**Unit - 9**

**Multimedia Databases**

**The nature of multimedia data and applications**

**Multimedia databases** provide features that allow users to store and query different types of multimedia information, which includes *images* (such as photos or drawings), *video clips* (such as movies, newsreels, or home videos), *audio clips* (such as songs, phone messages, or speeches), and *documents* (such as books or articles). The main types of database queries that are needed involve locating multimedia sources that contain certain objects of interest. For example, one may want to locate all video clips in a video database that include a certain person, say Michael Jackson. One may also want to retrieve video clips based on certain activities included in them, such as video clips where a soccer goal is scored by a certain player or team.

The above types of queries are referred to as **content-based retrieval**, because the multimedia source is being retrieved based on its containing certain objects or activities. Hence, a multimedia database must use some model to organize and index the multimedia sources based on their contents. *Identifying the contents* of multimedia sources is a difficult and time-consuming task. There are two main approaches. The first is based on **automatic analysis** of the multimedia sources to identify certain mathematical characteristics of their contents. This approach uses different techniques depending on the type of multimedia source (image, video, audio, or text). The second approach depends on **manual identification** of the objects and activities of interest in each multimedia source and on using this information to index the sources. This approach can be applied to all multimedia

sources, but it requires a manual preprocessing phase where a person has to scan each multimedia source to identify and catalog the objects and activities it contains so that they can be used to index the sources. An **image** is typically stored either in raw form as a set of pixel or cell values, or in compressed form to save space. The image *shape descriptor* describes the geometric shape of the raw image, which is typically a rectangle of **cells** of a certain width and height. Hence, each image can be represented by an *m* by *n* grid of cells. Each cell contains a pixel value that describes the cell content. In black-and-white images,

pixels can be one bit. In gray scale or color images, a pixel is multiple bits. Because images may require large amounts of space, they are often stored in compressed form. Compression standards, such as GIF, JPEG, or MPEG, use various mathematical transformations to reduce the number of cells stored but still maintain the main image characteristics. Applicable mathematical transforms include Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT), and wavelet transforms.

A **video source** is typically represented as a sequence of frames, where each frame is a still image. However, rather than identifying the objects and activities in every individual frame, the video is divided into **video segments,** where each segment is made up of a sequence of contiguous frames that includes the same objects/activities. Each segment is identified by its starting and ending frames. The objects and activities identified in each video segment can be used to index the segments. An indexing technique called *frame segment trees* has been proposed for video indexing. The index includes both objects, such as persons, houses, cars, and activities, such as a person *delivering* a speech or two people *talking.*

A **text/document source** is basically the full text of some article, book, or magazine. These sources are typically indexed by identifying the keywords that appear in the text and their relative frequencies. However, filler words are eliminated from that process. Because there could be too many keywords when attempting to index a collection of documents, techniques have been developed to reduce the number of keywords to those that are most relevant to the collection. A technique called *singular value decompositions* (SVD), which is based on matrix transformations, can be used for this purpose. An indexing technique called *telescoping vector trees,* or TV-trees, can then be used to group similar documents together.

**Audio sources** include stored recorded messages, such as speeches, class presentations, or even surveillance recording of phone messages or conversations by law enforcement. Here, discrete transforms can be used to Audio characteristic features include loudness, intensity, pitch, and clarity.

**Mobile Databases:**

Recent advances in wireless technology have led to mobile computing, a new dimension in data communication and processing. The mobile computing environment will provide database applications with useful aspects of wireless technology. The mobile computing platform allows users to establish communication with other users and to manage their work while they are mobile. This feature is especially useful to geographically dispersed organizations. Typical examples might include traffic police, taxi dispatchers, and weather reporting services, as well as financial market reporting and information brokering applications.

**Geographic information systems (GIS)** are used to collect, model, store, and analyze information describing physical properties of the geographical world. The scope of GIS broadly encompasses two types of data: (1) spatial data, originating from maps, digital images, administrative and political boundaries, roads, transportation networks; physical data such as rivers, soil characteristics, climatic regions, land elevations, and (2) nonspatial data, such as census counts, economic data, and sales or marketing information. GIS is a rapidly developing domain that offers highly innovative approaches to meet some challenging technical demands.

