

2016 年贵安 IBM 大数据学院置换课程

一、“大数据技术在主要行业的应用” 公选课程介绍

2014 年,《贵州省大数据产业发展应用规划纲要(2014-2020 年)》发布,贵州将成为西部地区重要的、全国有影响力的战略性新兴产业基地。为了达成规划,贵州省将需要更多了解各行业的大数据人才,也更需要了解大数据的各行业精英。大数据应用更是无处不在,这不仅仅是程序员与科学家的专利,也是每个身在大数据时代的天之骄子应该了解和掌握创新之道。大数据作为中国官方重点扶持的战略性新兴产业已逐步从概念到落地,90%的企业都在应用大数据。

2016 年下半年,坐落在我校的贵安 IBM 大数据学院,将提供课程“大数据技术在主要行业的应用”的公选课程。本课程将介绍在汽车及制造业,石油化工业,能源业,零售业,通信业,银行业,保险业,高新技术行业,交通旅游业,医疗保健业以及政府行业中的大数据应用的特点,现状,前景以及实际案例,帮助大家走进大数据,站在时代的前沿,抓住发展的热点,成就更好的自己。

教学大纲

章节	章节名称	摘要	学时	目标
第 1 章	大数据分析概述	本章节将用通俗易懂的语言来介绍大数据分析原理以及主流技术框架	4	了解
第 2 章	大数据分析应用-汽车及制造业	本章节介绍大数据在汽车及制造业的应用以及案例	2	掌握
第 3 章	大数据分析应用-银行业	本章节介绍大数据在银行业的应用以及案例	4	掌握
第 4 章	大数据分析应用-石油化工	本章节介绍大数据在石油化工的应用以及案例	2	掌握
第 5 章	大数据分析应用-能源业	本章节介绍大数据在能源业的应用以及案例	2	掌握
第 6 章	大数据分析应用-政府行业	本章节介绍大数据在政府行业的应用以及案例	4	掌握
第 7 章	大数据分析应用-保险业	本章节介绍大数据在保险业的应用以及案例	2	掌握
第 8 章	大数据分析应用-零售业	本章节介绍大数据在零售业的应用以及案例	2	掌握

第 9 章	大数据分析应用-通信业	本章节介绍大数据在通信业的应用以及案例	2	掌握
第 10 章	大数据分析应用-交通旅游	本章节介绍大数据在交通旅游的应用以及案例	2	掌握
第 11 章	大数据分析应用-医疗保健业	本章节介绍大数据在医疗保健业的应用以及案例	2	掌握
第 12 章	大数据分析回顾与展望	本章节将回顾大数据在各行业的应用，并展望其在未来发展的趋势。	2	了解

二、“商业智能基础知识及数据仓库设计” 公选课程介绍

大家知道人民银行他们每天的业务量产生的数据可以刻满 3.7 亿张光碟吗？大家知道一家银行，一个保险公司，一家上市公司他们这些珍贵的数据是怎样存储的吗？数据库呀，大家也许会这样想。有一个说法，数据库就像是一个市场，所有的菜都摆在一起，而数据仓库就像是一家超市，蔬菜是一类，水果是一类，零食又是一类，而这样的“分门别类”，在银行业等各个领域的重要性以及应用，大家这门课上就能有福的听到啦。

数据仓库这家大型超市是怎么对它的“产品”进行分类的——这关键的数据建模的知识，也会给大家有所涉及，小伙伴们准备好“健胃消食片”呀，这门全是干货的课，一定会让大家欲罢不能收获颇多的。

大量货物堆在仓库里，海量的数据放在哪里呢？没错，就是数据仓库。

数据仓库，是企业所有级别的决策制定过程，提供所有类型数据支持的战略集合，为需要业务智能的企业，提供指导业务流程改进、监视时间、成本、质量以及控制。

本课程主要介绍商业智能，及数据分析的数据需求知识，以及如何设计和建立高效、可持续发展且可扩展的数据仓库的方法，并介绍建立各种数据模型的方法。

教学大纲

章节	章节名称	摘要	学时	目标
第一章	概论	数据仓库的基本概念，数据仓库与传统数据库的区别，数据建模基本知识	2	了解
第二章	数据建模	与数据建模的概念以及数据模型的流程	7	掌握
第三章	关系型数据模型	关系型数据模型以及建模过程	7	掌握
第四章	维度数据模型 1	维度数据模型概念，雪花模型，星型模型概念	7	掌握
第五章	维度数据模型 2	维度数据模型建模过程	7	掌握

