

**Download Solved Mid/Final Term
Past Papers by Waqar Siddhu
in PDF form**

VUNET.BLOGSPOT.COM



Facebook.com/vunets



http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

Favorites Suggested Sites Web Slice Gallery

Examination VU Examination System (...)

Live Search

Home RSS Print Page Safety Tools

Question No : 41 of 52

Marks: 2 (Budgeted Time 4 Min)

Do you think it will create the problem of non-standardized attributes, if one source uses 0/1 and second source uses 1/0 to store male/female attribute respectively? Give a reason to support your answer.

Answer ([Please click here to Add Answer](#))

100%

Normal Arial 12 B I U

yes it is a major problem which makes complexity during DTS as we learn in lab lectures

PAGE NO 405

Each of the campuses uses different attribute name for the identification or primary keys e.g. Lahore uses *SID* while Peshawar uses *Reg#* and so on.
Different conventions for representing Gender across the campuses e.g. Lahore campus uses 0/1 while Islamabad uses 1/0 for representing male and female respectively.

Start Time: 5:29 PM

118:00

Time Left



41



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

Examination VU Examination System (...)

Home RSS Print Page Safety Tools

Question No : 42 of 52

Marks: 2 (Budgeted Time 4 Min)

What is the "intrinsic skew" problem in Hash based join? Give an example?

read all

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

Hash-Based Join: Intrinsic Skew
Intrinsic skew can become a problem for hash, as well as sort-merge join.
The skew is in data, NOT due to hash function.
Example: Many non-CS majors registering for CS-101 instead of CS students in summer.
Intrinsic skew occurs when attributes are not distributed uniformly; it is also called attribute value skew. For example a basic Computer Science (CS) course being offered in summer, and taken by many non-CS majors who want to know about computers. The course taken by few CS-majors who missed it or got an incomplete (i.e. I) grade during the regular semester due to one reason or another. Ironically, intrinsic skew effects the performance of both hash and sort-merge joins. Sort-merge join works best when the join attributes are the primary key of both tables. This property guarantees absence of duplicates, so that a record in the left-hand-side of the relation will join with at most one record in the right-hand-side of the relation, thus avoiding the intrinsic skew

Start Time: 5:29 PM

117:00

Time Left

42

MADE BY WAQAR SIDDHU

Question No : 43 of 52 Marks: 2 (Budgeted Time 4 Min)

List down the two ways to simplify the ER Model.

Answer (Please click here to Add Answer)

Rich text editor toolbar with icons for Bold, Italic, Underline, Bulleted List, Numbered List, Indent, Outdent, Undo, Redo, and a 100% zoom level.

Normal Arial 12

PAGE 103
two ways to simplify ER Modle

- Two general methods:
- De-Normalization
- Dimensional Modeling (DM)

Correct answer solved by hadi
Cell No:03228043306
Email: usmanraj20@gmail.com

Start Time: 5:29 PM
117:00
Time Left

43



MADE BY WAQAR SIDDHU

Question No : 44 of 52 Marks: 2 (Budgeted Time 4 Min)

List down any four Dynamic Attributes recorded by the scouts in Agriculture Data Warehouse Case Study.

Answer ([Please click here to Add Answer](#))

Rich text editor toolbar with icons for Bold, Italic, Underline, Bulleted List, Numbered List, Indent, Outdent, Undo, Redo, and a 100% zoom level.

Normal Arial 12

PAGE 342	
static attributes	dynamic attributs
1 Farmer Name	1 Date of Visit
2 Farmer Address	2 Pest Population
3 Field Acreage	3 CLCV
4 Variety(ies) Sown	4 Predator Population
5 Sowing date	5 Pesticide Spray Dates
6 Sowing method	6 Pesticide(s) Used

Start Time: 5:29 PM
117:00
Time Left

44 Navigation icons: Info, Previous, Next, Previous, Next, Print

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

Examination VU Examination System (...)

