**目录**

[1.引言 1](#_Toc491679881)

[1.1 实习目的和要求 1](#_Toc491679882)

[1.2 实习内容 1](#_Toc491679883)

[1.3 实习单位概况 2](#_Toc491679884)

[2.专题讲座心得 3](#_Toc491679885)

[2.1“大数据分析：问题、技术及挑战”专题讲座 3](#_Toc491679886)

[2.2“工业控制网络安全空间”专题讲座 4](#_Toc491679887)

[2.3“下一代互联网：机遇与挑战”专题讲座 4](#_Toc491679888)

[3.北京天河自行车销售模块数据分析 5](#_Toc491679889)

[3.1 项目概述 5](#_Toc491679890)

[3.1.1 公司背景 5](#_Toc491679891)

[3.1.2 项目背景 5](#_Toc491679892)

[3.1.3 业务需求 5](#_Toc491679893)

[3.1.4 主要目标 6](#_Toc491679894)

[3.1.5 功能要求 6](#_Toc491679895)

[3.2 总体设计 6](#_Toc491679896)

[3.3 数据源分析 7](#_Toc491679897)

[3.4 数据仓库设计与实现 7](#_Toc491679898)

[3.4.1 总体设计 7](#_Toc491679899)

[3.4.2 详细设计 8](#_Toc491679900)

[3.5 ETL的开发与设计 12](#_Toc491679901)

[3.5.1 表导入过程 12](#_Toc491679902)

[3.5.2 将数据导入表 17](#_Toc491679903)

[3.6 OBIEE的开发与设计 21](#_Toc491679904)

[3.6.1 准备数据（将数据从数仓导入OLAP） 21](#_Toc491679905)

[3.6.2 报表开发过程 22](#_Toc491679906)

[3.6.3 分析报表 27](#_Toc491679907)

[4. 实习总结 35](#_Toc491679908)

[4.1 实习内容的复杂性评价 35](#_Toc491679909)

[4.2 实习体会、收获与建议 36](#_Toc491679910)

[5.参考文献 37](#_Toc491679911)

[附录 实习日记 38](#_Toc491679912)

# 1.引言

## 实习目的和要求

(1) 实习目的：

专业生产实习是本科教学计划中非常重要的实践性教学环节，是专业培养目标和教学计划、课程设置的有机组成部分，是理论教学的完善和补充。通过生产实习，使学生通过实践了解和掌握计算机的应用的开发、设计、运行等环节。通过生产实习培养学生掌握从事计算机系统开发的技能，加深对已经学习过的专业理论知识的理解和认识，提高在生产实践中调查研究、观察问题分析问题以及解决问题的能力和方法。为学生后续学习打下基础。

(2) 实习要求：

邀请专家做专题报告，了解计算机专业各领域发展方向和前景，主要研究内容和应掌握的理论与技术。

参观东软集团，了解东软主要发展历程，企业发展方向。学习他们的创业、敬业精神。

了解商务智能概念以及所涉及的学科知识，练习并熟练使用SQL语句，掌握ETL技术，初步接触PL/SQL、Informatica、OBIEE并在指导下完成一个项目，最终独立完成销售模块的分析。

要求学生运用数据库等相关软件，对源数据进行分析，建立数据仓库。并完成对数据的提取、转换以及加载，将转换过后满足指定格式的数据加载进数据仓库，数据仓库会周期不断地从源数据库提取清洗好了的数据。同时建立能够直接访问数据仓库功能的前端应用，即符合逻辑，界面友好人性化的报表。要充分利用网上的现有资源，学习并熟练应用SQL语句以及各种开发工具，尤其是OBIEE；以小组为单位，互相协作，分工明确；严格出勤；认真完成生产实习报告，并提交程序。

## 实习内容

本次实验的实习时间、地点和相应内容如表1.1所示：

表1.1 实习内容

|  |  |  |
| --- | --- | --- |
| **实习时间** | **实习地点** | **实习内容** |
| 7.3 | 信息学馆A205 | 了解商务智能 |
| 7.4-7.9 | 信息学馆A205 | 练习使用PL/SQL,熟练掌握SQL语句 |
| 7.5上午 | 生命学管A101 | 专题讲座 |
| 7.10-7.16 | 信息学馆A205 | 分析数据源，设计并建立数据仓库 |
| 7.11上午 | 生命学管A101 | 专题讲座 |
| 7.16上午 | 生命学管A101 | 专题讲座 |
| 7.17-7.23 | 信息学馆A205 | ETL的开发与设计，并在指导下完成简单项目熟悉流程 |
| 7.18上午 | 东软集团 | 参观 |
| 7.24-7.30 | 信息学馆A205 | 对销售模块进行实战演练分析，并完成前段报表开发与设计 |

**1.3实习单位概况**

东北大学计算机学院实验教学中心简介：东北大学计算机学院实验教学中心承担着计算机学院计算机类专业和电子信息类专业的实验教学任务。此外，还承担各类课程设计、生产实习和毕业实习等实践教学任务。2015年东北大学计算机实验教学中心入选国家级实验教学示范中心。

实验教学中心设有面向计算机科学与技术、物联网工程、通信工程和电子信息工程4个专业的10个实验室。其中，有面向计算机硬件的计算机网络、计算机组成原理、接口技术和嵌入式的2个硬件实验室，面向计算机软件的2个实验室，面向通信工程专业的3个实验室，面向电子信息工程的2个实验室（微控制器/微处理器实验室），面向物联网专业的1个实验室。

实验教学示范中心现有实验用房约3084平方米，仪器设备3400台套，其中台式计算机2100余套。中心现有实验技术人员26人，其中，高级实验师8名，实验师2名；具有硕士学位的11人，学士学位的14人。实验课程设置紧密围绕本科生培养计划制定的课程体系，以培养学生复杂工程问题的解决能力和提高动手能力为主旨的实验课程项目体系建设。

# 2.专题讲座心得

在生产实习期间，我们聆听了多位专家学者以及企业主管的讲座和报告。对此，感谢学校让我们有机会现场聆听各位专家学者对于计算机专业各领域的研究成果和个人见解，同时给我们提供了一个学术交流的平台，这些演讲将我们平时的学习成果和现实中的应用结合在一起，使我们受益颇多。

首先，是于戈教授的大数据分析：问题、技术及挑战专题讲座；接着是姚羽教授的工业控制网络安全空间专题讲座；然后是王兴伟教授的下一代互联网：机遇与挑战专题讲座。以下是我听取这些讲座的心得。

## 2.1“大数据分析：问题、技术及挑战”专题讲座

大数据分析给现代社会带来了新的机遇与挑战。讲座中，于戈教授用生动形象的语言和简单易懂的事例为我们介绍解释了各大数据发展已经现状各种问题。

一方面，与传统研究侧重于揭示事物的共性不同，大数据研究将有助于人们发现事物的个体特性，并针对每一个体的特性给出个体化的解决方案。同时，大数据研究也将使人们能够从大量个体的差异变化中，揭示其中存在的难以察觉的规律。另一方面，大数据的海量样本规模和高维数特征也引入以下显著特性：数据搜集的偏差性、数据产生的异母体性、计算成本、噪音的累积叠加、假关联性、外生性，以及测量误差等等。为了应对这些挑战，需要引入新的计算和统计方法。

对于现代社会的发展、人们生活的便利性而言，大数据的作用都得到了各行各业的认可，大数据本身并不是一个具体的东西，在不同的行业，大数据起到的作用也是不一样的，因此，在不同的行业背景下，对于大数据的定义也不相同。对于大数据的定义，也不是一蹴而就的，随着时代的发展，对于其定义也一直在补充完善。作为新时代计算机人，我们应该要不断努力。

其发展到今天，覆盖的范围越来越大，在各行各业产生的影响也越来越大，发展的速度在不断地加快，但是，从整体上来看，目前大数据的发展还是处于初期阶段，对于大数据的研究还要不断的进行深入，从基础理论上要进行深入，一些重要的技术上，进行深入，如何运用大数据上也要深入研究。路漫漫其修远兮，吾将上下而求索！

这次讲座让我更深入的了解了大数据，感谢于戈老师。

## 2.2“工业控制网络安全空间”专题讲座

最近几年网络安全问题越来越严重，甚至是是国家重要的基础设施受到了威胁，这个对于社会的安定和社会的进步会产生非常大的影响。姚羽教授为我们举了大量生动的例子，如伊朗核试验、斯诺登事件，从中学到了很多。

互联网的安全挑战是非常多的，例如源地址不做认证，路由会产生非常多的问题。另外，大规模的域名劫持和假冒，发现突然去了不希望去的网页，这些都是域名劫持产生的，国家安全法公布以后，很多软件提供商有所收敛。再比如，主干网上信息路由的劫持，也是对运营商有非常大的危害。

为什么互联网有这么多安全问题？根本的原因是互联网是个开放的网络，在这么大的开放网络中，所有访问是不可信的，你的每一个分组访问，对你的源地址不验证。如果我们能解决这个问题，我相信互联网的安全等级会大大提高。

并不是有了物理隔离以后就安全了，虽然确实可以降低网络安全风险。但是只要采用了TCP/IP的体系结构，入侵畅通无阻，和公网的安全势态是一样的，采用统一的互联网体系结构是没有安全的。

我国2014年成立了网信办，把网络强国作为我们国家发展互联网的战略目标，习总书记的4·19讲话特别谈到要尽快在核心技术上取得突破，这些对信息化和网络安全工作都有非常重要的指导意义。美国对安全也是非常关注，发表了一系列计划和行动纲要，恐怖分子不断利用互联网和其对抗，美国希望全球合作解决这个问题。

通过这次专题讲座，我对网络有了更深入的了解。我会更加学好专业知识，拓展自己的知识面，让自己能更好地适应时代的潮流。

## 2.3“下一代互联网：机遇与挑战”专题讲座

下一代互联网是一个建立在IP技术基础上的新型公共网络，能够容纳各种形式的信息的综合业务网络。1996年、1997年IPv6诞生以后，很多国家都发起了下一代互联网研究计划，他们想在新的网络空间里掌握更多的主动权。对于我们国家来说更加迫切，因为我国早期由于技术上的落后，没有申请到大量的IP地址，很多单位都是用私有地址，出口上再用公有地址转换，有了IPv6协议给我们带来了非常大的机会。

王兴伟教授提到和IPv6IPv4是一个改朝换代的机会，因为用了不同的传输格式以后，互联网很多的网络设备、网络应用都会发生重大的变化。如果在这个机会里面能够抓住的话，我们可以掌握更多的主动权，这是一个很好的机会。

