



# 教育大数据分析研究及其典型应用

西安交通大学 郑庆华  
大数据算法与分析技术国家工程实验室

2017年3月



## CONTENTS

01

一、大数据时代背景与战略必争

二、陕西省高等教育大数据平台

02

03

三、MOOC中国平台及其应用

四、课堂教学质量监测大数据平台

04

# 国家战略需要

国务院通过：《关于促进**大数据**发展的行动纲要》



十八届五中全会决定：实施国家**大数据**战略



大数据具有大价值：**社会科学的新方法论、科学研究的新范式、高新科技的新领域和社会进步的新引擎**；从大中发现新知识，创造新价值，形成新理念，是认知世界与改造世界的能力（**国家创新驱动发展的核心驱动力**）

# 大数据：时代特征，时代任务

信息技术与经济社会活动的交融

→ 大数据时代

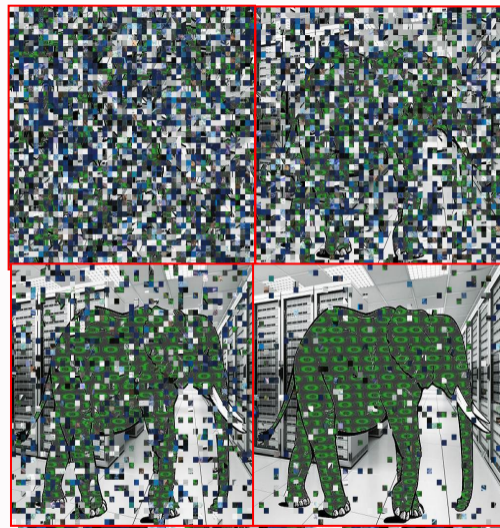
数据的积累“量变→质变”+ 计算能力的突飞猛进

→ 大数据智能分析

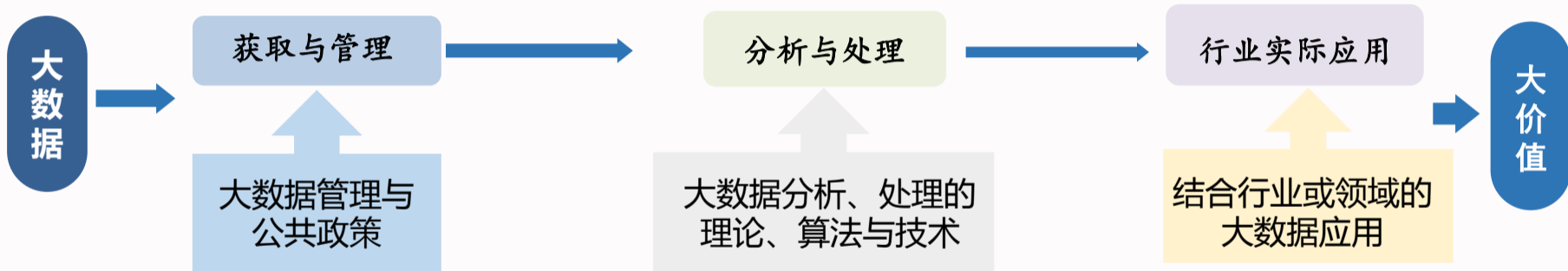
分析各领域的大数据

→ 正成为各行各业的新需求

- 拥有大数据是**时代特征**，分析大数据是**时代任务**，应用大数据是**时代机遇**
- 信息技术已经跨越了**数字化与网络化**阶段，进入了以数据深度挖掘与知识融合为特征的**智慧化**阶段。



# 大数据产业：政—产—学—研—用联动



数据是基础，平台是支撑，分析是核心，效益是根本

# 核心技术：必须抢占大数据分析技术的制高点



Center for Visualization and  
Data Analytics (CVADA)

