



云聚精英 创新创造
与聪明人一起做精彩的事!

高校大数据 教学科研解决方案

All-in-One Solution for Big Data
Teaching and Research



南京云创大数据科技股份有限公司
Nanjing Innovative Data Technologies, Inc.

股票代码: 835305
网站: <http://www.cstor.cn>
电话: 4008855360 025-83700385
地址: 南京市白下高新技术产业园中国云计算创新基地A栋9层

南京云创大数据科技股份有限公司
Nanjing Innovative Data Technologies, Inc.

云创大数据

南京云创大数据科技股份有限公司成立于2011年3月，是专业从事大数据、云计算、云存储技术领域研发及产品销售的高新技术企业。公司依托革命性的技术创新迅猛发展，被选为中国云计算创新基地理事长单位，被授予“2012年度第二批江苏省科技型中小企业”称号。公司于2015年12月28日在新三板挂牌，股票简称：云创数据，股票代码：835305，网址：<http://www.cstor.cn>，微信公众号：cStor_cn。

云创大数据以“与聪明人一起做精彩的事”为企业文化，拥有一支实力雄厚、协作高效的研发团队。公司董事长张真当选南京十大经济风云人物，荣获南京市五一劳动奖章。公司总裁刘鹏教授曾当选全军十大学习成才标兵、南京十大杰出青年、清华大学学术新秀等，是2002年全球大数据处理比赛PennySort的世界冠军和2003年全国挑战杯比赛总冠军。

云创大数据是国际上云计算产品线齐全的企业之一，针对爆炸式增长的大数据需求，研发了自主知识产权的cStor云存储、cProc云处理、cVideo云视频、cTrans云传输等产品线，每条产品线都让人眼前一亮。

云创大数据以知识产权为核心战略，目前已申请专利、著作权、商标近百项，通过了ISO9001质量管理体系认证，并与Intel、联想、神州数码、曙光、浪潮等多家知名企业开展了深入的合作。目前云创大数据的产品已经在环保、地震、公安、教育、广电等10多个领域建立了200多个成功案例。

云创大数据自成立以来，得到了国家工信部、江苏省政府、南京市政府等各级政府的大力支持和协助，国家和省市几百位领导先后视察了公司。云创大数据先后荣获国家科技部民营科技发展贡献奖、国家工信部“中国中小企业首选服务商”、江苏省高新技术企业、南京服务业名牌，牵头成立了江苏省信息产业云存储技术联合研发中心。由于过去几年增长速度位居江苏省数据和信息管理行业前列，近期被德勤会计师事务所和江苏省经信委评为“江苏高科技高成长30强”。

“与聪明人一起做精彩的事”，期待与您携手合作！

目录 Contents

2~9
大数据实验一体机



10~13
DeepRack
深度学习一体机



14~17
云创网盘高校应用方案



手机APP



我的PM2.5



同声译



我的南京



科技头条

微信公众号



云创大数据



刘鹏看未来



中国大数据



云计算头条



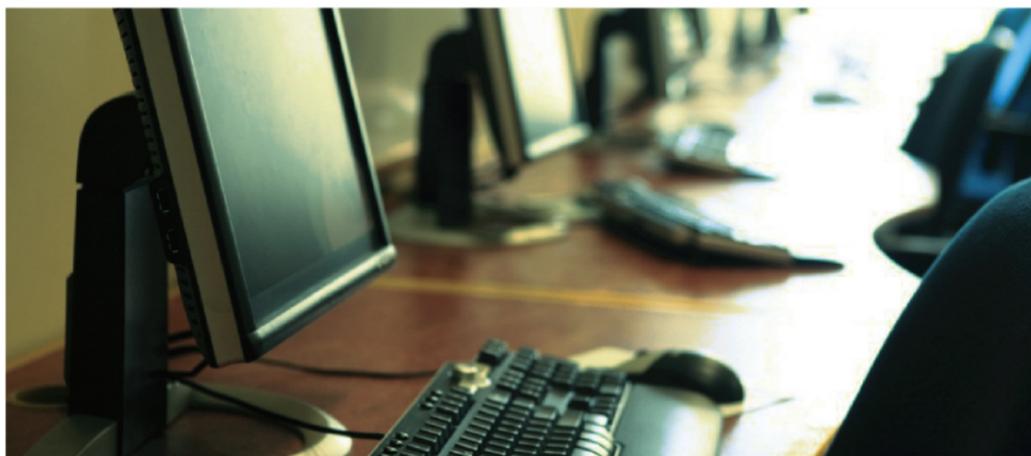
中国云计算创新基地
China Cloud Computing Innovation Base

