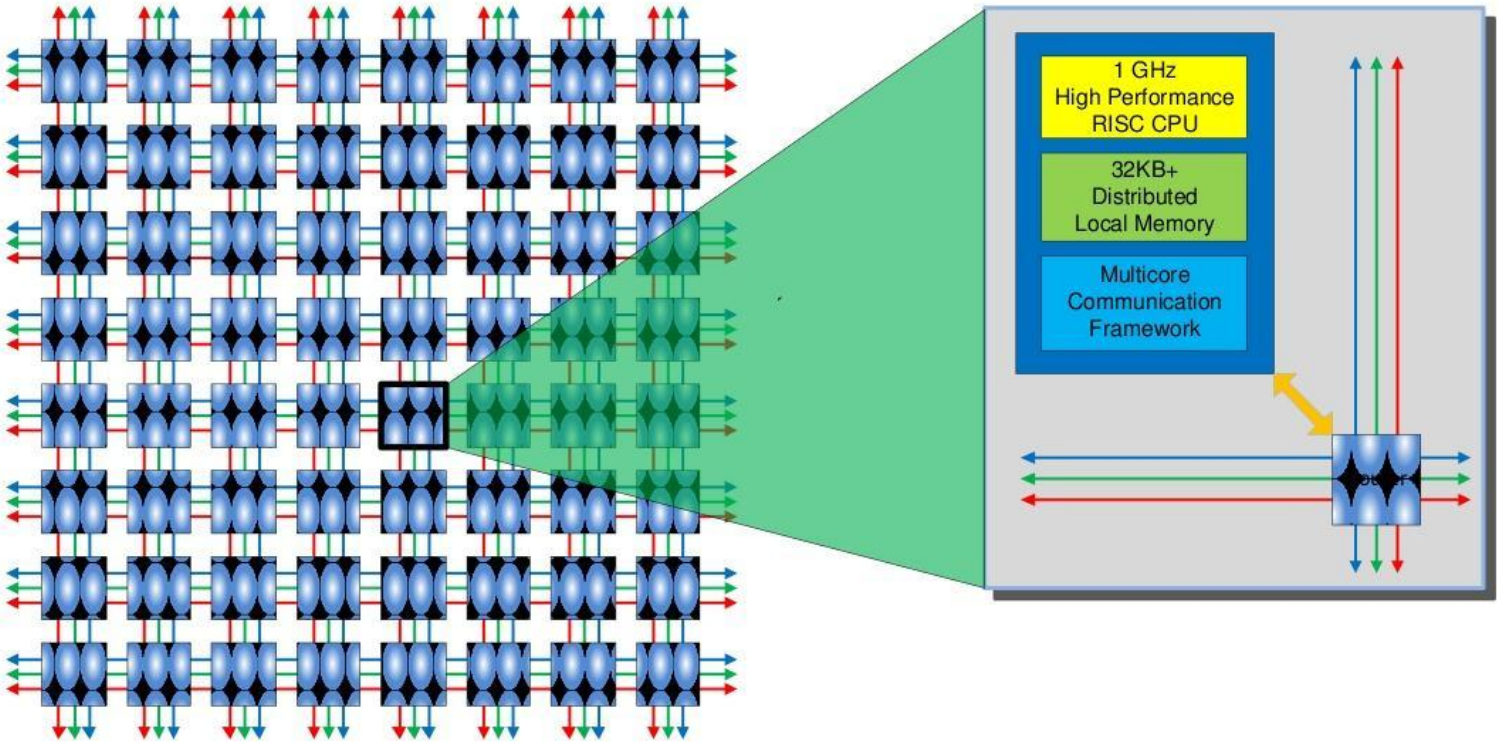
1.1 大数据时代

2.CPU处理能力大幅提升

CPU晶体管数目随时间变化情况



2.CPU处理能力大幅提升

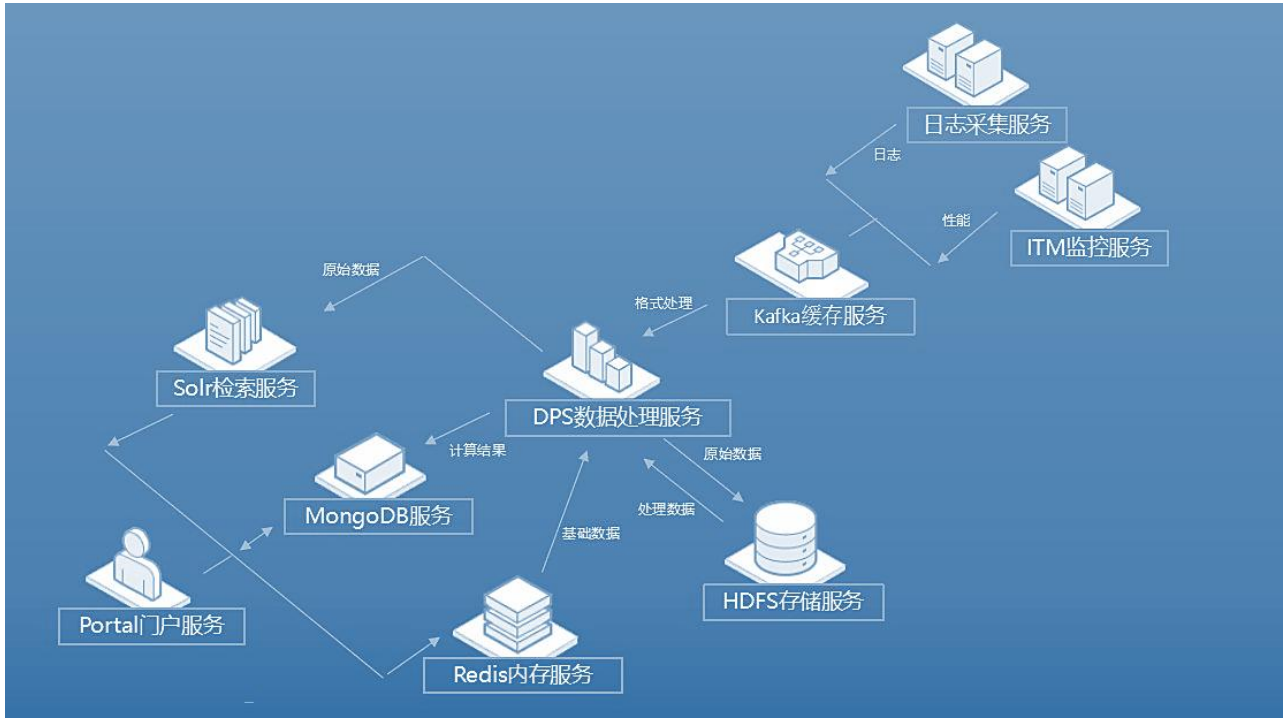




1.1 大数据时代

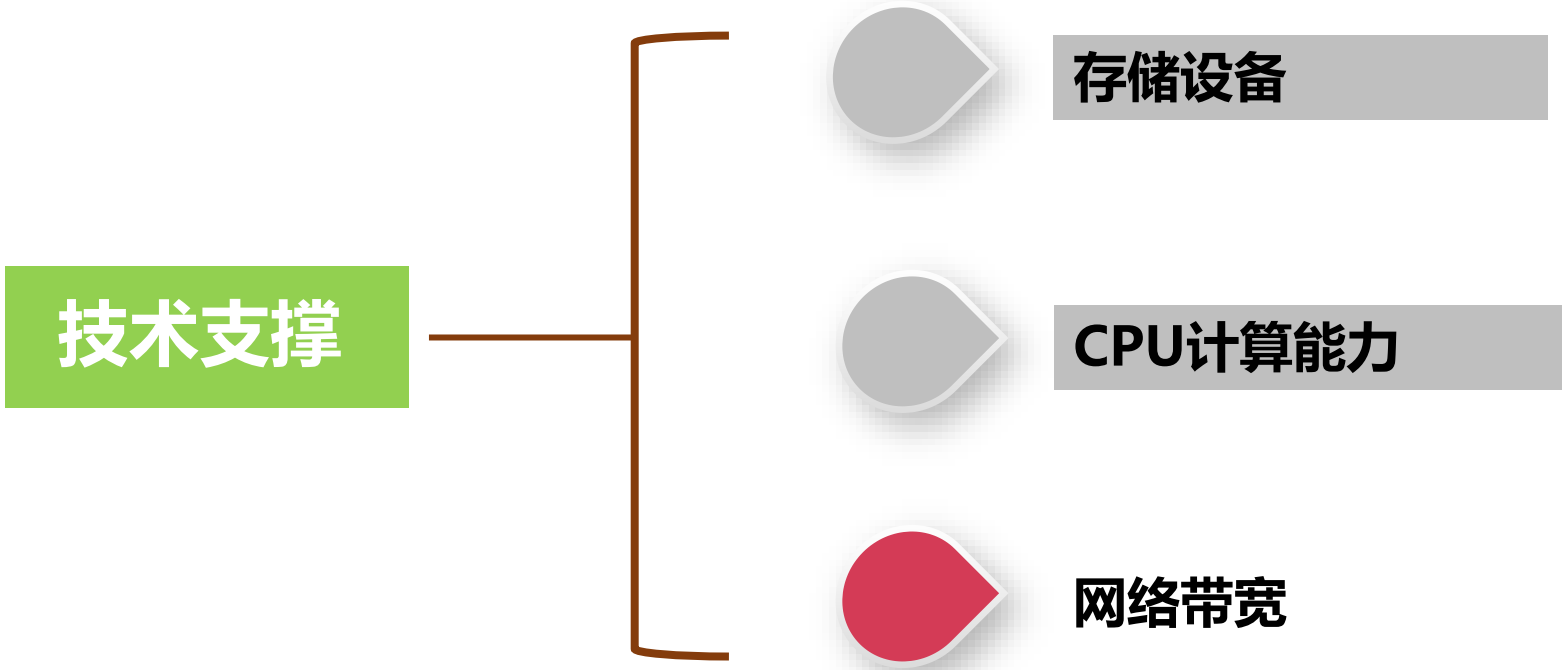


