# An experiment in plagiarism

Answer the following questions, either truly "in your own words", or by plagiarizing. (I have included some typical sources for you to copy from for each topic. If you have your smartphone or laptop with you, feel free to surf for more material!) You may plagiarize by cut & pasting, or by paraphrasing. I'm looking for 'short' answers of 10 sentences or less (so, not an essay.)

## Question #1: Explain how Popper's concept of falsifiability can be used as a demarcation criterion for science, using a simple example to illustrate.

# Falsifiability

From Wikipedia, the free encyclopedia

*"Falsify" redirects here. For other uses, see Falsifiability (disambiguation).*

**Falsifiability** or **refutability** is the logical possibility that an assertion can be shown false by an observation or by a physical experiment. That something is "falsifiable" does not mean it is false; rather, that *if* it is false, then this can be shown by observation or experiment. The term "testability" is related but more specific; it means that an assertion can be falsified through experimentation alone.

For example, "all men are mortal" is unfalsifiable, since no finite amount of observation could ever demonstrate its falsehood: that one or more men can live forever. "All men are immortal," by contrast, is falsifiable, by the presentation of just one dead man. Not all statements that are falsifiable in principle are falsifiable in practice. For example, "it will be raining here in one million years" is theoretically falsifiable, but not practically so.

Falsifiability is an important concept in science and the philosophy of science. The concept was made popular by Karl Popper, who, in his philosophical analysis of the scientific method, concluded that a hypothesis, proposition, or theory is "scientific" only if it is falsifiable. Popper asserted that unfalsifiable statements are non-scientific, but not of zero importance. For example, meta-physical or religious propositions have cultural or spiritual meaning, and the ancient metaphysical and unfalsifiable idea of the existence of atoms has led to corresponding falsifiable modern theories. A falsifiable theory that has withstood severe scientific testing is said to be corroborated by past experience, though in Popper's view this is not equivalent with confirmation and does not lead to the conclusion that the theory is true or even partially true.

Popper invented the notion of metaphysical research programs to name such ideas. In contrast to positivism, which held that statements are senseless if they cannot be verified or falsified, Popper claimed that falsifiability is merely a special case of the more general notion of criticizability, even though he admitted that refutation is one of the most effective methods by which theories can be criticized.

## Question #2: Explain the main advantage of the Kimball method of data warehousing, using a simple example to illustrate.

# Data warehouse

From Wikipedia, the free encyclopedia

A **data warehouse** is a repository of an organization's electronically stored data, designed to facilitate reporting and analysis [1].

This definition of the data warehouse focuses on data storage. However, the means to retrieve and analyze data, to extract, transform and load data, and to manage the data dictionary are also considered essential components of a data warehousing system. Many references to data warehousing use this broader context. Thus, an expanded definition for data warehousing includes business intelligence tools, tools to extract, transform and load data into the repository, and tools to manage and retrieve metadata.

Data warehousing arises in an organization's need for reliable, consolidated, unique and integrated reporting and analysis of its data, at different levels of aggregation.

The practical reality of most organizations is that their data infrastructure is made up by a collection of heterogeneous systems. For example, an organization might have one system that handles customer-relationship, a system that handles employees, systems that handle sales data or production data, yet another system for finance and budgeting data, etc. In practice, these systems are often poorly or not at all integrated and simple questions like: "How much time did sales person A spend on customer C, how much did we sell to Customer C, was customer C happy with the provided service, Did Customer C pay his bills" can be very hard to answer, even though the information is available "somewhere" in the different data systems.

Another problem is that enterprise resource planning (ERP) systems are designed to support relevant operations. For example, a finance system might keep track of every single stamp bought; When it was ordered, when it was delivered, when it was paid and the system might offer accounting principles (like double bookkeeping) that further complicates the data model. Such information is great for the person in charge of buying "stamps" or the accountant trying to sort out an irregularity, but the CEO is definitely not interested in such detailed information, the CEO wants to know stuff like "What's the cost?", "What's the revenue?", "did our latest initiative reduce costs?" and wants to have this information at an aggregated level.

…

## Top-down versus bottom-up design methodologies

### [edit]Bottom-up design

Ralph Kimball, a well-known author on data warehousing,[4] is a proponent of an approach to data warehouse design which he describes as *bottom-up*.[5].

In the *bottom-up* approach data marts are first created to provide reporting and analytical capabilities for specific business processes. Though it is important to note that in Kimball methodology, the bottom-up

process is the result of an initial business oriented Top-down analysis of the relevant business processes to be modelled.

Data marts contain, primarily, dimensions and facts. Facts can contain either atomic data and, if necessary, summarized data. The single data mart often models a specific business area such as "Sales" or "Production." These data marts can eventually be integrated to create a comprehensive data warehouse. The integration of data marts is managed through the implementation of what Kimball calls "a data warehouse bus architecture".[6]. The data warehouse bus architecture is primarily an implementation of "the bus" a collection of conformed dimensions, which are dimensions that are shared (in a specific way) between facts in two or more data marts.

The integration of the data marts in the data warehouse is centered on the conformed dimensions (residing in "the bus") that define the possible integration "points" between data marts. The actual integration of two or more data marts is then done by a process known as "Drill across". A drill-across works by grouping (summarizing) the data along the keys of the (shared) conformed dimensions of each fact participating in the "drill across" followed by a join on the keys of these grouped (summarized) facts.

Maintaining tight management over the data warehouse bus architecture is fundamental to maintaining the integrity of the data warehouse. The most important management task is making sure dimensions among data marts are consistent. In Kimball's words, this means that the dimensions "conform".

Some consider it an advantage of the Kimball method, that the data warehouse ends up being "segmented" into a number of logically self contained (up to and including The Bus) and consistent data marts, rather than a big and often complex centralized model. Business value can be returned as quickly as the first data marts can be created, and the method gives it self well to an exploratory and iterative approach to building data warehouses. For example, the data warehousing effort might start in the "Sales" department, by building a Sales-data mart. Upon completion of the Sales-data mart, The business might then decide to expand the warehousing activities into the, say, "Production department" resulting in a Production data mart. The requirement for the Sales data mart and the Production data mart to be integrable, is that they share the same "Bus", that will be, that the data warehousing team has made the effort to identify and implement the conformed dimensions in the bus, and that the individual data marts links that information from the bus. Note that this does not require 100% awareness from the onset of the data warehousing effort, no master plan is required upfront. The Sales-data mart is good as it is (assuming that the bus is complete) and the production data mart can be constructed virtually independent of the sales data mart (but not independent of the Bus).

If integration via the bus is achieved, the data warehouse, through its two data marts, will not only be able to deliver the specific information that the individual data marts are designed to do, in this example either "Sales" or "Production" information, but can deliver integrated Sales-Production information, which, often, is of critical business value. An integration (possibly) achieved in a flexible and iterative fashion.