1 大数据实验一体机

大数据人才荒，各大高校打造大数据人才高地

1. 人才缺口高达150万，实用型大数据人才抢手

- 1) 在未来5-10年，我国大数据市场规模年均增速将超过30%
- 2) 大数据人才平均月薪达1.5万，5年内国内大数据人才缺口将突破150万
- 3) 在BAT发布的招聘职位中，大数据人才需求超过60%



2. 高校启动大数据人才培养计划，专业建设困难重重

◆ 课程体系待完善，大数据教学科研入门难

- 1) 各大高校的大数据专业处于起步阶段，人才培养课程体系缺乏系统性
- 2) 大数据教学科研资源匮乏，可配置和指导实验环境的专业师资不足

◆ 基础实验环境薄弱，大数据实训难开展

- 1) 面向大数据的分布式系统，高校无法为每个学生提供多机实验环境
- 2) 大数据分布式软件系统安装、配置难度大，实验环境容易被破坏
- 3) 实验数据、实验教案、实验手册不足，实验开展束手束脚

◆ 项目实训缺失，教学效果与就业率打折扣

- 1) 缺少大数据项目实训，难以培养实用型人才所需的专业项目能力
- 2) 高校大数据专业学习与实际应用脱轨，教学效果与就业率打折扣

大数据实验一体机——搭建一站式的大数据教学科研平台

针对大数据专业建设的三大难题，云创大数据为各大高校量身定制了大数据软硬件一体化的教学科研平台——大数据实验一体机。大数据实验一体机通过应用容器技术，以少量机器虚拟大量实验集群，可供大量学生同时拥有多套集群进行试验，而每个学生的实验环境不仅相互隔离，方便高效地完成实验，而且实验彼此不干扰，即使某个实验环境被破坏，对其他人也没有影响，一键重启就可以拥有一套新集群，大幅度节省了硬件和人员管理的投入成本。



目前，高校大数据的教学、科研不仅缺少实验设备，同时对于具体实验内容和实验步骤等也不知何去何从。对此，大数据实验一体机提供了一套完整的实验教案，一方面通过大数据教材、讲义PPT、视频课程等理论学习，帮助学生建立从大数据监测与收集、存储与处理、分析与挖掘直至大数据创新的完整知识体系，另一方面，搭配教学组件安装包及实验数据、实验手册、专业网站等一系列资源，使高校可在42个大数据实验中任意挑选并轻松完成实验，大幅度降低大数据课程的学习门槛。

此外，作为一个可供大量学生完成大数据与云计算实验的集成环境，该平台同步提供了配套的培训服务，对于教学组件的安装、配置，教材、实验手册等具体应用提供一站式服务，有助于高校更好地满足课程设计、课程上机实验、实习实训、科研训练等多方面需求，并在一定程度上缓解大数据师资不足的问题。对于各大高校而言，即使没有任何大数据实验基础，该平台也能助其轻松开展大数据与云计算的教学、实验与科研。

具体而言，大数据实验一体机从以下三个方面解决了高校大数据的教学科研难题：

1.完整的大数据课程体系及配套资源，一步解决入门难

◆ 内容充实的课程体系

在《实战Hadoop2.0——从云计算到大数据》和实验手册的指导下，大数据实验一体机解决方案涵盖大数据算法、接口、工具、平台等多方面内容，从大数据监测与收集、大数据存储与处理、大数据分析与管理直至大数据创新，帮助高校构建完善的大数据课程体系。