三、“软件测试的艺术”公选课程介绍

甲：说一个包工队接一好活儿，盖一个 70 多米高的大烟囱！

乙：还真不错！

甲：起早贪黑把活干完了，人家来一验收，死活不给我们工钱，还叮当五四打一顿！

乙：为啥啊，质量不行？

甲：把图纸拿倒了，人家让挖口井！

我只想说，如果你有一个做测试的好基友，分分钟烦得你挖出一个完美的深井
作为一对好基友，开发和测试在漫长的纠葛中，相爱相杀。

测试是一种技术。然而事实上，测试是一种哲学，一种思想，思想的背后是一个人的眼界和世界观。做一个傲娇全能的好测试，不遗巨细，保持对开发的全面碾压。做一个懂测试的好开发，在弱势中逆袭，用完美的代码征服世界。

实践是验证真理的唯一途径，测试也是验证代码的唯一途径。通过软件测试，我们可以促进鉴定软件的正确性、完整性、安全性和控制软件质量。

本课程主要通过实际案例展示测试的重要性，并介绍通过人工操作或程序运行进行软件测试的各种方法以及步骤。

教学大纲

章节	章节名称	摘要	学时	目标
第 1 章	绪论	本章节将用通俗易懂的语言来介绍软件测试的背景，以及实质	2	了解
第 2 章	测试基础	如何解读产品说明书形成测试用例，代码检查，黑白盒测试	4	掌握
第 3 章	测试技术	配置测试，兼容性，安全性，易用性测试，以及测试文档	4	掌握
第 4 章	补充测试	缺陷轰炸与 beta 测试	4	掌握
第 5 章	测试文档	计划测试工作，编写和跟踪测试用例，报告发现的问题，测试成效评价	4	掌握
第 6 章	自动化测试	自动化测试及工具	3	掌握
第 7 章	敏捷测试	基于敏捷开发的面向对象测试过程	3	掌握

四、“SQL 高级应用及数据库高级设计与实例讲解” 公选课程介绍

我有一个衣柜，数了一下，里面有 15 件上衣，8 条裤子，它们中有 2 件红色，2 件紫色，3 件蓝色，4 件黄色，4 件绿色，4 件白色，4 件黑色，我要拿出蓝色上衣，白色裤子，正好配新买的棕色皮鞋。数以亿计的数据放在数据库里，就像放到衣柜里的衣服，一条条数过去，数完我的头发也就白了啊，更别提精准找到符合需求的数据。这时候，SQL 语言作为我们的好帮手，让你不再发愁，一键解决所有烦恼。来来来，让我们牵起 SQL 语言的小手，让友谊的小船永远向前。

数据作为大数据应用的灵魂，怎么样存储数据，读取数据也就成为了重中之重。

本课程主要介绍关系型数据库的基本原理、逻辑实现、数据存取方法，SQL 语言的高级应用，包括 SQL 调优，数据库配置优化，以及数据中心设计的基础知识。培养学生对数据库系统的设计、分析与操作能力，为数据库相关的编程打下基础。

教学大纲

章节	章节名称	摘要	学时	目标
第一章	概论	关系型数据与 sql 语言的发展史，数据库的工作原理，当前主流数据库的逻辑结构，数据库到数据仓库的演变历程。	4	了解
第二章	查询与子查询	限定性查询与数据排序，单行函数的使用，多表连接查询，组函数与分组统计，子查询与嵌套子查询	6	掌握
第三章	数据操纵	如何对数据库进行数据插入，数据更新以及数据删除，事务的概念，事务的提交与回滚	6	掌握
第四章	数据库对象定义	数据库有什么对象，如何创建数据库对象，不用的数据库对象的删除，create/drop table，create/drop view，create/drop index，为表添加约束	6	掌握
第五章	数据库对象访问权限控制	数据定义语言 grant，数据库的用户概念，组以及角色概念，如何对数据库授予访问权限	4	掌握
第六章	查询分析器与执行计划	什么是查询分析器，什么是执行计划，查询分析器的使用，如何生成以及查看 sql 的执行计划，如何进行 sql 调优	4	掌握

