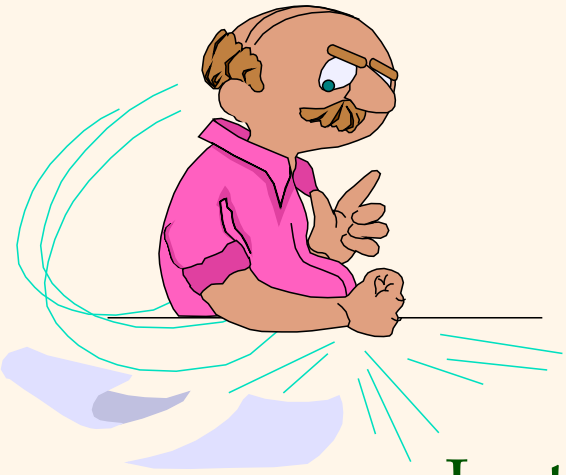
# *DBMS Models and Implementation Techniques*
# *Course Overview*



Instructor: Sharma Chakravarthy

sharmac@cse.uta.edu

The University of Texas at Arlington

# *Lectures & Office hours*

- Lectures are from 3:30 pm to 4:50 pm. WH 308
  - Will be NOT be recorded
- TA: Mr. Anamitra Roy
- Email: axr9563@mavs.uta.edu
- Office hours (in-person)
  - Mon/Wed 2 pm to 3 pm + by appointment (in ERB 501)
- TA will deal with the project, discussion on Canvas, homework, grading tests, …
- Please send all correspondence to both TA and me
  - Will not be answered otherwise
- Instructor Office hours: Tu/Th 2pm to 3pm + by appointment (send email) + on Teams if needed
  - In ERB 632 in person until further notice

# *Course Details*

- Combined class: CSE 5331-001 and CSE 4331-001
  - On Canvas

- Canvas (uta.instructure.com) will be used for everything -- lecture notes, assignments, announcements, and project submissions as well as project discussions; No email submissions; will NOT be graded
  - Course letter grades are NOT posted on Canvas
  - IGNORE course letter grades if shown by Canvas
  - Course letter grades posted only on MyMav
  - Scores of tests/projects will be posted on Canvas

- Include course number in email header; NO UTA id please
  - Send email to me and TA (otherwise WILL NOT be answered!)

- It is your responsibility to check for material (announcements, notes, homework, and quiz/exam details) added to the course on Canvas! Understand settings on Canvas

- My Research web site: http://itlab.uta.edu/sharma
- My contact: through Canvas (sharmac@cse.uta.edu)

# *Tests, projects, and HW Breakdown*

- We will have 3 tests 40%
- We will have 3 projects (Ia, Ib, II, and III) 45%
- Pop quizzes over the semester (3) 15%

TOTAL 100%

Class participation is important (asking questions, understanding concepts, class discussion, project presentation)

Oracle and SQL will be used for project I

- Typically, ½ to 1 std deviation above and below overall class average is a B
  - 1 (one) std deviation above class average is a guaranteed A
  - ½ to 1 (one) std deviation below class average is likely to be a C. and so on
  - 50% is passing grade
- Class stats will be communicated after each test/project in class & on Canvas

- Grades WIL NOT be posted on Canvas

# *Very Important*

- Cheating, collusion, and plagiarism will be seriously dealt with (an automatic Fail grade)
- You will also be reported to UTA as per rules

  And will go on your record
- If you have difficulty, come see us but do not resort to the above
- We know that you search Github and other sources for project code.
  - The fact is WE can also search and we will
  - If caught using them, it will be considered plagiarism
- We use a sophisticated AI-based (what else) plagiarism detection tool for submitted code

# *Let us talk about the Elephant in the room*

- Can anyone guess what I am referring to?
- ChatGPT
  - I am sure all of you are familiar with it and are thinking how that can be leveraged
  - If you can make use of it positively to learn and improve your grade, more power to you
  - Unfortunately, we are also familiar with it and are thinking how to make students learn, not memorize, and not use tools without understanding what they are doing!
- You are not allowed to use ChatGPT for this course projects, tests, and pop-quizzes!
  - If you use and are caught, it is treated as plagiarism!