◆ 多元化的大数据资源包

1) 国内早期Hadoop编程书籍的升级版——《实战Hadoop2.0》及其配套PPT

Hadoop编程书籍的全新升级版

- 知名云计算专家刘鹏教授与叶晓江编著
- 国内早期Hadoop书籍全新升级
- 全面系统，一网打尽前沿技术
- 数年实战经验，倾囊相授

2) 综合42个大数据实验的实验手册及配套高清视频课程

涵盖原理验证、综合应用、自主设计及创新的多层次实验内容

每个实验呈现详细的实验目的、实验内容、实验原理和实验流程指导

配套相应的实验数据和高清视频课程，参照手册即可轻松完成每个实验

实验手册大纲

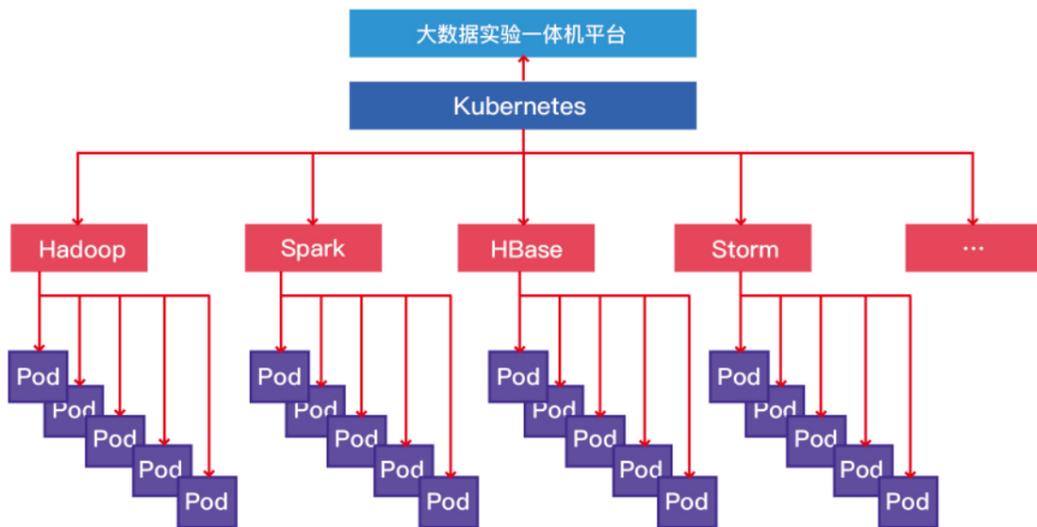
实验一	基本操作实验
实验二	HDFS实验：部署HDFS
实验三	HDFS实验：读写HDFS文件
实验四	YARN实验：部署YARN集群
实验五	MapReduce实验：单词计数
实验六	MapReduce实验：二次排序
实验七	MapReduce实验：计数器
实验八	MapReduce实验：Join操作
实验九	MapReduce实验：分布式缓存
实验十	Hive实验：部署Hive
实验十一	Hive实验：新建Hive表
实验十二	Hive实验：Hive分区
实验十三	Spark实验：部署Spark集群
实验十四	Spark实验：SparkWordCount
实验十五	Spark实验：RDD综合实验
实验十六	Spark实验：Spark综例
实验十七	Spark实验：Spark SQL
实验十八	Spark实验：Spark Streaming
实验十九	Spark实验：GraphX
实验二十	部署ZooKeeper
实验二十一	ZooKeeper进程协作
实验二十二	部署HBase
实验二十三	新建HBase表
实验二十四	部署Storm
实验二十五	实时WordCountTopology
实验二十六	文件数据Flume至HDFS
实验二十七	Kafka订阅推送示例
实验二十八	Pig版WordCount
实验二十九	Redis部署与简单使用
实验三十	MapReduce与Spark读写Redis
实验三十一	MongoDB实验：读写MongoDB
实验三十二	LevelDB实验：读写LevelDB
实验三十三	Mahout实验：K-means
实验三十四	使用Spark实现K-means
实验三十五	使用Spark实现SVM
实验三十六	使用SPARK实现FP-Growth
实验三十七	综合实战：车牌识别
实验三十八	综合实战：搜索引擎
实验三十九	综合实战：推荐系统
实验四十	综合实战：环境大数据
实验四十一	综合实战：智能硬件大数据托管
实验四十二	综合实战：贷款风险评估

