**考点4：表格结构数据特征（2%）**

**表格结构数据（了解）**

1、表格结构数据：使用电子表格工具，以单元格为基本处理单位。工具：Excel，WPS，Numbers

2、表结构数据：使用数据库查询语言，以字段为基本处理分位。工具：数据库，ETL工具，可视化工具

**表格结构数据特征（掌握）**

两个对象具有父子级关系需要满足的条件：

1、一个父级对象下包含多个不同子级对象

2、一个子级对象只能属于某一个特定的父级对象

例如（满足）：省份与城市，月份与日期；例如（不满足）：月份与星期

一、数据层级（父子级关系）：单元格 ∈ 单元格区域 ∈ 工作表 ∈ 工作簿

二、数据类型：

1、数值型：整数-1 0 100 小数0.5 -9.67

2、文本型：文字a B 我 符号& ： 数字“9”“-6.9”（字符串）

3、逻辑型：真值TURE 1 假值FALSE 0

三、单元格格式属性：

1、数字格式：决定显示形式。如：数字、百分比、货币、日期、时间

2、显示格式：决定显示效果。如：字体、单元格填充色、图标集（红黄绿）、显示位置（文本默认居左，数值默认居右）

**考点5：表格结构数据获取、引用、查询与计算（3%）**

**表格结构数据获取方法（了解）**

一、企业后台数据库系统：用户表/品牌表/产品表/订单表——分析人员让数据库管理人员SQL数据库查询——电子表格工具支持的数据文件

二、前端操作平台：CRM、ERP、财务系统——电子表格工具支持的数据文件

优点：高效方便

缺点：只能导出设定好的数据内容，设定逻辑外的数据还是需要从数据库中获取

三、企业外部渠道：文本文件（.csv、.txt）、Excel文件（.xlsx）、WPS表格文件（.et）

**单元格区域的特征（掌握）**

1、由连续的单元格构成

2、是一个方形区域

**表格结构数据引用（应用）**

一、单元格值

1、同一工作表单元格：“=行号+列号”如“=A1”“=XFD1048576”

2、不同工作表单元格：“=表名!+列号+行号”如“=Sheet1!A1”（! 相当于 的）

二、单元格区域值

1、相同工作表区域：“=左上单元格：右下单元格”如“=A1:C4”

2、不同工作表区域：“=表名!左上单元格：右下单元格”如“=Sheet1!A1:C4”

3、引用多行区域：“=上边行号：下边行号”如“=1:5”

4、引用多列区域：“=左边列号：右边列号”如“=A:D”

**表格结构数据查询（应用）**

一、使用表格工具搜索功能ctrl+F（修改参数不方便）

二、使用查询函数=VLOOKUP(查询条件E2,查询范围A:B,返回值2,匹配模式FALSE)

**表格结构数据计算（应用）**

一、直接计算

1、算术运算符：加+，减-，乘\*，除/，幂^，返回数值

2、比较运算符：>，>=，＜，<=，不等于<>，返回TRUE或者FALSE

3、文本运算符：&，如“你”&“是好人”，返回“你是好人”

