Slide 1:

Slide 2:

Slide 3:

Slide 4: 빅데이터 계산 모드란?

오늘 우리가 다룰 주제는《빅데이터 계산 모드》입니다.

"다른 자물쇠에는 다른 열쇠가 필요하다"는 말을 들어보셨나요? 🔐
빅데이터 분야도 마찬가지입니다:

- 다른 계산 요구사항, 다른 응용 시나리오에는 다른 계산 모드와 도구를 사용해야 합니다!

마치 헬스장에 가는 것과 같습니다:

- 체지방을 줄이고 싶다면 러닝머신을 사용 🏃
- 근육을 키우고 싶다면 덤벨을 사용 💪
- 자고 싶다면... 그냥 오지 마세요 😂

Slide 1:

Slide 2:

Slide 3:

Slide 4:什么是大数据计算模式？

今天我们讲的主题是《大数据计算模式》。

你有没有听过这句话："不同的锁(suǒ)要配不同的钥匙(yàoshi)"？🔐
在大数据领域也是一样：

- 不同的计算需求、不同的应用场景，就要用不同的计算模式和工具！

就像你去健身房：

- 想减脂(jiǎnzhǐ)用跑步**pǎobù**机🏃
- 想增肌(zēngjī) 用哑铃(yǎlíng) 💪
- 想睡觉...别来了 😂

Slide 5: 세 가지 기본 계산 모드

이 그림은 세 가지 주류 계산 모드를 간단히 요약합니다:

1. **배치 처리 (Batch processing)**
2. **실시간 계산 / 스트림 계산 (Stream processing)**
3. **대화형 계산 / 쿼리 분석 (Interactive or Analytical)**

각 계산 방식은 적합한 시나리오가 다릅니다. 배달 음식 주문에 비유하면:

- **배치 처리**: 하루에 한 번 통일된 배송, 마치 구내식당처럼
- **실시간 계산**: 주문하면 바로 배송, 마치 배달 앱처럼
- **대화형**: "오늘 뭐 먹을까?"라고 물으면 바로 추천 😋

Slide 5: 三种基本计算模式

这张图简单总结了三种主流计算模式：

1. 批处理（Batch processing）
2. 实时计算 / 流计算（Stream processing）
3. 交互式计算 / 查询分析（Interactive or Analytical）

每种计算方式适合的场景不同，就像点外卖：

- 批处理：每天统一配送一次，像食堂
- 实时计算：你点了就马上送，像美团外卖
- 交互式：你问它"今天吃什么最好"，它马上推荐 😋

Slide 6: 전형적인 네 가지 빅데이터 계산 모드

오늘 우리는 이 "4대 천왕"에 중점을 둡니다:

1. **배치 처리 계산 (Batch)**
2. **스트림 계산 (Stream)**
3. **그래프 계산 (Graph)**
4. **쿼리 분석 계산 (Query & Analysis)**

📌 각 모드는 핵심 작업에 대응합니다:

- 대량의 빅데이터 일괄 계산
- 데이터 스트림에 실시간 대응
- 소셜/네트워크 구조 처리
- 쿼리 요청에 빠른 응답

Slide 6:典型的四种大数据计算模式

今天我们重点讲这"四大金刚(jīn gang)"：

1. 批处理计算（Batch）
2. 流计算（Stream）
3. 图计算（Graph）
4. 查询分析计算（Query & Analysis）

📌 每种模式都对应一种核心任务：

- 批量计算大数据
- 实时应对数据流
- 处理社交/网络结构
- 快速响应查询请求

Slide 7: 배치 처리 계산

먼저 가장 전통적이고 일반적인 배치 처리 계산을 살펴보겠습니다.
대표적인 기술은 **MapReduce**입니다!
예를 들어, 타오바오의 "광군제(11월 11일)" 하루 전체 거래 데이터를 분석하고 싶다면, 밤에 배치 처리 작업을 실행해서 한 번에 모든 계산을 완료할 수 있습니다.

장점:

- 안정적이고 대용량 데이터에 적합
- 높은 장애 허용성

단점:

- 응답이 느리고, "지금 당장" 필요한 작업에는 부적합

Slide 7:批处理计算(pī chǔ lǐ)

我们先看最传统也最常见的：批处理计算

它的代表是—MapReduce！
比如你想分析淘宝"双十一"一整天的交易数据，那你可以晚上跑(pǎo)个批处理任务，一次性全部算完。 光棍节 Guānggùn Jié

优点：

- 稳定、适合海量数据
- 容错强

缺点：

- 响应慢，不适合"现在立刻马上"的任务

Slide 8: 스트림 계산의 필요성

다음은 현재 매우 인기 있는 "스트림 계산"입니다!

