

**BIG DATA**

## **2.科学研究的四种范式**

# 什么是范式

---

• 范式 (Paradigm) 由托马斯·库恩《科学革命的结构》提出, 在某一个指定时间内一系列限定某一个科学学科的活动。他在其书中定义科学范式为:

- what is to be observed and scrutinized
- the kind of questions that are supposed to be asked and probed for answers in relation to this subject
- how these questions are to be structured
- what predictions made by the primary theory within the discipline
- how the results of scientific investigations should be interpreted
- how an experiment is to be conducted, and what equipment is available to conduct the experiment.

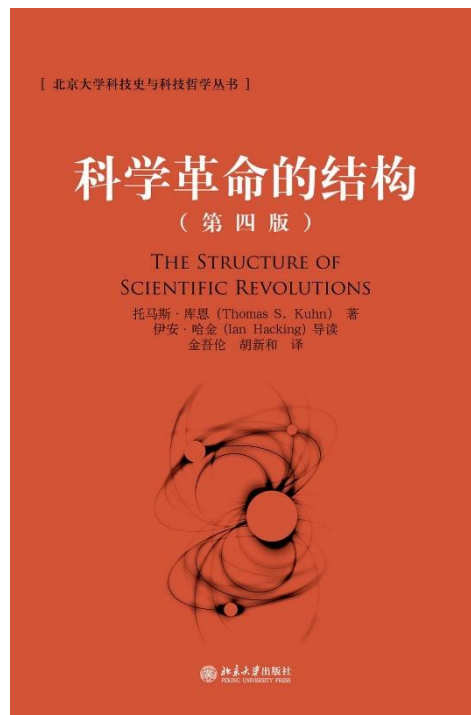
# 什么是范式

---

- 牛津百科全书中对范式的定义是：
- 一种科学学派或学科的哲学或理论框架，在这个框架内来制定理论(theories)、规律(laws)、概括(generalizations)以及为支持它们而进行的实验。
- 广义上说：任何种类的哲学或理论框架。

# 科学革命的结构

- 《科学革命的结构》（The Structure of Scientific Revolutions），1962年出版，于1970年再版（英文简称为SSR），是美国科学史家与科学哲学家托马斯·库恩的著作。
- 库恩挑战当时对“常态科学”历程的主流观点：常态科学的历程被认为是由公认事实和理论“累积而发展”，而他认为**相对于常态科学这样累积连续概念的期间中，另外有一段被革新科学所中断的模式。**
- 在科学革新历程中发现的“异常现象”导出了新的范式，然后以此范例质疑旧数据的新问题，超过之前范式单纯的“解谜”，改变研究规则并指导新研究的“地图”。



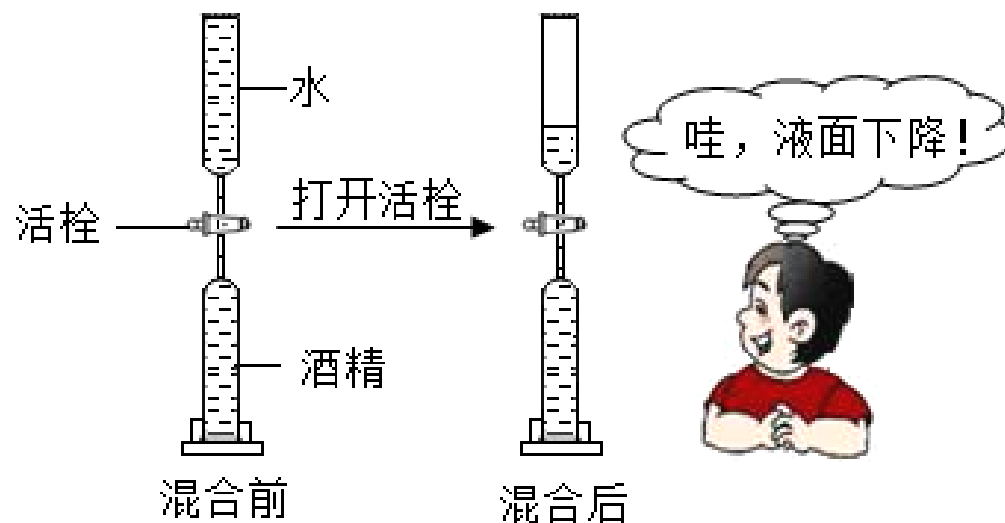
# 科学研究和发现的过程

---

- 在库恩此书的出版之前，关于科学研究和发现的过程，已经有许多想法被提出了。
- 路德维希·弗莱克在他的书中提到：**科学事实的起源与发展，提出了科学知识社会学的第一个系统**。他表示想法的交流会引导一个集体思想的建立，当其充分发展时，会分化为深奥（专业）和通俗（外行）的圆。
- 卡尔·波普尔爵士的“证伪” (falsificationism)：它强调区分什么是科学的或不科学的，是否能被证伪是最重要的标准。