五、“ETL 原理及实现” 公选课程介绍

如果说数据仓库的模型设计是一座大厦的设计蓝图，数据是砖瓦的话，那么 ETL 就是建设大厦的过程，任务繁重，却不可或缺。

怎样去解读这三个字母呢？经典解释是 Extract, Transform, Loading，表示数据的抽取，转换，和加载。大家知道银行有多少的数据量吗？而怎样生动而合理的安排这些珍贵的“砖瓦”，是每个做 ETL 的数据工程师最重要的事情。其实 ETL 三个字母的还有另一种解读，Essential, Technological, Lovely，一群由可爱的人做的重要的工作。如果你想对它有更深的了解，那么本课程就详细的介绍了 ETL 的原理以及 IBM 主流的 ETL 工具 InfoSphere DataStage，带大家走进 ETL 的世界，获益匪浅。

ETL 是数据仓库建设中数据抽取、转换、装载到模型的过程。作为大数据分析的数据准备阶段，数据的优质处理将为分析结果打下坚实的基础。

本课程介绍了 ETL 的工作原理，设计开发实现 ETL 过程的基础知识以及如何应用 IBM 主流的 ETL 开发工具 InfoSphere DataStage 进行可视化 ETL 过程开发。

教学大纲

章节号	章节名称	摘要	学时	目标
第一章	概论	ETL 的定义介绍，ETL 的一些前提的原则，ETL 的模式比较，ETL 的同构与异构	2	了解
第二章	ETL 原理与方法论	ETL 的过程，什么是数据的数据的抽取、清洗、转换与加载，ETL 开发设计规范，ETL 开发的异常处理，以及常见的问题诊断	4	掌握
第三章	DataStage 产品介绍	DataStage 开发工具介绍，DataStage 产品的安装与部署，介绍 DataStage 的各个功能组件	2	掌握
第四章	Administering DataStage	datastage 管理控制台介绍，介绍 datastage 的用户和组以及用户和组的创建方法，介绍如何给用户和组分配角色以及权限，用户认证管理，datastage 的工程管理，datastage 的环境设置	2	掌握
第五章	DataStage Designer	介绍如何登陆 datastage designer, designer 工作区的介绍，怎么设计 datastage 的 jobs, 导入和导出 datastage 对象，导入和导出表定义。	2	掌握
第六章	创建并行的 job	掌握什么是 Parallel jobs,怎么在 job 中加入 stage 和链接不同的 stage,了解 peek Stage 的作用和里面要设置的选项的意义,掌握 job 里面参数的设置,了解如何编译并运行做好的 job,怎么在 Director 察看运行的日志和 job 运行的状态,掌握 Director 的其它功能和设置 Director,介绍怎么在命令行中运行完成的 job	2	掌握

