Slide 1:

Slide 2:

"Spark 플랫폼 구축과 응용" 강의 소개

여러분, 안녕하세요!
이번 학기에는 **Spark 빅데이터 기술**을 체계적으로 학습하게 되며, 본 강의는 기초 원리부터 고급 응용까지의 완전한 지식 체계를 아우르는 총 8개의 핵심 모듈로 구성되어 있습니다.

강의 주요 내용은 다음과 같습니다:

1. **Spark 빅데이터 기술 개요**
   - Spark의 발전 배경, 기술적 특징, 생태계 소개
   - Hadoop과 Spark의 구조적 차이와 장점 비교

2. **Spark의 설계와 실행 원리**
   - Spark의 핵심 아키텍처, 작업 스케줄링 및 실행 과정 심층讲解
   - 메모리 계산, 탄력적 분산 데이터셋(RDD) 등 핵심 개념 분석

3. **실험 환경 구축과 사용 방법**
   - Spark 로컬 및 클러스터 환경을 직접 구축해 봄
   - Spark Shell, Jupyter Notebook 등 개발 도구 사용법 학습

4. **RDD 프로그래밍**
   - RDD의 생성, 변환, 액션 연산 숙달
   - 분산 데이터 처리 프로그래밍 논리 구현

Slide 1:

Slide 2:

《Spark平台搭建与应用》课程介绍

各位同学，
本学期我们将系统学习**Spark大数据技术**，本课程共分为八大核心模块，涵盖(**hán gài)**从基础原理到高级应用的完整知识体系。

课程主要内容如下：

1. **Spark大数据技术概述**
   - 介绍Spark的发展背景(**bèijǐng)**、技术特性与生态系统
   - 对比Hadoop与Spark的架构差异与优势
2. **Spark的设计与运行原理**
   - 深入讲解Spark的核心架构、任务调度与执行流程
   - 解析内存计算、弹性分布式数据集（RDD）等核心概念

3. **实验环境搭建和使用方法**
   - **手把手带领**大家搭建Spark本地与集群环境
   - 学习使用Spark Shell、Jupyter Notebook等开发工具
4. **RDD编程**
   - 掌握RDD的创建、转换与行动操作
   - 实现分布式数据处理的编程逻辑

5. **Spark SQL**

- DataFrame과 Dataset을 이용한 구조화 데이터 처리 학습
- Hive, JSON, Parquet 등 다양한 데이터 소스 통합

6. **Spark Streaming**

- 기존의 전통적인 스트림 처리 아키텍처와 API 이해
- 참고: 본 부분은 Structured Streaming으로 대체되어, 여기 강의에서는 생략함

7. **Structured Streaming**

- 현대적인 스트림 처리 기술을 중점적으로 학습
- 실시간 데이터 스트림 처리와 분석 구현

8. **Spark MLlib**

- Spark 머신러닝 라이브러리 입문
- 특성 공학, 모델 훈련, 평가 등 실전 과제 수행

5. **Spark SQL**

- 学习使用DataFrame和Dataset进行结构化数据处理
- 整合Hive、JSON、Parquet等数据源

6. **Spark Streaming**

- 了解传统的流处理架构与API
- 本部分内容已由Structured Streaming替代，本课程不再讲解

7. **Structured Streaming**

- 重点学习现代流处理技术
- 实现实时数据流的处理与分析

8. **Spark MLlib**

- 入门Spark机器学习库
- 完成特征工程、模型训练与评估等实战任务

페이지를 넘기지 말고 아래 내용 계속하세요!!!

강의 목표:
본 강의를 통해 여러분은 다음과 같은 능력을 갖추게 될 것입니다:
- Spark 환경을 독립적으로 구축하고 구성할 수 있음
- Python을 사용하여 빅데이터 처리 작업을 완수할 수 있음
- 구조화 데이터, 스트림 데이터, 머신러닝 응용 프로그램의 개발 방법을 숙지함
- 기업급 Spark 응용 프로그램의 기본 개발 능력을 보유함

궁금한 사항이 있으면 수업 중이나 수업 후 언제든지 질문해 주세요.

함께 Spark 빅데이터 기술 학습의 여정을 시작해 보겠습니다!

Slide 3:

Slide 4:

Slide 5: 빅데이터란 무엇인가?
- 빅데이터란 무엇인가?
- 왜 우리는 지금을 "빅데이터 시대"라고 말하는가?
- 그것은 어떻게 생겨났는가?
- 그리고 우리의 삶에 어떤 영향을 미칠 것인가?

课程目标：

通过本课程，大家将能够：

- 独立搭建和配置Spark环境
- 使用Python完成大数据处理任务
- 掌握结构化数据、流数据与机器学习应用的开发方法
- 具备企业级Spark应用的基本开发能力

如果大家有任何问题，欢迎随时在课堂或课后提出。

让我们一起开启Spark大数据技术的学习之旅！(**lǚ)**

Slide 3:

Slide 4: PPT보면서 목록을 이야기 한다.

Slide 5: **大数据时代**

什么是大数据？
为什么我们说现在是"大数据时代"？
它是怎么来的？
又会对我们的生活带来哪些影响？

Slide 6: 빅데이터의 특징과 영향

빅데이터는 단순한 기술 혁명이 아니라, 우리가 세계를 이해하는 방식을 바꾸는 사고의 혁명입니다.

핵심 사고방식의 3가지 변화

1. **전체 데이터 분석 (More)**: 샘플링에서 벗어나 가능한 모든 데이터를 분석하는 능력
2. **정확성보다 양 중시 (Messy)**: 완벽하게 정확한 작은 데이터보다 약간 부정확하더라도 대량의 데이터를 수용
3. **인과관계에서 상관관계로 (Correlation)**: 왜 그런지 원인을 찾는 것보다 무엇이 일어나고 있는지 패턴과 상관관계에 집중

Live, Work, Think 각 영역의 변화

- **Live (생활)**: 구글 검색이 독감 확산을 의료 전문가보다 더 정확하게 예측하는 것처럼, 일상생활에서 데이터 기반 예측과 개인화 서비스가 일반화
- **Work (업무)**: 기업들이 N=전체 접근법을 통해 완전한 데이터셋을 분석할 수 있게 되어, 추정이나 표본에 의존하지 않고 전체 그림을 파악
- **Think (사고)**: 문제 해결 방식이 근본적으로 변화 - "왜?"를 묻는 대신 "무엇이?"에 집중하며, 정확한 원인보다는 유용한 결과를 우선시

Slide 6.