# 酒精和水会反应吗？

• 十八世纪的科学家认为，同质的溶液是化学合成物。因此水和酒精的混合，一般被分类为**化合物**。水和酒精不会自动地分离，蒸馏过程也不能完全分离(它们是共沸混合的)。水和酒精能以任何比例结合使用。



在这种范式下科学家认为，化学反应（诸如水和酒精的组合）并不需要固定的比例就可发生。

但这种信念最后被道尔顿的原子论推翻，它宣称原子仅能在简单的，整数比例下才能结合。根据这个新的范式，没有在固定比例发生的反应，不能成为化学过程。如今它被认为是一种溶液，已无理由去质疑它是不是一种化合物。

在科学社群间，这种类型的世界观转变，展示了库恩所谓的**范式转移**。

# 哥白尼革命

---

- 哥白尼提出的日心说在开始时，和托勒密的地心系统比较，对天文事件如行星的轨道位置，并没有更精确的预测。
- 而是呼吁一些研究人员未来将可以发展出更好，更简单的解决方案基础。
- 库恩将某一优越革新的核心概念，称为它的“典范”，而使得这个词语在20世纪下半叶被广泛地类比使用。
- 库恩坚持认为典范转移是研究社群，热忱和科学承诺的综合，而不是一确定的逻辑程序，他的说法引起了一片哗然。

# 本章内容

---

- 2.1 科学研究第一范式
- 2.2 科学研究的第二范式
- 2.3 科学研究的第三范式
- 2.4 科学研究的第四范式



## 2.1 科学研究第一范式

科学研究第一范式产生于几千年前，当人们不满足仅在自然条件下观察对象，而要求对被研究对象进行更积极的干预时，就导致科学实验的产生。科学实验主要描述自然现象，以观察和实验为依据的研究，可称之为经验范式。科学实验是人们为实现预定目的，在人工控制条件下，通过干预和控制科研对象而观察和探索科研对象的规律和机制的一种研究方法。它是人类获得知识、检验知识的一种实践形式。科学实验是观察的一种形式。基于铁球比羽毛更快落地的观察，亚里士多德总结说越重的物体降落越快，这就是最典型的基于经验的科学发现范式。

严格意义上的科学实验是从近代开始的，其根本原因在于工业生产得到了长足的发展，实验方法的运用形成了近代自然科学的主要特点。

事实不但提供了大量可供观察的材料，而且也提供了与以往完全不同的实验手段，并使新的工具的制造成为可能。近年来，科学实验经历了很大发展，科学实验的社会性也逐步提高。科学实验的规模越来越大，科学实验再也不是科学家个人的事业，而成为整个社会事业的一部分。科学实验不仅是搜集科学事实、获得感性材料的基本方法，也是检验科学假说，形成科学理论的实践基础，二者互为补充。但实验在认识自然中，具有独特的认识功能。多种仪器的使用，使获得的感性材料更为丰富与精确，而且能排除次要因素的干扰，更快而准地揭示被研究对象的本质。

# 第一范式的第一个阶段

---

- 第一阶段（17世纪以前）不区分自然科学与社会科学，对社会现象的观察较笼统，把所有知识大一统于“自然哲学”的体系之内，为自然和社会现象提供同一套解释系统。
- 对社会认知进行哲学思辨，建立了朴素的唯物主义和唯心主义理念论、早期辩证法、演绎法、三段论与归纳证明、有机论的自然观和经验论等。
- 对推动后来的科学发展起到了巨大的作用，许多知识至今仍然是我们认识社会的出发点和基本准则。