　 但王教授也提到，我们国家在互联网技术使用上还是比较落后的，当然这个落后是由于种种原因产生的，一是互联网来我们国家比较晚。第二，技术落后，申请地址就会受限。像教育网地址非常丰富，但是国家整体上七亿网民只有三亿地址，平均两个人一个地址，在美国是一个人有五个地址。中国的互联网之路任重而道远。

王兴伟教授同时介绍了互联网+，大众创业万众创新的一些概念，受益匪浅。

# 3.北京天河自行车销售模块数据分析

## 3.1 项目概述

### 3.1.1 公司背景

北京天河自行车制造有限公司（简称天河自行车），是一家自行车生产制造及销售的企业，目前公司的业务已经覆盖了全国25个省市的超过1000家企业。

公司主营业务为自行车的制造及销售，公司的业务流程包括了，材料的采购、铸造、自行车成品及备件的加工、制造等多个环节。在中国自行车制造行业处于领军地位。

### 3.1.2 项目背景

公司自2010年采用ERP系统以来积累了大量采购、生产、销售、以及公司内部运营等数据，随着业务的不断扩充，数据量的增多，使各部门的数据分析也越来越复杂。随着公司业务发展的扩充以及向海外发展的长远目标，公司各阶层领导已经逐步意识到经验促使决策的片面性和不确定性。

信息部门结合公司的当前业务运营管理需求与健康扩张（做大的同时，更需要做强）的战略，系统地设计与开发各类关键基础运营（原始）数据的管理分析报表体系与相应的基础运营数据集中管理平台。

### 3.1.3 业务需求

天河自行车集团组织架构丰富、业务多样、业绩增长快速。公司决策层面及各业务负责人、部门负责人在决策管理方面的需求具有多样化，各部门业务间的需求差异性大，根据决策管理需要，天河自行车数据决策支持平台从各业务、部门等维度出发，将各部门具体业务需求模块化，在实施过程中分模块进行需求调研与分析。

### 3.1.4 主要目标

本期主要进行业务分析决策平台的开发，平台功能需达到使公司的供应商管理、门店绩效管理、各商品品类业绩考核建立在可靠的业务数据基础上，使公司的相关职能部门员工与管理者能够及时看到他们所关心的业务数据，从而为企业决策者正确决策提供准确、及时的依据，并使企业能更好地为供应商和顾客服务，进而提高企业的经营管理水平，提升企业整个供应链的竞争力。

### 3.1.5 功能要求

依据ERP系统中销售模块数据收集，以及公司销售订单业务流程，进行公司贸易整合。商贸系统要求如下：

(1) 销售订单

主要功能包括接收、更新、保存来自于ERP系统的销售订单数据，为后续销售业务提供基础数据。系统支持对销售订单的实时跟踪，以更好的把控整个销售流程。

(2) 产品库存

为了完成产成品出厂发运任务，支持对产成品信息的统一和全程管理。在物料管理模块的支撑下，收集生产厂厂内成品的产出信息，并且集成厂外库的仓库信息，形成统一的产成品库信息。成品库存及原料库存信息主要来源于生产管理系统的实时物料库存信息。

(3) 销售出库

为指导销售出厂，系统接收来自ERP的发货计划（即发货提单），并转化成商贸系统的物流任务，并由发货单位在物料模块中执行具体的发货任务。

(4) 销售结算

根据发货出库明细形成销售结算清单，作为销售结算需要结齐的单据，并且可形成出库明细汇总后的形式发票，上传给ERP系统完成销售结算流程。

**3.2 总体设计**

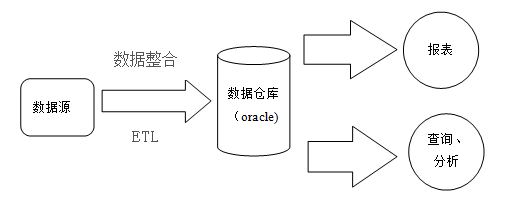


图3.1系统总体设计图

**3.3 数据源分析**

想要分析自行车销售数据，首先应该从分析数据源入手。该销售模块的数据源存放在SQL Server中，数据多而杂，所以我们要根据销售模块的需求从源数据库中查找和销售有关的分析数据。经研究发现：

在OM\_So\_Headers表中存有订单的基本信息，包括时间、销售客户、销售员等信息；

在OM\_So\_Lines表中存有订单的详细信息，包括销售产品、销售量、销售额、运输方式等信息，

其中HEADER\_ID和OM\_So\_Headers表中的ID相关联；

在CONSUMER\_Detail表中存有客户的详细信息，包括名称、所在地等信息，

在CONSUMER\_type\_New表中存有地区信息，

此外还有专门存储运输方式的表CONSUMER\_SENDPAY,

因此后续工作就以这几个表为基础展开。

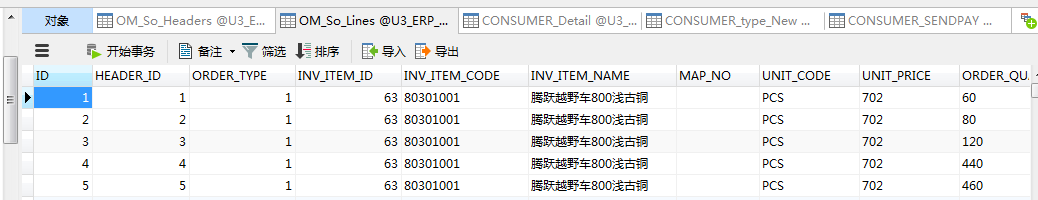


图3.2 数据源情况分析图

## 3.4 数据仓库设计与实现

### 3.4.1 总体设计

根据以上数据源分析，我们发现，要分析销售情况，可以以销售额和销售量为度量进行分析。围绕这两个指标，可以先从时间上来分析，从总体上观察每一年的销售变化情况。除了时间，还可以从地区来分析，观察不同地区的销售情况有什么差异，此外，不同的客户之间也会存在销售差异，还可以比较不同产品的销售好坏，最后从企业内部可以查看不同销售员的业绩。

因此，我们决定从时间、地区、客户、产品、销售员这五个维度进行分析。

表3.1 数据仓库表清单

|  |  |
| --- | --- |
| 名称 | 表名称 |
| 时间维度表 | TIME\_TABLE\_6 |
| 地区维度表 | LOC\_TABLE\_6 |
| 产品维度表 | PRODUCT\_6 |
| 客户维度表 | CUSTOMER\_6 |
| 销售人员维度表 | SALER\_6 |
| 事实表 | SALE\_TABLE\_6 |