In a general sense, the term describes any information system that integrates, stores, edits, analyzes, shares, and displays geographic information for informing decision making. GIS applications are tools that

allow users to create interactive queries (user-created searches), analyze spatial information, edit data in maps, and present the results of all these operations.[2] Geographic information science is the science underlying geographic concepts, applications, and systems.

**Data Mining and Data Warehousing**

The increasing processing power and sophistication of analytical tools and techniques have resulted in the development of what are known as data warehouses. These data warehouses

provide storage, functionality, and responsiveness to queries beyond the capabilities of transaction-oriented databases. Accompanying this ever-increasing power is a great demand to improve the data access performance of databases.

we defined a *database* as a collection of related data and a *database system* as a database and database software together. A data warehouse is also a collection of information as well as a supporting system. However, a clear distinction exists. Traditional databases are transactional (relational, object-oriented, network, or hierarchical).*Data warehouses* have the distinguishing characteristic that they are mainly intended for decision-support applications. They are optimized for data retrieval, not routine transaction processing.

W. H. Inmon characterized a **data warehouse** as *a subject-oriented, integrated, nonvolatile, time-variant collection of data in support of management’s decisions*. Data warehouses provide access to data for complex analysis, knowledge discovery, and decision making. They support high- performance demands on an organization’s data and information. Several types of applications—OLAP, DSS, and data mining applications are supported. We define each of these next.

|  |
| --- |
|  |

The goal of a data warehouse is to support decision making with data. Data mining can be used in conjunction with a data warehouse to help with certain types of decisions. Data mining can be applied to operational databases with individual transactions. To make data mining more efficient, the data warehouse should have an aggregated or summarized collection of data. Data mining helps in extracting meaningful new patterns that cannot be found necessarily by merely querying or processing data or metadata in the data warehouse. Data mining applications should therefore be strongly considered early, during the design of a data warehouse. Also, data mining tools should be designed to facilitate their use in conjunction with data warehouses. In fact, for very large databases running into terabytes of data, successful use of database mining applications will depend first on the construction of a data warehouse.

As the term connotes, data mining refers to the mining or discovery of new information in terms of patterns or rules from vast amounts of data. To be practically useful, data mining must be carried out efficiently on large files and databases. Although some data mining features are

being provided in RDBMSs, data mining is not well-integrated with database management

systems.

**Applications of Data Mining**

Data mining technologies can be applied to a large variety of decision-making contexts in business. In particular, areas of significant payoffs are expected to include the following:

**• Marketing**—Applications include analysis of consumer behavior based on buying patterns; determination of marketing strategies including advertising, store location, and targeted mailing; segmentation of customers, stores, or products; and design of catalogs, store layouts, and advertising campaigns.

• **Finance**—Applications include analysis of creditworthiness of clients, segmentation of account receivables, performance analysis of finance investments like stocks, bonds, and mutual funds; evaluation of financing options; and fraud detection.

• **Manufacturing**—Applications involve optimization of resources like machines, manpower, and materials; optimal design of manufacturing processes, shop-floor layouts, and product design, such as for automobiles based on customer requirements.

• **Health Care**—Applications include an analysis of effectiveness of certain treatments; optimization of processes within a hospital, relating patient wellness data with doctor qualifications; and analyzing side effects of drugs.

**Data Mining versus Data Warehousing**

The goal of a data warehouse is to support decision making with data. Data mining can be used in conjunction with a data warehouse to help with certain types of decisions. Data mining can be applied to operational databases with individual transactions. To make data mining more efficient, the data warehouse should have an aggregated or summarized collection of data. Data mining helps in extracting meaningful new patterns that cannot necessarily be found by merely querying or processing data or metadata in the data warehouse. Therefore,

data mining applications should be strongly considered early, during the design of a data warehouse. Also, data mining tools should be designed to facilitate their use in conjunction with data warehouses. In fact, for very large databases running into terabytes and even petabytes of data, successful use of data mining applications will depend first on the construction of a data warehouse.

