**61。当今企业经营的环境越来越复杂，商业环境因素可以分为四大类。这些类别是什么？**

答：

•市场

•消费者需求

•技术

•社会的

**62。列出明茨伯格的四个管理者决策角色。**

答：

•企业家：在组织及其环境中寻找机会，并发起改进项目以带来变革；监督某些项目的设计

•干扰处理者：当组织面临重要的、意外的干扰时，负责采取纠正措施。

•资源分配者：负责各种组织资源的分配；实际上，负责制定或批准所有重要的组织决策。

•谈判代表：负责代表组织进行重大谈判。

**63。管理者通常通过遵循四个步骤来做决定。步骤是什么？**

答：

1. 定义问题（即，可能处理某些困难或机遇的决策情况）。

2. 构建描述实际问题的模型。

3. 确定建模问题的可能解决方案并评估解决方案。

4. 比较、选择并推荐问题的潜在解决方案。

**64。列出有助于促进决策支持和分析增长的三个发展。**

答：

•集团沟通与协作

•改进了数据管理

•管理大型数据仓库和大数据

•分析支持

•克服处理和存储信息的认知限制

•知识管理

•随时随地提供支持

**65。描述可用于结构化、半结构化和非结构化决策的计算机支持类型。**

答：

•结构化决策：反复遇到的结构化问题具有高层次的结构。因此，可以将它们抽象、分析和分类为特定的类别，并使用科学的方法使这类管理决策的某些部分自动化。

•半结构化决策：半结构化问题可能涉及标准解决程序和人类判断的组合。管理科学可以为结构化的决策问题部分提供模型。对于非结构化部分，DSS不仅可以提供单个解决方案，还可以提供一系列备选解决方案及其潜在影响，从而提高决策所依据的信息的质量。

•非结构化决策：这些决策只能部分地得到标准计算机定量方法的支持。通常需要开发定制的解决方案。然而，这些解决方案可能受益于公司或外部数据源生成的数据和信息。

**66。商务智能（BI）系统的四个主要组件是什么？**

答：

1. 数据仓库及其源数据；

2. 业务分析，用于操作、挖掘和分析数据仓库中数据的工具集合；

3. 用于监视和分析性能的业务性能管理（BPM）；以及

4. 用户界面（如仪表板）。

**67。列出并描述三个层次或类别的分析，这些分析通常被视为连续的和独立的，但有时也被视为重叠的。**

答：

•描述性分析或报告分析是指了解组织中发生的事情，了解这些事件的一些潜在趋势和原因。

•预测分析旨在确定未来可能发生的情况。这种分析是基于统计技术以及属于数据挖掘一般范畴的其他最近开发的技术。

•规范分析能够识别正在发生的事情以及可能的预测，并做出决策以实现可能的最佳绩效。

**68。Amazon.com如何使用预测分析来响应客户的产品搜索？**

答：

亚马逊使用聚类算法将客户分为不同的集群，以便能够针对特定的促销活动。该公司还使用关联挖掘技术来估计不同采购行为之间的关系。也就是说，如果一个客户购买了一种产品，那么该客户还可能购买什么？这有助于亚马逊推荐或推广相关产品。例如，在amazon.com上的任何产品搜索都会导致该零售商同时建议其他可能引起客户兴趣的类似产品。

**69。描述和定义大数据。为什么搜索引擎是一个大数据应用程序？**

答：

•大数据是不能存储在单个存储单元中的数据。大数据通常是指以许多不同的形式到达的数据，无论它们是结构化的、非结构化的，还是流中的。这些数据的主要来源是来自网站的点击流、Facebook等社交媒体网站的帖子，或来自交通、传感器或天气的数据。

•像谷歌这样的网络搜索引擎需要搜索和索引数十亿个网页，以便在几秒钟内为您提供相关的搜索结果。虽然这不是实时完成的，但是生成Internet上所有网页的索引并不是一项简单的任务。

**70。谷歌为大数据开发了什么存储系统和处理算法？**

答：

•Google作为Apache项目开发和发布了Hadoop分布式文件系统（HDFS），用于以分布式方式存储大量数据。

•Google作为Apache项目开发并发布了将计算推送到数据的MapReduce算法，而不是将数据推送到计算节点。

第11章 大题

61. A relatively new approach to supporting decision making is called automated decision systems (ADS), sometimes also known as decision automation systems (DAS). Give a definition of an ADS/DAS in simple terms?