Home RSS Print Page Safety Tools

Question No : 45 of 52

Marks: 3 (Budgeted Time 6 Min)

Identify the given statement as correct or incorrect and justify your answer in either case.

"The problem of Referential Integrity always occurs in traditional OLTP system as well as in DWH".

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 189

Referential Integrity Validation
Example:
How many outstanding payments in the DWH without a corresponding customer_ID in the customer table?
While doing total quality measurement, you measure RI every week (or month) and hopefully the number of orphan records will be going down, as you will be fine tuning the processes to get rid of the RI problems. Remember, RI problem is peculiar to a DWH, this will not happen in a traditional OLTP system.

Start Time: 5:29 PM

117:00

Time Left



45



MADE BY WAQAR SIDDHU

http://localhost/YUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

Examination VU Examination System (... x)

Home RSS Print Page Safety Tools ?

Question No : 46 of 52

Marks: 3 (Budgeted Time 6 Min)

Identify the given statement as correct or incorrect and justify your answer in either case.

“One-way Clustering gives a local view whereas Two-way Clustering gives a global view”.

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 271

1. One-way Clustering-means that when you clustered a data matrix, you used all the attributes. In this technique a similarity matrix is constructed, and then clustering is performed on rows. A cluster also exists in the data matrix for each corresponding cluster in the similarity matrix.

2. Two-way Clustering/Biclustering-here rows and columns are simultaneously clustered. No any sort of similarity or dissimilarity matrix is constructed. Biclustering gives a local view of your data set while one-way clustering gives a global view. It is possible that you first take global view of your data by performing one-way clustering and if any cluster of interest is found then you perform two-way clustering to get more details. Thus both the methods complement each other.

Correct answer solved by hadi
Cell No:03228043306
Email: usmanraj20@gmail.com

Start Time: 5:29 PM

117:00

Time Left

46

MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

Favorites Suggested Sites Web Slice Gallery

Examination VU Examination System (...)

Home RSS Print Page Safety Tools

Question No : 47 of 52

Marks: 3 (Budgeted Time 6 Min)

We can identify the Session in Word Wide Web by using "Time-contiguous Log Entries" however there are some limitations of this technique. Briefly explain any two limitations.

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 365

Limitations

- This method breaks down for visitors from large ISPs because different visitors may reuse dynamically assigned IP addresses over a brief time period.
- Different IP addresses may be used within the same session for the same visitor.
- This approach also presents problems when dealing with browsers that are behind some firewalls.

Notwithstanding these problems, many commercial log analysis products use this method of session tracking, which requires no cookies or special Web server features.

Start Time: 5:29 PM

117:00

Time Left

47



MADE BY WAQAR SIDDHU

Question No : 48 of 52 Marks: 3 (Budgeted Time 6 Min)

Give an example of the company which is using Data Warehouse, for each of the following fields in Pakistan?

- 0. Financial service/insurance
- 1. Telecommunications
- 2. Transportation

Answer (Please click here to Add Answer)

Rich text editor toolbar with icons for Bold, Italic, Underline, Bulleted List, Numbered List, Indent, Outdent, Undo, Redo, and a 100% zoom level.

Normal Arial 12

- Financial service/insurance.
 - Union Bank
 - State Bank of Pakistan
- Telecommunications.
 - Ufone
 - PTCL
 - PAKNET
- Transportation.
 - PIA
- Government.
 - NADRA

PAGE 323

Correct answer solved by hadi
Cell No:03228043306
Email: usmanraj20@gmail.com

Start Time: 5:29 PM

117:00
Time Left

48

Navigation icons: Information, Previous, Home, Next, Last

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

Examination VU Examination System (... x)

Home RSS Print Page Safety Tools ?

Question No : 49 of 52

Marks: 5 (Budgeted Time 10 Min)

Describe how business rules are validated using student databases in LAB lectures?