3) 百度搜索排名前列的网站——中国大数据、中国云计算、中国存储等提供全线支持，一网打尽各类优质资源。



2.安全可靠的实验环境，大幅度提升大数据技能

- 1) 基于容器Docker技术，大数据实验一体机可瞬间创建随时运行的实验环境
- 2) 使用几台机器即可虚拟出大量实验集群，方便上百学生同时使用
- 3) 采用Mesos+ZooKeeper+Mrathon架构管理集群，实验集群完全隔离
- 4) 实验环境互不干扰，如果实验环境破坏，一键重启即可建立新集群
- 5) 内置数据挖掘等教学实验数据，可导入高校各学科数据进行教学、科研，校外培训机构同样适用



3.热门实战项目贯穿始终，进一步提高教学效果与就业率

大数据实验一体机解决方案采用理论与实践相结合的人才培养模式，帮助教师提高教学水平，促使学生完善大数据知识体系。

- 1) 基于真实的企业基地实训经验，提供丰富的项目实训案例
- 2) 结合高校各专业实际情况进行行业数据研究，培养实用型人才的专业项目能力

4.更多潜在效益，同步增强高校的硬实力和影响力

- 1) 大数据上升为国家战略，发改委明确组建13个国家级大数据实验室，大数据实验一体机有助于高校大数据实验室建设以及高层次大数据人才的深度培育
- 2) 大数据实验一体机解决方案将理论与实践双管齐下，帮助提升了高校信息化管理水平和实验项目研究水平
- 3) 大数据产业迎来发展黄金期，以大数据实验一体机提高大数据专业就业率，可进一步增强高校的硬实力和影响力

在2016年暑期全国高校大数据培训中，云创大数据利用大数据实验一体机搭建了Docker容器云，为每个学员分配5套虚拟服务器集群，为其提供了简洁易用的上机操作环境，得到了学员的一致好评。在理论讲解的基础上，讲师通过这一实践平台，为学员提供准确到每一步的操作指导，真正做到了学思结合、知行统一，所有学员的大数据应用能力均得以提升，并获得了相应的大数据能力等级证书。



详细的规格参数，助力高校选配个性化的私人定制方案

目前，大数据实验一体机具有经济型、标准型与增强型三种规格，通过发挥实验设备、理论教材、实验手册等资源的合力，可满足数据存储、挖掘、管理、计算等多样化的教学科研需求。

配套/型号		经济型	标准型	增强型
大数据实验一体机	管理节点	1台	3台	3台
	处理节点	6台	8台	15台
上机人数		30人	60人	150人
理论教材		《实战Hadoop2.0》 50本	《实战Hadoop2.0》 80本	《实战Hadoop2.0》 180本
实验教材		《大数据实验手册》 50本	《大数据实验手册》 80本	《大数据实验手册》 180本
配套PPT		有	有	有
配套视频		有	有	有
培训		提供现场实施及 3天技术培训服务	提供现场实施及 5天技术培训服务	提供现场实施及 7天技术培训服务