# *Let us talk about the Elephant in the room*

- Use of ChatGPT (from Syllabus)

  - **Use of chatGPT for this course**: Use of chatGPT in any of its avatars is explicitly prohibited for solving projects. **You can benefit from chatGPT use for understanding concepts that are not clear and confirm them by asking questions in the class**. *You need to clearly understand that chaptGPT gives answers to any question you ask and it is your responsibility to determine whether it is correct or not. Blindly assuming it to be correct can be a costly mistake!*

# Academic Honesty

## What Constitutes Scholastic Dishonesty?

### 1. Cheating

- Copying another's test or assignment.
- Communication during an exam or assignment (i.e. written, oral or otherwise).
- Giving or seeking aid from another when not permitted by the instructor.
- Possessing or using unauthorized materials during the test.
- Buying, using, stealing, transporting, or soliciting a test, draft of a test, or answer key.

# Academic Honesty

## What Constitutes Scholastic Dishonesty?

### 2. Plagiarism

- Using someone else's work in your assignment without appropriate acknowledgement (for papers)
  - Does not apply to code from git or any other source
- Making slight variations in the language and failing to give credit to the source.
- Copying materials from the Internet without citing the source.
- Using code/material from previous years without acknowledging the source

- Note that acknowledging does not give permission to plagiarize!

# Academic Honesty

## What Constitutes Scholastic Dishonesty?

### 3. Collusion

- Without authorization, collaborating with another (team or outside entities) when preparing an assignment or homework or other requirements of the course
- You can discuss the project on discussion board, but cannot submit the same code or slightly modified code taken from elsewhere
- .
- Make sure your code base is different!

# *Overview*

- This is a second course on Database management systems at the CSE department at UTA

- The first course (CSE 5330) discusses DBMS from a *user or application developer* viewpoint
  - In Project Ia and Ib, you will use SQL on a Oracle database
- The emphasis of this course is on understanding the *internal modules* of a DBMS, their dependencies, system aspects and implementation techniques
  - Hence 3 hands-on implementation projects

# *Organization of the course*

- 4 modules
  - File structure, Storage, and indexing
  - Transaction Management
    - Concurrency control and
    - recovery
  - Cloud computing, map/reduce, and NoSQL (may not be all)
  - Query optimization overview
- 3 implementation projects (C/C++/Java) No python!
- 3 tests (in-person)

- Practice problems are assigned as homework (and checked if submitted) to help prepare for quizzes/exams (no grade)

# *How to do well on the course*

- Attend all the lectures (will also help do well on pop quizzes)
- Go through my slides before coming to class. They are all on Canvas
- Do follow up reading before and immediately after the lecture (not 1 day before the exam)
- Come prepared and ask questions in the class
- Make the class interactive
- There are NO dumb or trivial questions in my class; all questions are important and will be answered
- Solve all practice problems yourself and you are welcome to check it with us
- Make use of my (and TA's) office hours
- Challenge yourself and us!

# *Project Teams Rules*

- You can form teams of at most 2 students (not a requirement!) for doing the project (self-subscribe on Canvas)
  - Self service on Canvas; finish by next Tuesday (Aug 26)
- You are solely responsible for choosing your partner (you can also do the project alone)
- Once you choose, you CANNOT change the partner (you cannot have a different partner for each project)
- **Both members of the team will get the SAME grade, no matter what!**
- **I will not entertain any complaints against each other EXCEPT plagiarism**
- **If one is caught cheating, both will be reported to UTA or take other judicial action as dictated by UTA**
  - **BOTH will get the SAME penalty**
- **So, choose your project partner wisely!**