什么是大数据？

大数据有三大特征

1.数量巨大（Volume）：不再依赖抽样，直接分析全部数据。

2.速度极快（Velocity）：实现实时或近实时的数据处理。

3.数据多样（Variety）：涵盖**(hán gài)** 了结构化数据和非结构化数据（如文本、图像、位置等半，结构化数据。

大数据不仅是技术革命，更是改变我们理解世界方式的思维革命。

核心思维模式的三大转变：

- **全数据分析：** 摆脱**(bǎi tuō)**抽样分析，转向分析全部可用数据的能力。
- **重数量轻精确：** 接受大体量但略有**(lüè yǒu)**瑕疵**(xiá cī,** 하자)的数据，而非完全精确的小量数据。
- **从因果到相关：** 关注"发生了什么"的模式和相关性，而非执着于 **(zhí zhuó yú)**探究 **(tàn jiū)** "为什么"。

生活、工作、思维各领域的变化：

- **生活：** 数据驱动的预测和个性化服务普及化，例如谷歌搜索**(sōu suǒ)**能比医疗专家更准确地预测流感传播。
- **工作：** 企业能够通过"N=全部"的方法分析完整数据集，从而掌握**(zhǎng wò)**全局，不再依赖估算或样本。

- **思维：** 问题解决方式发生根本性变化——更关注"是什么"而非追问"为什么"，优先考虑有用的结果而非精确的原因。

一本很好地阐述(chǎnshù)了这些内容的书，正是维克托·迈尔-舍恩伯格（Viktor Mayer-Schönberger）所著的《大数据时代：一场将改变我们生活、工作和思维的革命》。
感兴趣的同学不妨 **(bùfáng)**读一读。

Slide 7: 제3차 정보화 물결

이 페이지는 "빅데이터 시대"가 제3차 정보화 물결의 핵심이라는 내용입니다.

- **제1차 물결**: 개인용 컴퓨터(PC)의 보급
- **제2차 물결**: 인터넷의 부상
- **제3차 물결**(현재): "빅데이터", "클라우드 컴퓨팅", "사물인터넷"을 대표로 하는 새로운 시대

여러분도 느끼시나요? 우리의 삶은 이미 이러한 기술들을 떠날 수 없게 되었습니다.

특히 최근에는 AI 기술까지 융합되면서, 아침에 눈을 뜨는 순간부터 밤에 잠들 때까지, 일상생활이 모두 이 기술들과 밀접하게 관련되어 있음을 실감합니다.

스마트폰으로 음식을 주문하고, 클라우드에 사진을 저장하고, 스마트 홈 기기까지... 이러한 기술들은 이미 우리 삶에 깊이 뿌리내렸습니다.

이제는 정말로 "디지털 디톡스"가 필요한 정도입니다.

Slide 7: 浪潮 **"làng cháo"** 普及 **"pǔ jí"**

这一页讲的是"大数据时代"是第三次信息化浪潮的核心。

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，也就是现在，就是以"大数据"、"云计算"、"物联网"为代表的新时代。

你们有没有感觉到，我们的生活已经离不开这些技术了？

尤其是最近连AI技术也融合进来，
从早上睁眼(zhēng yǎn) 到晚上入睡，
深切感受到日常生活都与这些技术密不可分。

用智能手机点餐、云端存储的照片、
智能家居设备...
这些技术已经深深扎根(**zhāgēn**)于我们的生活，

现在真的到了需要"数字排毒(**páidú: Detox / Detoxification**)"的程度呢。

Slide 8: 빅데이터 시대의 시작

우리가 말하는 2010년경은 빅데이터 시대의 진정한 시작입니다.

비록 이 개념은 이미 오래전부터 있었지만, 실제로 응용되기 시작하고 영향을 미치기 시작한 것은 이 시기부터입니다.

그리고 **2013년**은 중국의 빅데이터 원년으로 널리 인정받고 있습니다. 즉, 그때 중국이 비로소 이 기술을 진정으로 중시하기 시작했습니다.

Slide 8:

我们说2010年左右，
是大数据时代的真正开始。

虽然这个概念早就有了，
但真正开始应用并产生影响，
是从这个时候开始的。

而2013年，被普遍认为是中国的大数据元年，
也就是说，那时中国才真正开始重视这项技术。

Slide 9: 정보화의 발전 역사

이제 정보화의 발전 과정을 회고해봅시다.

제1차 정보화 물결은 1980년대로, PC의 보급과 함께 우리는 컴퓨터를 사용하여 정보를 기록, 저장 및 처리하기 시작했습니다.

과거에는 종이와 펜으로 정보를 기록했지만, 지금은 Excel, Word, 심지어 AI를 사용합니다. 이것이 바로 정보화가 가져온 변화입니다.

Slide 9:

我们现在回顾一下信息化的发展历程。

第一次信息化浪潮是1980年代，随着PC的普及(pǔ jí)，我们开始用电脑来记录、存储(cún chǔ)和处理信息。

过去我们用纸和笔记录信息，现在我们用Excel、Word甚至AI。这就是信息化带来的变化。

Slide 10: 정보 처리 방식의 진화

이 그림은 정보 처리 방식의 진화를 보여줍니다.

처음에는 수작업으로 기록했습니다. 예를 들어 장부, 일기 등이었고, 나중에는 자동화 처리로 발전했습니다.

즉, 컴퓨터가 우리의 두뇌를 대신해서 대량의 데이터를 처리하게 되었습니다.

여러분은 집에서 가계부 소프트웨어나 건강 데이터 모니터링을 사용해본 적이 있나요?

이것이 사실 자동화의 한 표현입니다.

Slide 11: 정보 자동화 처리 시대

이제 우리는 정보 자동화 처리 시대로 진입했습니다.

예전에는 종이와 펜을 사용했고, 지금은 컴퓨터를 사용하는데, 미래는 어떨까요?

아마도 AI가 보고서를 작성하고, 우리는 "보고" "수정"하는 역할만 맡게 될 것입니다. 어떻게 생각하시나요? 😄

자동화의 장점은 효율성을 높이고 오류를 줄이며, 대규모 데이터를 처리할 수 있다는 것입니다.