第七章	Accessing Sequential Data	datastage 数据文件的类型，怎么访问 Sequential 数据及此数据的特点，介绍 Sequential File Stage 的规则，Sequential Stage 数据源的一些的设置，介绍 Data Set 数据类型，在 job 中怎样用 DataSet Stage 读取 Data Set 数据，Data Set Management 的功能。	2	掌握
第八章	Platform Architecture	学习并行的结构，描述管道的并行度是什么，描述并行聚集的算法，描述 Parallel job 的配置文件，介绍 datastage 下调 OSH 脚本	2	掌握
第九章	Combining Data	掌握怎么制造 Combining Data 数据，了解 Lookup Stage 的特点以及类型，lookup Stage 的例子，lookup 由于不能匹配关键列时，可以采取的行动，如何设计有范围的 lookup job，了解 join Stage 以及设置，了解 Merge Stage 以及设置，比较 joins,lookup,Merge Stage，了解 Funnel Stage。	2	掌握
第十章	Sorting and Aggregating Data	介绍了如何用想要的 stage 做数据排序与聚集处理	2	掌握
第十一章	Transforming Data	介绍了如何用 stage 进行数据的转换，在 stage 中如何使用变量，如何定义约束	2	掌握
第十二章	Repository Functions	Stage 内置功能介绍，如何进行查找，如何使用高级查找，怎么比较 2 个表的定义，怎么比较写好的 2 个 job	2	掌握
第十三章	Working With Relational Data	介绍 datastage 的数据源的类型，怎么创建不同类型的数据源，各个数据源 stage 的区别及其使用，通过数据源来导入表定义	2	掌握
第十四章	Job 调度	介绍了如何使用 job 调度器去创建 job 以及控制 job，如何在不同的 job 之间用变量传递信息，如何定义用户变量，job 的启动与重启，job 的错误与异常处理	2	掌握

六、“数据可视化原理及操作” 公选课程介绍

给你讲个笑话，如果每天我们面对的是四方格子里的 abc123 的话，那我们每天需要看的书就不只是《C 语言编程指南》，而是《颈椎病康复指南》和《近视眼康复指南》了。

你也许用过柱状图或者饼图来展示 2016 年班级成绩分布情况，可如果有一天我们手边的数据数量动辄千万记，大小动辄以 TB 为单位，我们就需要更精准，更多元的方式来展示数据，更何况，这一天早就已经来了。

想象一下用肉眼去看千万条数据，是不是想想就觉得身体被掏空，哈哈，这就是数据可视化的意义所在，在大数据充斥的今天，让数据接收更加直观，让数据分析更加便捷。美其名曰大数据，那海量的数据必不可少，如何把数据库里面的海量数据以及大数据分析结果易读有趣的呈现给读者？这就需要数据可视化解决方案。

本课程会更详尽的介绍数据可视化的更多干货，通过 IBM COGNOS BI 所带来的数据可视化解决方案，帮助大家认识这个大数据的时代，别再葛优躺啦，让我们一起抓住大数据峰会的余温，走进数据可视化的大门吧。

IBM Cognos 作为商业智能核心平台上，以服务为导向进行架构的一种数据模型，给我们数据可视化提供完美的解决方案。本课程介绍数据可视化的设计开发实施知识，以及如何运用 IBM Cognos 实现数据可视化解决方案

教学大纲

章节号	章节名称	摘要	学时	目标
第一章	概论	简单介绍商务智能，数据可视化的基本概念。IBM Cognos BI 概述，架构，数据流，开发流程介绍，Cognos 家族成员介绍。	2	了解
第二章	Framework Manager	介绍了 Metadata model 的设计过程，常用数据结构的特点，Mapping 过程中关系基数的设置。	6	掌握
第三章	Report Studio1	介绍了 Reprot Studio 的用法，如何创建 list report	4	掌握
第四章	Report Studio2	介绍 Cognos Crosstab Report 的使用方法	6	掌握
第五章	Report Studio3	介绍 Cognos 报表里面的 Gauge Chart	6	掌握
第六章	Report Studio4	介绍 Report Studio 里面的过滤与按条件格式化以及变量等得使用	6	掌握

七、“大数据项目实训”公选课程介绍

不以结婚为目的的谈恋爱是要流氓！

不以实践为目的的学理论，当然也是要不得的！

本课程采用真实项目需求为基础，从数据业务分析，数据建模，ETL 数据整理，到最后的数据可视化实施，按照实际项目开发实施流程，帮助大家练习理解所学大数据相关知识，达到融会贯通。本课程采用实际项目需求为基础，从数据建模，ETL 数据整理，数据分析，到最后的数据可视化，按照实际项目开发实施流程，帮助学生练习理解所学大数据相关知识，达到融会贯通。

教学大纲

章节号	章节名称	摘要	学时	目标
第一章	项目实训需求	描述项目需求，提供原始数据	90	掌握