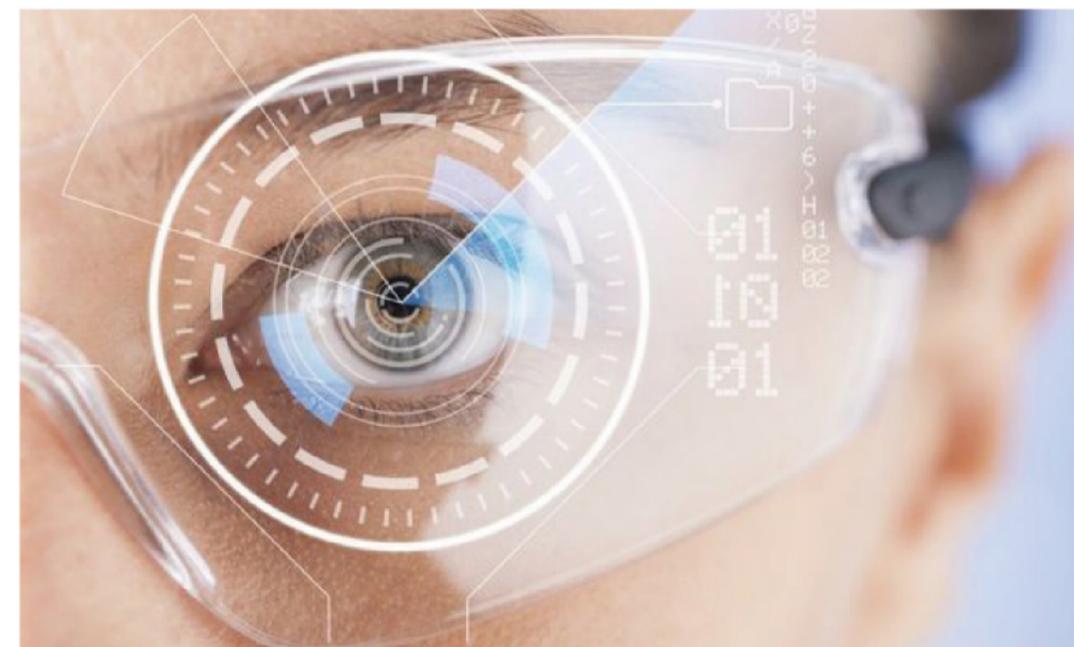
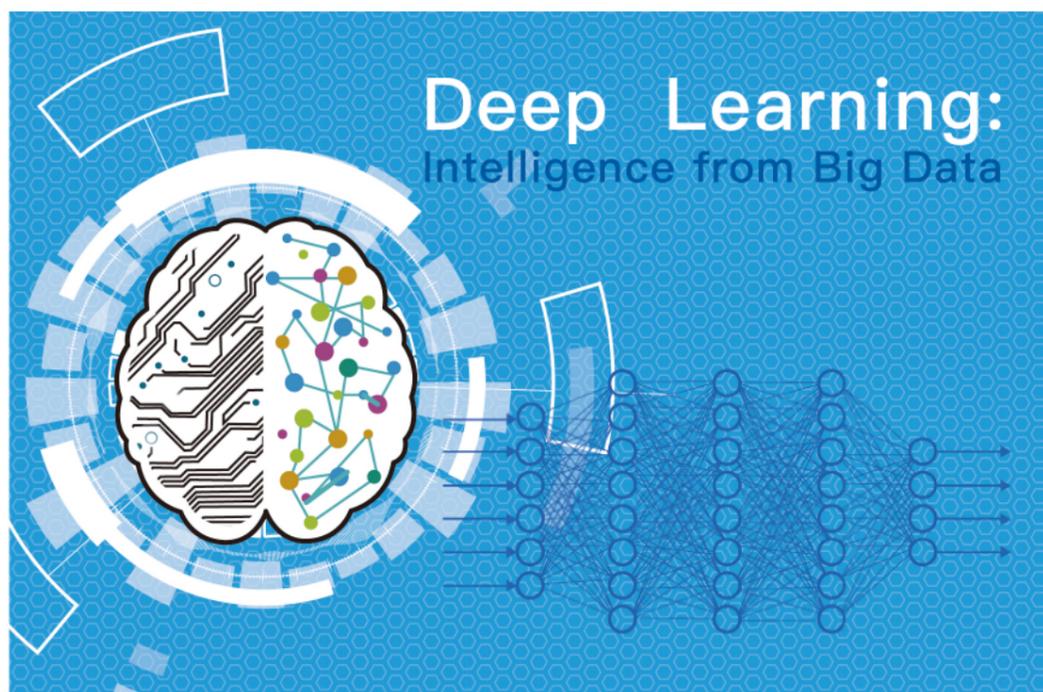
软件配置方面，各大高校可根据具体应用，选择Hadoop、HBase、Ambari、HDFS、YARN、MapReduce、ZooKeeper、Spark、Storm、Hive、Pig、Oozie、Mahout、R语言等不同的大数据应用组件。