1.1 大数据时代

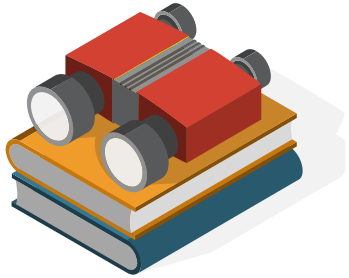


大数据时代特性 分布式处理

1.1 大数据时代



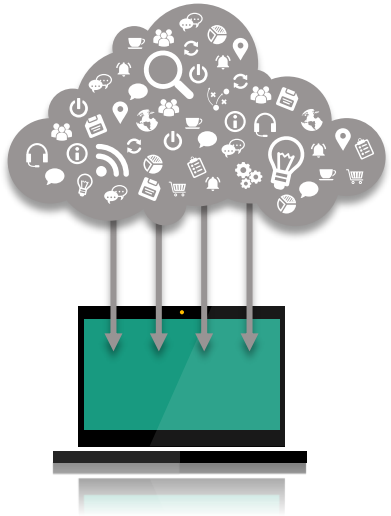
1.1 大数据时代



存储海量数据



管理海量数据



分析海量数据

大数据相关应用

1.1 大数据时代





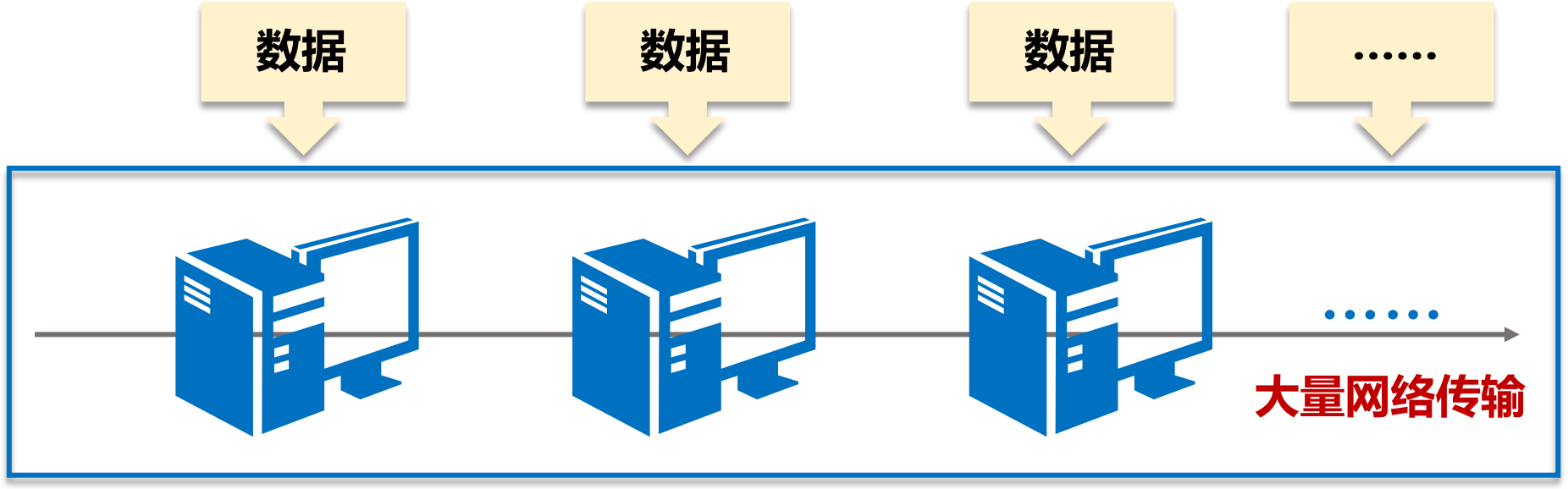
单机不能够完成海量数据存储和处理

1.1 大数据时代



借助网络分布式的集群运算

1.1 大数据时代

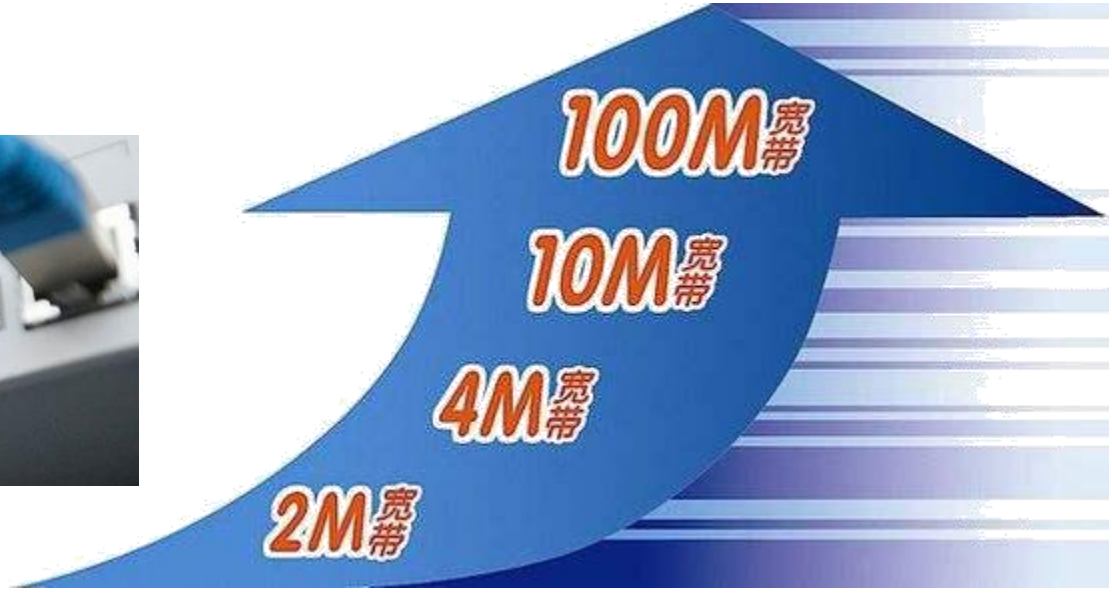




1.1 大数据时代