二、函数计算



**考点6：表结构数据特征（5%）**

**理解主键的意义（掌握）**

单字段主键：由一个字段构成的主键

多字段联合主键：由多个字段构成的主键

一、物理意义：

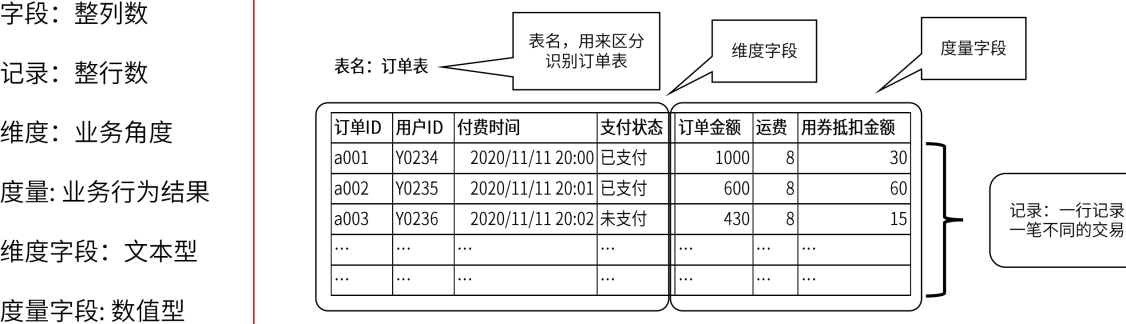
1、非空，不重复

2、定位记录行，字段名+主键值定位具体数值

3、多以“xxID”“XXNO”等名称命名

二、业务意义：表的业务记录单位，在一个数据表中的所有非主键字段都要围绕主键展开

**理解维度及度量的意义（掌握）**



维度表：只包含维度信息的表

事实表：既包含维度信息又包含度量信息的表

**理解缺失值（掌握）**

根据数据类型以及生成信息重要程度的不同，使用不同方法处理缺失值

1、文本型字段：影响不大：不处理，或者用无实际业务含义的文本进行替换；影响大：替换或删除

2、数值型字段：综合考虑该数值型字段的度量意义与针对该数值型字段进行汇总计算的方式，来最终决定具体处理方法

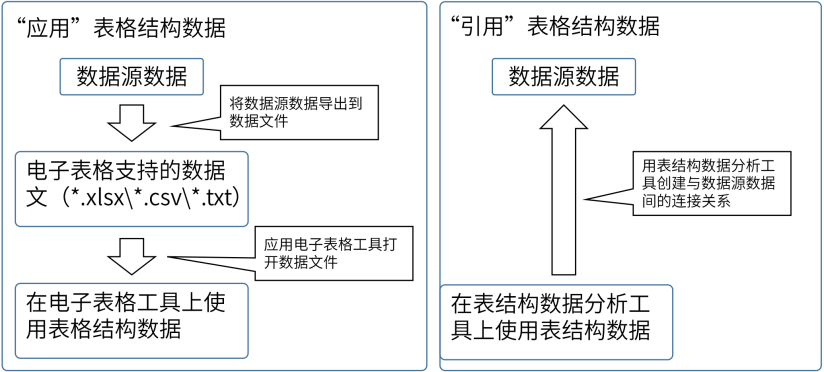
**表结构数据特征（掌握）**

1、以字段或记录作为数据的引用、操作及计算的基本单位

2、所有字段记录行数相同（方形结构）

3、1个表只有1个主键

**表结构数据与表格结构数据差异（掌握）**



**考点7：表结构数据获取、加工与使用（5%） **

**表结构数据获取（了解）**

1、关系型数据库管理系统RDBMS（结构化数据）：企业业务数据的存储、检索、访问与共享

多层级结构，在线事务处理OLTP，可量化、结构化数据，提供大部分数据源，不善于分析

2、商业智能BI系统：为企业决策者快速提供完整、准确、深入的数据分析结果，帮助企业决策者实现商业洞察

善于分析，多功能模块构成，企业级（大型）/敏捷型（中小型），多维数据集，所见即所得

3、数据仓库DW（结构化数据+非结构化数据）：用来存储分析所需要的不同数据源上的所有相关数据信息

OLAP技术：连接信息孤岛、创建多维数据模型

**ETL作用、特征（掌握）**

1、Extract抽取：创建不同数据源之间的连接关系，对这些数据源中的数据进行“引用”

2、Transform清洗转换：筛选过滤不完整、错误及重复的数据记录

①对“粒度”不一致的数据进行转换 ②对业务规则不一致的数据进行转换

3、Load加载：将抽取出来的数据经过清洗与转换后加载到数据仓库中进行存储与使用

**表结构数据连接（掌握）**

一、横向合并：将不同表中的字段信息合并到同一个表中使用

1、通过公共字段匹配：拥有相同记录值的字段

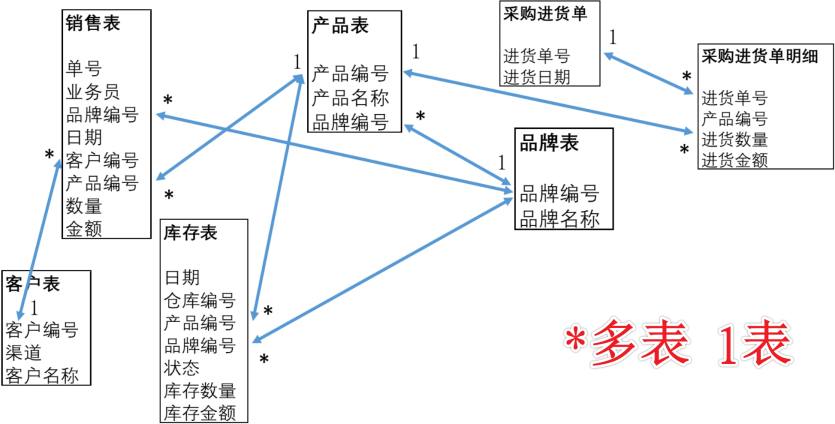
2、左表与右表：连接命令左侧的表为左表，右侧的表为右表

3、连接方向：决定表的主附关系，主要使用“左连接”“右连接”“内连接”

4、对应关系：决定连接结果行数是对应项乘积的结果

实际工作中，选择一对多对应关系连接：1表的主键join多表的非主键（字段有重复值为“多”表，非空不重复为“1”表）

5、E-R图（实体关系图）：多表连接的鸟瞰图



二、纵向合并：多表中记录信息合并到同一个表中

1、字段个数相同

2、相同位置字段的数据类型相同

3、去重合并UNION，全合并UNION ALL

**表结构数据汇总（掌握）**

数据透视：对零散数据进行汇总分析。

1、维度：业务观测角度

2、度量：业务行为结果

3、汇总计算规则：衡量业务行为结果好坏程度的测量仪

维度筛选度量，度量被维度筛选（分组依据）汇总计算规则（聚合规则）

数据透视规则（汇总计算规则）

1、合计规则：将相同维度值下对应的多个度量值相加，SUM函数

2、计数规则：对相同维度值下的度量个数进行计数、COUNT非空计数、DISTINCTCOUNT去重计数

3、平均规则：用合计规则的结果除以计数规则的结果，平均=合计/计数，AVERAGE函数

4、最大值规则：求相同维度下的最大度量值，MAX函数

5、最小值规则：求相同维度下的最小度量值，MIN函数

数据分析的业务意义：数据分析是连接零散数据与人类认知间的桥梁。零散数据——数据透视分析/数据挖掘分析——人类认知