### 3.4.2 详细设计

#### 3.4.2.1 时间维度表设计

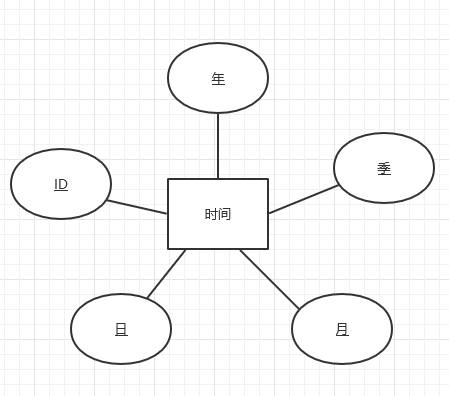


图3.3 时间维ER图

时间维度，通过时间维度我们可以比较不同年份、季度、月份的销售情况，来分析销售的淡旺季。

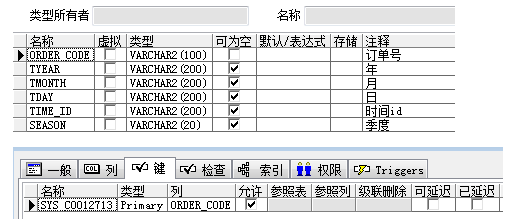


图3.4 时间维表结构图

#### 3.4.2.2 地区维度表设计

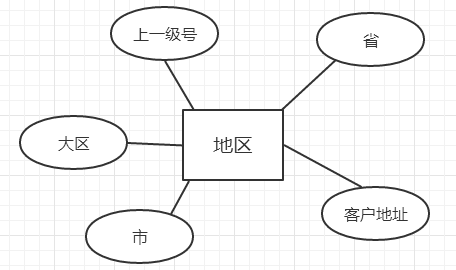


图 3.5 地区维ER图

通过地区维度我们可以比较不同地区、省、市的销售情况，来分析那些地方需求量大，进而该地区多投入生产，需求量小的次之。

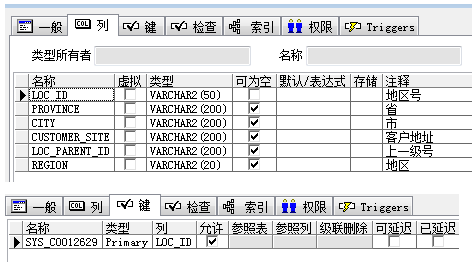


图 3.6 地区维表结构图

**3.4.2.3 产品维度表设计**

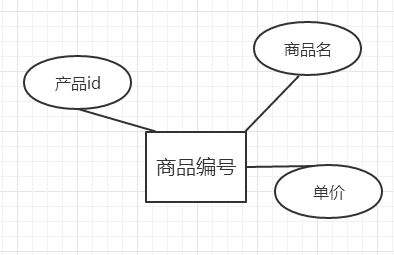


图 3.7 产品维ER图

通过产品维度表我们可以比较不同类型、型号、颜色的自行车销售情况，来分析什么车型卖的好，什么车型卖的差。

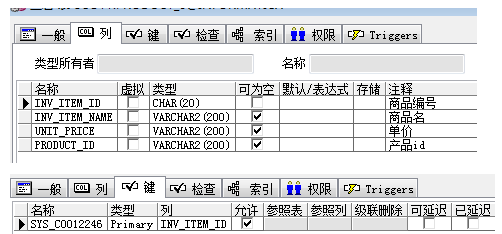


图3.8 产品维表结构

#### 3.4.2.4 客户维度表设计

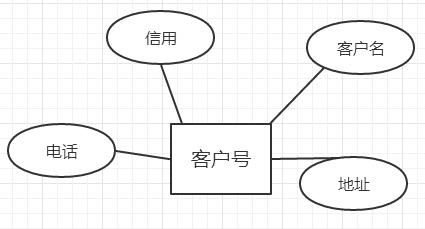


图 3.9 客户维ER图

通过客户维度表我们可以比较不同级别的不同客户的成交额，来分析那些客户是要重点发展的，哪些次之。

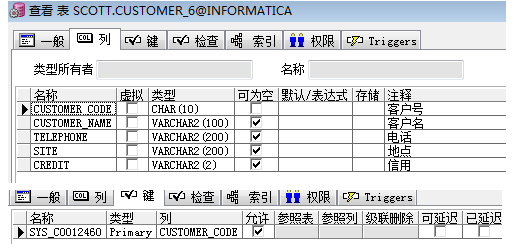


图3.10 客户维表结构

#### 3.4.2.5 销售人员维度表设计

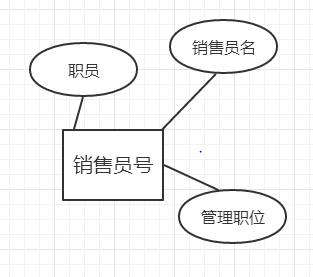


图 3.11 销售员维ER图

销售人员维度，这个维度可以评估销售人员的绩效。

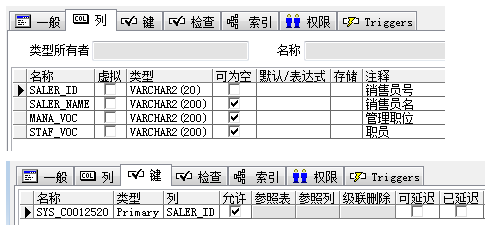


图 3.12 销售员维表结构图

#### 3.4.2.6事实表设计

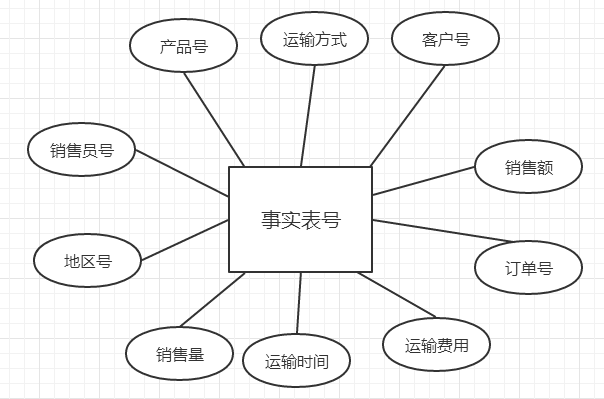


图 3.13 事实表ER图



图3.17 事实表表结构

## 3.5 ETL的开发与设计

本人在小组中承担地区维度的开发，故该部分主要对地区表进行操作。

### 3.5.1表导入过程

这次项目用到两个源数据库，ORACLE数据库和SQL数据库，以SQL数据库为例，介绍将SQL数据库中的地区表CONSUMER\_Detail导入到informatica PowerCenterDesigner生成源表CONSUMER\_Detail的过程（第一次导入需要添加新的数据源，以后就不需要了）

导入过程：

① 点击源，选择从数据库导入

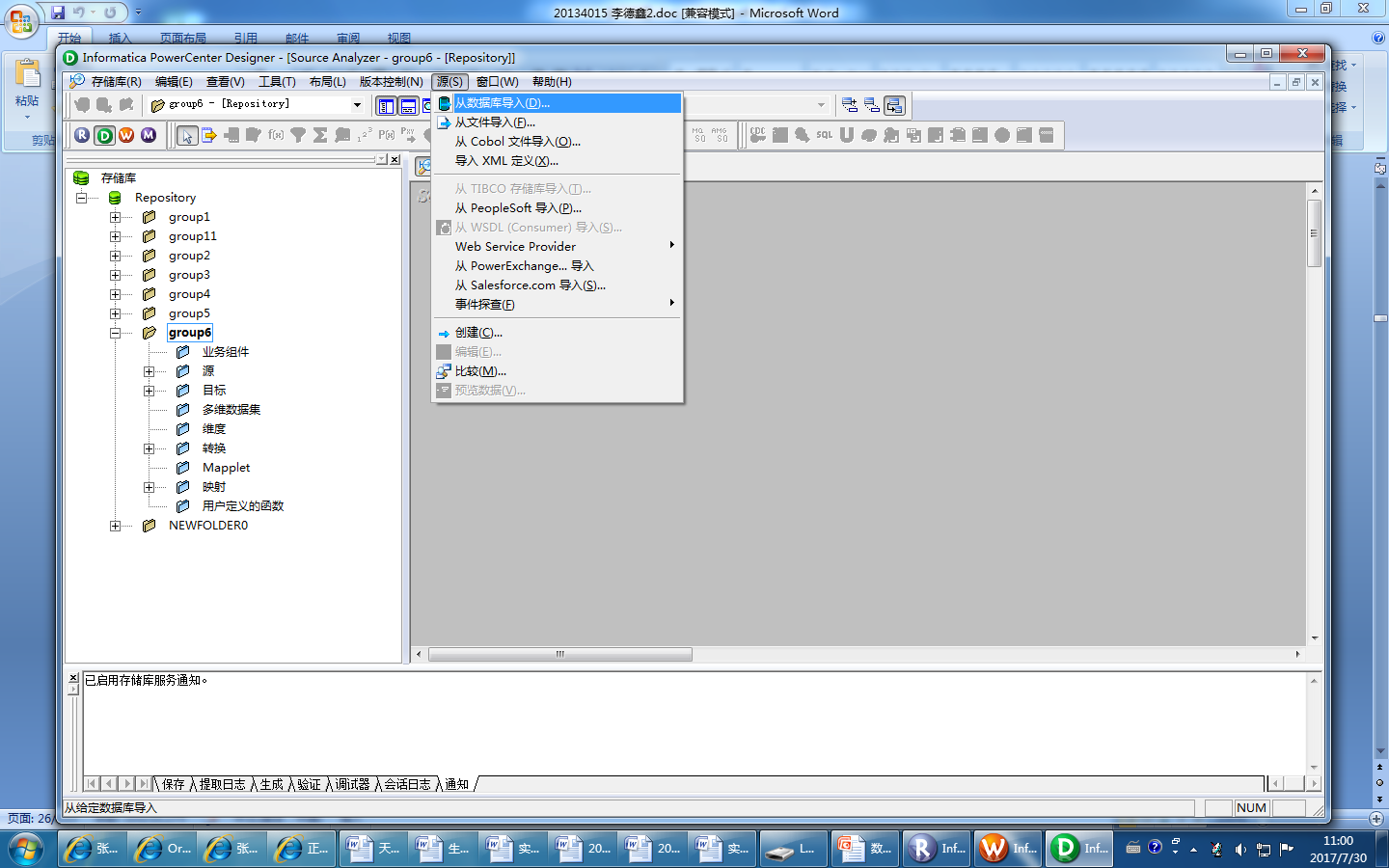


图3.18 选择源

② 选择SQL Server，点击完成

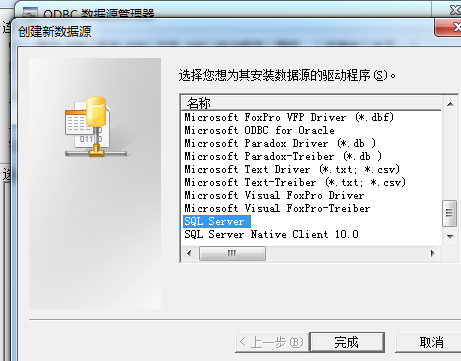


图3.19 选择数据库

③ 输入命名数据源的名称

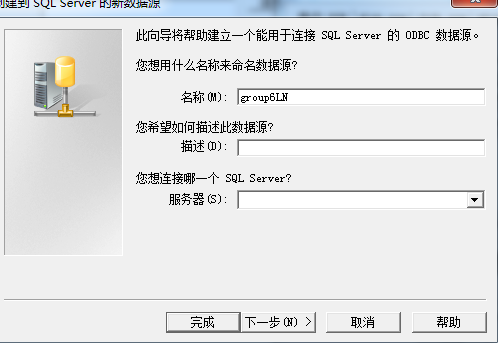


图3.20 命名

④ 验证方式选择“使用用户输入登录ID和密码的SQL Server验证（S）”

选择默认数据库U3\_ERP\_01并测试连接成功

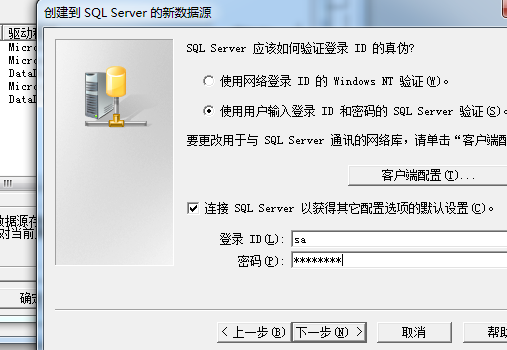


图 3.21 验证

⑤ 添加新建的源

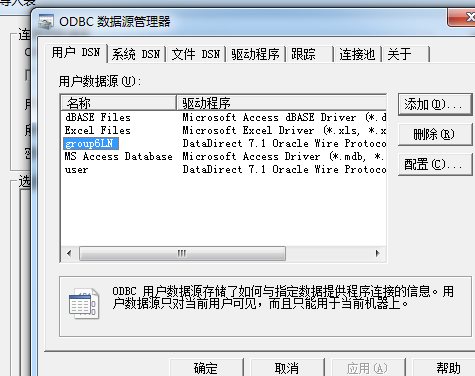


图 3.22 选择新建的源

⑥ 选择新添加的源，填写用户名密码，选择导入的表，点击确定

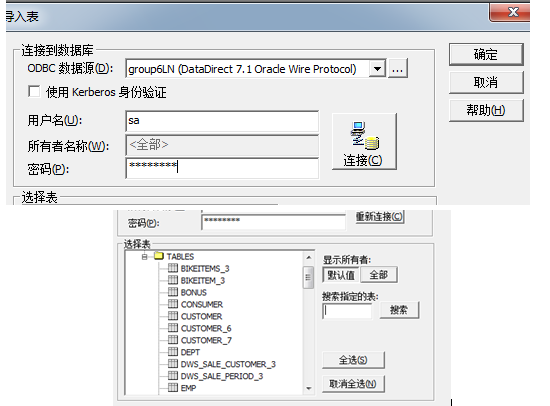


图 3.23 选择所需要的表

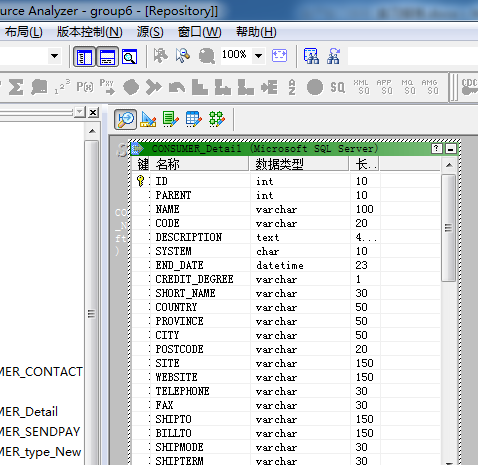


图3.24 导入地区表结果图

⑦ 将源表以及目标表依次导入

不再一一赘述，最终结果：

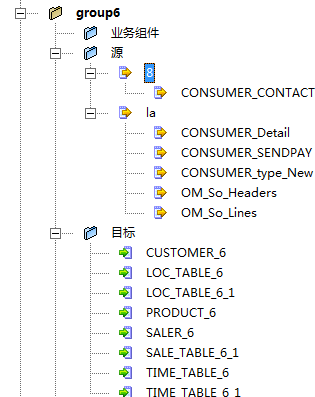


图3.25 最终结果图

### 3.5.2 将数据导入表

依然以地区表数据导入为例

①将SQL数据库中的地区表CONSUMER\_Detail导入到informatica PowerCenterDesigner生成源表CONSUMER\_Detail。将ORACL数据库中目标表LOC\_TABLE-6-1导入Informatica PowerCenterDesigner生成目标表。