支持决策的一种相对较新的方法称为自动决策系统（ADS），有时也称为决策自动化系统（DAS）。 用简单的术语定义ADS / DAS？

回答：简单来说，ADS是一种基于规则的系统，通常在一个行业，一个功能区域内提供解决特定重复管理问题的解决方案。

62. What are the various components of an airline revenue management system? Describe the function of each one.

航空公司收益管理系统的各个组成部分是什么？ 描述每部分的功能。

回答：

•定价和会计系统：处理票证数据，公布的票价和定价规则。

•飞机调度系统：根据客户需求处理航班时刻表。

•库存管理系统：用于处理出发数据的预订，取消和更改。

63. Describe, with examples, the two basic ideas most experts agree that artificial intelligence (AI) is concerned with.

通过实例描述大多数专家都赞同的人工智能（AI）所关注的两个基本思想。

回答：

•研究人类思维过程（了解智力是什么）

•机器（例如计算机，机器人）中这些思维过程的表示和重复

64 List five disciplines of artificial intelligence.

列出人工智能的五个学科。

回答：

•哲学

•社会学

•心理学

•人类认知

•语言学

•生物学

•模式识别

•信息系统

•机器人

•管理科学

•计算机科学

•数学

65. List five applications of artificial intelligence.

列出人工智能的五个应用。

回答：

•专家系统

•游戏

•计算机视觉

•自动编程

•言语理解

•自主机器人

•智能辅导

•智能代理

•自然语言处理

•语音识别

•神经网络

•遗传算法

•模糊逻辑

•机器学习

66. Describe the Turing test for determining whether a computer exhibits intelligent behavior.

描述用于确定计算机是否具有智能行为的图灵测试。

回答：

根据该测试，只有当人类访问者在与看不见的人和看不见的计算机交谈时无法识别计算机时，才能认为计算机是智能的。

67。What are three components that may be included in an expert system in addition to the three major components found in virtually all expert systems?

除了几乎所有专家系统中都有的三个主要组件外，专家系统中还可能包含哪三个组件？

回答：

•知识获取子系统

•黑板（工作场所）

•说明解释子系统（自适应）

•知识提炼系统

68．What is knowledge engineering?

什么是知识工程？

回答：

知识工程是集中活动的集合，包括从人类专家（和其他信息源）获取知识并将这些知识转换为存储库（通常称为知识库）。

69。Name and describe three problem areas suitable for expert systems.

命名并描述适用于专家系统的三个问题区域。

回答：

•解释：从观察中推断情境描述。

•预测：推断给定情况的可能后果。

•诊断：通过观察推断系统故障。

•设计：在约束下配置对象。

•规划：制定实现目标的计划。

•监控：将观察与计划进行比较并标记异常。

•调试：规定故障的补救措施。

•修复：执行计划以管理规定的补救措施。

•说明：诊断，调试和纠正学生的表现。

•控制：解释，预测，修复和监控系统行为。

70 The development of expert systems is often described as a tedious process. What activities does it typically include?

专家系统的开发通常被描述为一个繁琐的过程。它通常包括哪些活动？

回答：

•确定适当的专家

•获取知识

•选择构建工具

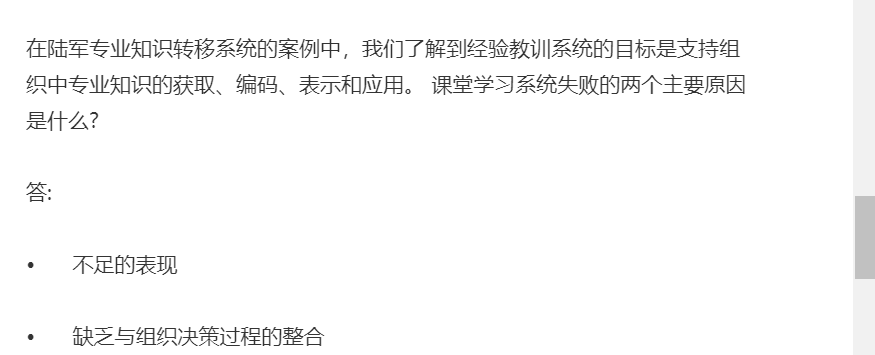
•编码系统

•评估系统

12 章

61.

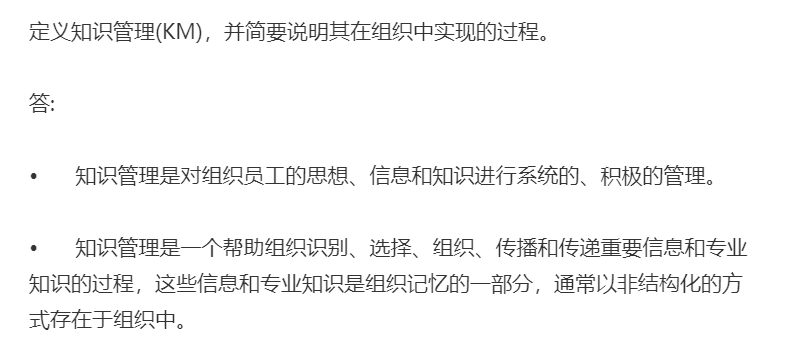
In the Army expertise transfer system case, we learn that the objective of lessons-learned systems is to support the capture, codification, presentation, and application of expertise in organizations. What are the two main reasons why lesson-learned systems have been a failure?