# *Project advise*

- Please start on the project immediately
  - **You have plenty of time if you pace yourself**
  - No extensions; only late penalties! (up to 5 days)
- Set milestones for the project and follow them (I am not going to do this for a graduate and elective course)
- Use debugging tools as needed. Cannot debug by using only print statements
  - Good design and time spent on it will reduce debugging time significantly
  - Time spent on design will save a lot of headache
- Learn to use makefile; these projects are not one file compilations
- Need to understand the system into which you are adding your code; similar to what you will encounter in a work place

# *What is assumed*

- Relational model
- Relational algebra, set theory
- User level understanding of a DBMS
- SQL
  - To understand query optimization, query plans, and use of indexing for query processing
- Overview of how SQL queries are processed
- Why Transaction management is needed
- Java and C/C++ familiarity for projects
  - Java for Map/Reduce project III
  - C++ for transaction management (project II)

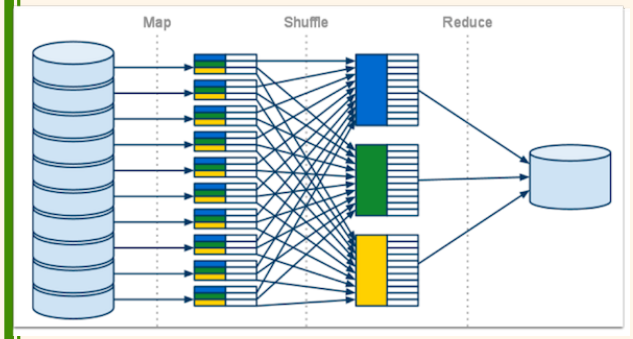
# *Beyond this course …*

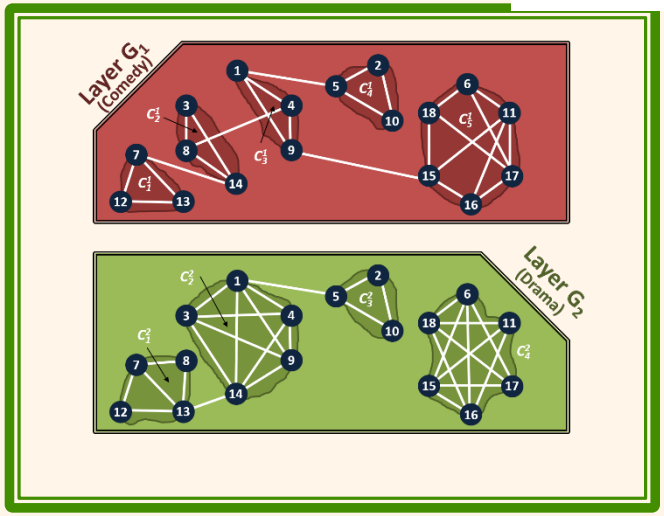
- If you get excited about databases and related research, there are a number of courses you can take beyond this course (e.g., CSE 6331)
- If you are interested in doing a thesis (MS/PhD) in the general areas of Databases, social network analysis, cloud computing, information integration, mining, machine learning, complex event and stream processing, information security – stop by and talk to me.
- If you intend to do a thesis, start right away. Don't try to do after 1 or 2 semesters.

# Sharma Chakravarthy

Professor
Information Technology Lab (IT Lab)
Ph.D. (University of Maryland, College Park, 1985)

https://itlab.uta.edu




**Scalablity using Map/Reduce**

**Video Situation Analysis**

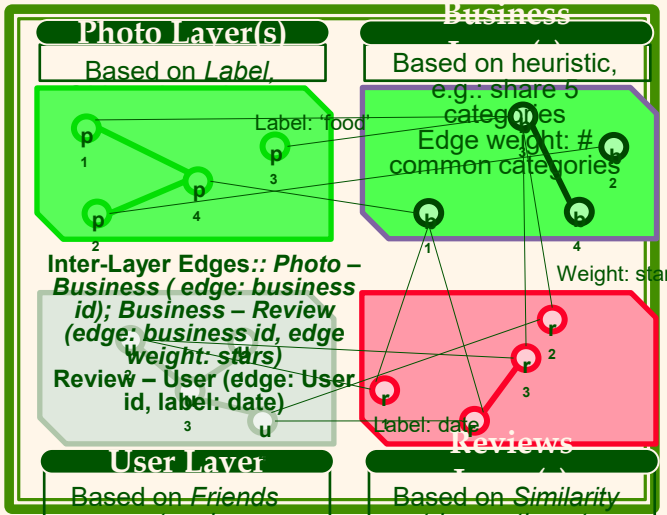
**Multilayer Network Analysis/visualization**