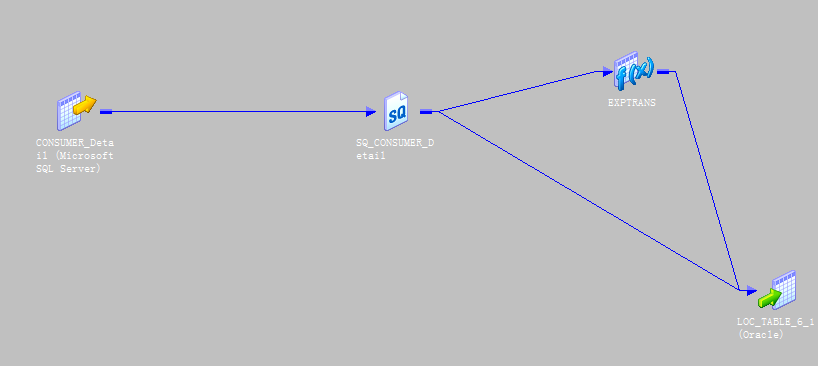
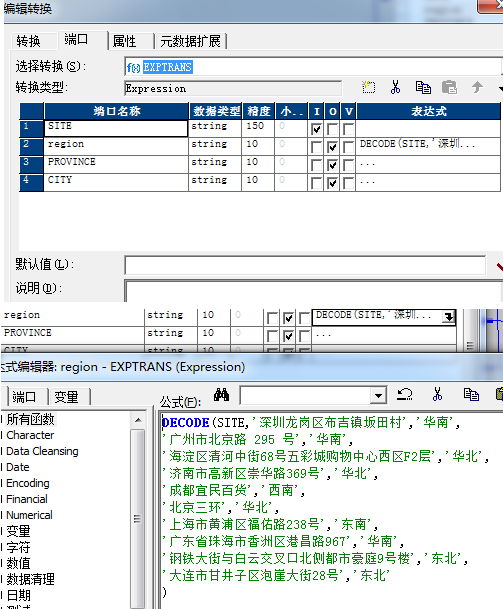


图3.26 建立映射

②依次设置大区、省（直辖市）、市



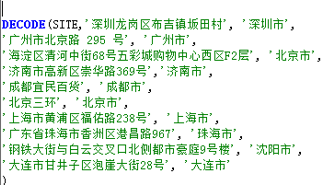
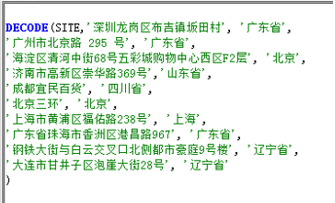