Slide 10:

这张图展示了信息处理方式的演变。

最初我们是手工记录，比如账本(zhàng běn)、日记等，后来发展到自动化处理。

也就是说，电脑代替了我们的大脑来处理大量的数据。

大家有没有用过家里的记账软件(**jì zhàng ruǎn jiàn)**、健康数据监测((jiàn kāng jiān cè) ？

这其实就是自动化的一种表现。

Slide 11:

现在我们进入信息自动化处理的时代。

以前我们用纸和笔，现在用电脑，未来呢？

也许就是 AI 写报告，我们只负责"看"和"批改(**pī gǎi)**"，你们觉得怎么样？😄

自动化的好处就是提高效率，减少错误，而且可以处理大规模的数据。

Slide 12: 사회 생산력의 발전

자동화를 말하자면, "사회 생산력의 발전"을 언급하지 않을 수 없습니다.

기술이 있으면 우리의 업무 효율이 점점 더 높아집니다.

상상해보세요. 만약 지금 배달 앱이 없다면, 우리는 매일 요리를 해야 하고, 공부할 시간이 전혀 없을 것입니다!
과학기술의 진보는 사실 인류가 게으름을 피우는 지혜의 결정체입니다~ 😎

Slide 12:

说到自动化，就不得不提"社会生产力的发展"。

有了科技，我们的工作效率越来越高。

想象一下，如果现在没有外卖软件，
我们可能要每天做饭，根本没时间学习啊！

科技进步，其实就是人类偷懒(tōulǎn) 的智慧结晶(jiéjīng) ~ 😎

偷懒 (tōulǎn) → "게으름 피우다"
智慧结晶 (zhìhuì jiéjīng) → "지혜의 결정체

Slide 13: IBM의 변화 - "코끼리도 춤출 수 있다"
이 분은 IBM 전 CEO **루이스 거스너(Louis Gerstner)**입니다.

그는 "**코끼리도 춤출 수 있다**"는 말을 한 적이 있습니다.

의미는: 전통적인 대기업이라도 유연하고 효율적으로 변할 수 있다는 것입니다.

우리는 이 말에서 빅데이터의 영혼이 사실은 "**변화**"라는 것을 느낄 수 있습니다.

Slide 13

:这位就是IBM前CEO 路易斯·郭士纳（Louis Gerstner）。

他曾经说过一句话："大象能跳舞"。

意思是：即使是传统的大公司，也可以变得灵活、高效。

我们从这句话中可以感受到，大数据的灵魂，其实是"改变"。

Elephant: IBM 같은 거대 기업
Dance: 디지털 전환, 스타트업 같은 민첩성

Slide 14: 정보화 3차 물결의 중요 시점

여러분은 이 세 가지 중요한 시점을 기억하세요:

- **제1차 물결**: 1980년경, PC 보급
- **제2차 물결**: 1995년경, 인터넷
- **제3차 물결**: 2010년경, 빅데이터, 사물인터넷, 클라우드 컴퓨팅!

딱 **15년 간격**이지 않나요?

각 세대가 역사를 새로 쓰는 것 같지 않나요?

이 규칙에 따르면, **2025년경에 제4차 물결이 나타날까요?** 아마도 양자 컴퓨팅과 범용 AI의 융합일 것입니다.

Slide 15: 제2차 정보화 물결 - 인터넷

이 페이지에서는 인터넷의 부상에 주목합니다.

처음에는 사실 미국 군부가 통신을 위해 발명한 기술이었고, 나중에 민간에 개방되었습니다.

그러고 나서 지금의 여러분이 탄생했습니다 - "**숏폼 비디오를 보는 고수들**"!

누군가 첫 번째 중국어 웹사이트가 무엇인지 아시나요? 😉

> **중국 최초의 웹사이트**는 1994년 4월 중국과학원 고에너지물리연구소(IHEP)에서 구축한 www.ihep.ac.cn입니다.

Slide 14: 信息化第三次浪潮的重要时点

请大家记住这三个重要的节点：

- 第一次浪潮：1980年左右，PC普及。
- 第二次浪潮：1995年左右，互联网。
- 第三次浪潮：2010年左右，大数据、物联网和云计算！

是不是刚好隔15年？

像不像每一代人都在刷新(**shuā xīn)**历史(**lì shǐ)**？

按此规律，2025年左右是否会出现第四次浪潮？可能是量子计算+通用AI的融合？

Slide 15:第二次信息化浪潮——互联网

这一页，我们关注互联网的兴起。

最早其实是美国军方为了通讯(tōng xùn)发明的技术，后来开放给民用，

然后就有了现在的你们——"刷短视频的高手"！

有没有人知道第一个中文网站是什么？😉

中国第一个网站是由中国科学院高能物理研究所（IHEP）于1994年4月建立的 www.ihep.ac.cn。

Slide 16-17: 군용과 민간용 수요의 차이

군부의 수요는 안전, 안정, 속도입니다.

Slide 17:
민간용 수요는요? 하하, 저렴하고, 편리하고, 게임도 할 수 있어야 합니다!

이 관점에서 보면, 기술의 보급은 "일반 대중"의 지원을 떠날 수 없습니다.

Slide 18: 인터넷이 해결한 문제

여러분 생각해보세요 인터넷이 도대체 어떤 문제를 해결했을까요?

답은 사실 - "**정보 전송**"입니다.

우리는 사진, 텍스트, 음성을 1초 안에 지구 반대편으로 보낼 수 있습니다. 위챗 음성처럼 편리합니다.

하지만 "**정보 과부하**"의 문제도 가져왔습니다.

여러분 중에 매일 휴대폰을 새벽까지 보는 분 계신가요? 🙈

Slide 16:

军方的需求是安全、稳定、快速。

Slide 17:
民用的需求呢？哈哈，是便宜、方便、还能打游戏！

从这个角度看(jiǎo dù kàn)，技术的普及，离不开"老百姓"的支持(zhī chí)。

Slide 18:

大家想一想，互联网到底解决了什么问题？

答案其实是——"信息传输(**chuán shū)**"。

我们可以把照片、文字、声音，在1秒钟内发送到地球的另一边。
就像微信语音一样方便。

但也带来了"信息过载"的问题，

你们有没有每天刷手机 (shuā shǒu jī)到凌晨 (dào líng chén)的？🙈

Slide 19: 제3차 물결의 도래

2010년 전후로, 우리는 "**빅데이터+클라우드 컴퓨팅+사물인터넷**"의 시대를 맞이했습니다.

세 가지 용어, 한 가지 공통점: 모두 "데이터"를 생성하고 처리합니다.