A DHS Science and Technology Center of Excellence

**BDC**  
BERLIN BIG DATA CENTER

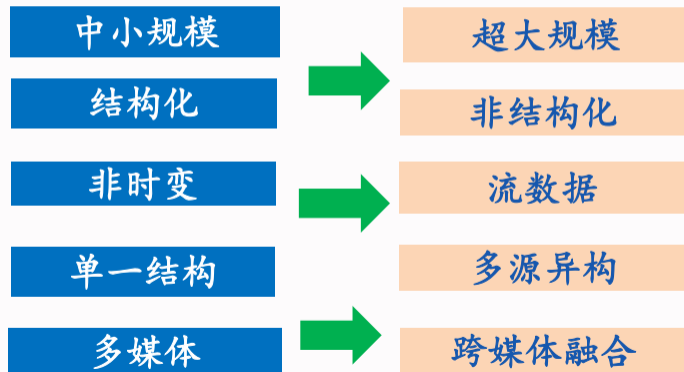
- **英国政府**：2014年成立阿兰图灵研究所（Alan Turing Institute），综合剑桥大学、爱丁堡大学、牛津大学、华威大学、伦敦大学学院和EPSRC专家，致力于大数据分析及应用研究。
- **美国**：以普渡大学为首建立“可视化和数据分析卓越中心”（CVADA），探索识别、分析、理解海量复杂动态信息的大数据科学技术。
- **德国**：于2014年在柏林理工大学设立了柏林大数据中心。

**BDIC**  
BIG DATA INDUSTRY  
CONFERENCE 2017

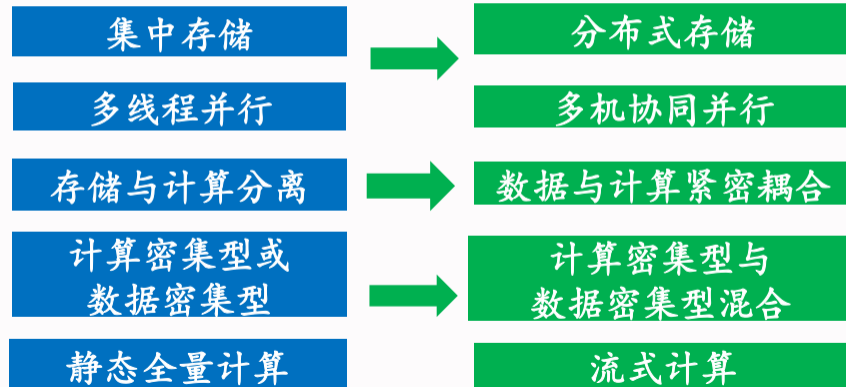
# 挑战1：传统计算理论和技术难以完全适用

(MIT Technology Review, 2015)

## 数据特征突变



## 计算平台和计算模式的转变



## 挑战2：大数据分析处理的核心技术严重缺乏

### 多源异构数据分析缺乏自主工具

- 缺乏大数据开发平台工具方面的自主核心技术
- 缺乏业界认可的可扩展分析处理平台构架技术
- 缺乏针对异构数据的模块化一次处理技术

Sparknotebook, Jupyter notebook, Eclipse, SBT

Hadoop, Spark, Storm, Hbase, Cassandra...

NPL, OpenCV, OpenSmile, Scrapy, Tableau...

## 挑战3：难以适配工程化应用的需求

### 理论与方法论的冲击

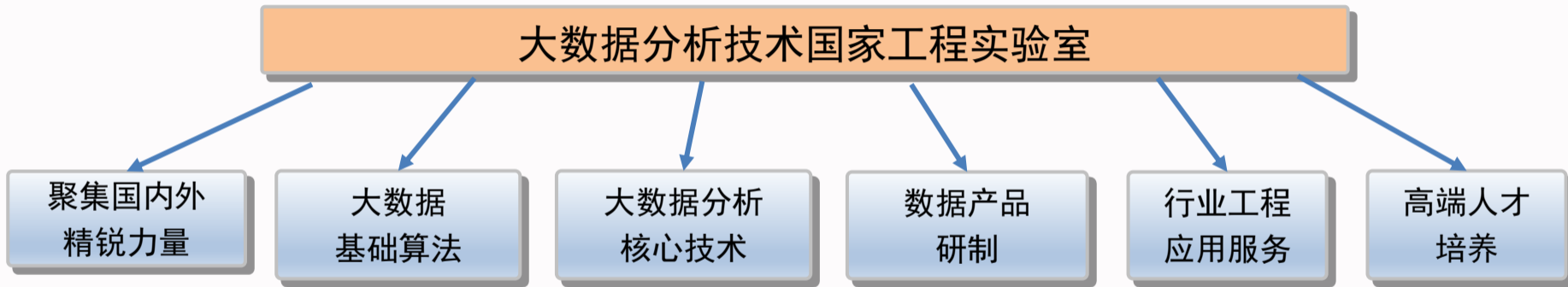
- 大数据算法工程化技术尚属空白
- 缺乏适用于大数据分析的工程化工具及快速开发平台
- 核心技术与产业需求存在缝隙，缺乏面向行业的智能大数据决策支持工具及解决方案