图3.27 设置函数

③ 建立工作流

右键映射，选择生成工作流

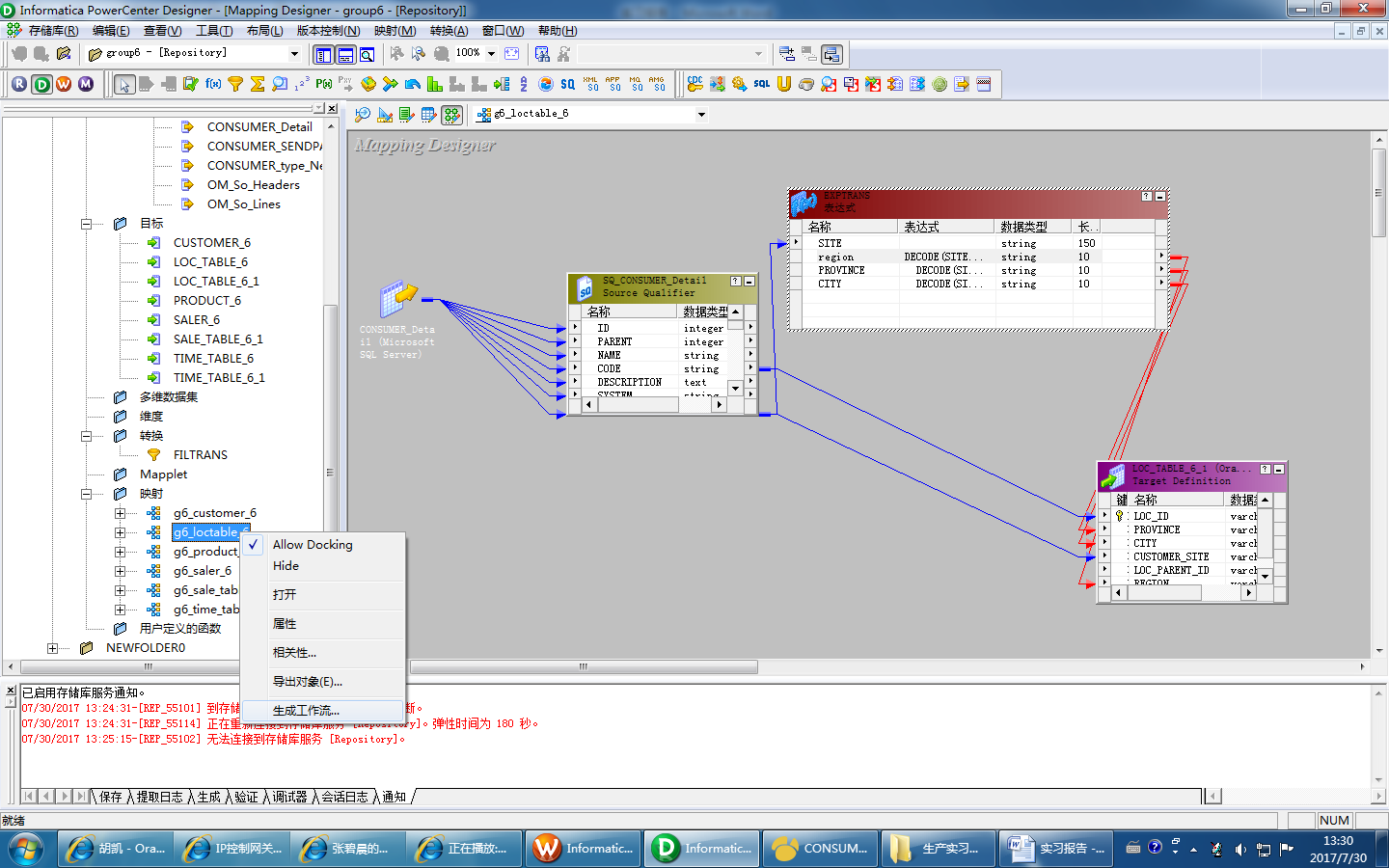


图 3.28 生成工作流

将源表的连接对象改为odbc\_sqlserver，目标表的连接对象改为group2



图 3.29 设置工作流

④生成工作流

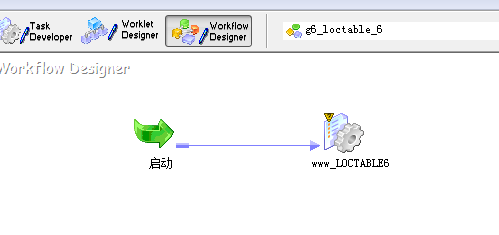


图 3.30 工作流示意图

工作流运行成功

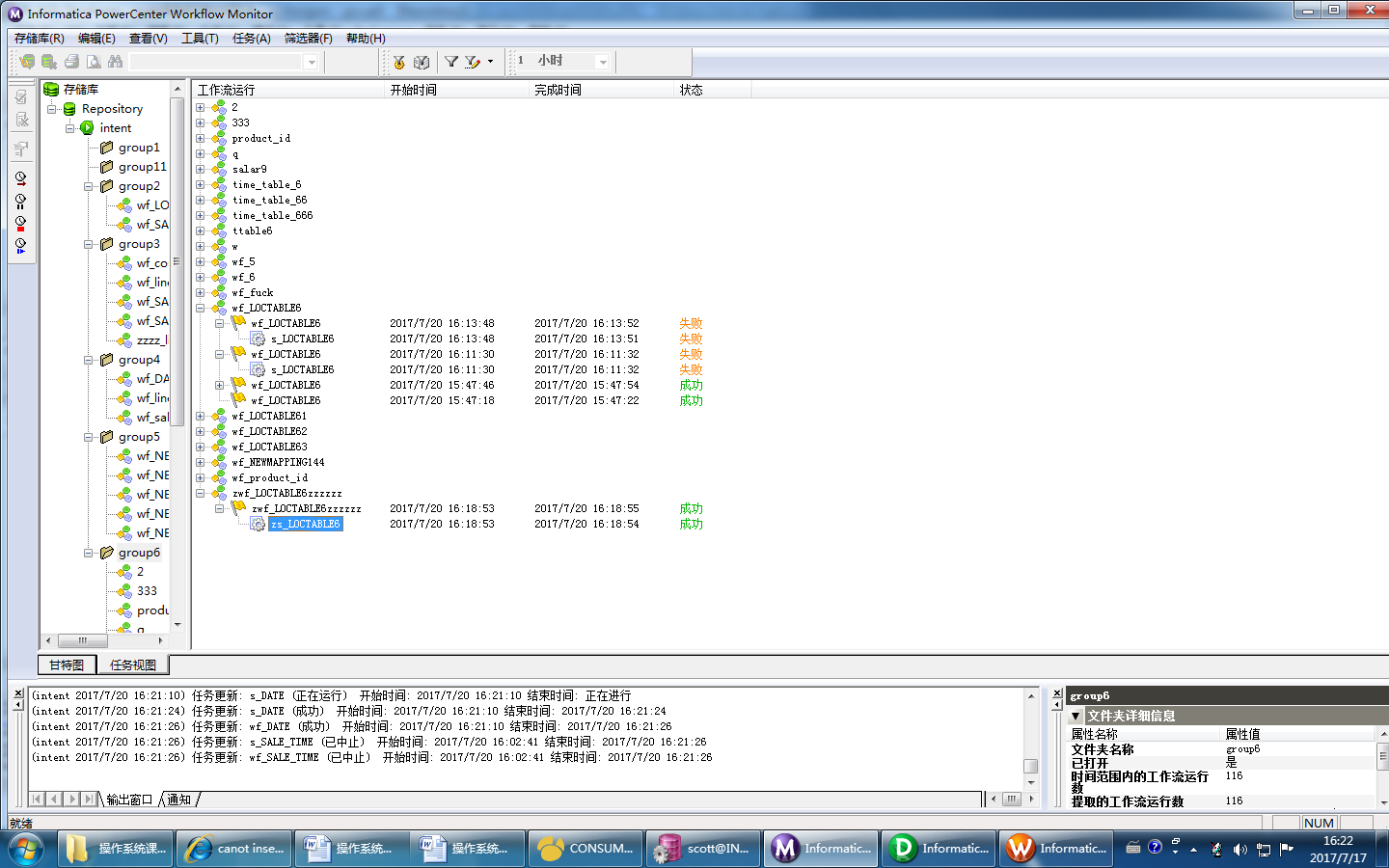


图 3.31 工作流运行图

⑤ 检测导入结果

目标表如图，导入结果正确

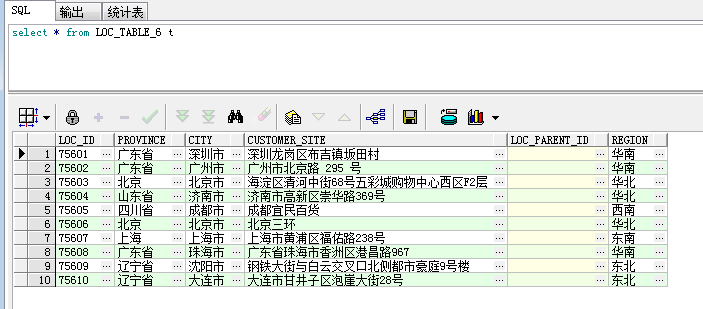


图 3.32 地区表数据图

其他维度均已导入，与此步骤相同，不再一一赘述，最终将数据全部导入目标表，经审查，数据导入无误

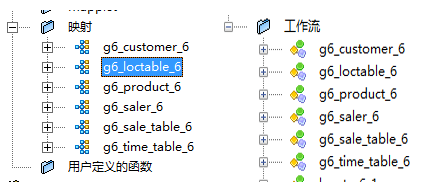


图 3.33 最终结果图

## 3.6 OBIEE的开发与设计

Oracle BIEE是一种提供了整套的分析和报表功能的企业级商业智能产品。OBIEE支持针对详细商业数据的深入分析和数据挖掘，提供对做出关键决策的真实有用信息。我们使用OBIEE来实现报表开发。

### 3.6.1 准备数据（将数据从数仓导入OLAP）

首先在物理层导入维度表和事实表，其次在业务层设定层级，并最终确认关联是否正确。

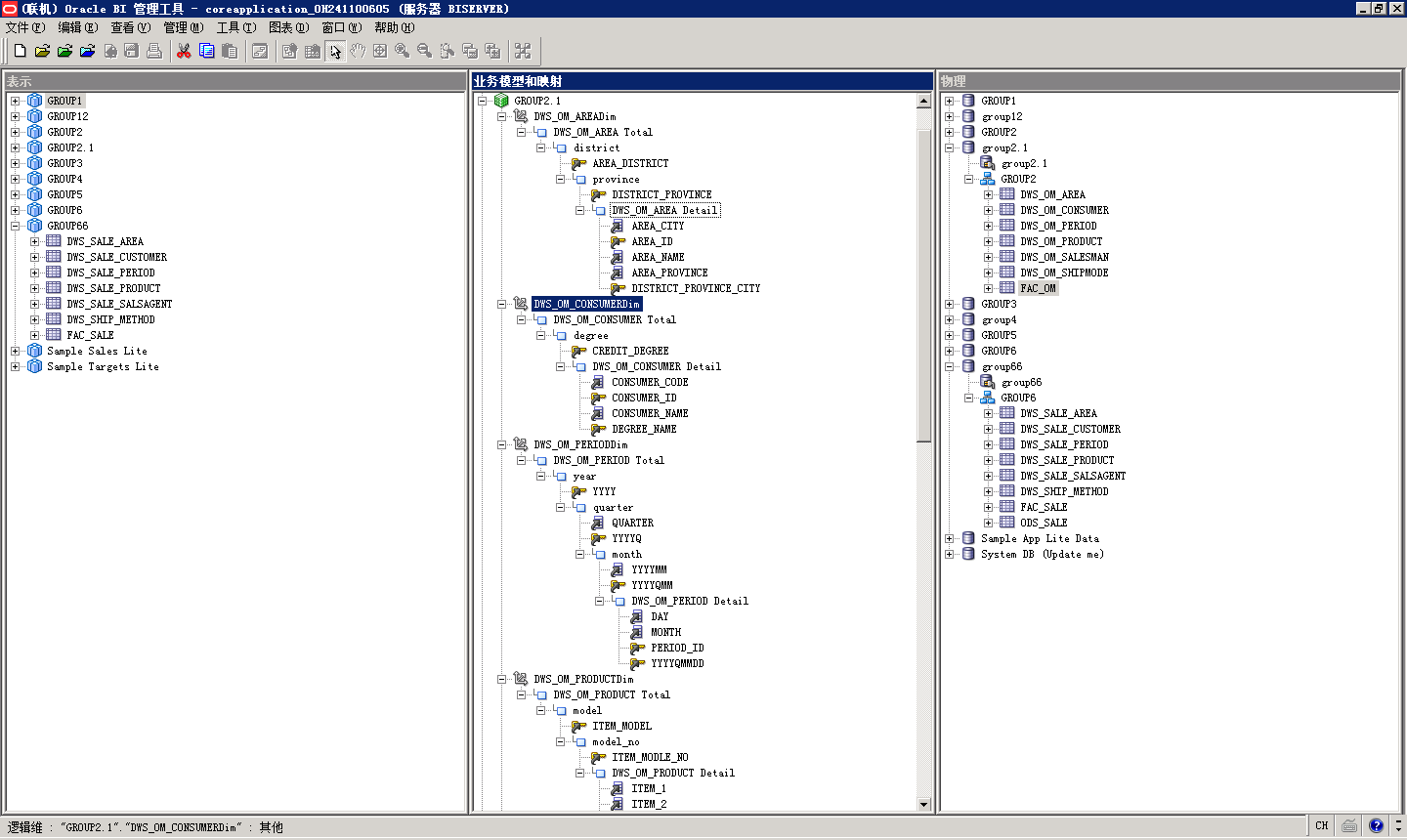


图 3.34 物理层导入图1

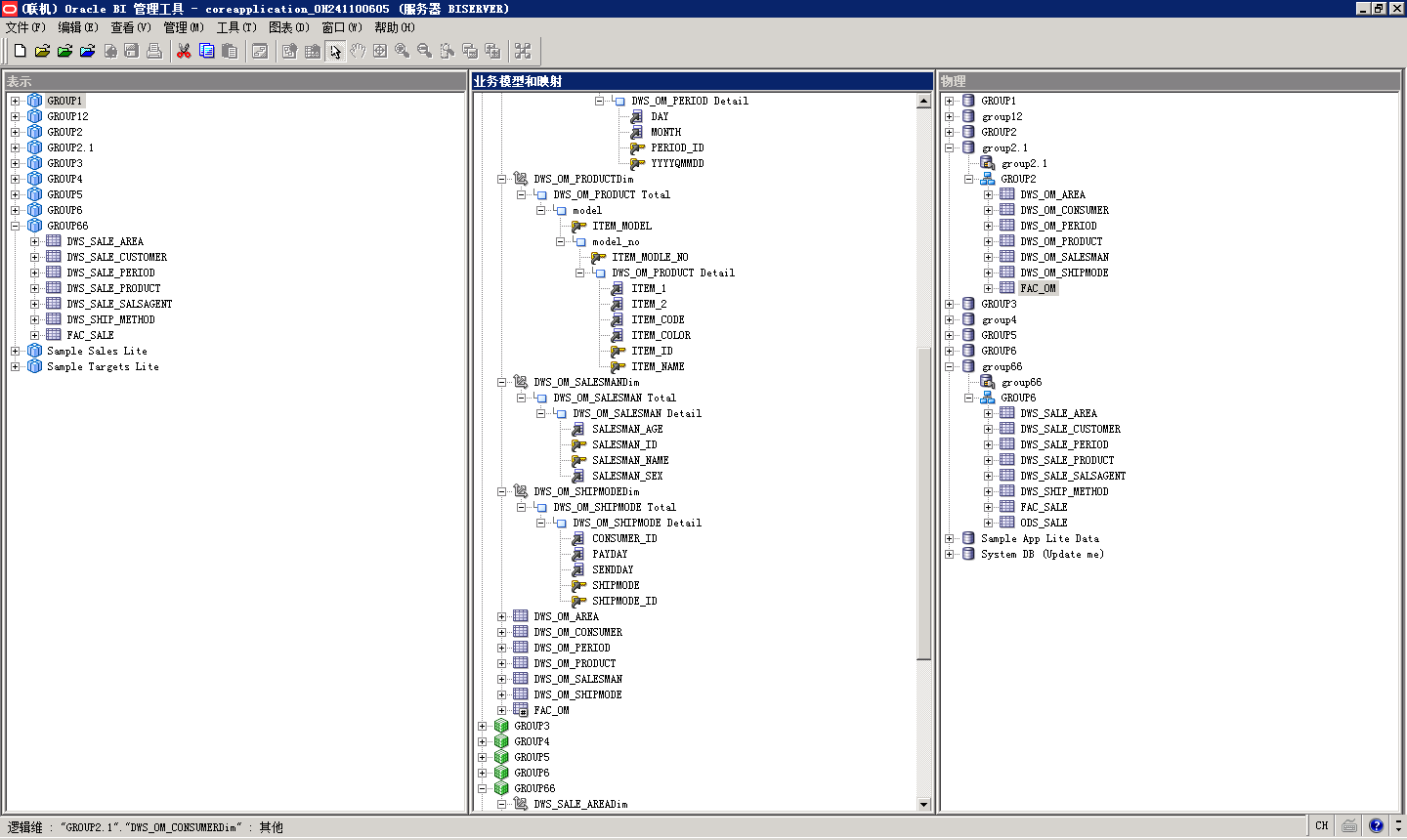


图 3.35 物理层导入图2

物理层：是负责和物理对象联系的一端，所以记录的是链接数据库等数据源的信息。物理层允许多个来源，包括市面上常见的数据库甚至包括 EXCEL 等文件信息，而且数据之间可以相互关联，而不用关心它的来源。主要功能是：数据导入、创建外连接(foreign key)。