학생 여러분은 몇년도에 태어났습니까?

그리고 여러분은 바로 이 시대의 "**원주민**"입니다!

빅데이터를 배우지 않으면 어떻게 이 세계를 장악할 수 있겠습니까?

Slide 19:

2010年前后，我们迎来了“大数据+云计算+物联网”的时代。

三个词，一个共同点：都在生成和处理“数据”。

同学们，你们是哪一年出生的？

而你们，就是这个时代的“原住民”！

不学大数据，怎么能驾驭(**jià yù**)这个世界？

Slide 20: 제3차 정보화 물결의 3대 표지

빅데이터, 클라우드 컴퓨팅, 사물인터넷은 제3차 정보화 물결의 3대 표지입니다.

쉽게 말하면:
- **빅데이터**는 "대량 정보"
- **클라우드 컴퓨팅**은 "슈퍼 두뇌"
- **사물인터넷**은 "모든 것이 연결됨"

미래에는 냉장고도 여러분에게 위챗을 보낼 것입니다: 우유 유통기한이 다가옵니다~ 🥛📱

Slide 20:

大数据、云计算、物联网是第三次信息化浪潮的三大标志(**biāo zhì)**。

通俗地说(**tōng sú de shuō**)：
- 大数据是“大量信息”
- 云计算是“超级大脑”
- 物联网是“所有东西都联网”

未来连冰箱都会给你发微信(**Wēixìn)**：牛奶快过期啦～ 🥛📱

Slide 21: 3차 정보화 물결의 특징 요약

이 표는 3차 정보화 물결의 특징을 요약합니다.

- **제1차 물결**: "정보 처리"가 주요 목적, 대표 기업은 IBM, 마이크로소프트
- **제2차 물결**: "정보 전송"이 주요 목적, 구글, 바이두, 텐센트 같은 기업
- **제3차 물결**: "폭발적 정보 증가"가 특징, 우리가 바로 이 단계에 있습니다!

여러분은 앞으로 어떤 새로운 기업들이 나타날지 상상할 수 있나요? 다음 "구글"은 어쩌면 여러분 중에 숨어 있을지도 모릅니다~ 😄

> 대표 기업: Apple, Microsoft, Lenovo, Dell, HP (제1차) / Yahoo, Google, Tencent (제2차)

Slide 21:

大家看看这个表格，总结了三次信息化浪潮的特点。

- 第一次浪潮：以"处理信息"为主，代表企业有 IBM、微软。
- 第二次浪潮：以"传输信息"为主，像谷歌、百度、腾讯。
- 第三次浪潮：以"爆炸(**bào zhà**)式信息增长(zēng zhǎng)"为特点，我们正处在这一阶段！

你们能想象以后会出现哪些新企业吗？下一家"谷歌"或许(**huò xǔ**)就藏在(**cáng zài**)你们当中哦～ 😄

Apple, Microsoft, Lenovo, Dell, HP
Yahoo, google, tencent

Slide 22: 빅데이터 시대의 형성 조건

빅데이터 시대는 갑자기 나타난 것이 아니라, 여러 조건이 함께 작용한 결과입니다.

마치 훠궈를 만드는 것과 같습니다. 냄비만 있으면 안 되고, 재료도, 불도, 먹을 사람도 있어야 하죠, 그렇지 않나요? 😆

그렇다면 이 "훠궈" - 빅데이터는 어떻게 만들어진 걸까요? 계속 봅시다.

Slide 22:

大数据时代不是突然冒出(**mào chū**)来的，它是多个条件共同作用的结果。

就像做火锅，不是光有锅，还得有菜、有火、有人吃，对不对？😆

那么，这个"火锅"——大数据，是怎么煮成(**zhǔ chéng**)的呢？我们接着看。

Slide 23: 빅데이터를 지지하는 3가지 기둥

빅데이터를 지지하는 세 개의 "기둥"은:

- **저장 장치**
- **컴퓨팅 능력(CPU)**
- **네트워크 대역폭**

이 세 가지는 여러분이 게임할 때의 "사양"과 같습니다:
- 메모리가 큰가?
- 렉이 걸리는가?
- 네트워크가 빠른가?

Slide 23:

支撑zhī chēng大数据的三根"柱子zhù zi"是：

- 存储设备
- 计算能力（CPU）
- 网络带宽

这三样东西，
就像你打游戏时的"配置"——

- 内存大不大？
- 卡不卡？
- 网快不快？

Slide 24: 저장 장치의 발전

우선, 저장 장치가 점점 더 저렴해지고 있습니다.

예전에는 1GB 하드디스크가 엄청나게 비쌌는데, 지금은 1TB 하드디스크가 몇백 위안밖에 안 합니다.

여러분의 휴대폰에 저장된 것도 1990년대 서버보다 많습니다!

예전에는 가난해서 "일기"만 쓸 수 있었는데, 지금은 클라우드에 직접 "자동 백업"됩니다. 😄

Slide 24:

首先，存储设备越来越便宜。

以前1GB硬盘(**yìng pán)**贵得吓人，
现在1TB硬盘只要几百块。

连你们的手机，存的东西都比1990年的服务器多！

以前人穷只能记"日记"，
现在直接云端"自动备份(**bèi fèn:backup)**"了。 😄

Slide 25: 1TB 하드디스크의 보편화
지금은 모든 컴퓨터가 기본적으로 1TB 하드디스크를 장착하고 있습니다.

여러분은 발견하셨나요? 하드디스크가 클수록 파일도 더 많아집니다.
휴대폰 앨범이 수만 장인데, 삭제하고 싶지도 않습니다! 😂

Slide 26: 기업 데이터의 폭발적 증가
개인 하드디스크만 많아진 것이 아니라, 기업도 더 많은 장비를 구매할 여유가 생겼습니다.

기업 데이터가 폭발적으로 증가하고 있습니다. 매일 수집하는 사용자 데이터, 거래 기록을 쌓으면 건물을 지을 수 있을 정도입니다.

미래에는 누가 이러한 데이터를 처리할 수 있는지가 비즈니스 비밀을 장악하는 것입니다!

여러분 기억하세요: "**데이터가 점점 더 많아진다**"는 것은 구호가 아니라 현실입니다.