62.

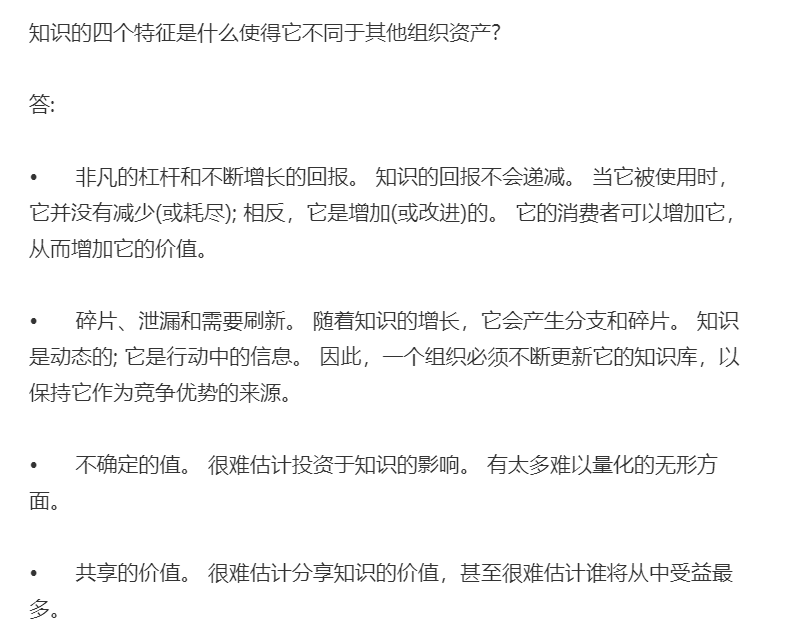
Define knowledge management (KM), and briefly explain the process through which it is implemented within an organization.

Answer:



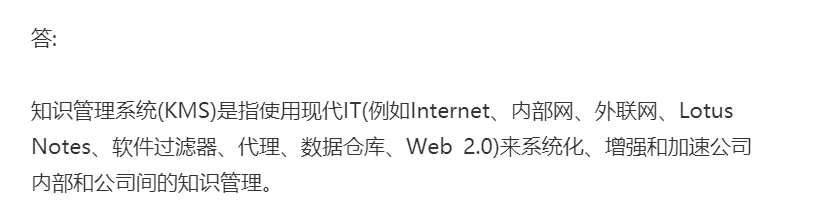
63.

What are the four characteristics of knowledge that make it unlike other organizational assets?



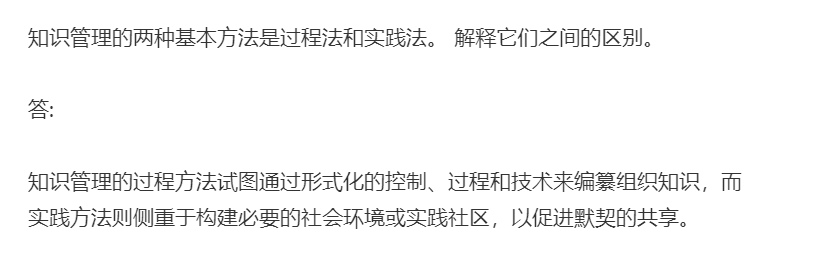
64.

One critical goal of knowledge management is to retain the valuable know-how that can so easily and quickly leave an organization. In this context, explain with examples what knowledge management systems refer to within an organization.



65.

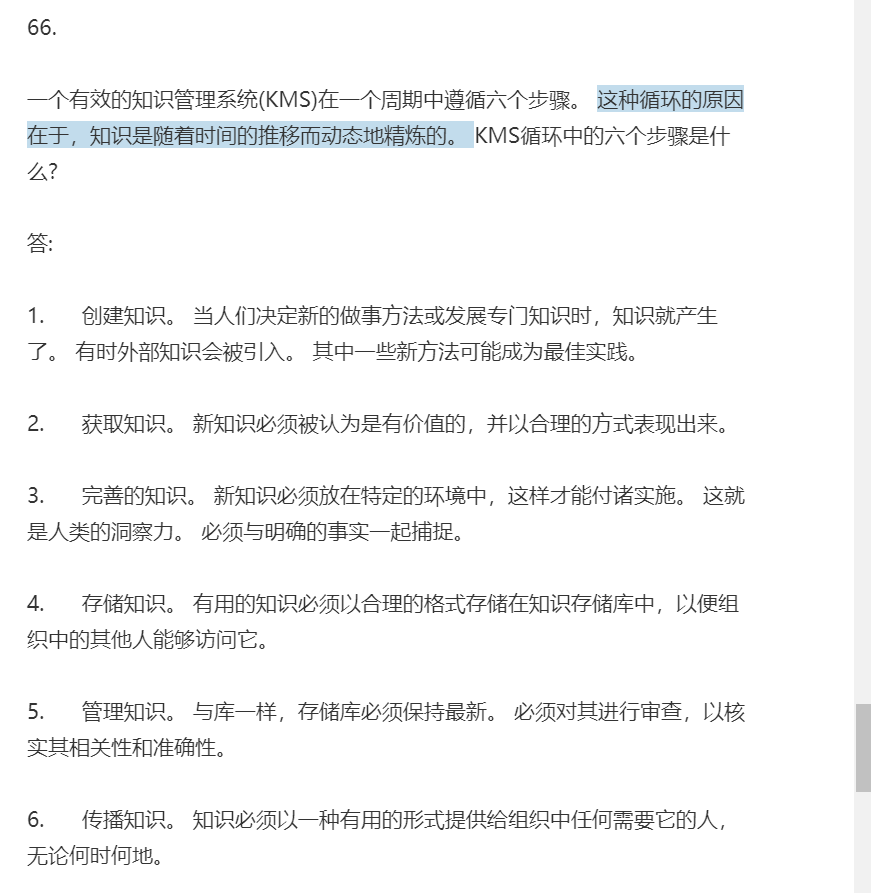
The two fundamental approaches to knowledge management are the process approach and the practice approach. Explain the differences between them.