逻辑层：即所称的业务模型和映射层。它是后台模型管理的重要部分，包括：建立维度、分析项、维度与分析项的关联、层次等基础操作，还有许多性能优化的操作如切割和聚合表使用等。

展现层：将处理好的数据展现出来。

**3.6.2报表开发过程**

**3.6.2.1 登录BI**

****输入GROUP2的账号及密码登录进入BI界面

图 3.36 登陆系统图

#### 3.6.2.2 报表详细设计

该部分依然以地区部分报表设计为例

① 新建分析并选择主题区域



图3.37 选择主题区域

② 选中所需要的列

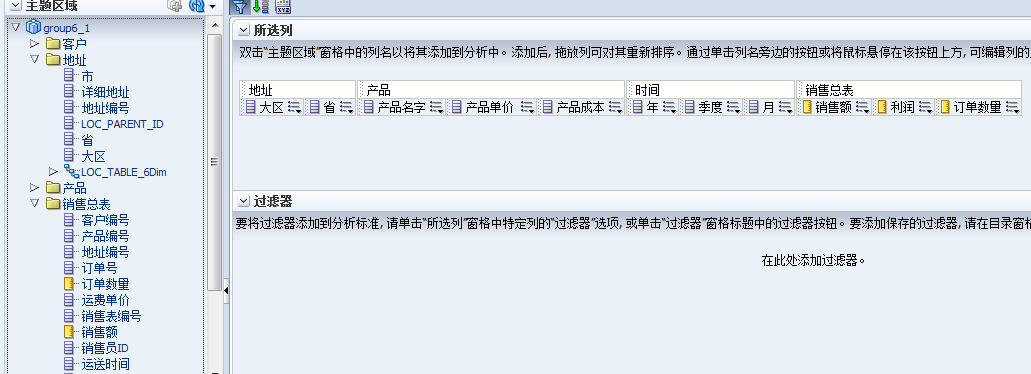


图3.38 选择列示意图

③ 对目标结果进行必要的转换操作：

如：成本（1号车成本为600元，2号车成本为800）

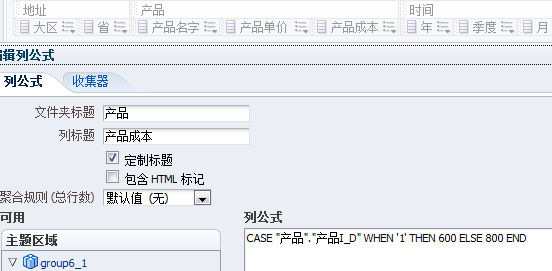


图3.39 成本计算

如：产品单价（将字符型转换为整型）



图3.40 单价转换

如：计算利润（产品单价\*订单数量-产品成本\*订单数量）

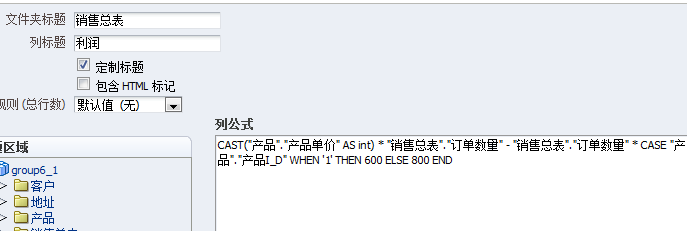


图3.41 计算利润

最终结果如图：



图3.42 最终结果

#### 3.6.2.3 选择透视图类型并对其进行编辑

布局中的区域、切片等内容均可通过拖动列来进行设计，不同类型的图形设计有所不同。图中已用红线框、红线圈进行标注。举例如下：

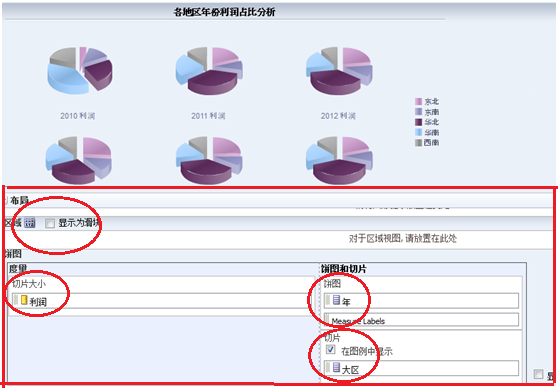


图3.43 饼图编辑界面



图3.44 瀑布图编辑界面



图3.45 条线图编辑界面

可以按照需求对图形的颜色等属性进行修改，或修改标题、标签



图3.46 属性修改示意图

其他分析同理，不再一一赘述，本人所完成部分如下：



图3.47 个人成果图

#### 3.6.2.4 生成仪表盘

① 选择新建仪表盘并选择主题区域：



图3.48 新建仪表盘图

② 依次选中左侧所需要的表，拖动至右侧适当位置松手

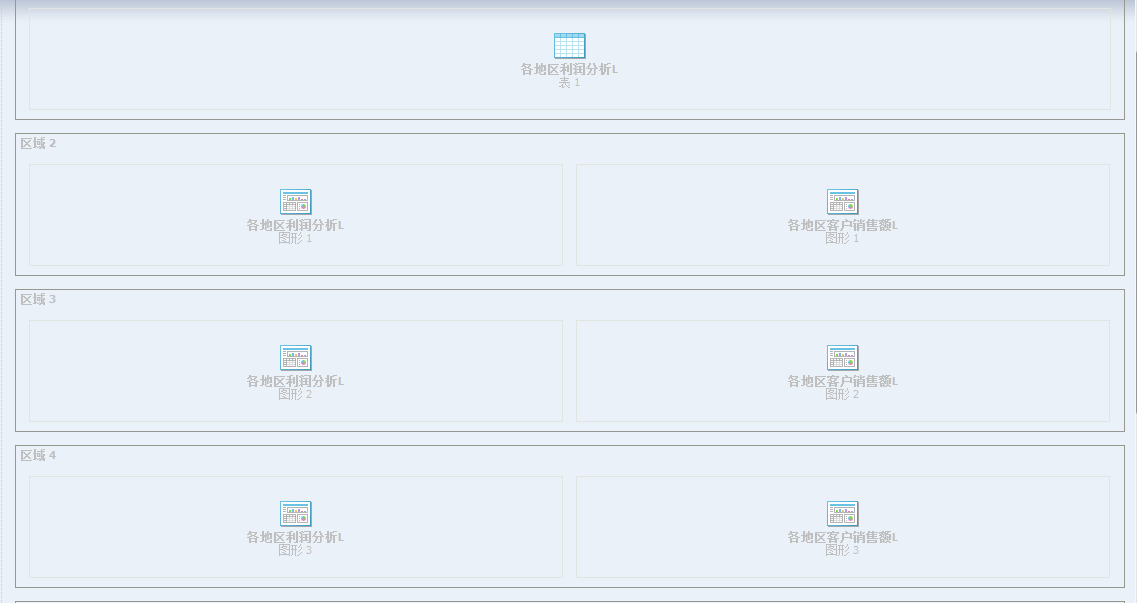


图3.49 编辑仪表图

③最终仪表盘如图



图3.50 成果展示图

### 3.6.3 分析报表

**3.6.3.1 产品分析**

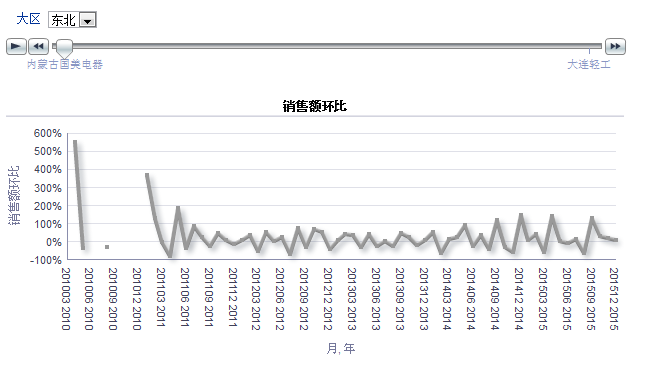
****

图3.51 地区-季度销售额环比图

****

图3.52 地区-季度订单数量环比图

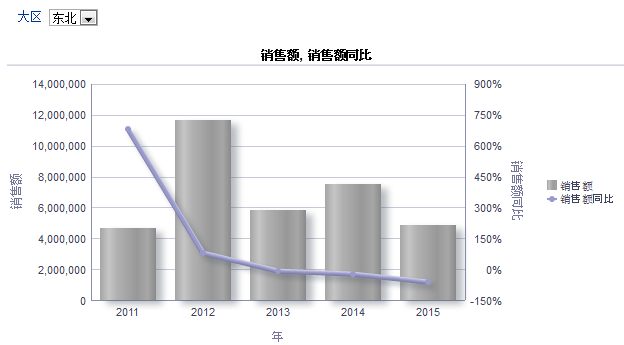
****

图3.53 地区-年份销售额同比图

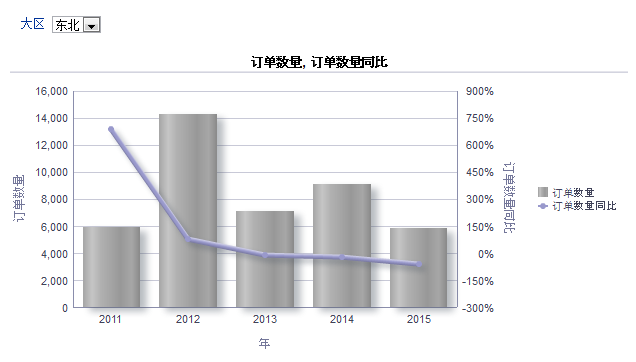
****

图3.54 地区-年份订单数量同比图

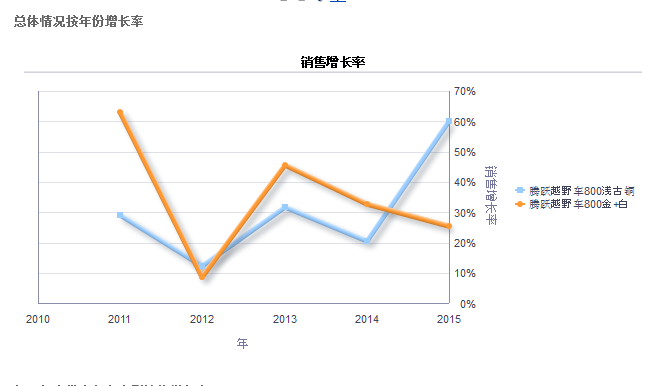


图3.55 产品-年份销售增长率图

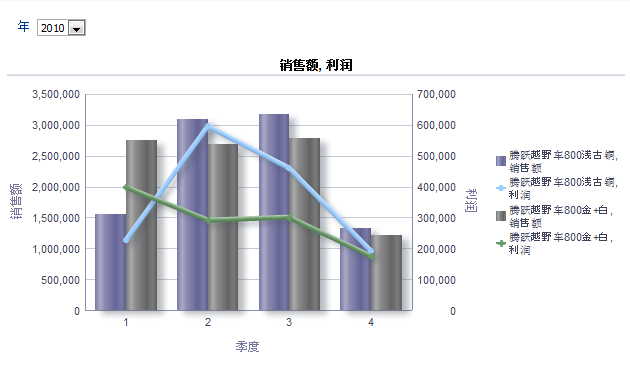


图3.56 产品-季度销售额、利润图

1. 整体销售额/销量分析

2010年至2015年，公司总销售额呈逐年递增趋势，其中2011年发展最为迅速，销售量同比增长46.30%；2012年发展较为缓慢，同比增长仅为10.28%；2013年至2015年公司恢复到快速发展状态中。直至2015年，公司年销量与2010年同比已增长299.34%。

