5-Years M.Tech Degree in Data Science

(with lateral entry)

1 Introduction

This is a proposal to start a 5-years M.Tech degree in Data Science with lateral entry from existing B.Tech/BS students after 4 semesters in the B.Tech/BS program. The concept is similar to the current Engineering Science program which is a 4-years Bachelor's degree.

In the first instance we propose to start with 15 seats in 2016-17-I semester. If the program has enough demand then the number of seats can be increased in the following years leading to a full fledged inter-disciplinary program (IDP) in Data Science with an added Ph.D program to the proposed 5-years MTech degree. It is possible that Data Science (like Computer Science) may become a full fledged department in the medium term future.

Currently, the departments/groups that are interested in participating are: CSE, Mathematics and Statistics (the Statistics group), IME, EE (Signal processing, communications and networks group), HSS (Economics group).

2 Justification

In the two sub-sections below we first discuss why data science is an important emerging discipline and then briefly summarize the status of data science programs in universities across the world and in India.

2.1 Why Data Science

Increasingly, human activity at work or leisure, transactional or recreational is mediated by digital data or leaves a digital footprint sometimes unintended. Aristotle posited the *luminiferous ether*, a hypothetical substance, spread throughout the universe through which electromagnetic radiation was supposed to propagate. A similar data ether now exists within which most activity happens. The luminiferous ether was shown not to exist but the data ether not only exists but its density in bits/person is increasing exponentially. As the *Internet of Things* (IoT) becomes mainstream and ubiquitous this data density will increase steeply. The need for scientists who understand data in all its aspects will continue to grow strongly.

Data science is concerned with the acquisition, storage, retrieval, processing and finally the conversion of data into knowledge where the quantum of data is very large. Two current disciplines that have strong overlap with data science are computer science and statistics.

An important difference between Data Science and Computer Science is that data science heavily uses the more continuous aspects of mathematics together with considerable statistics. It is different from statistics insofar as there is an intense concentration on efficient algorithms to handle very large data sets, in some cases streamed data, and their realization via different architectures and platforms.

Other disciplines that intensively use data science are the physical sciences (physics, astronomy, biology, chemistry), management, economics and finance, and medicine. There is a growing realization that in the data lies most of the knowledge. For example, lot of cutting edge science (LHC, radio astronomy, genetic medicine, genomics and proteomics, molecular simulation, organic synthesis), commerce (supply chains, e-commerce, finance), medicine (epidemiology, genetic medicine, drug discovery) would be impossible without data science. Consequently, there is a large amount of ongoing work in data science which is spread across disciplines. Given its importance it makes eminent sense to bring the different disciplinary participants together on a common academic platform. The proposed 5-years (with lateral entry) M.Tech in data science takes a first step in this direction. If it succeeds we can create an IDP (Inter Disciplinary Program) on Data Science that can mature into a department in the future.

2.2 Existing programs

A large number of programs related to "Big Data" have been started in colleges and universities around the world, more so in the US, in the last two to three years. Other than 'Data Science' they often go by names like: 'Data Analytics', 'Big Data Analytics', 'Business Analytics', 'Data Mining and Knowledge Generation', 'Applied Statistics' amongst others. An approximate count of the number of data science programs in colleges and universities across the world shows that there are over 500 such programs[1].

These programs are typically post graduate programs of 2-4 semesters duration leading to certificates, diplomas or MS degrees. In India, Shiv Nadar Univeristy has a 2-years MS program in Big Data Analytics and recently IIT Kharagpur, ISI Kolkata and IIM Kolkata have jointly started a 2-years program on Big Data. We expect many such programs to start in other institutes in India in the near future.

Almost all such programs have one of two goals a) provide mid-level and high-level managers with a basic understanding of the vocabulary, techniques and tools of data analysis and sensitize them to the power of data analytics to improve business performance. Typically, such courses are 1-2 semester courses and most of the material is covered at a 'bird's eye-view' level, b) equip technical people with basic skills required for data handling, storage, retrieval, analysis and decision making. Such courses are typically longer at 2-4 semesters.