66.

A functioning knowledge management system (KMS) follows six steps in a cycle. The reason for the cycle is that knowledge is dynamically refined over time. What are the six steps in the KMS cycle?

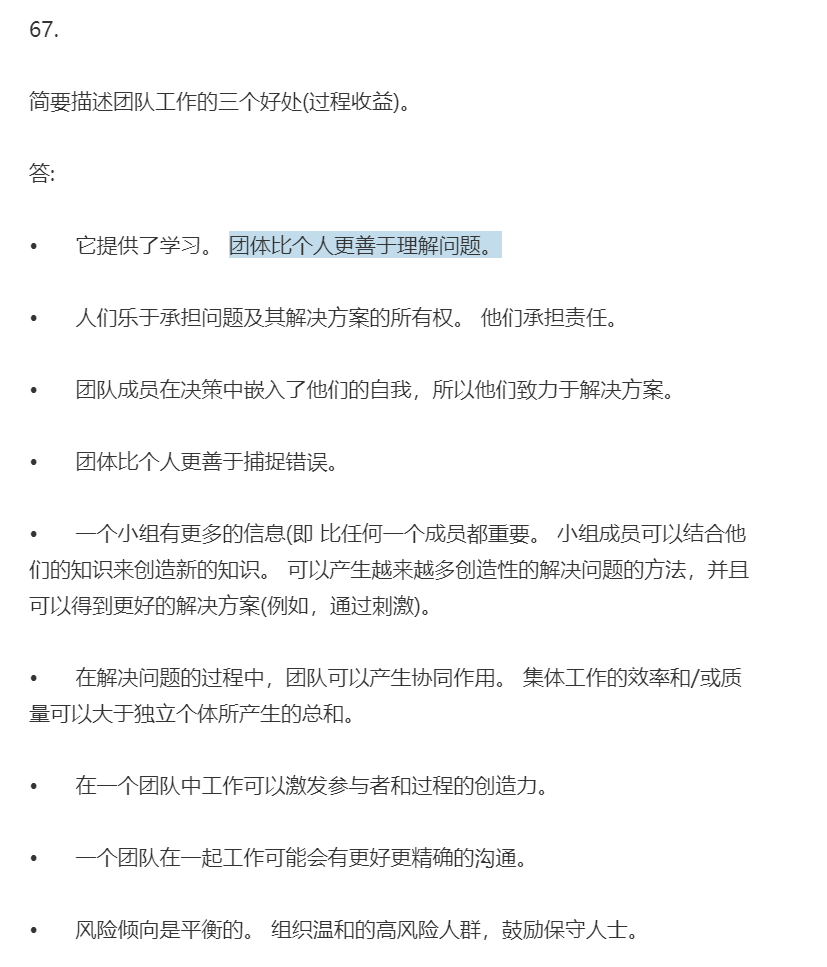
Answer:



67.

Briefly describe three benefits (process gains) derived from working in groups.

Answer:



68.

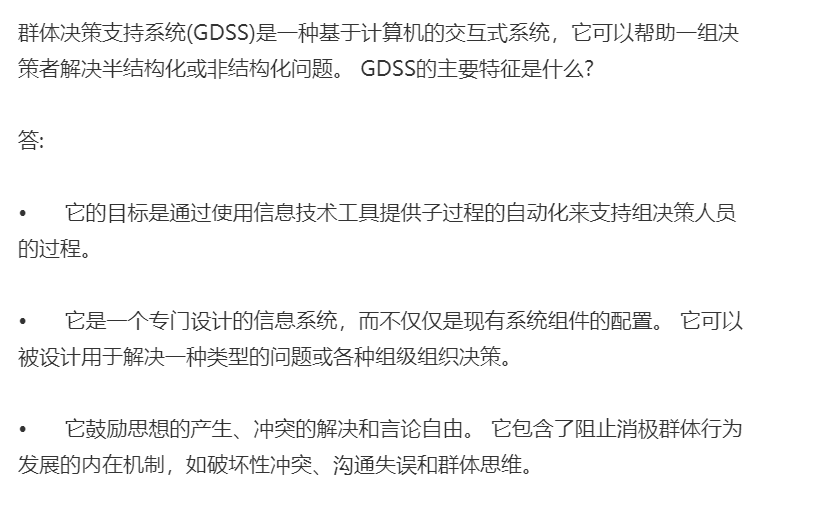
Groupware products provide a way for groups to share resources and opinions. Groupware implies the use of networks to connect people, even if they are in the same room.