여러분의 숏폼 비디오 시청 기록, 온라인 쇼핑 행동, 출퇴근 경로, 모두 데이터 소스입니다.
진짜로, 지금은 여러분이 말하지 않아도 데이터가 여러분을 "배신"할 수 있습니다~ 😈

> 예시: SNS에 강아지 사진만 올렸는데 → 광고로 사료·펫샵 할인 쿠폰이 쏟아집니다!

Slide 25:1TB硬盘变普遍
现在每台电脑基本都标配1TB硬盘**(yìng pán)**。

你们有没有发现，硬盘越大，文件越多？手机相册xiàng cè几万张，删都不想删(shān dōu bù xiǎng shān)！😂

Slide 26:
不仅个人硬盘多了，企业也有钱买更多设备。企业数据爆炸**(bào zhà)**式增长，每天收集的用户数据、交易记录，堆起来能盖房子了。

- **堆起来:** duī qǐ lái - 쌓아 올리다
- **能:** néng - ~할 수 있다
- **盖:** gài - (집을) 짓다
- **房子:** fáng zi - 집
- **了:** le - (문장 끝에서) 이미 ~한 상태를 나타냄

未来谁能处理这些数据，谁就掌握商业秘密！

请大家记住："数据越来越多"不是口号，而是现实。

你们的刷视频记录、网购行为、通勤轨迹**(tōng qín guǐ jì)**，都是数据来源。

讲真，现在就算你不说话，数据也能"出卖**(chū mài)**"你～ 😈

"我只是在社交网络(wǎngluò)发了狗狗照片 → 结果宠物粮**(liáng)**和宠物店的优惠券**(yōuhuìquàn)**广告铺天盖地 **(pūtiāngàidì)**而来！

Slide 28: 저장 장치의 가격 하락

이 그림이 우리에게 알려주는 것: 저장 장치가 점점 더 저렴해지고 용량은 점점 더 커지고 있습니다.

예전에는 하드디스크를 사는 것이 "투자"였는데, 지금은 하드디스크를 사는 것이 "소모품"입니다.

여러분 생각에는, 미래에는 직접 "뇌 속 칩 저장"이 될까요? 🤯

Slide 28:

这张图告诉我们：存储设备越来越便宜、容量越来越大。

以前买硬盘是"投资"，现在买硬盘是"消耗品(**xiāo hào pǐn)**"。

你们猜(cāi – 추측하다)，
未来是不是直接"脑中芯片(xīnpiàn – 칩)存储(cúnchǔ)"？🤯

Slide 29: 기술의 대중화

다시 한 번 강조합니다: 기술 발전이 가져온 두 글자 - "**대중화**".

예전의 서버 기능을 지금은 가정용 컴퓨터도 일부 할 수 있습니다.

이것이 의미하는 바는: 누구나 "빅데이터 플레이어"가 될 수 있다는 것입니다.

Slide 30: 3대 지원 기술의 역할

3대 지원 기술을 다시 살펴봅시다: 저장, CPU, 네트워크.

이 세 가지는 세 마리 말이 끄는 마차처럼 "빅데이터 마차"를 앞으로 나아가게 합니다.

여러분은 생각해보셨나요, 어떤 기술이 가장 중요할까요? 댓글 창에 의견을 남겨주세요~ 📢

Slide 31: 두 번째 지원 - 컴퓨팅 능력

두 번째 지원점은 **컴퓨팅 능력**, 즉 우리가 흔히 말하는 CPU입니다.

여러분 아시나요? 우리가 지금 사용하는 스마트폰의 성능은 1990년대 슈퍼컴퓨터보다 강력합니다!

여러분이 매일 게임을 하는 그 "칩"은 사실 대단한 능력을 가지고 있습니다. 다만 말을 하지 않을 뿐입니다. 😆

Slide 29:

再强调一次：技术发展带来两个字——"平民化"。

以前的服务器，现在家用电脑也能做到一部分功能。

这意味着：人人都可以成为"大数据玩家"。

Slide 30:

我们回顾一下三大支撑：存储、CPU、网络(**wǎngluò**)。

这三者像三驾马车"sān jià mǎ chē，拉动了"大数据马车"前进。

你们有没有想过，哪一个技术最关键？
在评论区打出来，我们一起看看～ 📢

Slide 31:

第二个支撑点是——计算能力，
也就是我们常说的CPU。

大家知道吗？我们现在的手机，性能比上世纪90年代的超级计算机还强！

你们天天打游戏的那颗"芯xīn:chip"，其实非常厉害，只是它不说话而已。😆

Slide 32: CPU 능력 배가 법칙 (무어의 법칙)

무어의 법칙에 따르면, CPU의 성능은 대략 18-24개월마다 두 배로 증가합니다.

이것은 마치: 여러분이 매년 기말고사 성적이 두 배로 오르는 것과 같습니다. 생각해보세요, 멋지지 않나요?

하지만 현실에서는 "칩"만 해냈고, 인류는 여전히 노력해야 합니다~

Slide 33: 트랜지스터 수량의 증가

이 페이지는 트랜지스터 수량의 증가를 보여줍니다.

간단히 말하면, 트랜지스터가 많을수록 CPU가 강력해집니다. 마치 여러분의 뇌세포가 많을수록 시험에서 고득점을 받기 쉬운 것과 같습니다.

하지만 칩은 밥도 먹지 않고 잠도 자지 않는데, 인류는 여전히 휴식을 많이 취해야 합니다~

Slide 34: CPU는 두뇌와 같다

계속 강조하자면, CPU는 "두뇌"와 같아서 정보를 처리하고, 프로그램을 실행하고, 게임을 하고, 모델링 분석을 할 수 있습니다...
여러분은 그것을 "사무실 주임"으로 생각할 수 있습니다. 모든 작업을 배정하고 각 업무를 조정합니다.

그리고 데이터는 그것이 매일 처리하는 "문서"입니다.

Slide 32:

第32页：CPU能力翻倍规律（摩尔定律）
根据摩尔定律，CPU的性能大约每18–24个月就会翻(fān)一番(**fān**)。두 배로 증가하다
这就像是：你每年期末考试成绩都能翻倍，想想是不是很爽shuǎng"?