📡 스트림 계산이 처리하는 것은 "실시간으로 생성되는 데이터 스트림"입니다.

예를 들어, 타오바오에서 상품을 탐색하고 있다면:

- 한 번 클릭하면 클릭 이벤트가 생성됩니다
- 한 번 스크롤하면 로그 기록이 생성됩니다
- 시스템은 1초 내에 새로운 상품을 추천해야 합니다!

📌 반응이 5초 늦으면 사용자는 이미 앱을 종료했을 수 있습니다. 그러면 이 데이터는 쓸모없어집니다 - 상업적 가치가 없습니다.

Slide 8:流计算的必要性

接下来是现在非常流行的"流计算"！

📡 流计算要处理的是—"实时产生的数据流"。

比如你正在淘宝浏览(**liú lǎn**)商品：

- 点一下，就产生一个点击事件
- 滑一下，又生成一个日志记录
- 系统要在1秒内给你推荐新商品！

📌 如果反应慢5秒，用户可能已经退出App了。
那这条数据就白费(**bái fèi**)了—没有商业价值。

Slide 9: 스트림 계산의 전형적인 도구

어떤 도구들이 스트림 계산을 지원할까요?

🔥 **대표 제품:**

- **Storm**: 분산 스트림 처리의 원조
- **Flume**: 로그 수집 및 스트리밍 전송 도구
- **S4**: 경량 스트림 계산 플랫폼 (Yahoo 제품)

Slide 10: 그래프 계산 소개

그래프 계산이란 무엇일까요?

👥 **그래프 = 노드 + 엣지**

예를 들어:

- 사람과 사람 사이는 "친구"
- 웹사이트와 웹사이트 사이는 "링크"
- 도시 사이는 "교통 노선"

📌 그래프 계산은 이러한 네트워크 구조의 데이터를 처리하는 데 특화되어 있습니다.

Slide 9:流计算的典型工具

有哪些工具支持流计算呢？

🔥 代表产品：

- Storm：分布式流处理鼻祖
- Flume：日志采集与流式传输工具
- S4：轻量级流计算平台（Yahoo出品）

Slide 10: 图计算简介

什么是图计算？

👥 图 = 点 + 边

比如：

- 人和人之间是"好友"
- 网站和网站之间是"链接"
- 城市之间是"交通线"

📌 图计算专门用来处理这种网络结构的数据。

Slide 11: 그래프 계산 응용 시나리오: GIS

그래프 계산은 지리정보시스템(GIS)에 매우 적합합니다:

🗺️ 예를 들어 지도 내비게이션:

- 도시는 노드이고, 도로는 엣지입니다
- 최단 경로 찾기, 최적 경로 계산, 정체 회피 등 모두 그래프 계산을 사용합니다!

📌 그래프 계산의 대표 도구: **Google Pregel**

Slide 11:图计算应用场景：GIS

图计算非常适合地理信息系统（GIS）：

🗺️ 比如地图导航：

- 城市是点，路是边
- 找最短路径、计算最优路线、避开(Bìkāi) 拥堵(yǒngdǔ)，都要用图计算！

📌 图计算的代表工具：Google Pregel

Slide 12: 소셜 네트워크 = 자연스러운 그래프 구조!

가장 흔한 예를 하나 더 들어보겠습니다:

📱 **소셜 네트워크 (웨이보, 틱톡, 모멘트)**

- 각 사람은 하나의 "노드"입니다
- 팔로우/좋아요/댓글은 하나의 "엣지"입니다

예를 들어, 제가 당신을 팔로우하고, 당신이 제 셀카에 좋아요를 누르면, 우리 사이에는 연결이 생깁니다 - 네트워크처럼 보이지 않나요?
따라서 소셜 플랫폼의 추천 알고리즘은 그래프 계산 없이는 불가능합니다!