# 第一范式的第二阶段

---

- 第二阶段（20世纪60年代以来），即现代社会科学定性分析。
- 通常涉及：
  - （1）观察和记录事实；
  - （2）分析、比较和分类；
  - （3）归纳概括事实间的关系；
  - （4）接受进一步检验四个步骤，是“自下而上”的研究路径。
- 定性研究在对定量研究的批判中逐渐发展起来，形成了独特的概念体系、具体方法和理论，开发了规范化的操作程序和研究工具，个案研究、扎根理论和叙事探究等定性研究设计类型也得以使用，并出现了“参与”和“倡导”实践。

## 2.1.1 科学实验特点

---

- 纯化观察对象条件
  - 排除干扰项
- 强化观察对象
  - 创造实验条件
- 可重复性
  - 在不同时间、不同地点，相同的实验条件可以得到相同的结论

## 2.1.2 科学实验步骤

---

- 观察
  - 观察事实和事件的详细记录
- 定义
  - 对问题进行定义，并且定义可操作
- 假设
  - 对一种事物或一种关系的暂时性解释
- 检验
  - 收集证据、假设检验
- 发表
  - 发表研究成果
- 建构
  - 建立理论（可证伪）

## 2.1.3 科学实验分类

---

- 基于实验目的分类：定性、定量、结构分析
- 基于手段作用分类：直接、间接、建模
- 基于对象性质分类：物理、化学、生命
- 基于预定目标分类：定性、定量、测量、对照、验证性
- 基于实验对象分类：黑箱、灰箱、白箱

## 2.1.4 科学实验的构成

---

- 实验者
  - 组织、设计和进行科学实验的人
- 实验对象
  - 可以是自然界的物体或现象，也可以是人造的物体及现象
- 实验手段
  - 实验仪器、设备等

## 2.1.5 科学实验的程序

---

- 1、准备阶段
  - 确立实验目的
  - 着手实验设计
  - 实验仪器、设备与材料的准备
- 2、实施阶段
  - 使用各类设备进行实验
- 3、结果处理阶段
  - 对实验结果进行分析



## 2.1.6 科学研究第一范式使用原则

---

- 1、掌握理论
- 2、提出假设
- 3、精心设计
- 4、做好准备
- 5、保持状态
- 6、控制因素
- 7、仔细观察
- 8、反复实验
- 9、核对结论

## 2.2 科学研究第二范式

---

- 以建模和归纳为基础的理论学科和分析范式。
- 特点：对某种经验现象或事实的科学解说和系统解释。
- 经验现象：
  - 可以直接感觉，或通过观测、计量等获得的信息证明
- 理论：

理论主要是解释事实，而不是描述事实。而且这种解释是对个别的、特殊的事件所做的一般的、概括的总结性归纳论述，这就可以通过一般性的定理、规则来认识个别事物。理论表现为一系列具有内在联系的概念、原理或基本命题，是系统化的观点体系。应说明的是，在真理面前任何科学理论都是试探性的、暂时的、猜测性的，而且是试探性的假说或真理的近似。

## 2.2.1 科学理论的特征

---

- 1、抽象性

抽象性是科学理论的重要特征，是对经验事实的简化与概括。因为任何事实或现象都受到多种因素的影响，在分析事实时，借助于理性思维的抽象对事实进行必要的简化，而使事实易于分析，并以纯粹的形态呈现出来。理论是关于经验事实的思想构造，并非现实但又合乎规律，用于解释具体的经验现象或直接用于指导具体问题的解决的是微观理论。在最为典型的意义上，一个陈述经验现象两个变量之间关系的命题就是一个微观理论。介于二者之间的中观理论是基于宏观理论并以某一类现象或某一方面的现象为研究对象，提供一种相对具体的分析框架。

## 2.2.1 科学理论的特征

---

- 2、逻辑性

科学理论建立在明确的概念、恰当的判断、正确的基于因果关系的推理与严密的逻辑证明基础之上。因此，科学理论具有严密的逻辑性。

## 2.2.1 科学理论的特征

---

### • 3、系统性

由于科学理论是一种系统化的逻辑体系，因此，系统性也是科学理论的重要特征。由于科学理论建立在对经验事实的简化，即一定的前提假设之上，所以任何科学理论都有其特定的边界或有效范围。换言之，科学理论只有当各种特定条件得到满足时，它才是有效的。为了使一个理论可以应用于某种给定情况，环境应合理地近似于理论的前提假设。显然，当一个特定的理论不适合某种特定情况时，表明这个理论不能应用于那种情况。