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

Slide 31

Summary: 'Date of Birth'

- Total dates found with inconsistent formats are 3
- Total invalid dates are 4
- Total Missing dates are 9
- Business rules are yet to be validated

PAGE 451

Slide shows the summary of data profiling of 'Date of Birth'. Business rule can be as follows:

At the time of joining for BS, student must be at least 16 years old

At the time of joining for MS, student must be at least 20 years old

Business rules are yet to be validated because to validate business rule we have to have date of registration as well.

Start Time: 5:29 PM

117:00

Time Left



49



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

Examination VU Examination System (...)

Home RSS Print Page Safety Tools

Question No : 50 of 52

Marks: 5 (Budgeted Time 10 Min)

Identify the given statements as correct or incorrect and justify your answer in either case.

1. "Median is an example of Distributive Aggregate".
2. "With the help of "current value field" method we can track as many changes as we wish".

Answer ([Please click here to Add Answer](#))

100%

Normal Arial 12 B I U

1 statement wrong
holistic:

- Require unbounded amount of information about each subgroup
- Examples: MEDIAN, COUNT DISTINCT
- Usually impractical for a data warehouses!

2nd statement wrong
we can track only recent 2 changes in current value field
statement which gives refer to **preserve history**
in which this statement is true but for **current value field** it is wrong

Correct answer solved by hadi
Cell No:03228043306
Email: usmanraj20@gmail.com

Start Time: 5:29 PM

117:00

Time Left



50



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

Examination VU Examination System (...)

Home RSS Print Page Safety Tools

Question No : 51 of 52

Marks: 5 (Budgeted Time 10 Min)

What is clustering? How it can be used in Telecom Industry for targeted sale promotion?

please pray for
everyone <3

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

Clustering: Identify outlier records using clustering based on Euclidian (or other) distance. Existing clustering algorithms provide little support for identifying outliers. However, in some cases clustering the entire record space can reveal outliers that are not identified at the field level inspection. The main drawback of this method is computational time. The clustering algorithms have high computational complexity. For large record spaces and large number of records, the run time of the clustering algorithms is prohibitive. PAGE 164 and 236

Cluster Index: Concept

- A Cluster index defines the sort order on the base table.
- Ordering may be strictly enforced (guaranteed) or opportunistically maintained.
- At most one cluster index defined per table.
- Cluster index may include one or multiple columns.
- Reduced I/O.

Start Time: 5:29 PM

117:00

Time Left



51



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

Favorites Suggested Sites Web Slice Gallery

Examination VU Examination System (...)

Home RSS Print Page Safety Tools

Question No : 52 of 52

Marks: 5 (Budgeted Time 10 Min)

A pilot project strategy is highly recommended in Data Warehouse Construction. In your opinion, what could be the reasons of this recommendation?

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 334

Will adopt a pilot project approach, because:

- A full-blown DWH requires extensive investment.
- Show users the value of DSS.
- Establish blue print for full-blown system.
- Identify problem areas.
- Reveal true data demographics.
- Pilot projects are supposed to work with limited data.

Start Time: 5:29 PM

117:00

Time Left



52



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 ::::: Fall...

Home RSS Print Page Safety Tools

Question No : 41 of 52

Marks: 2 (Budgeted Time 4 Min)

List down any two disadvantages of MOLAP.

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

Drawbacks of MOLAP:

- Long load time (pre-calculating the cube may take days!). PAGE 83
- Very sparse cube (wastage of space) for high cardinality (sometimes in small hundreds). e.g. number of heaters sold in Jacobabad or Sibi.

11.5 MOLAP Implementation Issues PAGE 84

Maintenance issue: Every data item received must be aggregated into *every* cube (assuming "to-date" summaries are maintained). Lot of work.

Storage issue: As dimensions get less detailed (e.g., year vs. day) cubes get much smaller, but storage consequences for building hundreds of cubes can be significant. Lot of space.

Scalability:
Often have difficulty scaling when the size of dimensions becomes large. The breakpoint is typically around 64,000 cardinality of a dimension.