What are three general features of groupware products that support commutation, collaboration, and coordination?



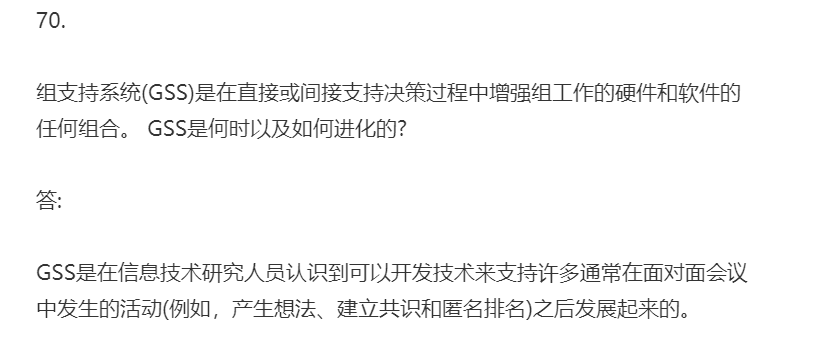
69.

A group decision support system (GDSS) is an interactive computer-based system that facilitates the solution of semi structured or unstructured problems by a group of decision makers. What are the major characteristics of a GDSS?

70.

A group support system (GSS) is any combination of hardware and software that enhances group work either in direct or indirect support of decision making. When and how did GSS evolve?

Answer:



**第十三章**

1.**欧洲核子研究组织或欧洲核研究组织收集的大数据的来源是什么**？

答：每秒四千万次，粒子在大型强子对撞机内发生碰撞，每次碰撞产生的粒子往往以复杂的方式衰变成更多的粒子。 LHC周围的精确电子电路记录每个粒子通过检测器作为一系列电子信号的通过，并将数据发送到CERN数据中心（DC）进行记录和数字重建。 数字化的数据摘要记录为“碰撞事件”。 每年生成15个PB左右的数字化摘要数据，由物理学家处理，以确定碰撞是否引发了任何有趣的物理过程。

2. **列出并描述表征大数据的三个主要“V”**

答：

•数量Volume：这显然是大数据最常见的特征。许多因素促成了数据量的指数增长，例如多年来存储的基于交易的数据，不断从社交媒体流入的文本数据，收集的传感器数据量增加，自动生成的RFID和GPS数据等等。

•多样性Variety：今天的数据有各种格式 - 从传统数据库到最终用户和OLAP系统创建的分层数据存储，到文本文档，电子邮件，XML，计量收集，传感器捕获数据，再到视频，音频和股票行情数据。据估计，所有组织的数据中有80％到85％是以某种非结构化或半结构化的形式存在的

•速度Velocity：这指的是数据的生成速度以及必须处理（即捕获，存储和分析）数据的速度以满足需求。 RFID标签，自动传感器，GPS设备和智能电表正在推动越来越需要近乎实时地处理数据流。

3. **列出并描述大数据分析的四个最关键的成功因素**

答：

•**明确的业务需求（与愿景和战略保持一致）**。商业投资应该是为了企业的利益，而不仅仅是为了技术进步。因此，大数据分析的主要驱动力应该是任何级别的业务需求 - 战略，战术和运营。

**•坚定，忠诚的赞助（执行冠军）**。众所周知，如果你没有强大的，坚定的高管赞助，那么很难（如果不是不可能）成功。如果范围是单个或几个分析应用程序，赞助可以在部门级别。但是，如果目标是企业范围的组织转型（大数据计划通常就是这种情况），则赞助需要处于最高级别和组织范围内。

•**业务和IT战略之间的一致性**。确保分析工作始终支持业务战略至关重要，而不是其他方式。分析应该在成功执行业务战略中发挥支持作用。

•**基于事实的决策文化**。在基于事实的决策文化中，数字而不是直觉，直觉或假设驱动决策。还有一种实验文化，看看哪些有效，哪些无效。为了创建一个基于事实的决策文化，高级管理层需要做到以下几点：认识到有些人不能或不会做出调整;是一个声音支持者;强调必须停止过时的方法;要求了解哪些分析可用于决策;将激励和补偿与期望行为联系起来。