The program being proposed differs from the programs mentioned above in a major way. Our aim is to produce graduates who are well versed in the science of data. This means they will have enough theoretical knowledge and the skill set to address all aspects of creating, managing, processing and converting data to knowledge. This requires thorough knowledge of several interdisciplinary areas like: programming, algorithms, operating systems, databases, signal processing, machine learning and data mining, tools and techniques of data science (CS, EE); probability and statistics, inference, regression, optimization, statistical simulation and data analysis, sampling theory (Math and Statistics); management decision analysis, decision models, game theory (IME, Economics). In addition we expect these students to be familiar with how data science is used in other disciplines like the physical sciences, management, economics, medicine etc.

Students graduating from the proposed program will have significantly more depth and breadth in the broad area of Data Science which will be approached as a rigorous science with its principles, tools and techniques, and applications much like other existing sciences. Since the program has a significant thesis component we expect a student to be very knowledgeable in the specific area and topic of his thesis and the program to have reasonable research output right from the beginning.

Since students will enter the program laterally from B.Tech/BS the quality of students is expected to be high. Compared to graduates of the other programs mentioned above we expect our graduates to be much better prepared to address problems in industry, in other disciplines that use data science intensively, as well as pursue higher studies in data science if they are so inclined.

[1] http://101.datascience.community/2012/04/09/colleges-with-data-science-degrees/

3 Template

The first 4 semesters will be spent as part of the student's undergraduate degree curriculum. During the first 4 semesters it is expected that the student planning to migrate will complete CS210/ESO207 (Data Structures), CS220 (Computer Organization), MSO201 (Probability and Statistics).

At the end of the fourth semester students can opt to laterally shift to the 5-years M.Tech program in data science. This shift can be treated in a manner similar to a branch change.

Assuming the students do approximately 210 credits in the first 4 semesters. The 5-years MTech program will be a program requiring 496 credits.

Sem	**Courses**	**Sem. Credits**
5	CS330(12), CS345(9), DS1(CS601)(9), CS771(9), MTH418(9), DS/CS300(3)	51
6	HSS3(9), DS2(6), CS315/CS365(9), MTH515(9), MTH506(10), MTH511(9)	52
7	HSS4(9), DS3(3), IME634(9), DE1(9), DE2(9), OE1(9)	48
8	HSS5(9), DE3(9), DE4(9), OE2(9), OE3(9), Thesis(9)	54
Summer	Internship/thesis(9)	9
9	Thesis(36)	36
10	Thesis(36)	36

DEx: stands for departmental elective; OEx: stands for open elective.

Course list for course numbers in template

CS210/ESO207: Data Structures and Algorithms
CS220: Introduction to Computer Organisation
DS1(CS601): Mathematics for CS
DS2: Data Science: Tools and Techniques (New Course)
DS3: Seminar: Data Science Verticals (lectures/seminars by domain specialists in various verticals like: physical sciences, health, governance, climate, finance, e-commerce, supply chains etc.)
DS300: Technical Communication
CS315/CS365: Principles of Data Base Systems/Artificial Intelligence
CS330: Operating Systems
CS345: Algorithms II
MTH418: Inference-I
MTH506: Optimization
MTH511: Statistical Simulation and Data Analysis
MTH515: Inference-II
IME634: Management Decision Analysis
CS771: Machine Learning: Tools, Techniques and Applications

Electives

All electives are from existing courses from the indicated disciplines.

CSE, EE: Probablistic Machine Learning, Online and Stochastic Learning, Data Mining, Computational Game Theory, Multi-agent Systems: Games, Algorithms, Evolution, Topics in Cryptography and Coding, Represenation and Analysis of Random Signals, Information and Coding Theory, Mathematical Methods in Signal Processing, Artificial Intelligence, Computer Vision, Machine Learning for Computer Vision.

Math and Statistics: Non-parametric Inference, Sampling Theory, Regression Analysis.

IME, Economics: Decision Models, Introduction to Game Theory, Introduction to Stochastic Process and Applications, Econometrics, Social Media Analytics. Financial Engineering, Marketing Research, Advanced Decision models, Predictive Analytics, Management of Risk in Financial Systems.

Notes:

The courses in the four semesters: 5, 6, 7, 8 have the distribution shown in the table below. There will be 90 credits of thesis. The table shows the number of courses and the credits in brackets.

HSS	Dep. Courses	Dept. Elec.	Open Elec.	Thesis
3(27)	12(106)	4(36)	3(27)	90

Course DS2 will be offered as a CSxxx courses. Course DS2 will be offered in the near future. DS3 is a seminar based course with presentations by spekers from data science application areas.