生态链软件、平台的选择与配置，  
算法选型，测试与评估

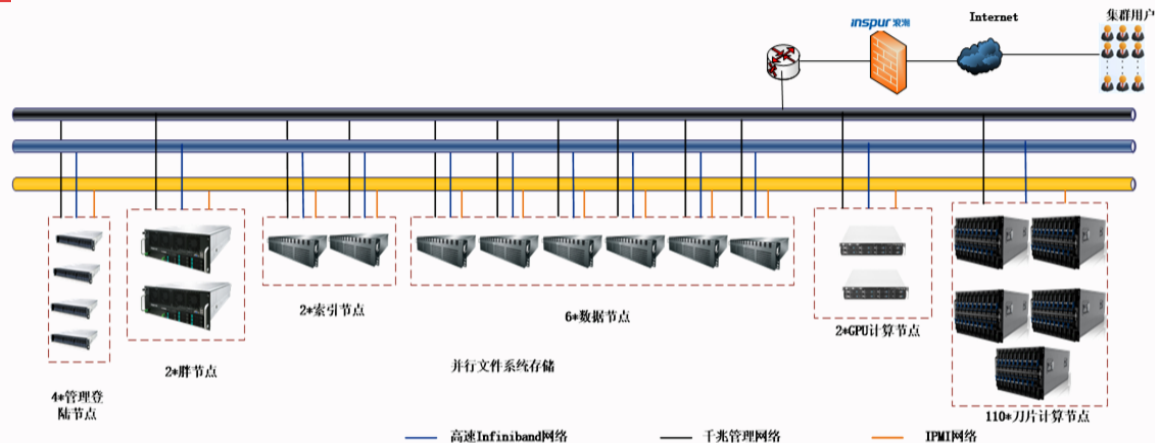
Weka, Tableau, SPSS, SAS

IBM InfoSphere BigINsights,  
Oracle Big Data Appliance,  
Google BigQuery

# 大数据分析技术国家工程实验室功能定位及使命



# 基础平台：建成230万亿次高性能计算平台



地位：西北高校最快超算系统

运算速率：232.79万亿次/秒

计算节点：双路刀片\*210台

异构节点：GPU机架\*4

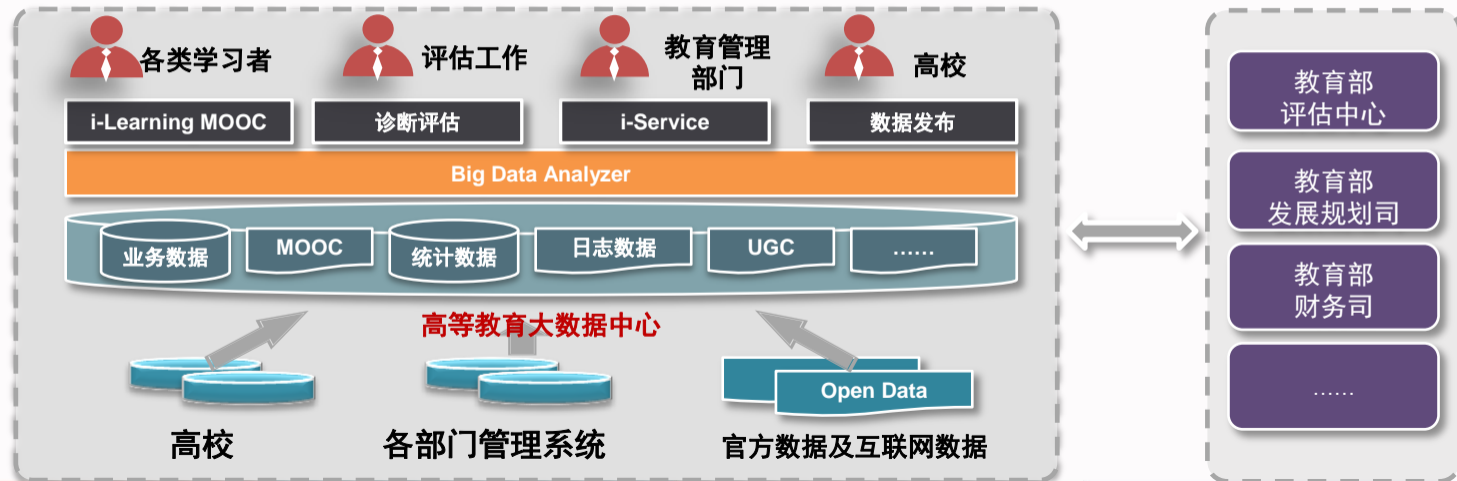
计算网络：112Gbps FDR 全线速

管理网络：万兆交换

存储系统：总计1240TB，5GB/s高带宽

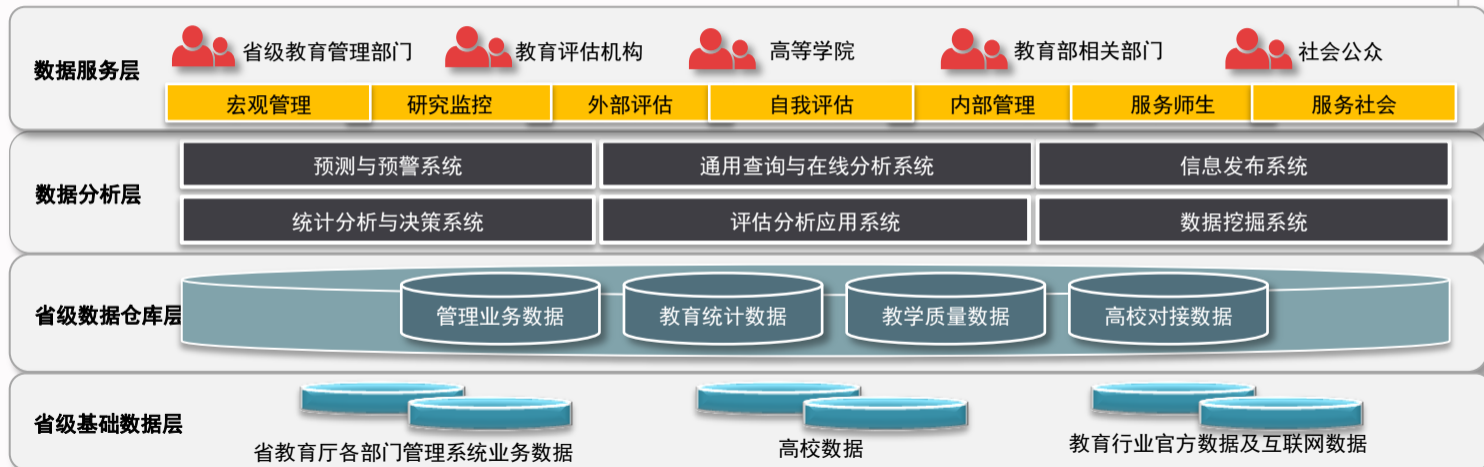


## 二、陕西省高等教育大数据平台

全省教育大数据汇聚、分析及应用的总体设计 <http://snedc.xjtu.edu.cn>

## 特点:

- 汇聚全省高等教育数据
- 打破数据孤岛，融合各方数据
- 横向关联比较，纵向历史分析
- 提供精准服务，支持科学决策



# 高校大数据汇聚与分析评估

序号	数据名称	数据类别	数据量（万条）	数据容量（M）	数据时间范围
1	管理业务数据	高教处、研究生处、科研处、财务处等11个业务部门，54类数据	1596	3547	1993-2015年
2	教学质量数据	与教育部评估中心数据接，采集全省普通高校的学校基本条件、教师信息、学科专业、人才培养等教学状态和质量数据	149	/	2014-2016年
3	统计类数据	高等院校的基本情况报表数据（高基表）：全省54所院校，涉及办学条件、教师、学生、其他等4类数据	552	1400	2005-2015年
		教育经费统计数据：全省113所大中专院校，学校基本情况、教育经费收入、教育经费支出、科研经费收支等4类	1962	5010	2000-2015年
		教育科研统计数据：全省高校、科研院所统计数据	44	/	2008-2015年
4	高校对接数据	通过省校对接接口，采集高校教职工、学生两类数据	4	/	实时采集
总 计			5237	12327	