1. 产品—利润分析

在销售额相差不大情况下，浅古铜车利润要高于金+白车. 15年浅古铜的销量超过金+白但利润却没有金+白高这一点我们可以推断该公司应该是进行低价促销手段来推广该型号的自行车，这才导致其销量遥遥领先但利润却并未水张船高的现象。

1. 产品—年销售额分析

金+白车型与整体销售情况类似。2011年时相比2010年销售额增长达到了63.02%，2012年时增长最为缓慢，增长率仅为8.72%；2013至2015年恢复快速增长。2015年相比2010年同比已增长329.90%。

浅古铜2011年时的增长速度为29.05%,远低于整体的46.30%和金+白的63.02%，2012年发展也十分缓慢，2013与2014也小于整体增长速度，在2015年，浅古铜车型得到了迅猛发展，同比增长达到了59.97%。2015年相比2010年同比已增长267.84%。

两种车型现有情况下发展势头正旺。虽然金+白稍好于浅古铜车型，但浅古铜车型销售情况更平稳。建议公司适当增加相应车型生产数量。

1. 产品—季度分析

根据季度销量柱状图与线形图可知，天和公司生产的两种自行车销售量均与季节相关，两种车均在第四季度销量最低，2、3季度销量较高。而对于第一季度，金+白颜色比浅古铜更受欢迎，且销售额基本与2、 3季度持平。建议公司根据季度适时调整生产策略。

#### 3.6.3.2 地区分析

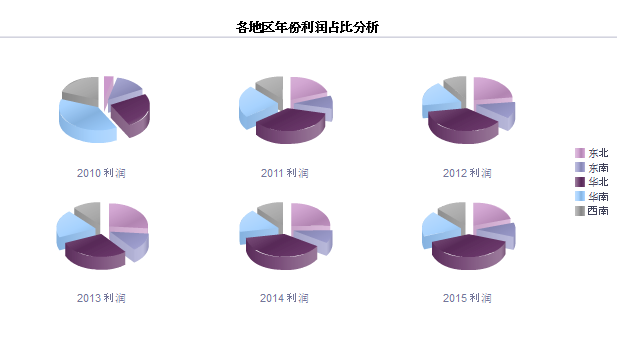


图3.57产品-季度销售额、利润图

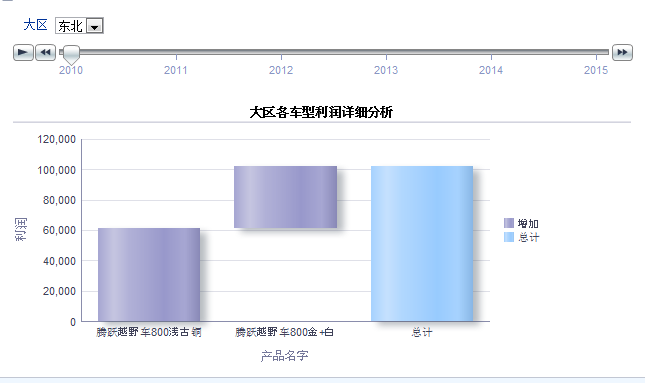


图3.58地区-年度产品利润瀑布图

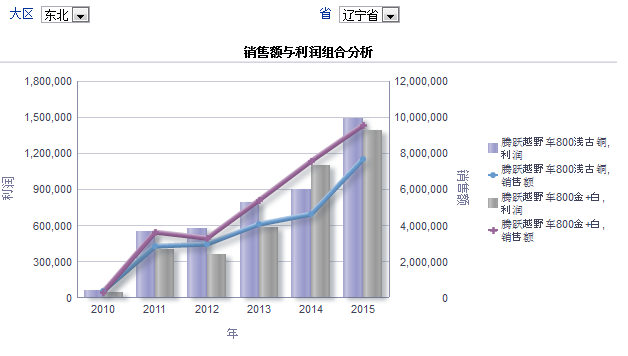
****

图3.59地区-年度销售额、利润条线图

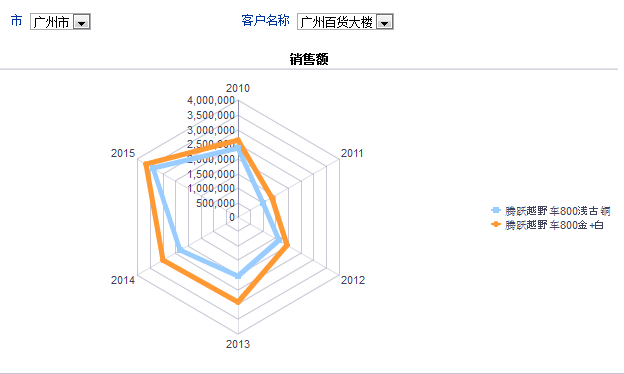
****

图3.60地区-年度销售额雷达图

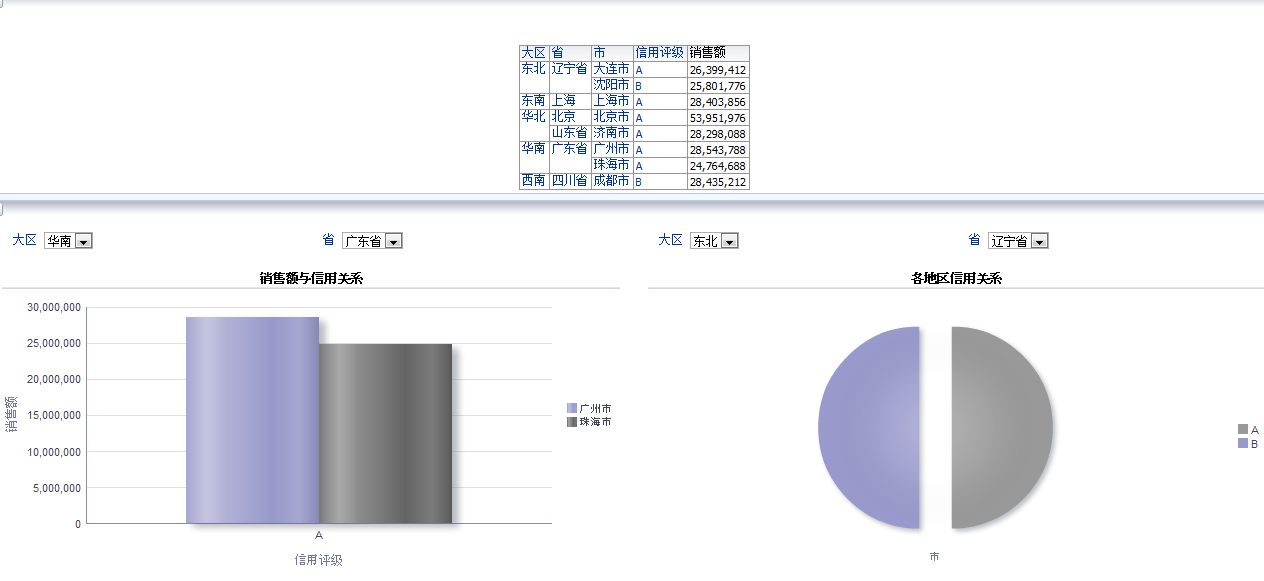
****

图3.61地区-销售额信用图

本人主要完成地区纬度的设计

1.区域分布

从天河公司区域销售占比图中可以看出，2010年之前，天河公司的主力销售区域为华南区和华北区，占据了全部销售量的58.27%，而东北地区仅有极少部分的客户，销售量仅占全部销售量的3.22%。2011年之后，随着天河公司区域转型战略的实施，呈现出了东北、华东、华北、华南四区鼎力的形势。所有区域销量在2011年之后皆呈现高速增长趋势，直至目前，天河公司区域战略实行正常，发展迅速。

2.重点区域分析

从利润以及销量占比分析可以看出西南地区发展缓慢，由此推断西南地区崇山峻岭，市民对于自行车的热情不大。东北、华北、华东、华南四个区域销量接近，皆为重点区域。为此该公司应适当抽调西南地区的人力、物力、财力来支持其它四区的产业建设以达到利益最大化。建议继续实施‘四区并行，兼顾其他’战略，保持快速稳定增长。

从信用水平来看，南方各地整体水平高于北方。且其它如地区等因素相同情况下，信用水平越高的客户销售额越高。因此对于信用等级高的地区客户要着重发展。

3.区域—产品分析

东北地区在2014年时浅古铜车型销量增长率有着明显的下降，但2015年时又迅速回暖，这和全国整体情况类似。可能原因是2014年并不流行此款颜色的车型除此之外每一年都是腾跃越野车800浅古铜销售量占比更大，建议可以在东北地区加大浅古铜车型的生产。