Start Time: 1:23 PM

117:00

Time Left

41

MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 ::::: Fall...

Home RSS Print Page Safety Tools

Question No : 42 of 52

Marks: 2 (Budgeted Time 4 Min)

What is Unsupervised Learning in Data Mining?

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

Chapter 31 PAGE 269

Supervised learning is when you are performing DM the supporting information that helps in the DM process is also available. What could be that information? You may know your data that how many groups or classes your data set contains. What are the properties of these classes or clusters?

When we will talk about unsupervised learning you will not have such known or a priori knowledge. In other words you can not give such factors as input to the DM technique which can facilitate your DM process. So wherever the user gives some input that is supervised else that is unsupervised learning.

Unsupervised

- One-way Clustering
- Two-way Clustering

Start Time: 1:23 PM

117:00

Time Left

42



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 :::: Fall...

Home RSS Print Page Safety Tools

Question No : 43 of 52

Marks: 2 (Budgeted Time 4 Min)

List down any four ways to Identify the Session in Word Wide Web.

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 364

Identifying the Session

- Web-centric data warehouse applications require every visitor session (visit) to have its own unique identity
- The basic protocol for the World Wide Web, HTTP, stateless so session identity must be established in some other way.
- There are several ways to do this
- Using Time-contiguous Log Entries
- Using Transient Cookies
- Using HTTP's secure sockets layer (SSL)
- Using session ID Ping-pong
- Using Persistent Cookies

Start Time: 1:23 PM

117:00

Time Left

43

MADE BY WAQAR SIDDHU

Question No : 44 of 52 Marks: 2 (Budgeted Time 4 Min)

Which authority is recording the Pest Scouting Data in Punjab?

Answer ([Please click here to Add Answer](#))

Rich text editor toolbar with icons for Bold, Italic, Underline, Bulleted List, Numbered List, Indent, Outdent, Undo, Redo, and others. The text area contains the following content:

PAGE 333

The pest scouting data is being constantly recorded by the Directorate of Pest Warning and Quality Control of Pesticides (DPWQCP), Punjab since 1984.

Start Time: 1:23 PM
117:00
Time Left

44



http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 ::::: Fall...

Home RSS Print Page Safety Tools

Question No : 45 of 52

Marks: 3 (Budgeted Time 6 Min)

Identify the given statement as correct or incorrect and justify your answer in either case.

"If defects are found in the process of Attribute Domain Validation then it is better to fix the errors in DWH and leave the data source as it is".

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 190

Step-3: Once the defects are found, go and track them down back to source cause(s). Point to be noted is that, if at all possible, fix the problem in the source system. People have the tendency of applying fixes in the DWH. This is a wrong i.e. if you are fixing the problems in the DW; you are not fixing the root cause. A crude analogy would clarify the point. If you keep cleaning the lake, and keep on flushing the toilet in the lake, you are not solving the problem. The problem is not being fixed at the source system, therefore, it will persist.

Start Time: 1:23 PM

117:00

Time Left



45



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

Favorites Suggested Sites Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 ::::: Fall...

Home RSS Print Page Safety Tools

Question No : 46 of 52

Marks: 3 (Budgeted Time 6 Min)

Identify the given statement as correct or incorrect and justify your answer in either case.

“In prediction, Test Data set is used to form the model and the associated rules.”

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 261

First of all a model is built using existing data. The existing data set is divided into two subsets, one is called the training set and the other is called test set. The training set is used to form model and the associated rules. Once model built and rules defined, the test set is used for grouping. It must be noted the test set groupings are already known but they are put in the model to test its accuracy.

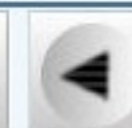
Start Time: 1:23 PM

117:00

Time Left



46



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 ::::: Fall...

Home RSS Print Page Safety Tools

Question No : 47 of 52

Marks: 3 (Budgeted Time 6 Min)

Briefly explain the three Major Approaches for accessing the information stored on the Web.