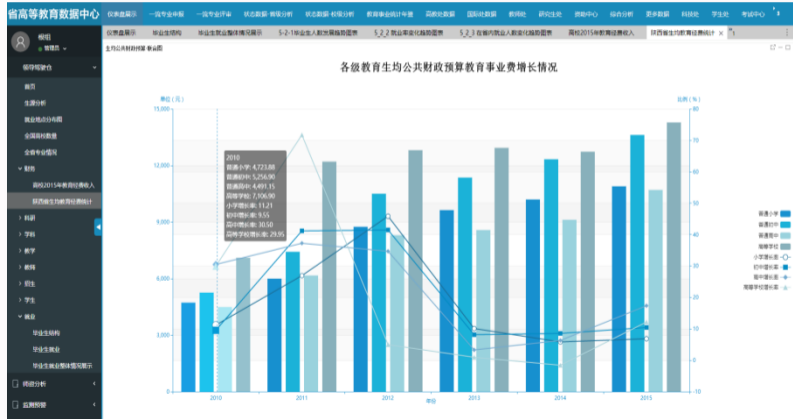




# 典型应用3：陕西省教育经费绩效分析与评估

- 建成陕西省教育经费统计分析应用服务
- 编制发布《2014年陕西省教育经费统计年鉴》

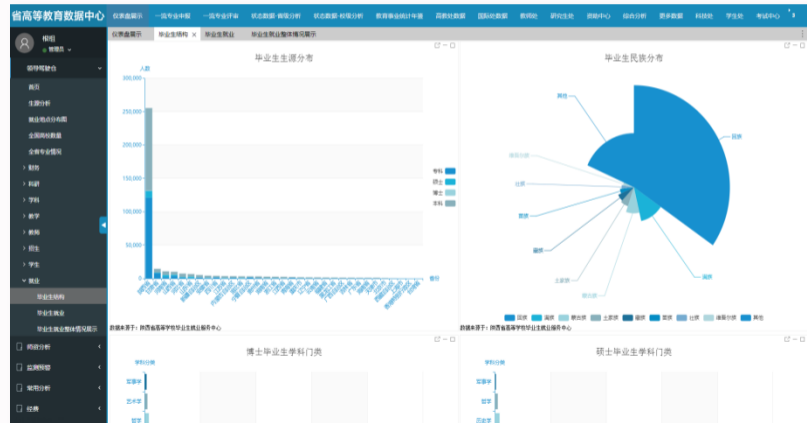
- 连续6年承担全省教育经费统计工作
- 提高数据在科学管理、决策支持和趋势预测方面的作用



# 典型应用4：陕西省毕业生就业质量跟踪评价

- 建成陕西省毕业生就业质量大数据分析应用服务
- 编制发布《2016年陕西省高校毕业生就业状况报告》，8.6万字

- 推进全省毕业生就业质量调查研究的常态化
- 面向全省25万学生发放调研问卷，重点采集对教育教学的反馈



内部资料 请勿外传

## 2016年陕西省高校毕业生就业状况报告

中国西部高等教育评估中心  
陕西省高等教育数据中心  
陕西省高等学校毕业生就业服务中心  
2016年12月



### 三、MOOC中国平台及其应用



### 背景与使命：

2015年元月，国内37所一流大学和奥鹏远程教育公共服务体系共同发起成立“MOOC中国”联盟，这是在经济进入新常态、创新驱动发展、教育综合改革的大背景下，为实现继续教育转型升级，探索互联网教育公共服务新模式，采取的一项重大举措，体现了联盟大学的共同价值和新的使命。