**Concept Big Data**

Big data is [data sets](https://en.wikipedia.org/wiki/Data_set) that are so voluminous and complex that traditional [data-processing](https://en.wikipedia.org/wiki/Data_processing) [application software](https://en.wikipedia.org/wiki/Application_software) are inadequate to deal with them. Big data challenges include [capturing data](https://en.wikipedia.org/wiki/Automatic_identification_and_data_capture), [data storage](https://en.wikipedia.org/wiki/Computer_data_storage), [data analysis](https://en.wikipedia.org/wiki/Data_analysis), search, [sharing](https://en.wikipedia.org/wiki/Data_sharing), [transfer](https://en.wikipedia.org/wiki/Data_transmission), [visualization](https://en.wikipedia.org/wiki/Data_visualization), [querying,](https://en.wikipedia.org/wiki/Query_language) updating, [information privacy](https://en.wikipedia.org/wiki/Information_privacy) and data source. There are five concepts associated with big data: *volume*, *variety*, *velocity* and, the recently added, *veracity* and *value.*

**Definition**

The term has been in use since the 1990s, with some giving credit to [**John Mashey**](https://en.wikipedia.org/wiki/John_Mashey) for coining or at least making it popular. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to [capture](https://en.wikipedia.org/wiki/Data_acquisition), [curate](https://en.wikipedia.org/wiki/Data_curation), manage, and process data within a tolerable elapsed time. Big data philosophy encompasses unstructured, semi-structured and structured data, however the main focus is on unstructured data. Big data "size" is a constantly moving target, as of 2012 ranging from a few dozen terabytes to many [petabytes](https://en.wikipedia.org/wiki/Petabyte) of data. Big data requires a set of techniques and technologies with new forms of integration to reveal insights from datasets that are diverse, complex, and of a massive scale.

A consensual definition that states that "Big data represents the information assets characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value".Additionally, a new V, *veracity*, is added by some organizations to describe it, revisionism challenged by some industry authorities. The three Vs (volume, variety and velocity) have been further expanded to other complementary characteristics of big data:.

Characteristics[[edit](https://en.wikipedia.org/w/index.php?title=Big_data&action=edit&section=2" \o "Edit section: Characteristics)]

Big data can be described by the following characteristics:

**Volume**

The quantity of generated and stored data. The size of the data determines the value and potential insight, and whether it can be considered big data or not.

* **Variety**

The type and nature of the data. This helps people who analyze it to effectively use the resulting insight. Big data draws from text, images, audio, video; plus it completes missing pieces through data fusion.

* **Velocity**

In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development. Big data is often available in real-time.

* **Variability**

Inconsistency of the data set can hamper processes to handle and manage it.

**Veracity**

* The [data quality](https://en.wikipedia.org/wiki/Data_quality) of captured data can vary greatly, affecting the accurate analysis.

Applications

[](https://en.wikipedia.org/wiki/File:2013-09-11_Bus_wrapped_with_SAP_Big_Data_parked_outside_IDF13_(9730051783).jpg)

Bus wrapped with [SAP](https://en.wikipedia.org/wiki/SAP_AG) Big data parked outside [IDF13](https://en.wikipedia.org/wiki/Intel_Developer_Forum).

Big data has increased the demand of information management specialists so much so that [Software AG](https://en.wikipedia.org/wiki/Software_AG), [Oracle Corporation](https://en.wikipedia.org/wiki/Oracle_Corporation), [IBM](https://en.wikipedia.org/wiki/IBM), [Microsoft](https://en.wikipedia.org/wiki/Microsoft), [SAP](https://en.wikipedia.org/wiki/SAP_AG), [EMC](https://en.wikipedia.org/wiki/EMC_Corporation), [HP](https://en.wikipedia.org/wiki/Hewlett-Packard) and [Dell](https://en.wikipedia.org/wiki/Dell) have spent more than $15 billion on software firms specializing in data management and analytics. In 2010.

Developed economies increasingly use data-intensive technologies. There are 4.6 billion mobile-phone subscriptions worldwide, and between 1 billion and 2 billion people accessing the internet. Between 1990 and 2005, more than 1 billion people worldwide entered the middle class, which means more people became more literate, which in turn led to information growth. The world's effective capacity to exchange information through telecommunication networks was 281 [petabytes](https://en.wikipedia.org/wiki/Petabytes) in 1986, 471 [petabytes](https://en.wikipedia.org/wiki/Petabytes) in 1993, 2.2 exabytes in 2000, 65 [exabytes](https://en.wikipedia.org/wiki/Exabytes) in 2007 and predictions put the amount of internet traffic at 667 exabytes annually by 2014. According to one estimate, one-third of the globally stored information is in the form of alphanumeric text and still image data, which is the format most useful for most big data applications. This also shows the potential of yet unused data (i.e. in the form of video and audio content).