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 350

Currently, users can choose from three major approaches when accessing information stored on the Web:

- (i) **Keyword-based search** or topic-directory browsing with search engines such as Google or Yahoo, which use keyword indices or manually built directories to find documents with specified keywords or topics;
- (ii) **Querying deep Web sources**—where information, such as amazon.com's book data and realtor.com's real-estate data, hides behind searchable database query forms—that, unlike the surface Web, cannot be accessed through static URL links; and
- (iii) **Random surfing** that follows Web linkage pointers. The success of these techniques, especially with the more recent page ranking in Google and other search engines shows the Web's great promise to become the ultimate information system.

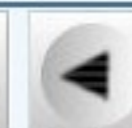
Start Time: 1:23 PM

117:00

Time Left



47



MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 :::: Fall...

Home RSS Print Page Safety Tools

Question No : 48 of 52

Marks: 3 (Budgeted Time 6 Min)

We can identify the Session in Word Wide Web by using "Time-contiguous Log Entries" however there are some limitations of this technique. Briefly explain any two limitations.

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 365

Limitations

- This method breaks down for visitors from large ISPs because different visitors may reuse dynamically assigned IP addresses over a brief time period.
- Different IP addresses may be used within the same session for the same visitor.
- This approach also presents problems when dealing with browsers that are behind some firewalls.

Notwithstanding these problems, many commercial log analysis products use this method of session tracking, which requires no cookies or special Web server features.

Start Time: 1:23 PM

117:00

Time Left

48

MADE BY WAQAR SIDDHU

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 ::::: Fall...

Home RSS Print Page Safety Tools

Question No : 49 of 52

Marks: 5 (Budgeted Time 10 Min)

In Most of the retail banks 50 % of the customers are unprofitable. By considering this situation, which particular application of DWH can help the banks to improve the business and how?

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 37**Profitability Analysis**

- Banks know if they are profitable or not.
- Don't know which customers are profitable.
- Typically more than 50% are NOT profitable.
- Don't know which one?
- Balance is not enough, transactional behavior is the key.
- Restructure products and pricing strategies.
- Life-time profitability models (next 3-5 years).

Start Time: 1:23 PM

117:00

Time Left



49

**MADE BY WAQAR SIDDHU**

http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

★ Favorites | ★ Suggested Sites | Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 ::... Fall...

Home RSS Print Page Safety Tools

Question No : 50 of 52

Marks: 5 (Budgeted Time 10 Min)

What is clustering? How it can be used in Telecom Industry for targeted sale promotion?

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

Clustering: Identify outlier records using clustering based on Euclidian (or other) distance. Existing clustering algorithms provide little support for identifying outliers. However, in some cases clustering the entire record space can reveal outliers that are not identified at the field level inspection. The main drawback of this method is computational time. The clustering algorithms have high computational complexity. For large record spaces and large number of records, the run time of the clustering algorithms is prohibitive. PAGE 164 and 236

Cluster Index: Concept

- A Cluster index defines the sort order on the base table.
- Ordering may be strictly enforced (guaranteed) or opportunistically maintained.
- At most one cluster index defined per table.
- Cluster index may include one or multiple columns.
- Reduced I/O.

Start Time: 1:23 PM

116:00

Time Left



50



MADE BY WAQAR SIDDHU

Question No : 51 of 52 Marks: 5 (Budgeted Time 10 Min)

We have seen many issues during Data Acquisition & Cleansing in Agriculture Data Warehouse Case Study. In your point of view why those issues have arisen? Briefly explain.

Answer ([Please click here to Add Answer](#))

Rich text editor toolbar with icons for Bold, Italic, Underline, Bulleted List, Numbered List, Indent, Outdent, Undo, Redo, and others. The text area contains:

chapter 38 page 340
read the context

Start Time: 1:23 PM
116:00
Time Left

51



http://localhost/VUTES/client/Instructions.aspx

File Edit View Favorites Tools Help

Favorites Suggested Sites Web Slice Gallery

VU Examination System (CLIENT) VUTES 6.5 ::... Fall...