1. 国家利益需要
2. 国际竞争需要
3. 学习者需要



做政府想做、社会愿做、单一高校做不了的事情



实现网络教育公共服务体系从1.0到2.0的蜕变



“互联网+教育”的探索及实践创新基地



117所高校



10000+  
视频课程



10000+  
注册用户

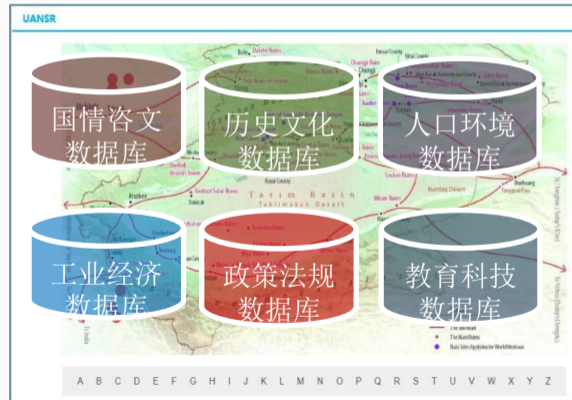


300  
精英讲师



900W+  
APP下载

- 创办“丝路学院”，成为中国工程院UNESCO国际工程科技知识培训基地；
- 与国家汉办合作在泰国建立“东盟中心”；
- 面向“一带一路”培养包含中国文化、语言、教育、技术等的特色资源，搭建国际共享平台。



与IKCEST联合举办“丝路工程科技发展专项培训”。以集中面授、参观学习、互动交流、实践应用等多种形式，对丝路沿线国家来华留学生、在华外向型企业技术及**管理骨干**等开展工程科技培训。面向巴基斯坦、印度、尼泊尔、约旦、伊朗、多哥、巴林等30余个国家。



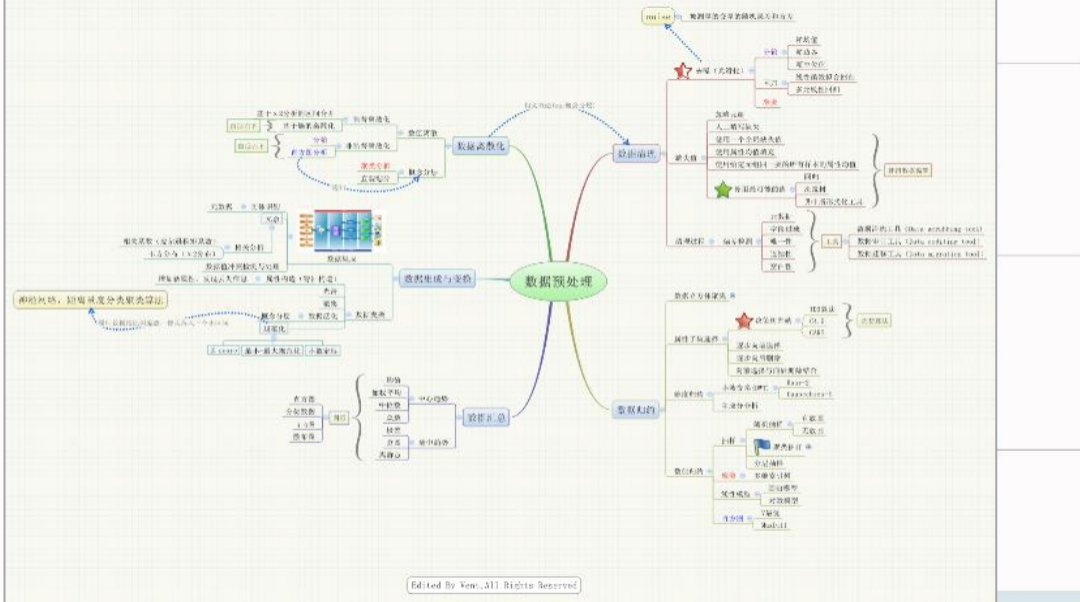
# 实现基于知识地图的导航学习

## 导航学习:

根据先验知识与学习目标在知识地图上为学习者规划一条由知识单元关联关系组成的学习路径。

## 特点:

- ◆ 将碎片化知识聚合成符合认知特点的知识森林，缓解学习迷航问题。
- ◆ 整合特定知识单元的资源，缓解碎片化问题；
- ◆ 依据关联关系快捷地获取与特定知识相关的学习资源