While many vendors offer off-the-shelf solutions for big data, experts recommend the development of in-house solutions custom-tailored to solve the company's problem at hand if the company has sufficient technical capabilities.[[60]](https://en.wikipedia.org/wiki/Big_data#cite_note-60)

### Government

The use and adoption of big data within governmental processes allows efficiencies in terms of cost, productivity, and innovation,[[61]](https://en.wikipedia.org/wiki/Big_data" \l "cite_note-61) but does not come without its flaws. Data analysis often requires multiple parts of government (central and local) to work in collaboration and create new and innovative processes to deliver the desired outcome.

### International development[[edit](https://en.wikipedia.org/w/index.php?title=Big_data&action=edit&section=7" \o "Edit section: International development)]

Research on the effective usage of [information and communication technologies for development](https://en.wikipedia.org/wiki/Information_and_communication_technologies_for_development) (also known as [ICT4D](https://en.wikipedia.org/wiki/ICT4D)) suggests that big data technology can make important contributions but also present unique challenges to [International development](https://en.wikipedia.org/wiki/International_development).[[62]](https://en.wikipedia.org/wiki/Big_data#cite_note-62)[[63]](https://en.wikipedia.org/wiki/Big_data#cite_note-63) Advancements in big data analysis offer cost-effective opportunities to improve decision-making in critical development areas such as health care, employment, [economic productivity](https://en.wikipedia.org/wiki/Economic_productivity), crime, security, and [natural disaster](https://en.wikipedia.org/wiki/Natural_disaster) and resource management.[[64]](https://en.wikipedia.org/wiki/Big_data#cite_note-HilbertBigData2013-64)[[65]](https://en.wikipedia.org/wiki/Big_data#cite_note-65)[[66]](https://en.wikipedia.org/wiki/Big_data#cite_note-66) Additionally, user-generated data offers new opportunities to give the unheard a voice.[[67]](https://en.wikipedia.org/wiki/Big_data#cite_note-67) However, longstanding challenges for developing regions such as inadequate technological infrastructure and economic and human resource scarcity exacerbate existing concerns with big data such as privacy, imperfect methodology, and interoperability issues.

### Manufacturing[[edit](https://en.wikipedia.org/w/index.php?title=Big_data&action=edit&section=8" \o "Edit section: Manufacturing)]

Based on TCS 2013 Global Trend Study, improvements in supply planning and product quality provide the greatest benefit of big data for manufacturing. Big data provides an infrastructure for transparency in manufacturing industry, which is the ability to unravel uncertainties such as inconsistent component performance and availability. Predictive manufacturing as an applicable approach toward near-zero downtime and transparency requires vast amount of data and advanced prediction tools for a systematic process of data into useful information.[[68]](https://en.wikipedia.org/wiki/Big_data#cite_note-68) A conceptual framework of predictive manufacturing begins with data acquisition where different type of sensory data is available to acquire such as acoustics, vibration, pressure, current, voltage and controller data. Vast amount of sensory data in addition to historical data construct the big data in manufacturing. The generated big data acts as the input into predictive tools and preventive strategies such as [Prognostics](https://en.wikipedia.org/wiki/Prognostics) and Health Management (PHM).

### Healthcare

Big data analytics has helped healthcare improve by providing personalized medicine and prescriptive analytics, clinical risk intervention and predictive analytics, waste and care variability reduction, automated external and internal reporting of patient data, standardized medical terms and patient registries and fragmented point solutions.[[71]](https://en.wikipedia.org/wiki/Big_data#cite_note-ref135-71) Some areas of improvement are more aspirational than actually implemented. The level of data generated within healthcare systems is not trivial. With the added adoption of mHealth, eHealth and wearable technologies the volume of data will continue to increase. This includes electronic health record data, imaging data, patient generated data, sensor data, and other forms of difficult to process data. There is now an even greater need for such environments to pay greater attention to data and information quality.[[72]](https://en.wikipedia.org/wiki/Big_data#cite_note-72) "Big data very often means `[dirty data](https://en.wikipedia.org/wiki/Dirty_data)' and the fraction of data inaccuracies increases with data volume growth." Human inspection at the big data scale is impossible and there is a desperate need in health service for intelligent tools for accuracy and believability control and handling of information missed.[[73]](https://en.wikipedia.org/wiki/Big_data#cite_note-Mirkes2016-73) While extensive information in healthcare is now electronic, it fits under the big data umbrella as most is unstructured and difficult to use.