•**强大的数据基础架构**。数据仓库为分析提供了数据基础架构。随着新技术的发展，这种基础设施正在改变并在大数据时代得到增强。成功需要将旧的与新的结合起来，以实现协同工作的整体基础设施。

4. **在考虑大数据项目和架构时，列出并描述设计人员应注意的五个挑战，以便使分析能力之旅的压力减轻。**

答：

•数据量：以可接受的速度捕获，存储和处理大量数据的能力，以便决策者在需要时可以获得最新信息。

•数据集成：能够组合结构或来源不相似的数据，并以合理的成本快速完成。

•处理能力：捕获数据时快速处理数据的能力。收集和处理数据的传统方式可能不起作用。在许多情况下，一旦捕获数据需要进行分析以利用最大价值。

•数据治理：能够跟上大数据的安全性，隐私，所有权和质量问题。随着数量，种类（格式和来源）以及数据速度的变化，治理实践的能力也应如此。

•技能可用性：使用新工具利用大数据，并以不同方式进行研究。缺乏具备完成工作技能的数据科学家。

•解决方案成本：由于大数据开辟了一个可能的业务改进世界，因此需要进行大量的实验和发现，以确定重要的模式和转向价值的洞察力。因此，为确保大数据项目的投资回报率具有良好的投资回报率，降低用于发现该价值的解决方案的成本至关重要。

5.**定义MapReduce**

答：正如Dean和Ghemawat（2004）所描述的那样，“MapReduce是一种编程模型，是处理和生成大型数据集的相关实现。以这种功能形式编写的程序自动并行化并在大型商用机器集群上执行。这使程序员在没有任何并行和分布式系统的经验的情况下，可以轻松利用大型分布式系统的资源。“

6. **什么是NoSQL用于大数据？ 描述其主要缺点。**

答：

•NoSQL是一种新型数据库，与Hadoop一样，可以处理大量的多结构化数据。 然而，尽管Hadoop擅长支持大规模的批量式历史分析，但NoSQL数据库的目标大部分（尽管有一些重要的例外），用于提供服务给存储在终端用户和自动化大数据应用程序中大量多结构数据的离散数据。 这种能力严重缺乏关系数据库技术，这种技术根本无法在大数据规模上维持所需的应用程序性能水平。

•当今大多数NoSQL数据库的缺点是它们在性能和可伸缩性方面交换ACID（原子性，一致性，隔离性，持久性）合规性。 许多人还缺乏成熟的管理和监测工具。

7. **什么是数据科学家，工作涉及什么？**

答：数据科学家是经常与大数据或数据科学相关联的角色或工作。 在很短的时间内，它已成为市场上最受欢迎的角色之一。 目前，数据科学家最基本的，当前的技能是编写代码的能力（在最新的大数据语言和平台中）。 一项更持久的技能是数据科学家需要用所有利益相关者都能理解的语言进行交流，并展示讲故事所涉及的特殊技能，包括口头，视觉或理想两者。 数据科学家结合其业务和技术技能，调查大数据，寻找改进当前业务分析实践的方法（从描述性到预测性和规范性），从而改善新业务机会的决策。

8. **为什么今天磁带备份工作负载的某些部分被重定向到Hadoop集群？**

答：

•首先，虽然在磁带上存储数据可能看起来很便宜，但真正的成本却来自于检索的难度。这些数据不仅是离线存储的，还需要数小时甚至数天来恢复，但磁带盒本身也会随着时间的推移而降级，从而使数据丢失成为现实，并迫使公司将这些成本考虑在内。更糟糕的是，磁带格式每隔几年就会发生变化，要求组织执行大规模数据迁移到最新的磁带格式，否则就有可能无法从过时的磁带恢复数据。

•其次，已经证明，保持历史数据在线和可访问是有价值的。与点击流示例一样，将原始数据保留在旋转磁盘上的时间更长，这使得公司可以在上下文更改并需要应用新约束时轻松重新访问数据。使用Hadoop搜索数千个磁盘比旋转数百个磁带快得多，也更容易。此外，随着磁盘密度每18个月持续翻倍，组织在HDFS中持有多年的原始或精炼数据变得经济可行。

9. **流分析和永久分析之间有什么区别？ 你什么时候使用其中一个？**

答：

•在许多情况下，它们同义使用。 但是，在智能系统的背景下，存在差异。 流分析涉及将事务级逻辑应用于实时观察。 适用于这些观察的规则考虑到以前的观察，只要它们发生在规定的窗口中; 这些窗口具有一些任意大小（例如，最后5秒，最后10,000个观察等）。 另一方面，永久分析会根据所有先前的观察结果评估每个进入的观察结果，其中没有窗口大小。 认识到新观察如何与所有先前的观察相关，使得能够发现实时洞察力。