Slide 13: 쿼리 분석 계산의 필요성

네 번째 계산 모드 - **쿼리 분석 계산**
이 모드가 목표로 하는 것은: 📊 대규모 데이터의 저장 관리 + 빠른 쿼리

예를 들어:

- 시스템에 "지난 7일간 방문량이 가장 많은 상품은 무엇인가?"라고 물어봅니다
- 또는 "특정 사용자의 쇼핑 선호도는?"

이런 시나리오는 초 단위 응답 + 유연한 쿼리가 필요합니다.

Slide 12:社交网络 = 天然图结构！

再来一个最常见的例子：

📱 社交网络（微博、抖音、朋友圈）

- 每个人是一个"节点"
- 关注/点赞/评论 就是一条"边"

比如我关注了你，你点赞了我发的自拍，
我们之间就有连接——是不是像个网？

所以，社交（**shè jiao)**平台的推荐算法，就离不开图计算！

Slide 13:查询分析计算的需求

第四种计算模式——查询分析计算
这种模式针对的是：
📊 大规模数据的存储管理+快速查询。

比如：

- 你问系统："过去7天访问量最多的商品是什么？"
- 或者"某个用户的购物偏好（**piān hào)**"？

这种场景需要的是秒级响应 + 灵活查询

Slide 14: 쿼리 분석 도구 대표

일반적인 도구들:

- **Google Dremel**: 분석 쿼리 엔진
- **Hive**: SQL을 MapReduce로 변환, 오프라인 데이터 웨어하우스에 적합
- **Cassandra**: 고가용성, 고확장성 컬럼 기반 NoSQL 데이터베이스
- 기타: Impala, Presto, ClickHouse 등

📌 이러한 도구들은 "스마트 고객 서비스 시스템"과 같습니다. 당신이 물으면 답변하지만, 실제로는 많은 기계가 뒤에서 작동하고 있습니다!

Slide 15: 네 가지 계산 모드 비교 요약표

이 페이지는 중요한 복습 차트입니다!

📌 다음 구호를 기억하세요:
"배치, 스트림, 그래프, 쿼리 –
네 가지 주먹으로 천하를 제패한다" 🥊

Slide 14:查询分析工具代表

常见的工具有：

- Google Dremel：分析性查询引擎
- Hive：将SQL转换为MapReduce，适合离线数据仓库
- Cassandra：高可用、高扩展的列式NoSQL数据库
- 其他还有
  Impala、Presto、ClickHouse 等

📌 这些工具就像"智能客服系统"，你一问，它一答，背后其实跑(**pǎo)**了很多机器！

Slide 15:四种计算模式对比总结表

这一页是重点复盘图表！

📌 记住一句口诀：
"大批流图查，四拳打天下" 🥊
dà pī liú tú chá, sì quán dǎ tiānxià.

Slide 16 ✅ **정답: C**
Slide 17 ✅ **정답: C**
Slide 18 ✅ **정답: C**
Slide 19 ✅ **정답: C**
Slide 20 ✅ **정답: D**
Slide 21 ✅ **정답: C**
Slide 22 ✅ **정답: B**
Slide 23 ✅ **정답: C**
Slide 24 ✅ **정답: D**
Slide 25 ✅ **정답: D**

Slide 26: 수업 마무리

좋습니다, 이번 수업은 여기서 마치겠습니다!
오늘 우리는 "빅데이터 시대"의 전말을 알아보았습니다. 정보화 물결, 세 가지 지원 기술부터 데이터가 어떻게 생성되고 어떻게 진화하는지까지 배웠습니다.

다음 수업에서는 "빅데이터의 핵심 기술"에 대해 계속 탐구하겠습니다.

마지막으로 질문 하나 드리겠습니다: 여러분은 매일 실제로 "데이터를 생산"하고 있다는 것을 오늘 발견하셨나요? 😉
감사합니다, 수업 마치겠습니다! 👏👏👏

Slide 16 ✅ **정답: C**
Slide 17 ✅ **정답: C**
Slide 18 ✅ **정답: C**
Slide 19 ✅ **정답: C**
Slide 20 ✅ **정답: D**
Slide 21 ✅ **정답: C**
Slide 22 ✅ **정답: B**
Slide 23 ✅ **정답: C**
Slide 24 ✅ **정답: D**
Slide 25 ✅ **정답: D**

Slide 26: 수업 마무리

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，从信息化浪潮、三大支撑技术，到数据如何产生、如何演变。

下节课我们会继续探讨"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！👏👏👏