3.网络带宽不断增加

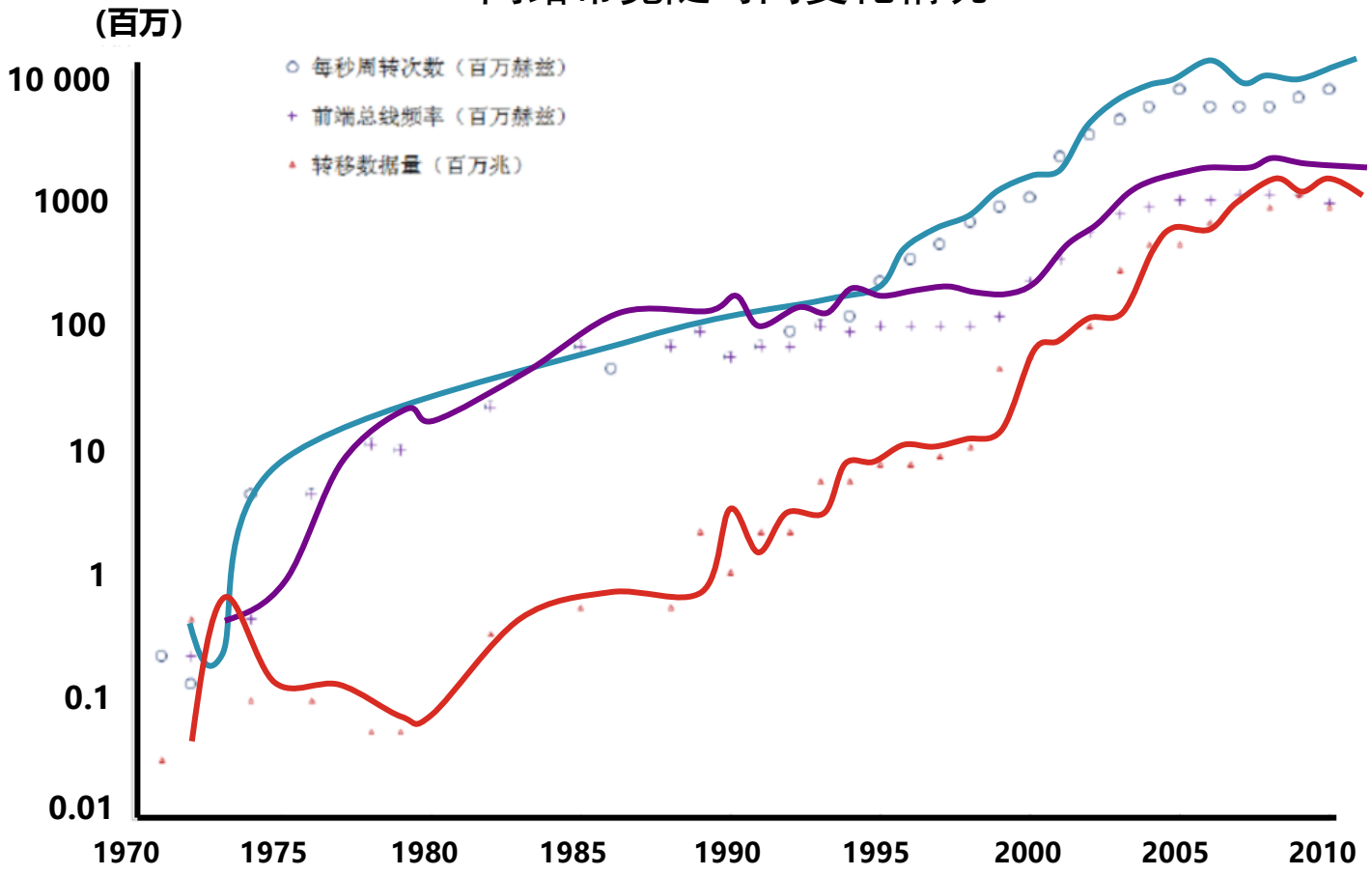




1.1 大数据时代

3.网络带宽不断增加

网络带宽随时间变化情况





1.1 大数据时代

数据产生方式的变革



1.1 大数据时代

运营式系统阶段实例——超市购物，在数据库系统中生成购物信息



企业零售系统数据库