## 四、教学质量实时监测大数据平台

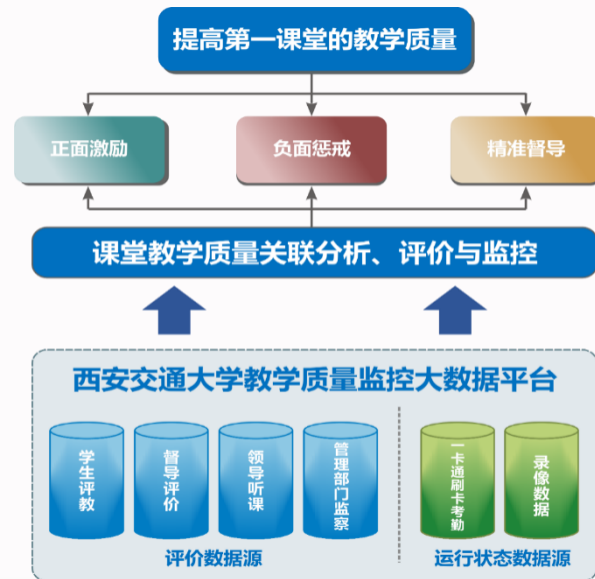
# 典型应用：课堂教学质量监测大数据平台

## 实现四大转变

模糊宏观 → 量化精准    部分随机 → 全面覆盖

每学期制 → 常态持续    事后评价 → 实时动态

本科教学质量综合评价系统应用作为学校教  
育教学五项重点工作之一



# 课堂教学状态数据的实时采集

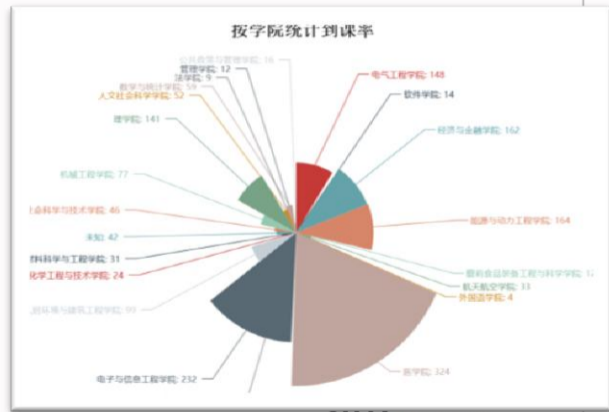
- 基于“物联网+云计算”技术的智慧教室管理，实时统计学生出勤率，教师到课率。课堂教学的全程录制。
- 解决数据精准采集问题，从而严肃教学秩序。

- 考勤数据：每天约187,726条，每周约1,314,083条。
- 课堂录像数据：目前采集的教室约110个，全程实时录制；同时每堂课截取12张采样照，每天约6,870张（600MB），每周4.8万张约4GB，每学期约80GB。



# 精准回答：学生去哪儿了？学生喜欢上哪些课？

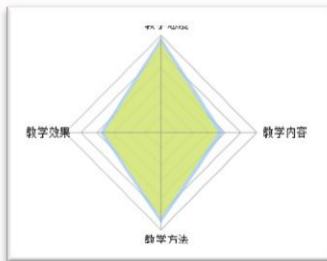
- **到课率：**实时反映和跟踪每节课的到课情况，未到课学生的去向分析
- **教师维度：**实时给出每位教师的课堂到课率
- **课程维度：**按课程统计并分析到课率
- **学院维度：**分学院统计并分析到课率
- **秩序建设：**可以自动精准地知会未按时到课的教师和学生，增强师生的纪律要求，敬畏课堂。





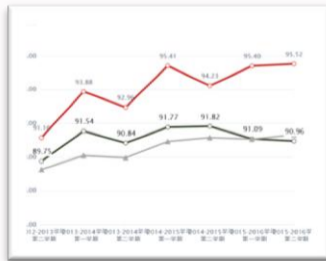
专家评价

➤ 教师课堂教学质量进行综合评价



学生评教

➤ 教学质量的历史比较



到课率

➤ 每一堂课到课率的演变情况



课堂现场

➤ 使用视频、快照方式还原课堂



# 大幅提升学生培养质量

2016届本科毕业生共**3290人**，考取国内外研究生**2242人**，占比达**68.15%**。其中：



**214名**优秀毕业生  
赴西部地区建功立业；



**678名**本科毕业生  
进入各类企业工作；



继续攻读硕、博士的人数为**2242人**，占**68.15%**。比去年增加近5个百分点其中，国内升学**1739人**，占**52.86%**；出国出境留学**503人**，占**15.29%**，比去年增加**123人**；



**15名**毕业生到外交部、税务总局等国家机关  
和事业单位工作；

# 使命与责任

扎根西部，瞄准教育大数据重大需求，聚集“产学研”精锐力量，将“大数据分析技术国家工程实验室”建成一流的大数据基础算法与核心技术研究、数据产品研发与推广应用以及高端人才培养与培训”的国家级基地。



谢谢大家！