**考试时间5月24日，上午9：00-11:00**

**地点：SY207题目来自于收集的照的照片，可能不太准确，还有一些看不清，大家看着复习。能整理出答案最好了^\_^~**

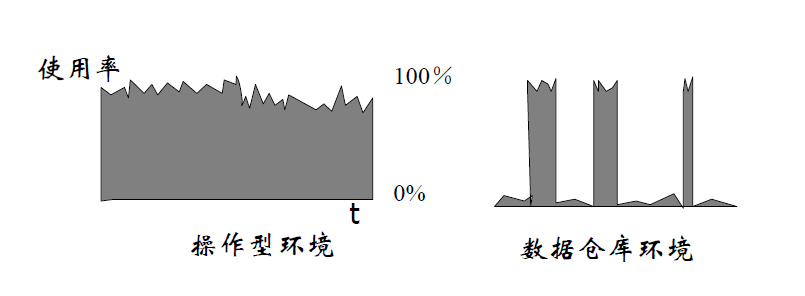
1. 概念题（40分）

请将如下概念串起来，描述概念之间的关系，每个概念不必单独解释。

企业信息系统架构、基础业务系统、数据仓库、数据挖掘、机器学习、RDBMS、HADOOP、Oracle、大数据、DB2、关系型数据库、多维分析引擎、数据集成、数据模型、元数据、决策支持系统

1. 简答题（共5道题，共计60分）

1.如下两个图标分别是操纵系统环境与数据仓库环境在一天内的硬件资源典型使用率模式，横坐标是时间，纵坐标是使用率（0-100%）



1. 请说明为什么这两种环境的硬件资源使用率会有如此不同？（5分）
2. 请说明这种现象对企业信息系统架构设计的影响？（5分）

2.在数据仓库环境的数据体系建设过程中，进行粒度设计时要考虑的因素有哪儿些？粒度分析为什么要与技术环境评估同时进行。（12分）

3.对于企业中常见的孤岛式信息系统架构和蜘蛛网信息系统利用架构，请完成如下要求：1）分别值出这两种架构产生的主要原因；2）指出这两种架构中存在的主要问题；3）给出解决这些问题的常见方案。（15分）

4.设有如下数据仓库的数据模型设计\*\*，请

（6）建立企业全局数据视图；（7）确定每个主题包含的实体集以及每个实体集的属性集；（8）确定粒度层次；（9）确定索引结构；（10）确定存储分配；（11）分析确定数据来源。（11分）

5.假设有3年的销售数据，有如下两种存储方案：

方案1：在数据库系统中用户看的表只有一个，但是有6个分区，每半年一个物理分区。

方案2：在数据库系统中有6个独立的表，分别存储近三年的数据。

1. 这两种方案哪个是数据库系统层分区，哪种是应用层分区？各有什么优缺点。（8）
2. 数据分区对数据归档、数据清除有什么影响？（4）