1.1 大数据时代



企业零售系统数据库

日常销售 : 表

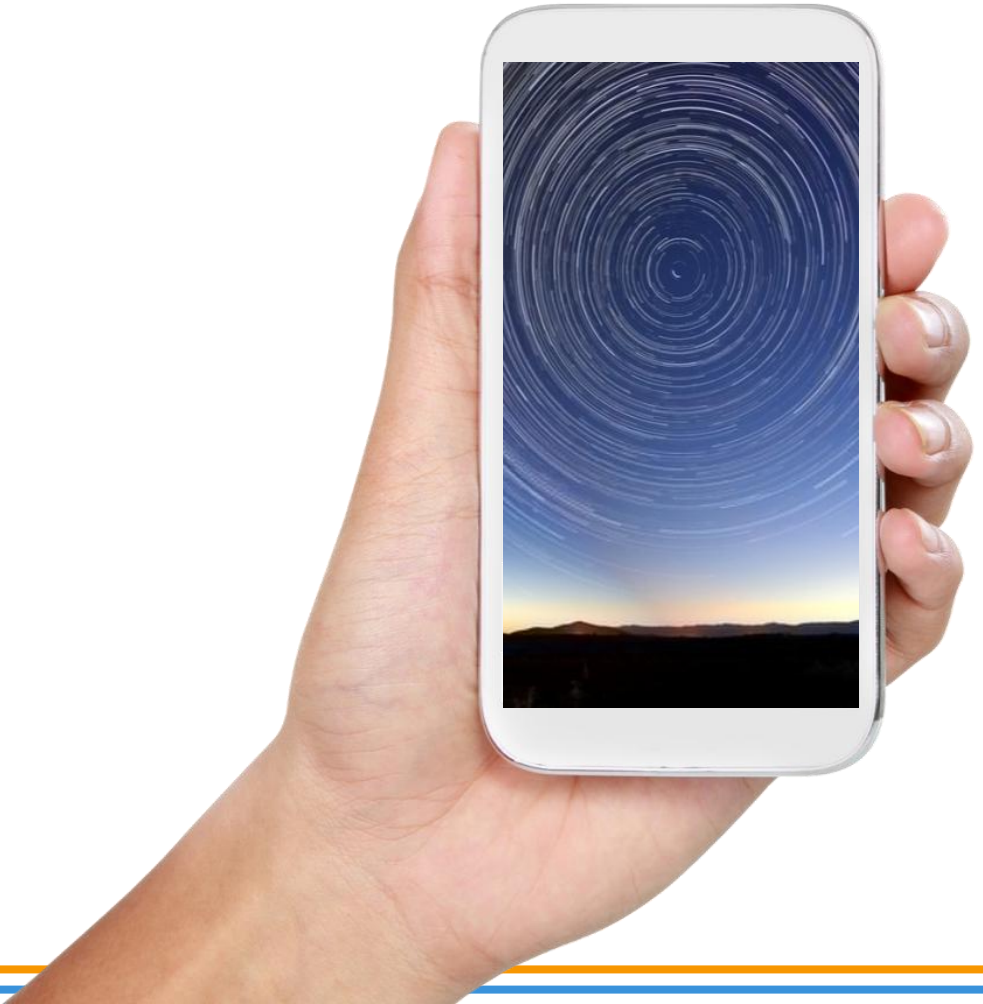
商品编号	品牌	类型	计量单位	数量	销售单价	总价	日期	ID
00004	XZA	帽子	顶	1	¥120.00		2013-2-1	1
00001	AAAA	上衣	件	1	¥680.00		2013-2-1	2
00004	XZA	帽子	顶	1	¥120.00		2013-2-1	3
01203	WEQ	裤子	条	3	¥320.00		2013-2-1	4
00004	XZA	帽子	顶	1	¥120.00		4694-2-1	5
01203	WEQ	裤子	条	1	¥350.00		2013-2-2	6
				0	¥0.00	¥0.00		(自动编号)