## 2.2.2 科学理论的结构

### 1. 概念

概念是人们对经验事实获得丰富而充分的感性认识基础上经反复抽象思维而形成的反映事物本质联系的逻辑形式，是经验事实在人们头脑中的重构。它纯粹是思辨性的、逻辑形态的，具有抽象特征。建立概念是理论建构必要的而且是最重要的环节。概念作为对经验事实的规定，是建立前提假设的一种形式。

### 2. 联系

一个科学理论所使用的概念往往不是单一的，而是一个甚至若干个概念群。这些概念按着固有的隶属关系、包含关系、并列关系、联结关系等，形成一个有序的概念网络或概念体系，并构成理论论述体系的关节点。理论论述体系就是在概念由抽象到具体的转化或运动中建立起来的，理论就是概念的展开。

### 3. 原理

原理是指以概念为基础，对事实或现象进行分类和分析，概括或假设它们之间的逻辑关系，并能给予合理的解释。原理是对经验事实基本关系的反映，是一种表现为科学判断的思维形式。原理一般用全称肯定判断来陈述。

基于概念和原理推演出来的逻辑结论，即各种具体的特殊规律和预见。其中，基本概念和基本原理或命题构成了理论的核心元素。尽管理论有大有小，有粗有精，但堪称理论的，一定包含概念与原理这两个核心元素，它们通常也是识别某一科学理论的主要判别标准。

## 2.2.3 科学理论的价值

---

- 解释和预测功能
- 1、为审视或界定问题提供了可选的背景知识
- 2、有助于将问题概念化或解构化
- 3、有助于思想过程的准确性
- 4、提供了确切的交流手段

## 2.2.4 科学理论体系建立的一般方法

---

- 1、从抽象上升到具体
- 2、公理化方法
- 3、逻辑与历史相统一的方法



# 从抽象上升到具体

## 1. 从抽象上升到具体的方法

从抽象上升到具体的方法是将科学研究已经获得的结果(概念、原理、规律等),按照从低级到高级、从简单到复杂、从抽象到具体的上升过程加以系统化,进而构造一个严密的科学理论体系的方法。从抽象上升到具体的方法特征如下所述。

### 1) 上升起点的简单性

构成理论体系的逻辑起点或初始细胞应是简单的抽象规定,但它并不是主观随意确定的空洞抽象,而是反映对象所共有的一般属性和关系。

### 2) 上升过程的层次性

上升过程的内在逻辑机制按序展开。从起点开始,思维程序应从简单内容到复杂内容,从较低层次上升到较高层次,从抽象规定推演到具体结果,依次展开,不能跳过中间环节和相邻层次。每上升一步,初始细胞便包含在上升的结果中,而结果又包含进一步上升的机制,直至理论体系的建立。

### 3) 上升结果的具体性

上升过程凭借逻辑演绎来推导和证明上升的结果,还需不断接触现实,补充实际例证,并用科学事实来验证每一步上升结果的真理性。这就是说,从抽象上升到具体,不仅是一个演绎过程,而且也是一个综合过程,这是科学的上升法不同于纯粹演绎法的地方。

# 公理化、逻辑与历史相统一

## 2. 公理化方法

用公理化方法构建科学理论体系是该方法在科学认识中的重要作用之一，从尽可能少的基本概念、公理、公设出发，运用演绎推理规则，推导出一系列的命题和定理，并依次排列建立整个理论体系的方法。

## 3. 逻辑与历史相统一的方法

无论自然本身的发展过程还是人对自然认识的发展过程都按照从低级到高级、从简单到复杂的方向运动，表明了逻辑与历史的统一，即从最简单上升到复杂这个抽象思维的进程符合现实的历史过程。一个成熟的科学理论体系，不仅要实现从抽象到具体的上升，同时也应达到历史和逻辑的统一。逻辑和历史相统一的方法就为构造科学理论体系提供了更加切实的方法论手段。对于逻辑和历史的统一，历史是第一性，逻辑是第二性的，是对历史的理论概括。历史的东西决定逻辑的东西，逻辑的东西是由历史的东西派生出来的，二者辩证统一。