Appendix: Course Details of Departmental Core

CS210: Data Structures and Algorithms

1. Random-access-machine model, concept of problem size, and asymptotic behaviour of time/space complexity.
2. Estimation of time/space complexity by smooth functions and order notations.
3. A simple example of worst-case time/space complexity analysis.
4. Elementary data-structures: arrays, lists, queues, stacks and their applications.
5. Binary search algorithm, binary trees, binary-search-tree data-structure.
6. Balanced binary-search-tree: Red-Black trees.
7. Hashing for insert, search, delete.
8. Heap data structure.
9. Efficient data structures, apart from those in items 6,7, and 8, for sets with the following group of operations: (i) insert, delete, membership, (ii) insert, delete, minimum, (iii) union, intersection, dierence, (iv) disjoint-set union, nd.
10. Sorting algorithms, including the average case analysis of quick-sort.
11. Greedy paradigm with examples.
12. Divide and conquer paradigm with examples.
13. Dynamic-programming paradigm with examples.
14. Denition of graphs, paths, trees, cycles. Data structures for graphs: adjacency lists, adjacency matrix.
15. Graph algorithms: Depth First Search, Breadth First Search, Minimum Spanning Tree.

    Additional topics based on time and interest may be selected from the following list:

16. Single-source shortest path computation, topological sorting of a partially ordered set, convex- hull computation, string matching algorithms, median computation, distributed algorithms.

CS220: Introduction to Computer Organization

Introduction, Overview of basic digital building blocks; truth tables; basic structure of a digital computer, Number representation, Integer -unsigned, signed (sign magnitude, 1s complement, 2s complement, rs complement); Characters-ASCII coding, other coding schemes; Real numbers-fixed and floating point, IEEE754, Assembly language programming for some processor, Basic building blocks for the ALU, Adder, Subtractor, Shifter, Multiplication and division circuits, CPU Subblock, Datapath - ALU, registers, CPU buses; Control path microprogramming (only the idea), hardwired, logic; External interface, Memory Subblock, Memory organization; Technology-ROM, RAM, EPROM, Flash etc. Cache; Cache coherence protocol for uniprocessor (simple), I/O Subblock, I/O techniques -interrupts, polling, DMA; Synchronous vs. Asynchronous I/ O; Controllers, Peripherals, Disk drives; Printers- impact, dot matrix, ink jet, laser; Plotters; Keyboards; Monitors; Advanced Concepts, Pipelining; Introduction to Advanced Processors.

CS315: Principles of Database Systems

1. Introduction: Database applications, purpose, accessing and modifying databases, need for transactions, architecture - users and administrators, data mining, information retrieval. iRelational Databases: relational model, database schema, keys, relational query languages, algebra, tuple and domain calculus example queries, (optional: equivalence of relational calculus and relational algebra).

2. SQL: Data definition, basic SQL query structure, set operations, nested subqueries, aggregation, null values, database modification, join expressions, views.

3. Database Design: E-R model, E-R diagram, reduction to relational schema, E-R design issues, database integrity, specifying integrity constraints in SQL: unique columns, foreign key, triggers.

4. Relational Database Design: features of good design, Functional Dependency theory, decomposition using functional dependency and normal forms, algorithms for decomposition, normal forms, (optional: multi-valued dependency and 4th normal form).

5. Storage and File structure: Overview of secondary storage, RAID and flash storage. Storing tables: row-wise, column database, database buffer. Indexing: concepts, clustered and non-clustered indices, B+-tree indices, multiple key access, hashed files, linear hash files, bitmap indices, Index definition in SQL, ++R-trees.

6. Query Processing: Overview, measures of query cost, selection, sorting, join processing algorithms-nested loops, merge-sort, hash join, aggregation.

7. Query Optimization: purpose, transformation of relational expressions, estimating cost and statistics of expression, choosing evaluation plans, linear and bushy plans, dynamic programming algorithms.

8. Transactions: Concept and purpose, ACID properties and their necessity, transactions in SQL. Problems with full isolation and levels of isolation.

9. Concurrency Control: lock-based protocols, 2-phase locking, deadlock handling, multiple granularity, timestamp based protocols, index locking, (optional: validation protocols, multi-version protocols, snap shot isolation, predicate locking, concurrency control for index structures).