1.1 大数据时代



1.1 大数据时代

用户原创内容阶段





1.1 大数据时代

数据产生方式的第二次变革



>>

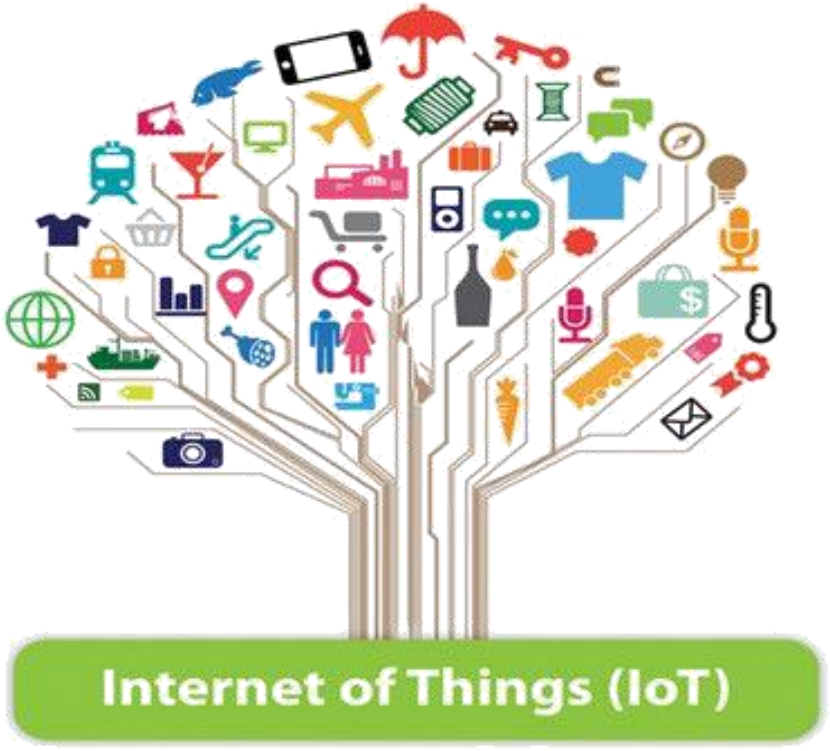


用户原创内容阶段信息的生成量

运营式系统阶段信息的生成量

1.1 大数据时代

感知式系统阶段



物联网（IOT）实现万物互联

1.1 大数据时代

感知式系统阶段







1.1 大数据时代



气象温、湿度传感器

1.1 大数据时代



大数据技术和视频监控的结合

1.1 大数据时代



物联网的兴起



1.1 大数据时代



大数据的发展历程

时间	阶段
上世纪90年代 至上世纪末期	萌芽期
本世纪前十年	成熟期
2010年以后	大规模应用期


AI
GoodBye