运用逻辑的和历史的统一方法建立科学理论体系方式如下：

(1) 按逻辑发展程序和自然本身的历史过程相一致的原则建立理论体系。一般说来，经验性较强的自然科学理论体系是采用这种方法建立的。

(2) 按逻辑发展程序和人类认识自然历史过程相一致的原则建立理论体系。一般说来，理论性较强的学科理论体系采用这种方法建立。

## 2.3 科学研究第三范式

科学研究第三范式产生于几十年前，计算机的诞生催生了第三范式的出现，阿兰·麦席森·图灵(Alan Mathison Turing, 1912.6.23—1954.6.7)是英国数学家、逻辑学家，他被视为计算机之父，如图 2-3 所示。

科学研究第三范式是以模拟复杂现象为基础的计算科学范式，又称之为模拟范式。程序模拟是一种能用来帮助用户在不确定条件下进行决策的方法。用户必须在不完全了解事件的发生及其影响如何的情况下，从若干方案中选出一种行动方案来。如果出现特殊事件，将会有何结果，这也有不确定性。

模拟有三种方法，数学模拟方法、程序模拟方法和物理模拟方法。其中数学模拟方法比较抽象，物理模拟方法的代价较高，程序模拟方法处在两者之间。数学模拟方法与程序模拟方法是常采用的方法。数学模拟方法是指以数学为工具的科学研究方法，用数学语言表达事物的状态、关系和过程，并经过推导形成解释和判断。这种方法的基本特征是高度抽象、高度精确，具有普遍意义。数学语言是科学研究的简洁精确形式化语言。



图 2-3 著名计算机科学家阿兰·席森·图灵



## 2.3.1 概述

当设计和构造复杂的系统时，或者研究自然界、人类社会中漫长的演变过程和不易重复试验的事物时，如果对研究对象本身进行试验，从时间、人力、物力等因素考虑要付出昂贵的代价，甚至不可能进行。因此，需要制造一个模型来进行各种试验。

为了对系统模拟，首先要确定或表达所要研究的系统。用数学模型能较方便地确定一个系统，全面地反映对系统的已有认识或需要验证的假设，但缺乏直观性，也不便于进行试验。在数学模型的基础上，可进一步作出实物模型，它体现人们所要求的真实系统有关的性质，但在形式和规模上不必与真实系统完全一致。用实物模型试验比较直观、可信，但仍不够经济和方便。

计算机出现以后，可以把数学模型编制成计算机程序，提供新的、通用的试验方法。计算机也可用于模拟与运筹有关的活动。它的应用领域很快就扩展到各种类型的系统，从规模巨大的系统一直到小型的系统，这些系统的数学描述常常非常复杂，要给出完全的解析解或精确的数值解非常困难。程序模拟通过反复试验，帮助人们了解系统的性能，检验预想的假设，进行系统分析、设计、预测或评估，还可提供相当逼真的环境，借以培养和训练人员。程序模拟已成为工程研制、自然学研究、经济和社会问题研究、教学训练活动、军事研究、组织管理等许多领域中的一个有力的工具。著名的四色问题就是利用程序模拟方法成功解决的。

## 2.3.2 系统模拟

---

- 离散系统
  - 数字信号
- 连续系统
  - 微分方程
- 模拟语言
  - 高级语言如java或python等

## 2.4 科学研究第四范式

图灵奖获得者、美国计算机科学家詹姆斯·尼古拉斯·吉姆·格雷 2007 年 1 月 11 日在计算机科学与电信委员会上的最后一次演讲中描绘了关于科学研究第四范式的愿景。这个范式成为由实验、理论与仿真所主宰的早期历史阶段的符合逻辑的自然延伸。

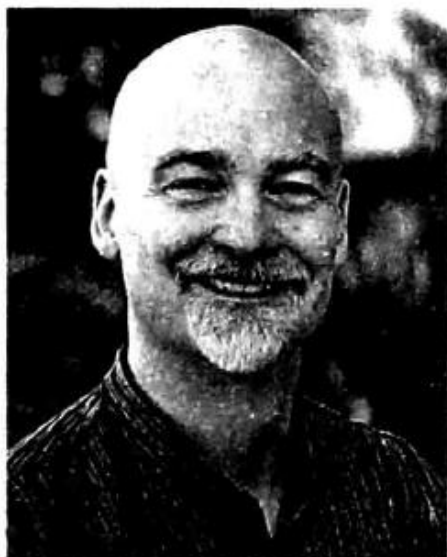


图 2-4 詹姆斯·尼古拉斯·吉姆·格雷博士

詹姆斯·尼古拉斯·吉姆·格雷博士生于 1944 年，如图 2-4 所示。2007 年 1 月 28 日在海上失踪，他因其作为一个程序员、数据库专家、工程师与研究者的工作而广受尊重。1969 年吉姆成为加利福尼亚大学伯克利分校第一个获得计算机科学博士学位的人，曾在多个著名高科技公司工作，包括贝尔实验室、IBM 研究院、天腾计算机公司、数字设备公司，最后加盟硅谷的微软研究院。