10. Recovery: Failures and their classification, recovery and atomicity, recovery algorithms, Undo-Redo with write ahead logging, no Undo no Redo and other combinations, buffer management, (optional: ARIES recovery).

    Optional/Advanced topics below covered at the discretion of instructor

11. Parallel Databases: Avenues for parallelism: I/O parallelism, interquery, inter-query and intra operation parallelism, databases for multi-core machines.

12. Distributed Databases: Distributed data storage, distributed transactions, commit protocols, concurrency control in distributed databases, heterogeneous and cloud-based databases.

13. Data Mining: Decision Support Systems, data warehousing, mining, classification, association rules, clustering.

14. Information Retrieval: relevance ranking using terms and hyperlinks,page rank, indexing of documents, measuring retrieval effectiveness.

15. XML and semi-structured data: necessity, XML document schema, querying: XPath and XQuery languages, applications.

CS330: Operating Systems

1. Introduction: review of computer organization, intoduction to popular operating systems like UNIX, Windows, etc., OS structure, system calls, functions of OS, evolution of OSs.

2. Computer organization interface: using interrupt handler to pass control between a running program and OS.

3. Concept of a process: states, operations with examples from UNIX (fork, exec) and/or Windows. Process scheduling, interprocess communication (shared memory and message passing), UNIX signals.

4. Threads: multithreaded model, scheduler activations, examples of threaded programs.

5. Scheduling: multi-programming and time sharing, scheduling algorithms, multiprocessor scheduling, thread scheduling (examples using POSIX threads).

6. Process synchronization: critical sections, classical two process and n-process solutions, hardware primitives for synchronization, semaphores, monitors, classical problems in synchronization (producer-consumer, readers-writer, dining philosophers, etc.).

7. Deadlocks: modeling, chararcterization, prevention and avoidance, detection and recovery.

8. Memory management: with and without swapping, paging and segmentation, demand paging, virtual memory, page replacement algorithms, working set model, implementations from operating systems such as UNIX, Windows. Current Hardware support for paging: e.g., Pentium/ MIPS processor etc.

9. Secondary storage and Input/Output: device controllers and device drivers, disks, scheduling algorithms, file systems, directory structure, device controllers and device drivers, disks, disk space management, disk scheduling, NFS, RAID, other devices. operations on them, UNIX FS, UFS protection and security, NFS.

10. Protection and security: Illustrations of security model of UNIX and other OSs. Examples of attacks.

11. Epilogue: Pointers to advanced topics (distributed OS, multimedia OS, embedded OS, real-time OS, OS for multiprocessor machines).

CS345: Algorithms-II

1. Amortized analysis.

2. Exposure to some advanced data structures (For example, Fibonacci heaps or augmented data structures or interval trees or dynamic trees).

3. As part of CS210/ESO211 course, three algorithm paradigms, namely, greedy method, divide and conquer, and dynamic programming are discussed. These algorithm paradigms should be revisited in CS345, but with advanced applications and/or emphasis on their theoretical foundations as follows.

   (a) Greedy method: theoretical foundations of greedy method (matroids) and other applications.
   (b) Divide and Conquer: FFT algorithm and other applications.
   (c) Dynamic Programming: Bellman Ford algorithm and other applications.

4. Graph algorithms: all-pairs shortest paths, biconnected components in undirected graphs, strongly connected components in directed graphs, and other problems.

5. Pattern matching algorithms.

6. Lower bound on sorting.

7. Algorithms for maximum flow and applications.

8. Notion of intractability: NP-completeness, reduction (the proof of Cook-Levin theorem may be skipped)

9. Exposure to some (one or more) topics from the following list :

   (a) Approximation algorithms.
   (b) Algebraic and number theoretic algorithms.
   (c) Computational Geometry.
   (d) Linear programming.
   (e) Parallel/distributed algorithms.
   (f) Randomized algorithms.

Artificial Intelligence

1. AI: Introduction
2. Brief history.
3. Agents and rationality, task environments, agent architecture types.
4. Search and Knowledge representation.
5. Search spaces
6. Uninformed and informed search.
7. Hill climbing, simulated annealing, genetic algorithms.
8. Logic based representations (PL, FoL) and inference, Prolog.
9. Rule based representations, forward and backward chaining, matching algorithms.
10. Probabilistic reasoning and uncertainty.
11. Bayes nets and reasoning with them.
12. Uncertainty and methods to handle it.
13. Learning.
14. Forms of learning.
15. Statistical methods: naive-Bayes, nearest neighbour, kernel, neural network models, noise and overfitting.
16. Decision trees, inductive learning.
17. Clustering - basic agglomerative, divisive algorithms based on similarity/dissimilarity measures.
18. Applications to NLP, vision, robotics, etc.