软件分类	软件名称
核心组件	HDFS
	MapReduce2
	YARN
	ZooKeeper
	HBase
基于MapReduce的数据分析组件	Spark
	Tez
	Hive
	Pig
	Mahout
	Tajo
数据库类组件	Kylin
	Drill
	Accumulo
	Cassandra
BSP(Bulk Synchronous Parallel)计算框架	Phoenix
	Hama
	Giraph
部分常用处理框架	Storm
	Reef
	Ignite
	Flink
工作流组件	Oozie
	Falcon
	Flume
ETL(Extract, Transform and Load)类组件	Kafka
	Chuwka
	Sqoop
序列化与持久化	Avro
	Gora
	Parquet
安全性组件	Knox
	Sentry

2 DeepRack深度学习一体机

深度学习成为风口，国内外高校为未来战略布局卡位

- ◆ 采用两种深度神经网络的AlphaGo已具有职业水准，仍在大量积累自对弈实现的深度学习以训练和提高其围棋实力
- ◆ 人工智能成B2O创新关键词，人工智能代表未来，任正非在会上表示“人类创造的新人类有可能取代我们真人类...”
- ◆ 计算机视觉、文本分析和语音识别等技术不断突破，谷歌、微软、IBM、Facebook、百度等科技巨头抢占技术制高点
- ◆ 麻省理工学院、牛津大学、加州伯克利大学、斯坦福大学等国外名校已抢占先机，其大数据挖掘与人工智能专业博士毕业生平均年薪可达到250万-300万美金



高校深度学习研究面临挑战

- ◆ 缺乏强劲的硬件支撑
 - 1) 深度学习网络模型复杂、计算量大
 - 2) GPU是提升计算性能的最关键选择，但不是所有GPU都适合深度学习
- ◆ 软件系统与应用项目难匹配
 - 1) Caffe、TensorFlow、MXNet等深度学习框架有着不同侧重
 - 2) 选择与研发项目匹配的软件框架耗时耗力
- ◆ 无法获取必备的训练数据
 - 1) 训练高准确率的模型需要海量训练数据
 - 2) 无法获取大规模数据成为应用者的瓶颈
- ◆ 缺少可靠的技术支撑
 - 1) 软件系统依赖于底层库，其部署与正常运行受多重因素影响
 - 2) 复杂的算法、频繁的参数调节等技术难题缺少可靠的技术支持

踏平高校深度学习研究的高门槛，DeepRack深度学习一体机震撼上线

DeepRack深度学习一体机作为云创大数据全力打造的深度学习软硬件平台，可为高校科研提供硬件、软件、数据以及技术等多方位支持，解决深度学习应用中遇到的障碍与困境。



◆ 硬件高规格

- 1) 服务器CPU选用最新的英特尔E5-2600系列至强处理器
- 2) 每台服务器最多可插入4块英伟达GPU卡
- 3) 提供最大每秒176万亿次的单精度计算能力
- 4) 四台服务器满配时相当于160台服务器的计算能力

◆ 软件国际化

- 1) 集成深度学习领域很流行的开源软件框架——Google开源的TensorFlow以及加州大学伯克利分校开发的Caffe
- 2) TensorFlow支持CNN、RNN和LSTM算法，这都是目前在Image, Speech和NLP很流行的深度神经网络模型
- 3) Caffe是纯粹的C++/CUDA架构，支持命令行、Python和MATLAB接口，可以在CPU和GPU64直接无缝切换

