**数据仓库与大数据工程 第1次作业题**

姓名：XXX 学号：XXX

**1.1.数据与信息之间的关系与区别**

①数据：

数据是针对社会生产生活的记录结果，是对客观事物的符号表示。（广义上）

数据是指所有输入到计算机中并被计算机程序处理的符号的总称。（狭义上）

②信息：

信息是物质存在的一种方式、形态或运动形态，也是事物的一种普遍属性，一般指数据、消息中所包含的意义，可以使消息中所描述事件中的不定性减少。

③关系与区别：

数据是反映客观事物属性的记录，是信息的具体表现形式。数据经过加工处理之后，就成为信息；信息需要经过数字化转变成数据才能存储和传输。

**1.2.信息与知识之间的关系与区别**

知识是对某个主题确信的认识，并且这些认识拥有潜在的能力为特定的目的而使用。知识是被验证过的、正确的、被人们相信的内容。

**1.3.intelligence（智能）和wisdom（智慧）**

智能：有了知识要有能力去行动，通过组织或系统去利用知识，服务于日常生活或业务。

智慧：有了知识要有行动，要有意识去利用知识，服务日常生活或业务中。

**2.1.给出决策、决策人员类别、决策层次、决策支持、决策支持系统的概念**

①决策：

指个人、集体或自动系统为解决某个问题，借助一定的科学手段和方法，从若干备选方案中选择或综合成一个满意合理的方案，并付诸实施的过程（多个可能选项中确定一个最终选项）。

②决策人员类别：

高级、中层、低层管理人员。

基础业务人员。

③决策层次：

重要决策、一般决策、不重要决策。

④决策支持：

为需要做决策的人提供支持的活动。

⑤决策支持系统：

决策支持系统（DSS）是企业信息系统或应用环境中的一大类重要的信息系统，这类系统 以数据为基础，通过数据统计、分析、挖掘、展现等手段为各层次决策者提供决策支持服务。