不过现实中，只有"芯片xīnpiàn(반도체칩)"做到了，人类还得努力~

Slide 33:

这一页展示了晶体管("jīngtǐguǎn, 트랜지스터)数量的增长。

简单来说，晶体管越多，CPU越强，就像你的脑细胞Nǎo xìbāo越多，考试就越容易得高。

不过，芯片(xīnpiàn)不吃饭也不睡觉，人类还是要多休息哦~

Slide 34:

我们继续强调，CPU就像"大脑(**dànǎo**)"，能处理信息、运行程序、玩游戏、建模分析……

你可以把它当成"办公室主任"，安排所有任务、协调(**xiétiáo**)各项工作。

而数据，就是它每天处理的"文件"。

Slide 35: 단일 코어에서 멀티 코어로

CPU가 단일 코어에서 멀티 코어로
발전한 것은, 한 사람이 일하는 것에서
여러 사람이 분업 협력하는 것으로 변한
것과 같습니다.

여러분의 과제 소그룹과 매우 비슷하지
않나요?

- 한 사람은 제목을 쓰고
- 한 사람은 이미지를 찾고
- 한 사람은 PPT 애니메이션을
담당합니다~ 🤣

멀티 코어 병렬 처리는 효율성의
대명사입니다.

Slide 35:

CPU从单核到多核的发展，
就像从一个人干活，
到一群人分工合作。

是不是很像我们写作业小组？

一个人写标题，
一个人找图 (zhǎo)이미지 검색，
一个人负责PPT动画~ 🤣

多核并行，就是效率的代名词。

Slide 36: 분산 처리

데이터가 너무 커서 한 대의 컴퓨터로는
처리할 수 없을 때 어떻게 할까요?

답은 바로 "**분산 처리**"입니다.

마치 설날에 만두를 빚는 것과 같습니다.
온 가족이 나서야 빨리 빚을 수 있습니다.
한 사람이 다 빚으면 식어버릴 것입니다~
😅

Slide 36:

当数据太大，
一台电脑处理不过来了怎么办？

答案就是——"分布式处理"。

就像过年包饺子，
全家上阵才包得快，
一个人包完估计都凉(**liáng**)了 ~ 😅

Slide 37: 3대 지원 기술의 발전

다시 한 번 이 세 가지 지원을 강조합니다: 저장 장치, CPU, 네트워크 대역폭.

이 세 가지가 계속 발전하여 빅데이터에 "비옥한 토양"을 제공했습니다.

그리고 우리는 이 "과학기술 농지" 위에 서 있는 농부로서, 데이터를 심고, 데이터를 수확하고, 데이터를 판매할 줄 알아야 합니다! 🌾📊

Slide 37:

再次强调这三项支撑：存储设备、CPU、网络带宽。

这三样东西不断发展，
为大数据提供了"沃土"。wò tǔ

而我们，就是站在这片"科技田地"上的农民，要会种数据、收数据、卖数据！🌾📊

Slide 38: 빅데이터 응용의 3가지 중점

빅데이터 응용의 세 가지 중점은:

- **해량 데이터 저장**
- **해량 데이터 관리**
- **해량 데이터 분석**

간단히 말하면: 먼저 모으고, 그다음 정리하고, 마지막으로 이해하는 것입니다!

여러분이 논문을 쓰는 것과 비슷합니다: 먼저 문헌을 찾고, 그다음 분류하고, 마지막으로 분석 부분을 씁니다~ 📝

Slide 38:

大数据应用的三个重点是：

- 存储海量数据
- 管理海量数据
- 分析海量数据

简单说：先收起来，再整理好，最后搞明白(**gǎo míngbai)**！

有点像你们写论文：先找文献、再分类、最后写分析部分～ 📝

Slide 39: 한 대의 컴퓨터로는 부족하다

이 페이지는 지적합니다: 한 대의 컴퓨터로는 이미 방대한 데이터량에 대응할 수 없습니다.

이것은 여러분 혼자 1000개의 햄버거를 먹으라고 하는 것과 같습니다. 현실적이지 않죠? 🍔😅

그래서 우리는 "클러스터", 즉 여러 대의 기계가 협동 작업하는 것이 필요합니다.

Slide 40: 네트워크를 통한 협업

어떻게 많은 컴퓨터가 함께 작업하게 할까요? 네트워크에 의존합니다.

데이터는 네트워크를 통해 서로 다른 기계 간에 전송됩니다. 마치 회의할 때 모두가 Zoom으로 원격 협업하는 것과 같습니다.

미래의 "슈퍼 공장"은 어쩌면 전 세계 컴퓨터로 구성된 "슈퍼 두뇌"일 것입니다!

Slide 41: 데이터 전송의 폭발적 증가

여러분 이 그림을 보세요. 데이터가 홍수처럼 각 곳에서 밀려오는 것을 보여줍니다.
과거에는 일주일에 한 번 도서관에 가서 자료를 찾았는데, 지금은 1초에 수억 건의 정보를 검색할 수 있습니다.

마치 우리가 정보의 "바다"에서 수영하는 것과 같습니다. 수영할 줄 모르면 "익사"하게 됩니다~ 🌊😅

Slide 39:

这一页指出：一台计算机已经无法应对庞大pángdà的数据量。

这就像让你一个人吃掉1000个汉堡，是不是不现实？🍔😅

所以我们需要"集群"，也就是多台机器协同工作。

Slide 40:

怎么让很多电脑一起工作？靠的是网络。

数据通过网络在不同的机器之间传输，就像开会时大家用Zoom远程协作。

未来的"超级工厂"，或许就是由全球电脑组成的"超级大脑"！

Slide 41:数据传输的爆炸bàozhà式增长

你们看看这张图，展示的是数据像洪水一样从各个地方涌来。

过去我们每周去一趟图书馆查资料，现在一秒钟能搜索(sōusuǒ)上亿(yì)条信息。

就像我们在信息的"大海"里游泳，不会游的就会被"淹没(yānmò) 익사하다"~ 🌊😅

Slide 42: 네트워크 대역폭 향상의 중요성

네트워크가 빠를수록 데이터 전송 효율이 높아집니다.

빅데이터를 "택배"에 비유한다면,
네트워크 대역폭은 "고속도로"입니다.
여러분은 시골길을 걷고 싶으신가요,
아니면 6차선 고속도로를 원하시나요?
🚙🧿

우리는 이제 텍스트를 보는 것에
만족하지 않고, 고화질 비디오, VR, 라이브 방송을 보고 싶어 합니다...

이 모든 것은 "넓고 넓고 넓은"
네트워크가 필요합니다.