采用传统的第一、第二、第三范式的方法，直接研究密集型数据本身已经无法进行模拟推演，无法通过主流软件工具在合理的时间内抽取、处理、

管理并整合成为具有积极价值的服务信息。正在出现的科学研究第四范式是以数据考察为基础，联合理论、实验和模拟一体的数据密集计算的范式，数据被捕获或者由模拟器生成，利用软件处理，信息和知识存储在计算机中，科学家使用数据管理和统计学方法分析数据库和文档，称之为数据密集型范式。



## 2.4.1 数据密集型计算

---

数据量的急剧增长以及对在线处理数据能力要求的不断提高，使海量数据的处理问题日益受到关注。源于自然观测、工业生产、产品信息、商业销售、行政管理和客户记录等的海量数据在信息系统中所扮演的角色正在从“被管理者”向各类应用的核心转变，并已经成为企业和机构的最有价值的资产之一。其典型特点是海量、异构、半结构化或非结构化。通过网络提供基于海量数据的各类互联网服务或信息服务，是信息社会发展的趋势。这一趋势为业界和学术界提出了新的技术和研究问题。这类新型服务的重要特征之一是它们都是基于海量数据处理的。在这种背景下，数据密集型计算作为新型服务的支撑技术引起广泛关注。

# 数据密集型计算

---

- 1、万维网应用
  - 无论传统应用还是新兴应用，都是以海量数据为基础，以数据处理为核心构建互联网服务
- 2、软件即服务
  - 通过提供公开的软件服务接口，使用户能够在公共的平台上得到定制的软件功能。
- 3、大型企业的商务智能应用
  - 跨区域分布的企业进行统一管理和全局决策。



# 数据管理

数据密集型计算系统中的数据管理问题是核心问题。其与传统的数据管理问题相比，在应用环境、数据规模和应用需求等方面有本质区别，体现在数据、处理技术，以及应用提供方式三个方面。

数据密集型计算处理的是海量、快速变化、分布和异构的数据，数据量一般是 TB 级甚至是 PB 级的，因此传统的数据存储和索引技术不再适用。地理上的分散性、模型和表示方式的异构性给数据的获取和集成带来了困难。数据的快速变化特性要求处理必须及时，而传统的针对静态数据库或者数据快照的数据管理技术已无能为力。

数据密集型计算中“计算”的含义是多元的。它既包括搜索、查询等传统的数据处理，也包括分析和理解等“智能”处理。数据密集型计算所需要的数据分析和理解不仅是单一的数据分析或挖掘算法，这些算法必须能够在海量、分布和异构数据管理平台上高效地实现。数据特性决定了不可能为每一个数据分析和理解任务从存储和索引开始开发新的算法。因此，数据密集计算需要的是与存储和管理平台紧密结合的、具有高度灵活性和定制能力的、易用的数据搜索、查询和分析工具。使用这一工具，用户可以构造复杂的数据分析甚至理解应用。由于数据密集型计算要求在海量存储和高性能计算平台上实现，因此数据密集型计算通常无法在本地提供服务。有效方式是以 Web 服务方式提供应用接口。然而和传统的高性能计算不同，用户的要求可能包括从数据获取到预处理再到数据的分析、处理的整个过程，可能涉及复杂的流程。因此，数据密集型计算应用的服务接口必须提供整体流程的描述功能，并提供良好的客户机与服务器之间的基于 Web 服务的交互功能。

## 2.4.2 格雷法则

---

对于大型科学数据集的大数据工程，吉姆·格雷制定了非正式法则或规则

- (1) 科学计算日益变得数据密集型；
- (2) 解决方案为横向扩展的体系结构；
- (3) 将计算用于数据，不是数据用于计算；
- (4) 以“20 个询问”开始设计；
- (5) 工作至工作。