**Information Integration and LLMs**

INFORMATION TECHNOLOGY LABORATORY @ UTA-CSE


https://itlab.uta.edu

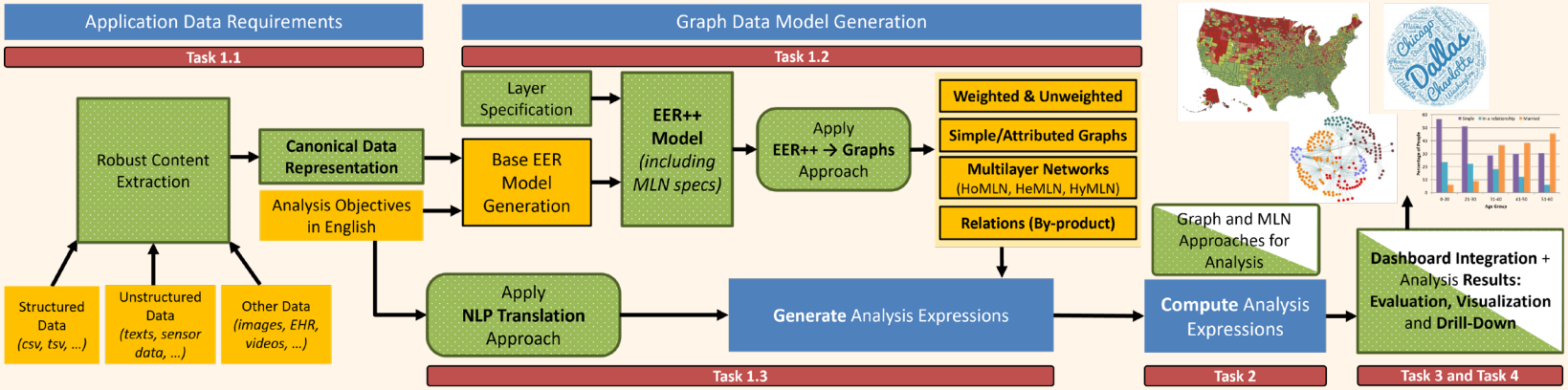
ERB 632
sharmac@cse.uta.edu



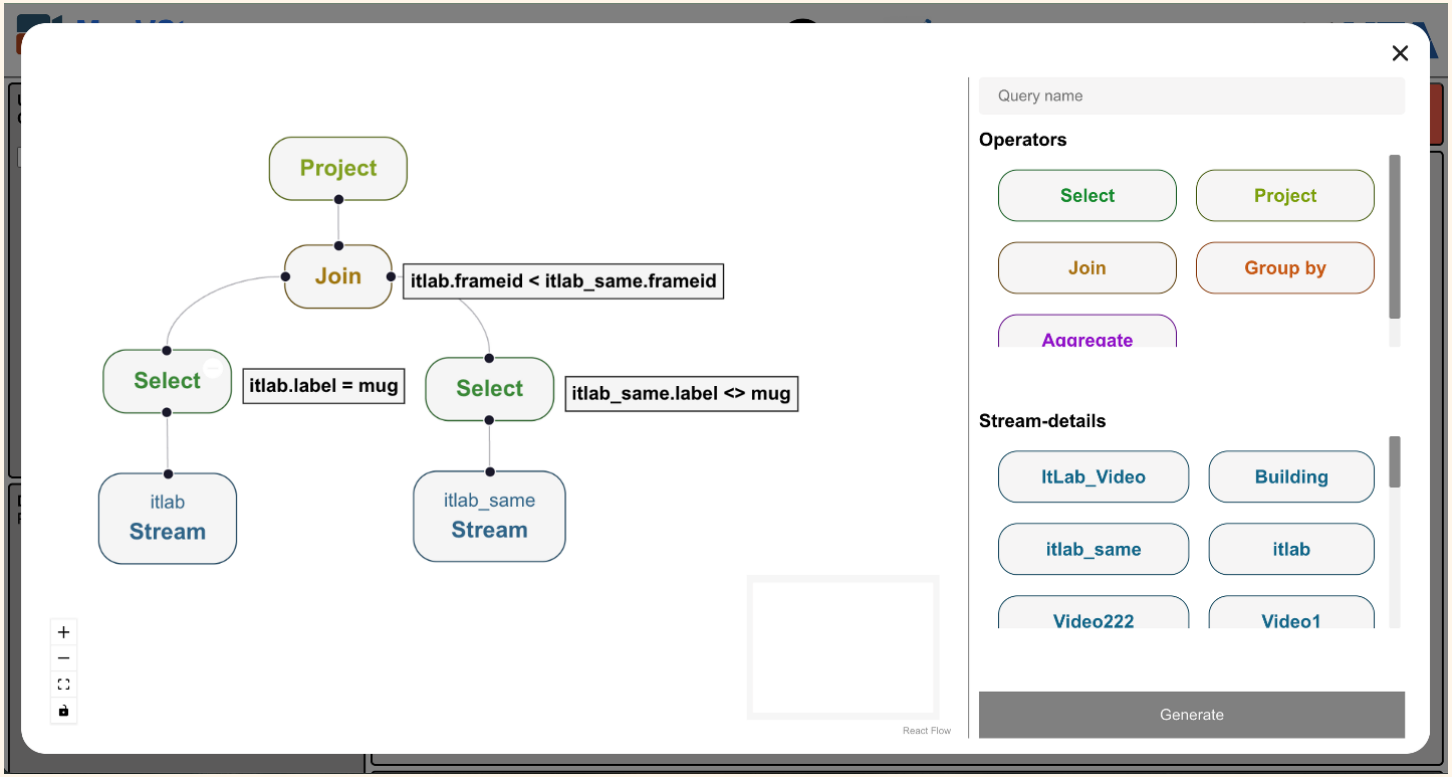
# *Dashboard developments*

- We have been developing a number of web-based dashboards for different projects
  1. Cowiz++ dashboard used multilayer networks and Covid data (https://itlab.uta.edu/Cowiz/)
  2. MLN dashboard for multilayer network analysis (https://itlab.uta.edu/mlndash-test/)
  3. VideoScoop dashboard for video situation analysis
     - A client server architecture
- Looking for undergrads and grads familiar with Python, React, flask, and other full stack packages

# *MLN-geeWhiz Dashboard*



# *VideoScoop Dashboard*



# Information Technology Laboratory (ERB 514)

**Prof. Sharma Chakravarthy (ERB 632)**

*Email:* sharmac@cse.uta.edu, *URL:* https://itlab.uta.edu/

**Funding Sources: NSF, Spawar, AFRL, Rome Lab, ONR, DARPA, TI, MCC**

## Select Projects

- ✓ **Multilayer Network Analysis & Visualization**
- ✓ **MavVStream: (querying video contents)**
- ✓ **Graph Mining scalability using Map/Reduce**
- ✓ **Several Dashbard developments**
  - ✓ **Mln-dashboard**
  - ✓ **Cowiz++ dashboard**
  - ✓ **Video query dashboard**
- ✓ **Expertise identification in Q/A community**
- ✓ **Ranking in web databases**
- ✓ **Mining:** Graph, Text, Assoc Rules