Slide 42:网络带宽提升的重要性

网络越快，传输数据的效率就越高。

如果把大数据比作"快递kuàidì
택배"，那么网络带宽就是"高速公路"。你们更愿意走乡村小路，还是六车道高速？
🚙🧿

我们现在已经不满足于看文字，要看高清视频、VR、直播……

这都需要"宽宽宽"的网络。

Slide 43: 네트워크 대역폭 역사 추세

그림에 표시된 것은 1970년부터
2010년까지 대역폭의 비약적 발전입니다.

예전에는 노래 한 곡 다운로드하는 데
5분이 걸렸는데, 지금은 앨범 전체를 1초 만에 끝낼 수 있습니다.

앞으로는 어쩌면 "눈 한 번 깜빡이면
영화 다운로드 완료"일 것입니다.
기대되지 않나요? 😎

Slide 43:网络带宽历史趋势图

图上显示的是从1970年到2010年，带宽的飞跃(**fēiyuè**)。

以前下载一首歌要5分钟，
现在一整张专辑1秒搞定。

以后可能是"眼睛一眨yǎnjing yī zhǎ，电影下载完"，是不是很期待？😎

Slide 44: 데이터 생성의 3단계

데이터는 어떻게 생성될까요? 3단계로 나뉩니다:

1. **운영 시스템 단계** (슈퍼마켓, 은행 등)
2. **사용자 제작 콘텐츠 단계** (웨이보, 틱톡 등)
3. **감지 시스템 단계** (사물인터넷, 센서)

여러분은 매일 얼마나 많은 데이터를 생성하고 있다고 생각하시나요? 먼저 말하지 마세요, 나중에 우리가 계산해봅시다~ 🧮

Slide 44:

数据是怎么产生的？分为三个阶段：

1. 运营式系统阶段（超市、银行等）
2. 用户原创内容阶段（微博、抖音等）
3. 感知式系统阶段（物联网、传感器）

你们觉得自己每天产生多少数据？
先别说，后面我们来算一算～ 🧮

Slide 45: 제1단계 - 운영 시스템

여러분이 슈퍼마켓에 가서 물 한 병을 사고 코드를 스캔하면, 시스템이 여러분의 구매 행동을 기록합니다.

이러한 데이터는 무엇에 사용될까요? 어쩌면 "여러분이 좋아할 만한 간식"을 추천하는 데 사용될 것입니다~

그래서 지금 물건을 살 때는 돈만 거래하는 것이 아니라 데이터도 거래하는 것입니다! 🛒💻

Slide 45:

你去超市买瓶水，刷个码，系统就记录下你的购买行为。

这些数据会被用来做什么？也许是推送"你可能喜欢的零食"～

所以现在买东西，不只是钱交易，还有数据交易哦！🛒💻

Slide 46: 기업 데이터베이스의 기록

기업의 데이터베이스는 여러분이 "무엇을 샀는지", "언제 샀는지", "어디서 샀는지"를 기록할 수 있습니다.

상상해보세요: 여러분이 밤새 배달 음식을 세 번 주문했는데, 다음 날 배달 앱이 헬스장 광고를 보냅니다... 배려심 있으면서도 무섭지 않나요? 😂

Slide 47: 데이터 생성 방식의 진화

이 페이지를 다시 봅시다: "사람이 데이터 입력"에서 "사람이 콘텐츠 생산", 그리고 "기계가 자동으로 데이터 생성"으로.

데이터는 이미 "사람이 기록"에서 "기계가 감지"로 변했습니다.

여러분이 아침에 몇 시에 일어나는지, 휴대폰이 이미 몰래 클라우드 서버에 알렸습니다~ 📱☁️

Slide 48: 제2단계 - 사용자 제작 콘텐츠

이 단계의 대표는 바로 **소셜 미디어**입니다!
모멘츠, 웨이보, 틱톡, 지후, 모두 사용자가 생산한 데이터입니다.

매번 여러분이 좋아요를 누르고, 댓글을 달고, 공유할 때마다, 사실은 플랫폼을 위해 "무료로 일"하는 것입니다~ 😆

Slide 46:

企业的数据库能记录你"买了什么"、"什么时候买"、"在哪里买"。

想象一下：你连夜点了三次外卖，第二天饿了吗就给你推健身广告……是不是很贴心也很恐怖(**kǒngbù**)？😂

Slide 47:

这一页我们再看一次：从"人输入数据"，到"人产出(**chǎnchū**)内容"，再到"机器自动生成数据"。

数据已经从"靠人记"变成"机器感知"。

你早上几点起床，手机已经偷偷(**tōutōu**)告诉了云服务器～ 📱☁️

Slide 48:

这一阶段的代表就是——社交媒体！

朋友圈、微博、抖音、知乎，全是用户生产的数据。

每次你点赞(**zàn**)、评论(**pínglùn**)、转发(**zhuǎnfā**)，其实都在给平台"免费打工"～😆

Slide 49: 사용자 데이터가 기업 데이터를 초월

이 페이지의 핵심은: **사용자 제작 데이터 > 기업 시스템 데이터**.

즉, 일반인이 매일 "수다 떠는" 내용이 기업 데이터베이스보다 더 풍부하다는 것입니다!
한 건의 불평 웨이보를 무시하지 마세요. 그 뒤에는 빅데이터 분석의 "황금 광산"이 있습니다!

Slide 50: 제3단계 - 만물 인터넷

제3단계로 진입: **만물 인터넷**.

냉장고는 우유가 부족하다는 것을 알고, 스마트워치는 어젯밤에 잘 못 잤다는 것을 알고, 변기까지도 "장 건강이 좋지 않다"는 것을 압니다~ 🤣

이것이 바로 "감지 시스템": 기기가 자동으로 정보를 수집하고 업로드합니다.

Slide 51: 감지 시스템의 다양한 응용

감지 시스템 단계에 진입한 후, 우리는 각종 기기를 사용하여 세계를 감지하기 시작했습니다.