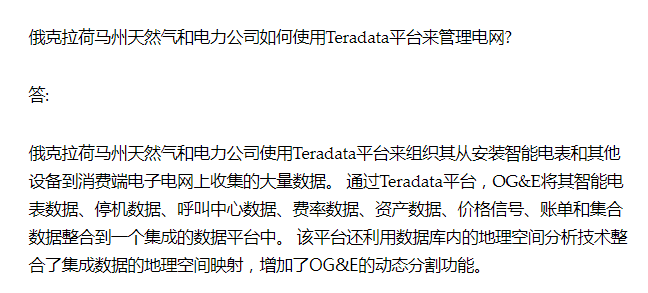
•当事务量很高且决策时间太短时，支持非持久性和小窗口大小，这转化为使用流分析。 但是，当任务至关重要且可以实时管理交易量时，永久分析就是更好的答案。

10. **描述数据流挖掘及其使用方法**

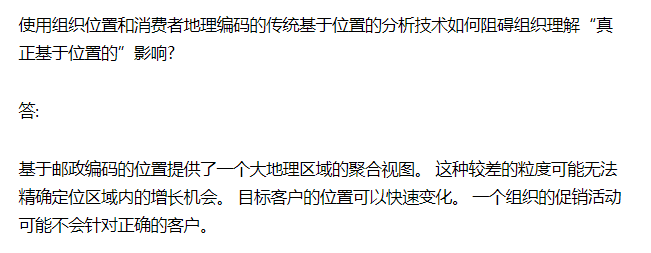
答：数据流挖掘作为流分析的支持技术，是从**连续，快速的数据记录**中**提取新模式和知识结构**的过程。 数据流是有序的实例序列的连续流，其在数据流挖掘的许多应用中可以使用有限的计算和存储能力仅读取/处理一次或少量。 数据流的示例包括传感器数据，计算机网络流量，电话对话，ATM交易，网络搜索和财务数据。 数据流挖掘可以被认为是数据挖掘，机器学习和知识发现的子领域。 在许多数据流挖掘应用程序中，目标是在给定**关于类成员资格或数据流中先前实例的值**的一些知识的情况下预测数据流中的**新实例的类或值**。

**第十四章**

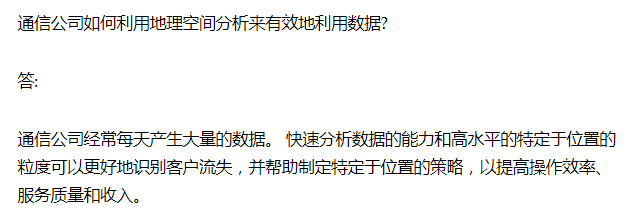
61. How does Oklahoma Gas and Electric use the Teradata platform to manage the electric grid？



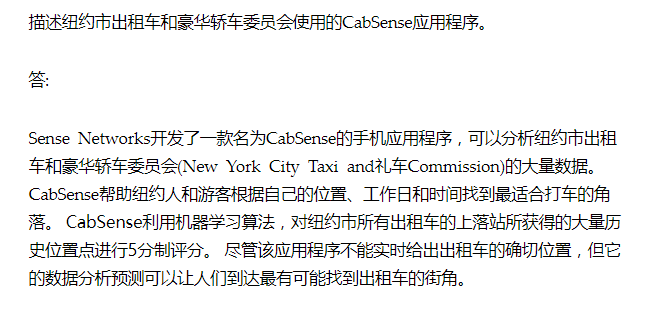
62. How do the traditional location-based analytic techniques using geocoding of organizational locations and consumers hamper the organizations in understanding "true location-based" impacts？



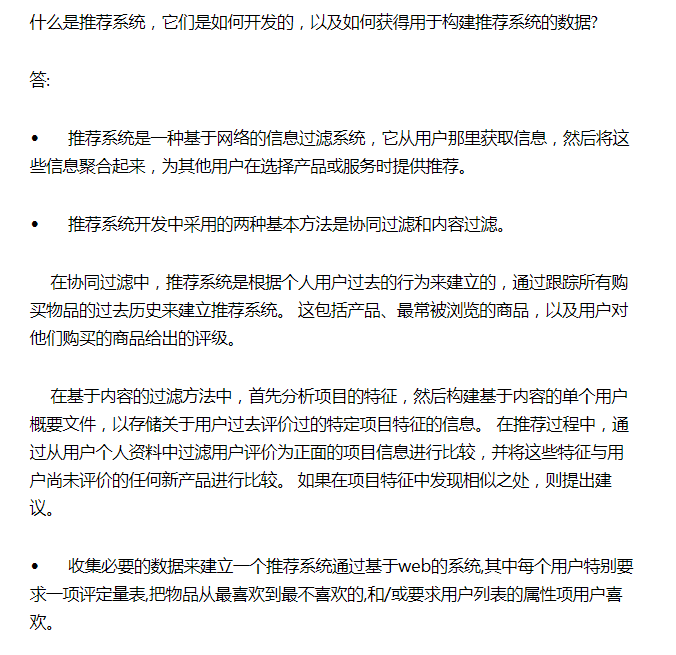
63．In what ways can communications companies use geospatial analysis to harness their data effectively?