Mathematics for Computer Science
The course will have roughly three equal parts: a) logic b) linear algebra c) probability. The mathematics should be linked up with actual applications in computer science.

Topics:
Logic:
Outcome: ability to express a problem in a suitable logical language and reasoning system and understand how it can be solved by a deductive process.

1. What is a proof? And proof methods.
2. Propositional logic syntax and semantics.
3. Tautologies, axiom system and deduction.
4. Soundness and completeness (proof sketch).
5. First order logic syntax, converting natural language into FoL wffs.
6. Structures, models, satisfaction and validity.
7. Axiomatization, soundness and completeness (no proofs).

8. Refutation and logic programming.

Algebra:
Outcome: ability to model problems through abstract structures and arrive at insights or solutions by manipulating these models using their properties.

1. Groups, rings, ideals, fields: basic understanding of these abstract structures and their properties.
2. Vectors spaces, examples, $\mathbb{R}^n$, $\mathbb{C}^n$; subspaces.
3. Linear independence, dependence and dimension.
4. Linear transformations.
5. Matrices, matrix algebra, determinants. Properties of matrices and determinants.
6. Systems of linear equations.
7. Eigenvalues, eigenvectors, eigenspaces, diagonalization and the spectral theorem.
8. Factorization and singular value decomposition.

Probability:
Outcome: ability to model uncertain phenomena using probability and understand and calculate the uncertainty in systems where such phenomena are a part of the system.

1. Sample spaces, events, axioms of probability.
2. Conditional probability and independence.
3. Random variables. Discrete and continuous random variables, densities and distributions.
4. Expectation and its properties.
5. Normal distribution and its properties.
6. Law of large numbers, central limit theorem.
7. Bounds on deviations: Chebyshev, Markov, Hoeffding, Chernoff.
8. Introduction to Markov chains, random walks.

CS771: Machine Learning: Tools, Techniques and Applications

1. Machine learning - what, how, where.
2. Supervised, unsupervised and semi-supervised learning.
3. Training, validation, testing, generalization, overfitting.
4. Features and feature engineering.
5. Decision trees, random forests.
6. Linear classifiers.
7. Kernel based methods and SVMs.
8. Nearest neighbour methods.

9. Hidden Markov models.
10. Neural and deep networks.
11. Ensemble methods - boosting, bagging, voting schemes.
12. Distance metrics and clustering.
13. Methods for semi-supervised learning.

MTH418 Inference-I

Parametric models, parametrs, random sample and its likelihood, statistic and its sampling distributions, problems of inference. Examples from standard discrete and continuous models such as Bernoulli, Binomial, Poisson, Negative Binomial, Normal, Exponential, Gamma, Weibull, Pareto etc. Concept of sufficiency, minimal sufficiency, Neyman factorization criterion, Fisher information, exponential families. Maximum likelihood estimators, method of moment estimators, percentile estimators, least squares estimators, minimum mean squares estimators, uniformly minimum variance unbiased estimators, Rao- 320Blackwell theorem, Cramer-Rao lower bound, different examples. Statistical Hyptheses-simple and composite, statistical tests, critical regions, Type-I and Type-II errors, size and power of a test, Neyman Pearson lemma and its different applications. Most powerful test, uniformly most powerful test, unbiased test and uniformly most unbiased test. Likelihood ratio test. Interval estimation, confidence intervals, construction of confidence intervals, shortest expected length confidence interval, most accurate one sided confidence interval and its relation to UMP test.

MTH506: Optimization Optimization Problem: various examples, Characterization of optimality and constrained optimal problems, Convex sets and convex functions and their properties, Non-linear programming theory - Kuhn-Tucker conditions, Lagranges theory, Duality theory, Search techniques - one variable and several variables, Pontryagins maximum principle and its applications, Dynamic programming and its applications.