예를 들어:
- **RFID(무선주파수 식별)**: 화물 위치 기록
- **카메라**: 얼굴 인식 가능
- **센서**: 온도, 습도, 동작 감지 가능

마치 세상의 모든 것이 "눈을 가진" 것 같은 느낌입니다~ 👀😅

Slide 49:

这一页的重点是：用户原创数据 > 企业系统数据。

也就是说，普通人每天"唠叨(**láodao)**"的内容，比企业数据库还丰富！

你们别小看一条吐槽 **(tùcáo)**微博，它背后可是大数据分析的"黄金矿**(huángjīn kuàng)**"！

Slide 50:

进入第三阶段：万物互联。

冰箱知道你缺牛奶，手表知道你昨晚没睡好，马桶 **(mǎtǒng)** 都知道你"肠胃不好**(chángwèi bù hǎo)**"~ 🤣

这就是"感知式系统"：设备自动收集并上传信息。

Slide 51:

进入感知系统阶段后，我们开始用各种设备来感知世界。

比如：
- RFID（射频识别）记录货物位置；
- 摄像头**(shèxiàngtóu)**可以识别人脸；
- 传感器可以感知温度、湿度、动作。

感觉好像世界上的每一样东西都"长了眼睛"~ 👀😅

Slide 52: 기상 센서

기상 센서는 매일 묵묵히 작업하며, 예를 들어 온도, 습도, 공기 질 등을 모니터링합니다.

만약 여러분이 날씨 앱을 본 적이 있다면, 그 데이터들은 사람이 수동으로 입력한 것이 아니라, 모두 이러한 "작은 기기들"이 여러분을 위해 수집한 것입니다!

여러분은 "오늘 흐림에서 맑음으로🌞"만 보지만, 그 뒤에는 빅데이터의 "군대 배치"가 있습니다~

Slide 52:

气象传感器每天都在默默工作，比如监测(**jiāncè**)温度、湿度、空气质量等。

如果你看过天气App，那些数据可不是人手动输入的，都是这些"小设备"帮你采集的！

你只看到"今天多云转晴🌞"，背后却是大数据的"排兵布阵"~

Slide 53: 비디오 모니터링 + 빅데이터

비디오 모니터링 + 빅데이터는 "스마트 시티"의 핵심 중 하나입니다.

예를 들어:

- 교통 위반 자동 식별
- 인파 밀집도 지능형 분석
- 보안 위험 예측

앞으로는 수업을 빼먹어도 AI의 눈을 피할 수 없을 것입니다~ 🎥🤖

Slide 53:

视频监控+大数据，就是"智慧城市"的核心之一。

比如：
- 自动识别交通违规；
- 智能分析人群密度；
- 预测安保风险。

以后逃课(táokè)可能都逃不过AI的眼睛了哦~ 🎥🤖

- **逃不过 (táobùguò): ~을 피해 나갈 수 없다**

Slide 54: 사물인터넷의 급속한 발전

기기가 네트워크에 연결되면서 사물인터넷(IoT)이 빠르게 발전하고 있습니다.

지금은 칫솔, 도어락, 에어컨도 네트워크에 연결할 수 있습니다! 마치 드라마 《블랙 미러》 속 미래 세계처럼 느껴지지 않나요? 🔪 📶

과학기술은 확실히 생활을 더 편리하게 만들었지만, "프라이버시"도 더 소중해졌습니다~

Slide 54:

随着设备的联网，物联网（IoT）飞速发展。

现在连牙刷(yá shuā,칫솔)、门锁(mén suǒ,도어락)、空调都可以联网了！是不是感觉生活像电视剧《黑镜:HēiJìng》里的未来世界？🔪 📶

科技确实让生活更方便了，但也让"隐私(**yǐn sī**)"变得更宝贵~

Slide 55: 빅데이터 발전의 3단계

빅데이터 발전의 3단계를 살펴봅시다:

1. **맹아기**: 1990년대, 개념이 막 등장
2. **성숙기**: 2000년대, 실천 시작
3. **대규모 응용기**: 2010년 이후, 전면 폭발!

지금 우리는 "전면 응용기"에 있으며, **데이터는 새로운 "석유"**입니다.

Slide 55:

我们来看一下大数据发展的三个阶段：

1. 萌芽(Méngyá)期：20世纪90年代，概念刚出现。
2. 成熟(Chéngshú)期：2000年代，开始实践。
3. 大规模应用期：2010年后，全面爆发(Bàofā)！

现在我们就处于"全面应用期"，数据就是新的"石油"。

Slide 56: 수업 마무리

좋습니다. 이번 수업은 여기서 마치겠습니다!
오늘 우리는 "빅데이터 시대"의 **전말**을 알아보았습니다. 정보화 물결, 3대 지원 기술부터 데이터가 어떻게 생성되고 어떻게 **진화**하는지까지 말입니다.

다음 수업에서는 계속해서 "빅데이터의 핵심 기술"을 **탐구**하겠습니다.
마지막으로 질문 하나 드립니다: 오늘 여러분은 사실 매일 "데이터를 생산"하고 있다는 것을 발견하셨나요? 😊

감사합니다. 수업 마치겠습니다! 👏👏👏

요약
이번 강의에서는 빅데이터의 기본 개념과 시대적 배경을 다루었습니다:

- 빅데이터의 3V 특성 (Volume, Velocity, Variety)
- 정보화의 3차 물결 (PC → 인터넷 → 빅데이터/클라우드/IoT)
- 빅데이터를 지지하는 3가지 기술 (저장장치, CPU, 네트워크)
- 데이터 생성의 3단계 (운영시스템 → 사용자 콘텐츠 → 감지시스템)
- 사고방식의 전환 (인과관계 → 상관관계, 샘플링 → 전수조사)

핵심 메시지: 빅데이터는 단순한 기술 혁명이 아니라, 우리가 세계를 이해하고 문제를 해결하는 방식을 근본적으로 바꾸는 사고의 혁명입니다.

Slide 56:

好了，我们本节课到这里就结束啦！

今天我们认识了"大数据时代"的来龙**去脉**(láilóngqùmài)，从信息化浪潮、三大支撑技术，到数据如何产生、如何**演变**(yǎnbiàn)。

下节课我们会继续**探讨(tàntǎo)** "大数据的关键技术"。
最后提问一下：你今天有没有发现，自己每天其实也在"生产数据"呢？😊

谢谢大家，下课！👏👏👏

总结
本课程涵盖了大数据的基本概念及其时代背景：

- 大数据的3V特性（数据量大、处理速度快、数据类型多样）
- 信息化的第三次浪潮（个人电脑 → 互联网 → 大数据/云计算/物联网）
- 支撑大数据的三大技术（存储设备、处理器、网络技术）
- 数据生成的三个阶段（业务系统 → 用户生成内容 → 传感采集系统）
- 思维模式的转变（从因果关系转向相关关系，从抽样分析转向全量分析）

核心启示**(qǐshì)**：大数据不仅是一场技术革新，更是从根本上改变我们认知世界和解决问题方式的思维革命。