## Select Publications

**1.** Santra, F. A. Irany, K. Madduri, S. Chakravarthy, and S. Bhowmick, Efficient Community Detection in Multilayer Networks Using Boolean Compositions, in Big Data Networks, Frontiers of Big Data journal, 2023

2. H. Billah, A. Santra, and S. Chakravarthy. Video situation monitoring to improve quality of life. In Advances in Databases and Information Systems, Barcelona, Spain. Springer-Verlag, Berlin, Heidelberg, 2023

3. Mukunda, K., Roy, A., Santra, A., and Chakravarthy, S. Stress centrality in heterogeneous multilayer networks: Heuristics-based detection. In IEEE Big Data Service, Athens, Greece, July 17-20, 2023

4. Santra, A., Mukunda, K., and Chakravarthy, S. Privacy and Anonymity For Multilayer Networks: A Reflection, In Big Data and Machine Learning with Privacy Enhancing Tech Workshop (IEEE Big Data Service), Athens, Greece, July 17-20, 2023

5. Mukunda, K., Santra, A., and Chakravarthy, S. The challenge of finding degree centrality nodes in heterogeneous multilayer networks. In SEBD 2023, Galzignano Terme, Italy, July 2-5, 2023

6. K Komar, A Santra, S Bhowmick, and S Chakravarthy: **EER → $MLN: EER Approach for Modeling, Mapping, and Analyzing Complex Data Using Multilayer Networks (MLNs).** ER 2020: 555-572

**7. Query processing on large graphs: Approaches to scalability and response time trade offs.** Data and Know Eng. 126: 101736 (2020)

8. Das, S., Chakravarthy, S. (2018). Duplicate Reduction in Graph Mining: Approaches, Analysis, and Evaluation. IEEE TKDE

## IT Lab Researcher

Dr. Abhishek Santra

## PhD Students

Ms. Umme Billah
Mr. Anamitra Roy
Ms. Ayomide Ayowole-Obi
Mr. Aman Gulati

## MS Thesis Students

## Undergrad Students

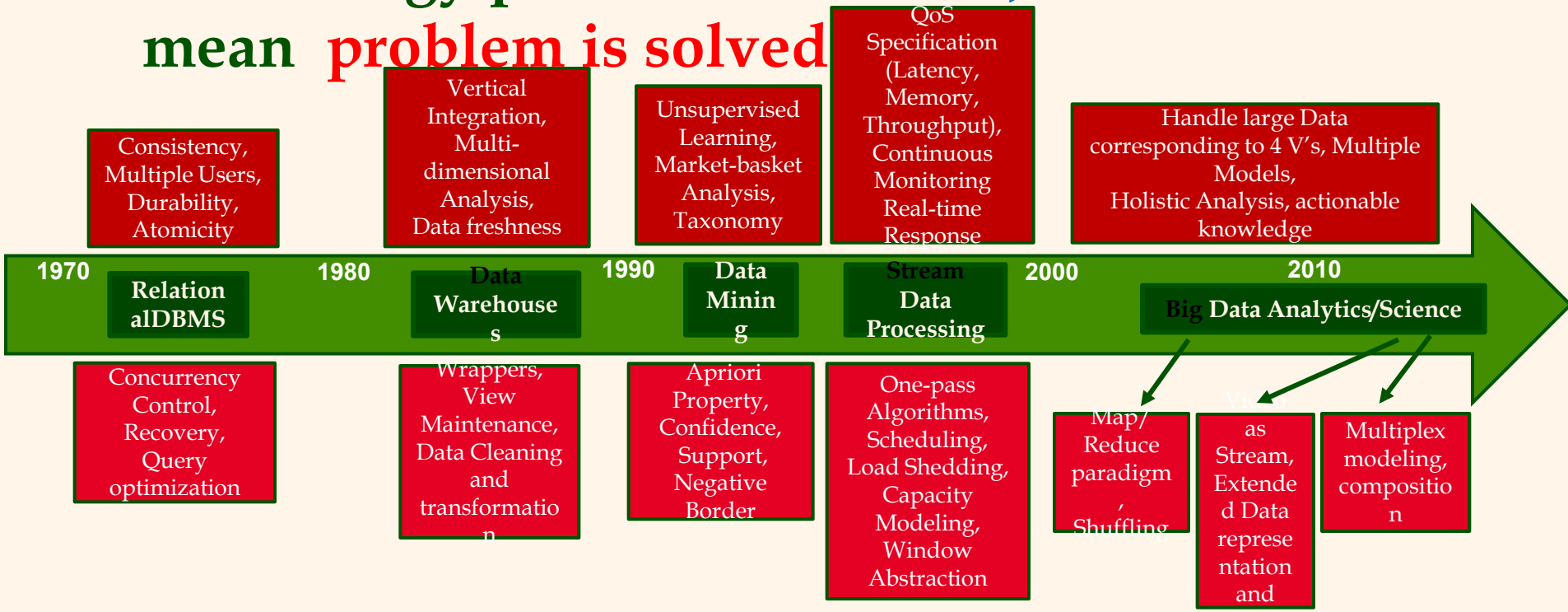
Mr. Sanuel vennali
Mr. Francisco martinez
Mr. Haresh Goutham
Mr. Tyagi siddarth