MTH511: Statistical Simulation and Data Analysis

Simulation of random variables from discrete, continuous, multivariate distributions and stochastic processes, Monte-Carlo methods. Regression analysis, scatter plot, residual analysis. Computer Intensive Inference Methods - Jack-Knife, Bootstrap, cross validation, Monte Carlo methods and permutation tests. Graphical representation of multivariate data, Cluster analysis, Principal component analysis for dimension reduction.

MTH515: Inference-II

Group families, the principle of equivariance, location family, scale family, location scale family. Minimum risk equivariance estimators, risk functions, admissibility, prior distribution, posterior distribution, geometric interpretation for finite parameter space, Bayes estimators, limit of Bayes estimators, minimax estimators and their relations. Review of convergence in probability and convergence in distribution. Consistency results of MLEs, and MMEs. Asymptotic relative efficiency. Consistent and Asymptotic Normal (CAN) estimators, invariance of CAN estimators under different transformations. CAN estimators obtained by moments and MLE methods in one parameter exponential family and multiparameter exponential family. Sequential Probability Ratio Tests and its applications in different practical problems. Invariant test and unbiased tests, Likelihood ratio test and its asymptotic distributions, Wald test, Rao's score test, Pearson χ^2 test for goodness of fit. Large sample tests and confidence intervals based on CAN estimators. Consistency of large sample tests and asymptotic powers of large sample tests.

Management Decision Analysis

Multi-objective decisions, Decisions under uncertainty, Statistical Decision Trees, Applications from Quality Control and Production Control.

DS2: Data Science: Tools and Techniques

This course builds hands-on skills in installing, tuning and using data stores and related system software that is widely used in data science applications. It also familiarizes students with basic statistical and machine learning tools and libraries used in data science.

Relational, document, key-value, object, graph databases. Install, create, populate, retrieve, tune and manage these databases. Hadoop file system and Map-reduce. Virtualization platforms. Cloud platforms - their installation and management.

R, Octave, Scilab. Python libraries: SciPy and sci-kitLearn, PyBrain, Pylearn2; Java libraries: Spark, JSAT, Mahout, Weka. PIG. Theano. C++ platforms: Vowpal Wabbit, Shogun.

The course will use open source software for all the above. The course will also require students to do a semester long project along with the lab sessions.

DS3: Data Science: Industrial Perspectives

This is a seminar course where specialists will give talk(s) on various verticals like: biology, health, governance, physical sciences, e-commerce, climate, supply chains, finance etc. This seminar series will provide the industry or domain specific context for advanced analytics. This will include the following topics:

- Industry/domain overview description, history, challenges faced, key players, industry trends.
- Key business processes (marketing, financial, sales, logistics, order fulfillment, procurement, executive reporting, customer facing).
- Key data elements for each major business process and systems of record. Critical metrics (KPI) for each business process and the definitions. Sources/location of the data, typical errors in data.
- Benchmarking information sources.
- Signal and noise in system information relevance.
- Classes of users and associated metrics/information needed/questions asked or would like to ask..potential for new apps.
- Historical access to information and impacts.
- Impact of social media, text analysis, web and other sources of information. Timeliness of data/information.
- Predictive Vs historical analysis (customer segmentation, churn etc).

Subject domains/industry:

- Politics/elections required reading `http://fivethirtyeight.com/`, Nate Silvers book (signal and the noise), invite party strategists from major parties. Crowd sourced analytics (www.kaggle.com).
- Healthcare invite CFO/CMO from AIIMS, Apollo and/or other large hospitals to discuss metrics for hospital operation including financial, operational and medical procedures efficacy.
- Supply chain management invite VP/Operations/Logistics from auto companies, Flipkart/Amazon to discuss supply chain analytics and issues.
- Weather/climate change National weather bureau chief.
- Agriculture Industry think tanks re crop yields, weather patterns, evolving trends.
- Corporate - HR, Finance, Sales, Marketing, Web marketing, IT.
- Customer services/support.
- Website analytics - Google.

- Financial Services Chief Marketing Officers/CEO to discuss customer analytics, services, service quality, innovation, profitability.
- Energy prospecting ONGC, Chevron, BP.
- Security communications, RAW, National Security agencies.
- Entertainment - movie/show launches.
- Retail Hindusthan Lever, Proctor & Gamble, Godrej.
- Telecom Airtel, BSNL, Vodafone.
- Education Teach for India, UNICEF, Gates Foundation.