东南区中，两种车型销量几乎持平，可以和全国整体战略接近，略微增加浅古铜车型的生产。

华北区两种车型销量无大区别，可维持现有生产比例。

华南区两种车型的销量波动较大，浅古铜车型销量发展势头迅猛，可加大投资。

西南地区由于地形影响，民众对于自行车需求没有其他四个地区需求量大，可在现有基础上适当减少两种车型的投放比重。

**3.6.3.3 销售员分析**

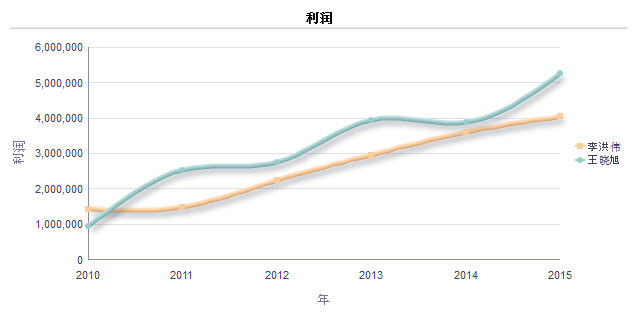
****

图3.61 销售员-年度利润图

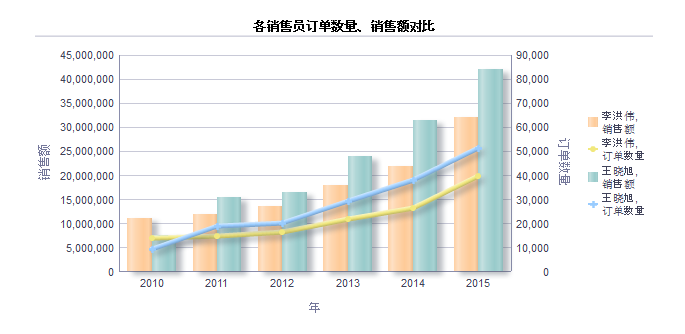
****

图3.62 销售员-年度订单数量、销售额对比图



图3.63 销售员-地区订单数量、销售额对比图

1. 个人总销量分析

根据销售员占比情况分析图可知，除2010年李销售员领先王销售员之外，2011至2015年王销售员销量均领先李销售员。按此发展速率，王销售员与李销售员之间的销量数量差距会越来越大，王销售员整体能力强于李销售员。但从利润图可以看出，王销售员年份利润增长不如李销售员稳定，公司可以对王销售员实施激励等必要措施。

1. 个人—产品分析

李销售员总销量中浅古铜的销量要高于金+白的销量，王销售员总销量中金+白的销量要高于浅古铜的销量。公司妥善分配任务，两位销售员发挥各自的优势,着重营销自己擅长的车型。

1. 个人—区域分析

根据个人—区域销售量雷达图可知，在2010年之前，李销售员的销售地区是华东区和华北区，在东北区则只有很少的业绩；王销售员的主要业务地区则是西南区和华南区，在华北区有少量业绩。在2011年之后，由于公司的地域转型战略，王销售员在东北地区开拓了市场，在接下来的几年中业绩便超过了李销售员。但总的来说，两位销售员各据南北。建议李销售员加强在华南区和华北地区的投入，王销售员需要保持高速增长的势头。

# 4. 实习总结

## 4.1 实习内容的复杂性评价

本次实习的工程复杂性评价如表4.1所示：

表4.1 工程复杂性评价表

|  |  |  |
| --- | --- | --- |
|  | **复杂工程问题特征** | **问题描述及解决方案（方法）** |
| 1 | 必须运用深入的工程原理，经过分析才可能得到解决 | 需要实际性的需求分析、总体设计、详细设计、项目实现等环节，了解项目的过程和内容。在项目过程中还利用E-R图、思维导图等方式分析项目，遵守软件工程的原则和过程。以及数据仓库技术、ETL技术和OLAP等共同作用，完成一整套的商务智能流程。 |
| 2 | 涉及多方面的技术、工程和其它因素，并可能相互有一定冲突 |  |
| 3 | 需要通过建立合适的抽象模型才能解决，在建模过程中需要体现出创造性 | 建立了数据仓库，并对现有维度建立星型模型，其中要较好的体现出各纬度间层级关系。 |
| 4 | 不是仅靠常用方法就可以完全解决的 | 项目中采用非程序的方法进行测试，现有条件下测试方法只能是人为对照数据正确性，图表合理性。 |
| 5 | 问题中涉及的因素可能没有完全包含在专业工程实践的标准和规范中 | 数据库建表为了满足真实的业务需求，若刻意满足范式约束，会大大加大问题复杂度，因此实际中往往为了真实业务需求而打破范式约束。 |
| 6 | 问题相关各方利益不完全一致 |  |
| 7 | 具有较高的综合性，包含多个相互关联的子问题 | 包含数据库、数据仓库、OLAP、报表开发等知识，它们涉及到计算机科学与技术（主体）、商务（销售业务）、数学（增长率等指标的计算）、设计（图表美化）等专业知识。 |

## 4.2 实习体会、收获与建议

实习是计算机专业学生培养的重要一环。这能让我们将理论知识有效地运用到实践中去，将我们在学校所学到的知识得到很好的组合。在实习期间，我认真地学习商务智能的知识，在周老师的讲解中了解了很多课堂以外的关于互联网业务和商务智能行业的知识，并对职业发展和专业前景进一步加深了了解。总结这近一个月的体会和收获，分以下三个方面：

第一，实践能力有较大提升

在整个实习过程中，我巩固了数据库中SQL语句的知识，理解并熟悉了Oracle PLSQL，并对其自带函数进行熟练应用；学会使用Informatica PowerCenter系列工具和OBIEE报表开发工具。

在本学期学习了数据库原理，自以为对数据库掌握的还不错，可是进入实战环节才发现并不熟练。在进入正式项目之前老师对我们进行了几天的数据库强化训练，有很大的帮助。另外，我对商务智能（BI）产生了浓厚的兴趣，并开始去主动了解数据仓库和数据挖掘这个研究方向，去寻找一些数据分析行业的故事。

第二，团队合作成效显著，提升自身的职业素养

为了更好地完成开发商务智能项目，我们进行了分组，我被分到了第六组。小组合作这种模式确实能让组内每个人受益。组内每个人都承担着自己的任务，能相互学习，相互补充，有漏洞和不足会及时提出并改进。组间讨论能帮忙寻找多个组之间的共同问题和解决方案，提高项目开发效率。在实战演练中，我们小组相互扮演需求人员和客户，对项目进行深入分析。在创建工作流时，发现一开始的数据仓库有问题，无法成功导入数据。我们分析

了整个ETL环节，并从新建表，从而使问题得到解决。

在工作过程中，我们肯定会碰到很多的问题，不懂的东西我们就要虚心向同事请教。同时，我们也不要怕犯错。每一个人都有犯错的时候，要及时调整确立明确的目标，并端正自己的态度。态度决定一切，即使自己以前没学过的知识也可以在工作中逐渐的掌握。

第三，能力仍需继续提升

在实习期间，虽然我完成了项目的开发，但在开发过程中遇到的问题正显露出了我在专业技能方面的知识欠缺。我的编程能力稍弱，在老师布置SQL语句的练习后不能迅速完成，往往需要老师的提醒才能解决问题。我对数据仓库的理解也不太足，在和同学们探讨数据仓库建立维度和事实表的时候我混淆了一些概念（如：成员和层次），对下钻、上卷等概念模糊不清。  
 深入了解到商务智能行业的生机与乐趣的同时也体会到业务的艰辛和曲折。在学习之余，周老师还为我们介绍了商务智能行业发展情况、如何做好项目经理等，这些都是平时课堂接触不到的。

本人各方面的能力都需要进一步提升，衷心感谢学校给予这次实习机会，有幸进入商务智能（BI）实习强化班，实习班里有着三十个认真、有激情、积极上进的同学，虽然只有短短的三周，但我确收获颇丰！感谢周老师，感谢同学们，使我看到了自己欠缺的技能与能力。接下来的时间里，我要踏踏实实锻炼自己，不断了解行业前沿，积极主动地探讨商务智能行业的知识，希望未来能在商务智能领域获得有效的能力提升。

# 5.参考文献

1. 张斌,郭军.软件工程及应用. 沈阳:东北大学出版社,2007.
2. 王珊,萨师煊.数据库系统概论. 北京:高等教育出版社,2006.
3. 李英杰，李厚勤. 基于数据仓库技术的决策支持系统的应用研究. 微型电脑世界，2000，16(11)：16～17.
4. 李琪，白彩英. 数据仓库中维的建模和查询. 计算机研究与发展，2002，39(5)：612～618.
5. 杜绍森.Informatica PowerCenter权威指南. 北京:电子工业出版社,2015

# 附录 实习日记

|  |  |
| --- | --- |
| 7月3日 | 老师介绍了商务智能的基本概念和概况，并初步了解了商务智能的体系和其典型案例 |
| 7月4日 | 巩固并加深了数据库的基础知识 |
| 7月5日 | 下午于教授的专题报告 |
| 7月6日 | 了解了PL/SQL的基本用法，学习了PL/SQL的基本组成部分，并在最后进行了练习 |
| 7月7日 | 练习PL/SQL，熟练掌握SQL语句 |
| 7月10日 | 初步了解数据仓库，分析数据源 |
| 7月11日 | 上午王教授专题报告 |
| 7月12日 | 分析了数据源并初步考虑如何设计维度表 |
| 7月13日 | 建立数据模型，确定所需要的维度 |
| 7月14日 | 建立数据仓库，创建各个维度表 |
| 7月16日 | 上午王教授讲座，下午继续完善最终完成数据仓库的搭建 |
| 7月17日 | 上午并初步了解Informatica Power Center相关知识  下午练习了如何使用Informatica Power Center中的各个组件 |
| 7月18日 | 上午参观东软集团  下午练习如何使用Power Center Designer |
| 7月19日 | 学习使用Power Center Workflow，练习建立会话和工作流，并成功将数据导入到维度表中 |
| 7月20日 | 对照目标数据库与源数据库，数据仓库运行与测试 |
| 7月22日 | 上午完善销售模块源数据，目的数据  下午老师验收初步成果 |
| 7月23日 | 介绍数据挖掘，了解数据挖掘在实际项目中的运用 |
| 7月24日 | 数据可视化介绍，主流厂商产品说明 |
| 7月25日 | 练习OBIEE使用，组件学习 |
| 7月26日 | 利用OBIEE对报表进行开发，小组成员确定主题分析方向 |
| 7月27日 7月28日 | 开发仪表盘，小组为单位完善报表并对报表进行分析  接收测试实施，老师验收；并对项目实施进行总结，BI技术总结归纳 |
| 7月29日 | 大数据分析案例介绍；并自行对项目进行拓展、开发和部署 |
| 7月30日 | 撰写实习报告 |