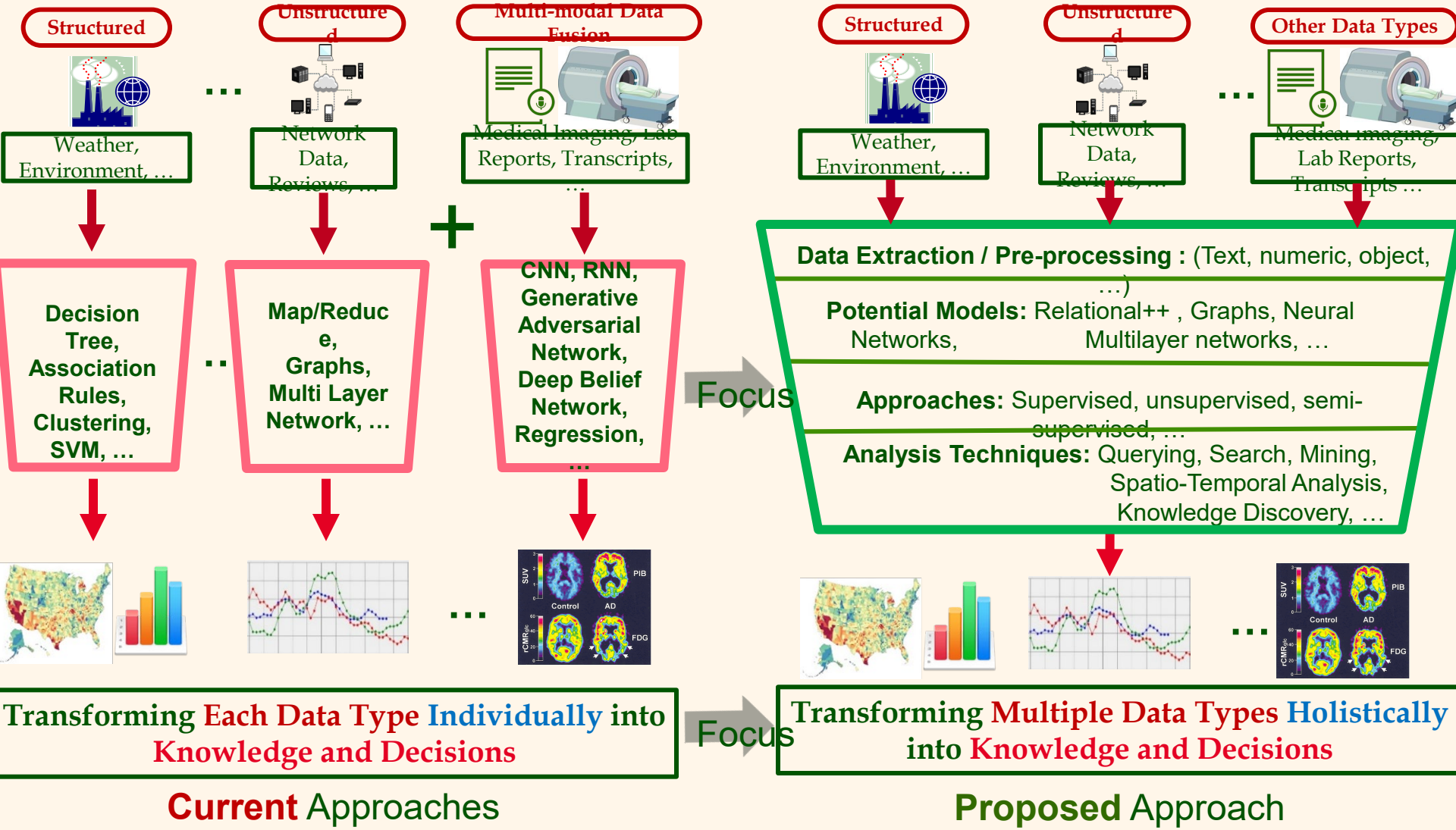
***Looking for more***

***Always Looking for good undergrad, MS, and PhD students***

# *Where are we headed?*

- **Without understanding the past, it is very difficult to appreciate the present and plan for the future!**
- **Technology provides solutions; it does *not* mean problem is solved**




Structured
Unstructured
Multi-modal Data Fusion
Weather, Environment, …
Network Data, Reviews, …
Medical Imaging, Lab Reports, Transcripts, …
Decision Tree, Association Rules, Clustering, SVM, …
Map/Reduce, Graphs, Multi Layer Network, …
CNN, RNN, Generative Adversarial Network, Deep Belief Network, Regression, …
Transforming Each Data Type Individually into Knowledge and Decisions
Current Approaches
Focus
Structured
Unstructured
Other Data Types
Weather, Environment, …
Network Data, Reviews, …
Medical Imaging, Lab Reports, Transcripts …
Data Extraction / Pre-processing : (Text, numeric, object, …)
Potential Models: Relational++ , Graphs, Neural Networks, Multilayer networks, …
Approaches: Supervised, unsupervised, semi-supervised, …
Analysis Techniques: Querying, Search, Mining, Spatio-Temporal Analysis, Knowledge Discovery, …
Transforming Multiple Data Types Holistically into Knowledge and Decisions
Proposed Approach
Focus
PIB
Control
AD
FDG
SUV

# CSE 5300 (DASC 5300)

- Foundations of Computing (new course)
- Covers Python programming, Problem solving, data structures, algorithms, role of OS
- Basic algebra, set/graph theory, Algorithms, and Big O complexity
- Role of DBMS in data science, SQL/OLAP, disk-based sort merge, speedup, scalability

# CSE 3330/5330

- Introductory database course
- From a user perspective
  - EER model for schema development
  - Database architecture
  - SQL
- Hands on project from application requirement to EER diagram to schema generation, data population, SQL querying, and use of JDBC/embedded SQL

# CSE 5334

- DATA MINING
- Preparing data for mining using preprocessing, data warehouses, OLAP
- Data mining primitives, languages, and system architecture;
- Data mining techniques including association rule mining, classification/prediction, and cluster analysis.

# CSE 6331 (and others)

- Advanced topics in Database systems
- The topics may vary from offering to offering based on the instructor.
- Deals with new/advanced topics that are currently being researched
- I offer graph mining, stream processing, and cloud computing in spring
- Topics such as web db & XML, DB and information exploration have been offered
- I have offered data warehousing, data mining, and event processing as part of this course

# CSE 6399 – Seminar course

- Advanced DB topics
- Typically a seminar course
- Reading and analyzing papers in new areas of research
- This semester I am offering this course on: Complex event & stream processing and information integration

# Thank You !!!

**For more information visit:**

**http://itlab.uta.edu**