## 2.4.3 第四范式的核心内容

---

- 1、科学研究范式的演化过程
- 2、数据密集型科学的基本活动
- 3、学科的发展



# 科学研究范式的演化过程

在漫长的科学研究范式进化过程中，最初只有实验科学范式，主要描述自然现象，以观察和实验为依据的研究，又称之为经验范式。后来出现了理论范式，是以建模和归纳为基础的理论学科和分析范式，科学理论是对某种经验现象或事实的科学解说和系统解释，是由一系列特定的概念、原理(命题)以及对这些概念、原理(命题)的严密论证组成的知识体系。开普勒定律、牛顿运动定律、麦克斯韦方程式等正是利用了模型和归纳而诞生的。但是，对于许多问题，用这些理论模型分析解决过于复杂，只好走上了计算模拟的道路，提出了第三范式。第三范式是以模拟复杂现象为基础的计算科学范式，又可称为模拟范式。模拟方法已经引领我们走过了上个世纪后半期的全部时间。现在，数据爆炸又将理论、实验和计算仿真统一起来，出现了新的密集型数据的生态环境。模拟方法正在生成大量数据，同时实验科学也出现了巨大数据增长。研究者已经不用望远镜来观看，取而代之的是通过把数据传递到数据中心的大规模复杂仪器上来观看，开始研究计算机上存储的信息。

毋庸置疑，科学的世界发生了变化，新的研究模式是通过仪器收集数据或通过模拟方法产生数据，然后利用计算机软件进行处理，再将形成的信息和知识存于计算机中。科学家通过数据管理和统计方法分析数据和文档，只是在这个工作流中靠后的步骤才开始审视数据。可以看出，这种密集型科学研究范式与前三种范式截然不同，所以将数据密集型范式从其他研究范式中区分出来，作为一个新的、科学探索的第四种范式，其意义与价值重大。

# 数据密集型科学的基本活动

---

数据密集型科学由数据的采集、管理和分析三个基本活动组成。数据的来源构成了密集型科学数据的生态环境，主要有大型国际实验、跨实验室、单一实验室或个人观察实验，个人生活，等等。各种实验涉及多学科的大规模数据，例如，澳大利亚的平方公里阵列射电望远镜、欧洲粒子中心的大型强子对撞机、天文学领域的泛 STARRS 天体望远镜阵列等每天能产生几个千万亿字节(PB)的数据。特别是它们的高数据通量，对常规的数据采集、管理与分析工具形成巨大的挑战。为此，需要创建一系列通用工具支持从数据采集、验证到管理、分期和长期保存等整个流程。

# 数据密集型科学的基本活动

---

关于学科的发展，格雷认为所有学科 X 都分有两个进化分支，一个分支是模拟的 X 学，另一个分支是 X 信息学。如生态学可以分为计算生态学和生态信息学，前者与模拟生态的研究有关，后者与收集和分析生态信息有关。在 X 信息学中，把由实验和设备产生的、档案产生的、文献中产生的、模拟产生的事实以编码和表达知识的方式都存在一个空间中，用户通过计算机向这个空间提出问题，并由系统给出答案。为了完成这一过程，需要解决的一般问题有：数据获取、管理 PB 级大容量的数据、公共模式、数据组织、数据重组、数据分享、查找和可视化工具、建立和实施模型、数据和文献集成、记录实验、数据管理和长期保存。可以看出，科学家需要更好的工具来实现大数据的捕获、分类管理、分析和可视化。

## 参考资料

---

- 大数据概论, 陈明, 科学出版社, 2015
- 托马斯·库恩著, 金吾伦、胡新和译, 科学革命的结构, 北京大学出版社, 2012
- 米加宁, 章昌平, 李大宇, & 林涛. (2018). 第四研究范式:大数据驱动的社会科学研究转型. *社会科学文摘*, No.28(04), 22-24.
- Jim Gray博客: <http://jimgray.azurewebsites.net/>

## 作业：在下面两个任务中任选一个完成

---

- 1、阅读《科学革命的结构》，写不少于500字的读后感
- 2、在Jim Gray的博客上找到其演讲PPT
  - eScience Talk at NRC-CSTB meeting Mountain View CA, 11 January 2007.
  - 阅读并写不少于500字的读后感



- 小结：

- 本章通过对四种科学范式的讲解，说明了大数据科学的发展对科学知识的影响