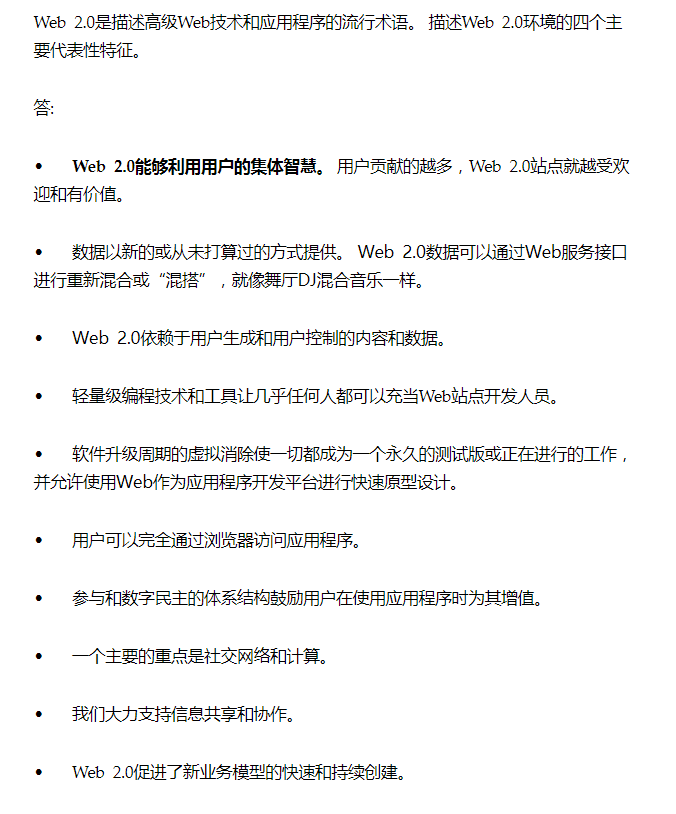
64. Describe the CabSense application used by the New York City Taxi and Limousine Commission。



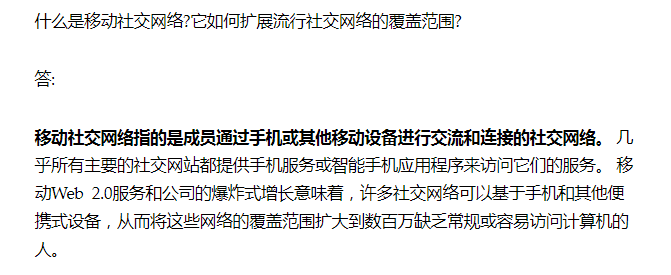
65. What are recommender systems, how are they developed, and how is the data used to build a recommendation system obtained。



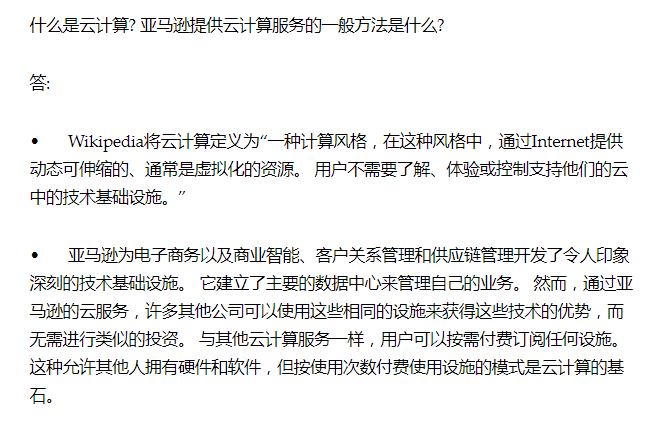
66. Web 2.0 is the popular term for describing advanced Web technologies and applications. Describe four main representative characteristics of the Web 2.0 environment



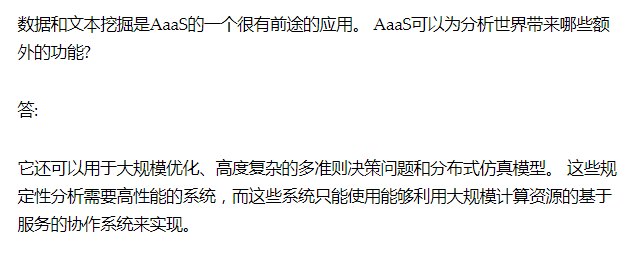
67. What is mobile social network and how does it extend the reach of popular social networks?



68. What is cloud computing? What is Amazon's general approach to the cloud computing services it provides?



69. Data and text mining is a promising application of AaaS. What additional capabilities can AaaS bring to the analytic world



70. Describe your understanding of the emerging term *people analytics*. Are there any privacy issues associated with the application.

