Course Wise Syllabus (Continued)

Distributed Data Systems Distributed File Systems - File System Models; Replication and Synchronization - Caching; Failure & Recovery; File System Security. Distributed Databases - Distributed Data Sources and Updates; Database Connectivity; Concurrency Control and Distribution mechanism; Distributed indexing schemes. Database security. Data on the Web - Web as a distributed data repository. Data Collection and Use Crawlers, Search Engines, and Indexing Schemes. Information Retrieval Techniques. Data Exchange - Hierarchical Data Models, XML, and query languages. Semi-structured / Unstructured data -querying and synchronization. Pervasive Data - Data distribution and access for non-computing devices, small computing devices, embedded computing devices and sensory devices. Data visualization and Interpretation Visualization as a Discovery tool, Visualization skills for the masses, The Visualization methodology, Visualization design objectives, Exploratory vs. explanatory analysis, Understanding the context for data presentations, 3 minute story, Effective Visuals, Gestalt principles of visual perception, Visual Ordering, Decluttering, Story Telling, Visualization Design; Taxonomy of Data Visualization Methods: Exploring Tableau, Dashboard and Stories, Bullet graphs, Pareto charts, Custom background images; Dashboard : Dashboard categorization and typical data, Characteristics of a Well-Designed Dashboard, Key Goals in the Visual Design Process; Power of Visual Perception: Visually Encoding Data for Rapid Perception, Applying the Principles of Visual Perception to Dashboard Design. Stream Processing and Analytics Real Time , Streaming Data & Sources, Real time streaming system architecture , Characteristics of a Real Time Architecture and Processing ; Configuration and Coordination Systems: Distributed State and Issues, Coordination and Configuration using Apache ZooKeeper; Data Flow Management : Distributed Data Flows , Various Data Delivery and Processing Requirements, N+1 Problem, Apache Kafka (High-Throughput Distributed Messaging); Processing Stream Data with Storm; Overview of Data Storage – Requirements: Need for long-term storage for a real time processing framework, In-memory Storage, No-Sql Storage Systems, Choosing a right storage solution; Visualizing Data :Requirements, Principles and tools; Bounds of Random variables, Poisson Processors, Maintaining Simple Statistics from Data Streams, Sliding Windows and computing statistics over sliding windows, Data Synopsis (Sampling, Histograms, Wavelets, DFT), Exact Aggregation, Timed Counting and Summation, Multi Resolution Time Series Aggregation, Stochastic Optimization; Statistical Approximation to Streaming Data: Probabilities and Distributions, Sampling Procedures for Streaming Data, Approximating Streaming Data with Sketching, Registers and Hash Functions, Working with Sets, The Bloom Filter, Distinct Value Sketches, The Count-Min Sketch; Clustering techniques for Streaming Data; Classification methods : Decision Tree (VFDT); Evaluating stream processing algorithms; Case Studies in Designing solutions to streaming data

Artificial and Computational Intelligence Agents and environments, Task Environments, Working of agents; Uninformed Search Algorithms: Informed Search. Local Search Algorithms & Optimization Problems: Genetic Algorithm; Searching with Non-Deterministic Actions, Partial Information and Online search agents, Game Playing, Constraint Satisfaction Problem, Knowledge Representation using Logics: TT-Entail for inference from truth table, Proof by resolution, Forward Chaining and Backward Chaining, Inference in FOL, Unification & Lifting, Forward chaining, Backward Chaining, Resolution; Probabilistic Representation and Reasoning : Inference using full joint distribution, Representation of Conditional Independence using BN, Reinforcement Learning; Difference between crisp and fuzzy logic, shapes of membership function, Fuzzification and defuzzification, fuzzy logic reasoning; Decision making with fuzzy information, Fuzzy Classification; Connectionist Models: Introduction to Neural Networks, Hopfield Networks, Perceptron Learning, Backpropagation & Competitive Learning, Applications of Neural Net: Speech, Vision, Traveling Salesman; Genetic Algorithms - Chromosomes, fitness functions, and selection mechanisms, Genetic algorithms:

crossover and mutation, Genetic programming. Machine Learning Introduction to Machine Learning, Various kinds of learning, Supervised Learning, Unsupervised Learning, Model Selection; Bayesian Learning, MAP Hypothesis, MDL Principle, Bias Variance Decomposition, Bayes Optimal Classifier, Naive Bayes Classifier; Linear Models for Regression, Linear Models for Classification; Non-Linear models, Decision trees; Instance Based Learning, KNN Algorithm, CBR Learning; Support Vector Machines, VC Dimension; Neural Networks, Perceptron Learning, Back Propagation Algorithm; Introduction to Genetic Algorithms. Applied Machine Learning Need for machine learning. Prediction and classification methods. Use cases in application domains. Interpretation of results. Limitations of various techniques. End to end Machine learning - data collection, data preparation, model selection. Dissertation A student registered in this course must take a topic in an area of professional interest drawn from the on the job work requirement which is simultaneously of direct relevance to the degree pursued by the student as well as to the employing / collaborating organization of the student and submit a comprehensive report at the end of the semester working under the overall supervision and guidance of a professional expert who will be deemed as the supervisor for evaluation of all components of the dissertation. Normally the Mentor of the student would be the Dissertation supervisor and in case Mentor is not approved as the supervisor, Mentor may play the role of additional supervisor. The final grades for dissertation are Non-letter grades namely Excellent, Good, Fair and Poor, which do not go into CGPA computation