Home RSS Print Page Safety Tools

Question No : 52 of 52

Marks: 5 (Budgeted Time 10 Min)

Briefly explain any two types of precedence constraints that we can use in DTS.

Answer ([Please click here to Add Answer](#))

Normal Arial 12 B I U

PAGE 395

Precedence constraints sequentially link tasks in a package. In DTS, you can use three types of precedence constraints, which can be accessed either through DTS Designer or programmatically:

Unconditional: If you want Task 2 to wait until Task 1 completes, regardless of the outcome, link Task 1 to Task 2 with an *unconditional precedence* constraint.

On Success: If you want Task 2 to wait until Task 1 has successfully completed, link Task 1 to Task 2 with an *On Success precedence* constraint.

On Failure: If you want Task 2 to begin execution only if Task 1 fails to execute successfully, link Task 1 to Task 2 with an *On Failure precedence* constraint. If you want to run an alternative branch of the workflow when an error is encountered, use this constraint.

Start Time: 1:23 PM

116:00

Time Left

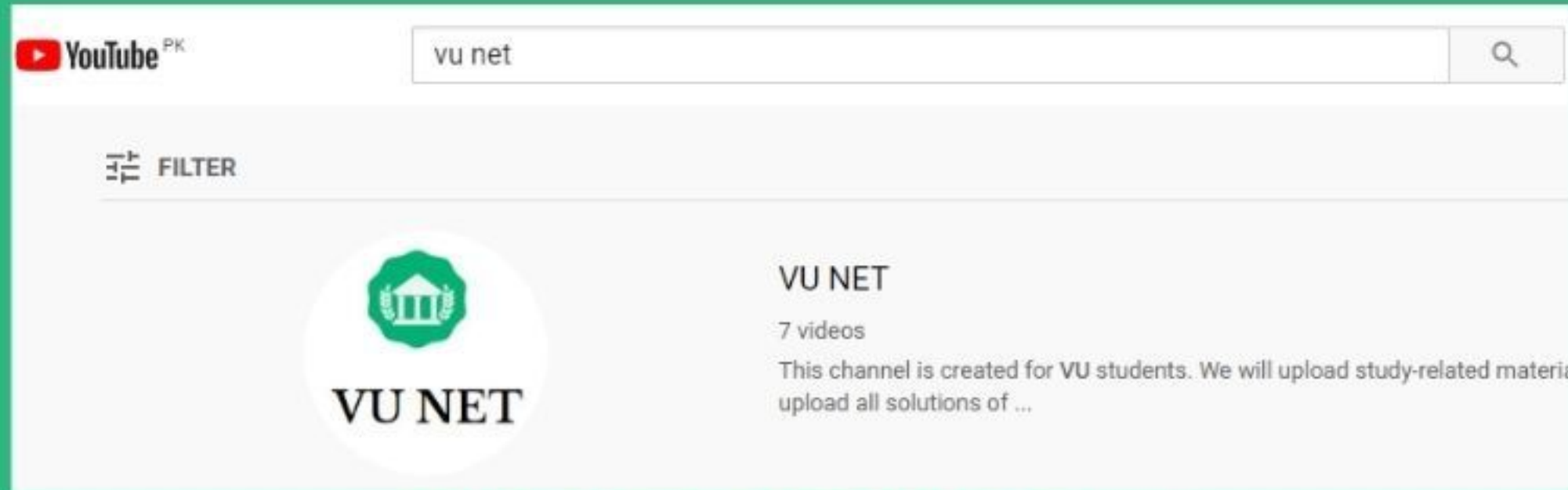


52



MADE BY WAQAR SIDDHU

For More Solutions or Short Lectures (Playlist) Visit our Youtube Channel



VUNET.BLOGSPOT.Com