**2.2.列出你个人曾经或经常需要做的决策案例，并将其归入不同层次，并思考分析说明你在做这些决策的时候，决策的依据有什么，相应的决策支持系统由什么组成？**

①决策案例1：晚上吃什么？

决策层次：不重要决策/一般决策

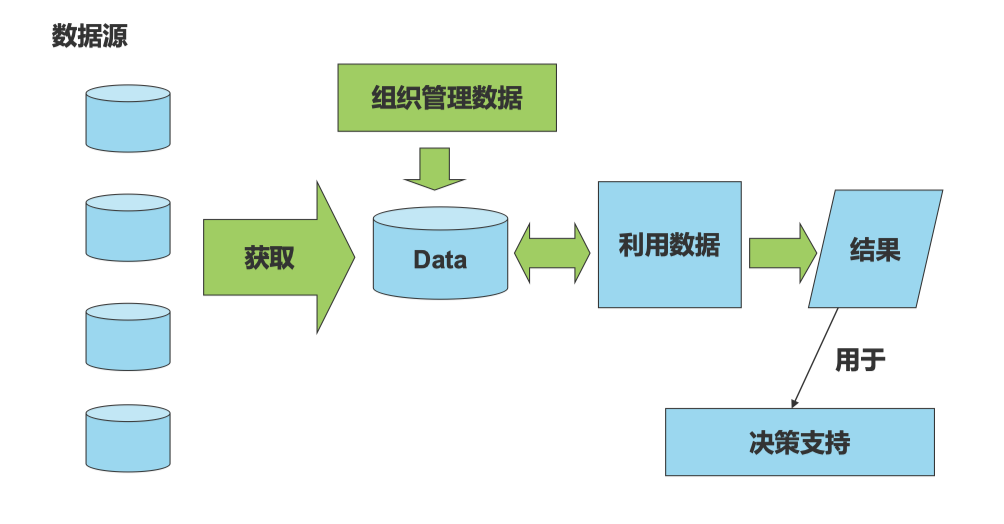
决策依据：资金/饮食喜好/之前晚上都吃了什么/最近有没有朋友推荐我吃什么/……

②决策案例2：XX系统给推荐什么商品广告？

决策层次：重要决策

决策依据：用户喜好/热点商品/合适价格/……

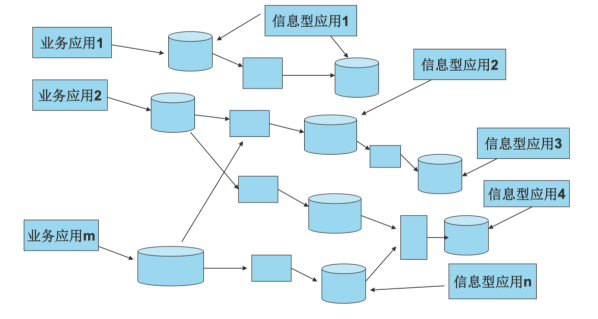
决策支持系统架构：如下图所示。



1. **请简要说明蜘蛛网架构、孤岛式架构、烟囱式架构的区别，给出各种架构中可能存在的问题。**

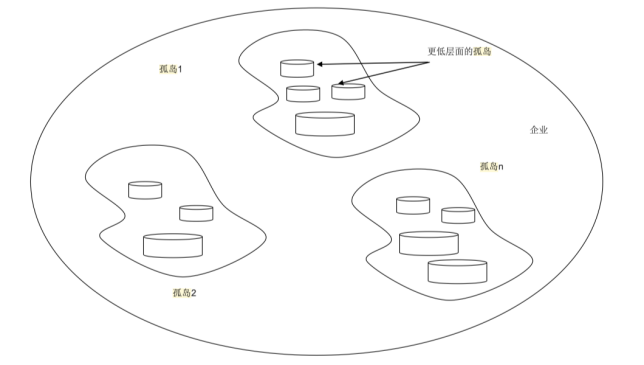
①蜘蛛网式架构：表示企业中数据存储、业务调用等方面比较分散的一重组织方式。

问题：不管从数据还是业务逻辑上都比较复杂，难以控制，数据会存在一致性等问题。



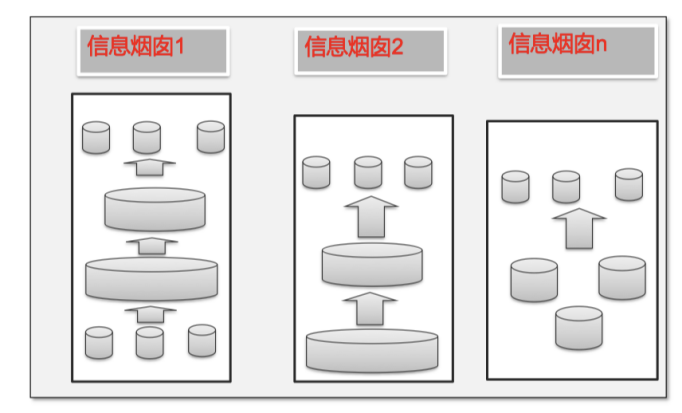
②孤岛式架构：表示企业中某些数据和业务之间彼此独立，没有连接。

问题：孤岛式架构使各个孤岛之间信息无法共享，数据组织也会存在可靠性问题。



③烟囱式架构：每个信息孤岛的演化最后会造成烟囱式架构，即：每个孤岛各自发展成比较庞大的系统，但彼此连接关系很弱。

问题：烟囱式架构是孤岛演化而来的，他使部门之间几乎独立，对业务发展不利，且会存在信息冗余等问题。



1. **对于自然演化信息系统架构存在的三个主要问题，请分别说明数据集成的必要性。**

①现有系统所产生的数据缺少集成。

②不同系统所保存的数据的时间跨度的不同

③各个系统的可用数据时间跨度无法满足DSS对数据时间跨度的要求。

面对种种问题如数据可信度低，信息型应用开发成本高，许多情况下难以将数据转化成有用的信息，所以数据集成十分必要，需从数据应用架构的方法论角度，设计出合理架构的数据仓库。

1. **请学习掌握Bounded Rationality的概念说明大数据驱动决策支持系统与该概念之间的潜在关系及有可能解决的问题。**

①OLT系统中要求业务处理系统必须具有很高的性能，要求数据库系统的负担不能过重。

②MIS或DSS系统的数据访问模式与OLTP大不相同，经常需要访问和处理大量的数据，这种不定时发生的数据处理工作对数据库系统的资源占用可能会很大。解决问题从OLTP系统的数据库中提取数据出来，单独构成用数据利用的系统。

1. **思考题：大规模企业特别是互联策支持系统面向的服务对象的层级下移（更多地面向中下层业务人员）及外部化（更多的面向最终客户）趋势给决策支持系统的数据平台的需求带来的影响。**

①决策支持对象层次的下移和外移：从面向高层的决策，向更多面向中下层决策演变；从向面人的决策更多地面向机器的决策；从面向内部人员的决策更多的面向外部客户的决策。

②决策支持时效性不断提升：从偶发性决策支持近实时和实时决策支持演变。

③整体系统架构从开环更多地演变到闭环：业务→数据→决策支持→决策→业务。

④从关系型+BI组件，到混合架构。