◆ 数据够丰富

提供大量ImageNet图片、人脸图片、环保数据用于模型训练

◆ 服务接地气

- 1) 从部署到使用提供完整的服务体系，参照操作手册就可以使用
- 2) 提供免费培训，帮学校打造一支能够应用深度学习的团队

详细配置参数

DeepRack深度学习一体机具有以下四种规格：极简型、经济型、标准型和增强型。参照NVIDIA GPU参数表，用户可选配合适的GPU型号，如NVIDIA TITAN X, GeForce GTX 1080、K80等。



	极简型	经济型	标准型	增强型
CPU	Dual E5-2620 V4	Dual E5-2620 V4	Dual E5-2650 V4	Dual E5-2697 V4
GPU	GeForce GTX 1080*4	Nvidia Titan X *4	Tesla K80 *2	Tesla K80 *4
硬盘	240G SSD+4T 企业盘	240G SSD+4T 企业盘	480G SSD+4T 企业盘	800G SSD+4T*7 企业盘
内存	64G	64G	128G	256G
计算节点数	1	4	4	4
单精度浮点计算性能	36万亿次/秒	176万亿次/秒	64万亿次/秒	128万亿次/秒
系统软件	全套Caffe、TensorFlow深度学习软件、样例程序，大量免费图片数据			
是否支持分布式深度学习系统	否	是		

GPU型号	GeForce GTX 1080	Nvidia Titan X	K80
Peak single precision floating point performance	9 Tflops	11 Tflops	8.73 Tflops
CUDA cores	2560	3584	4992
Memory size	8 GB	12 GB	24 GB

3 云创网盘高校应用方案

大数据催生云端教育，高校云平台建设面临“围墙”

- ◆ 云端教育强调共享优质教学资源，提供个性化学习空间
- ◆ 目前云平台建设广泛应用的公有云盘存在风险
 - 1) UC网盘、金山快盘等公有云盘相继关停，公有云盘面临关停风险
 - 2) 出于安全考虑，高校各种关键材料并不适合放在公有云盘

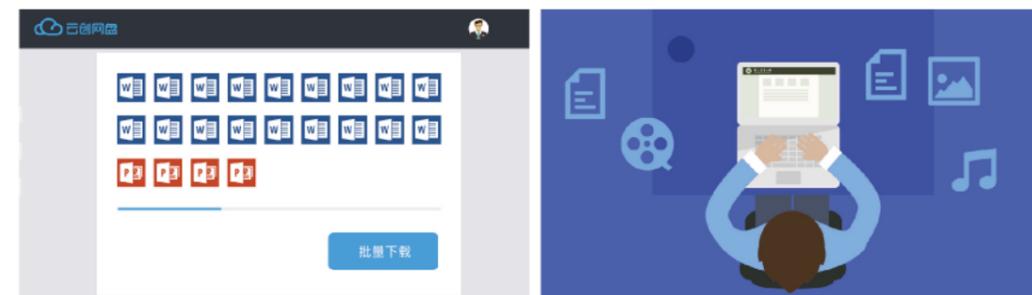


云创网盘——为高校量身定制的私有云盘，畅享公有云的便利与私有云的安全

- ◆ 优质资源班班通——与学校架构无缝对接，实现走动式管理
 - 1) 按照不同专业、班级和院系建立群组，方便快速选择用户分享课件资源



- 2) 创建公有群空间汇总作业，学生在家即可交作业，方便老师批量下载



- ◆ 网络空间人人有一专属的个人存储空间，你的地盘你做主
 - 1) 专属个人存储空间，校内访问不计流量
 - 2) 轻松上传下载照片、视频、文档、音乐、通讯录，对文件实现统一管理



- 3) PC端、网页端、手机端、平板等多终端同步文件，方便随时随地查看文件



- ◆ 分享动态，方便老师与学生随时随地互动
 - 1) 照片、电子书、电影一键分享，一秒拉近师生距离 (晒晒生活中的新鲜事，分享勤快的小伙伴还可获评分享达人)



