**Unit 12 – Lecturecast – The Future of Big Data Analytics**

Data management concepts and underlying technology is evolving rapidly and that increases emphasis on using data for decision making

**Emerging Trends in Databases**

Trends in databases business processes continue to change, which changes business requirements. For example, organisations need to search for unstructured data such as videos and data is being stored remotely or in the cloud. Various database technologies are being developed in response to organisational and business requirements:

**Distributed databases:** Distributed databases allow actual data to be spread across several smaller databases connected through a network.

**Object-oriented databases:** Object-oriented databases allow different types of data to be stored as objects. The objects can be retrieved and manipulated using the object-oriented database management system. It is possible to reuse and modify existing objects making it efficient to use for applications development.

**Image database:** An image database is also technology available to facilitate the storage of images. Some companies are making use of this technology. For example, banks can store scanned pay-in cheques instead of papers.

**Hypertext database:** Provide flexibility in defining relationships, it offers the possibility to search for unstructured content. Using hypertext databases, for example, it is possible for all cross-referenced information to be retrieved from the database by a click on a hypertext of a record.

**Hypermedia databases:** Hypermedia database is an extension of the hypertext database. It allows manipulation of all multimedia data types such as graphs, videos, sounds and many more.

**Graph databases:** A graph database contains a collection of nodes and edges. A node represents an object, and an edge represents the connection or relationship between two objects. In graph databases, it is possible to map multiple types of relational and complex data, which make graph databases flexible to implement rather than set tables and columns.

**Machine Learning**

Programming computers to optimise a performance criterion using sample data. Data parameters are defined, then a model is trained to learn the patterns in the data. The algorithms allow large datasets to be analysed with the aim of identifying new patterns and meaning. Results can help with strategic decision making. Machine learning helps solve complex data related problems.

***Applications:***

Learning associations – Retail - between what a customer buys to be able to target those that buy only some elements of what other similar customers buy, they are potentially still open to additional items being sold. For example, if customers bought products X and Y and there is a customer who bought X and did not buy Y, then we could draw an inference that a customer who bought product X but did not buy Y is still a potential buyer of product Y

**Classification –** Deciding on who to loan money to – Assessment of whether a customer can pay a loan back and bank will make a profit. A credit check of customer along with parameters such as age, occupation, income, debts, collateral, savings, any existing loans or whether previous loans or credits were paid back by the customer. Model can be created that then classifies that person based on these and other factors as to whether they are high or low risk and based on this risk, loan can be awarded or not.

**Regression Analysis – System for price of a used car. Based on factors about the car and knowledge of previous sale prices, can project future sales.** Based on past sales and transactions, we would be able to obtain training data for our machine learning model or algorithm where we can formulate a function (described as a fitness function) to make Y as a function of X.  This can be represented as y =   wx + wժ where w represents the weights/values assigned to the data variables and ժ the change in weights assigned to the training data x . Both regression and classification are supervised problems where there is an input X and an output Y - the task is to map the input to the output. In other words, if the relative weight of the different factors can be identified and plugged into the formula it predicts the outcome, so when we get a used car we can plug that into the formula and predict a price.

**Big Data Best Practice**

* Start small with big data
* Forward thinking and considering what end point would be
* Avoid: If we build, they will come, software will have all answers, forgetting lessons of past, not having the required business and analytical experience, treating it as a science experiment, promising and trying too much.

Define small high value opportunities and use those as starting points to build expertise to later deal with the big business questions. Think ahead about potential scalability of project.

* Simply deploying a system will not solve issues without understanding of requirements. A successful solution involves extensive business knowledge
* Building an analytics system is complex and big data projects need to start from scratch
* Collecting and analysing data is only the start, it musty be incorporated into business process.
* Do not overpromising and underdeliver – good way to cause organisation to walk away even if it’s working elsewhere.