### Education

A [McKinsey Global Institute](https://en.wikipedia.org/wiki/McKinsey_%26_Company) study found a shortage of 1.5 million highly trained data professionals and managers[[49]](https://en.wikipedia.org/wiki/Big_data#cite_note-McKinsey-49) and a number of universities[[75]](https://en.wikipedia.org/wiki/Big_data#cite_note-75) including [University of Tennessee](https://en.wikipedia.org/wiki/University_of_Tennessee) and [UC Berkeley](https://en.wikipedia.org/wiki/UC_Berkeley), have created masters programs to meet this demand. Private boot camps have also developed programs to meet that demand, including free programs like [The Data Incubator](https://en.wikipedia.org/wiki/The_Data_Incubator) or paid programs like [General Assembly](https://en.wikipedia.org/wiki/General_Assembly). In the specific field of marketing, one of the problems stressed by Wedel and Kannan is that marketing has several subdomains (e.g., advertising, promotions, product development, branding) that all use different types of data. Because one-size-fits-all analytical solutions are not desirable, business schools should prepare marketing managers to have wide knowledge on all the different techniques used in these subdomains to get a big picture and work effectively with analysts.

### Media

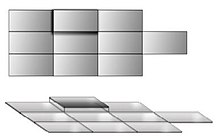
To understand how the media utilizes big data, it is first necessary to provide some context into the mechanism used for media process. It has been suggested by Nick Couldry and Joseph Turow that [practitioners](https://en.wiktionary.org/wiki/practitioner) in Media and Advertising approach big data as many actionable points of information about millions of individuals. The industry appears to be moving away from the traditional approach of using specific media environments such as newspapers, magazines, or television shows and instead taps into consumers with technologies that reach targeted people at optimal times in optimal locations. The ultimate aim is to serve or convey, a message or content that is (statistically speaking) in line with the consumer's mindset. For example, publishing environments are increasingly tailoring messages (advertisements) and content (articles) to appeal to consumers that have been exclusively gleaned through various [data-mining](https://en.wikipedia.org/wiki/Data-mining) activities.

No SQL :-

A NoSQL (originally referring to "non [SQL](https://en.wikipedia.org/wiki/SQL)" or "non relational")[[1]](https://en.wikipedia.org/wiki/NoSQL#cite_note-1) [database](https://en.wikipedia.org/wiki/Database) provides a mechanism for [storage](https://en.wikipedia.org/wiki/Computer_data_storage) and [retrieval](https://en.wikipedia.org/wiki/Data_retrieval) of data that is modeled in means other than the tabular relations used in [relational databases](https://en.wikipedia.org/wiki/Relational_database). Such databases have existed since the late 1960s, but did not obtain the "NoSQL" moniker until a surge of popularity in the early twenty-first century triggered by the needs of [Web 2.0](https://en.wikipedia.org/wiki/Web_2.0) companies such as [Facebook](https://en.wikipedia.org/wiki/Facebook), [Google](https://en.wikipedia.org/wiki/Google), and [Amazon.com](https://en.wikipedia.org/wiki/Amazon.com). NoSQL databases are increasingly used in [big data](https://en.wikipedia.org/wiki/Big_data) and [real-time web](https://en.wikipedia.org/wiki/Real-time_web) applications. NoSQL systems are also sometimes called "Not only SQL" to emphasize that they may support [SQL](https://en.wikipedia.org/wiki/SQL)-like query languages.

Motivations for this approach include: simplicity of design, simpler ["horizontal" scaling](https://en.wikipedia.org/wiki/Horizontal_scaling#Horizontal_and_vertical_scaling) to [clusters](https://en.wikipedia.org/wiki/Cluster_computing) of machines (which is a problem for relational databases), and finer control over availability. The data structures used by NoSQL databases (e.g. key-value, wide column, graph, or document) are different from those used by default in relational databases, making some operations faster in NoSQL. The particular suitability of a given NoSQL database depends on the problem it must solve. Sometimes the data structures used by NoSQL databases are also viewed as "more flexible" than relational database tables.

HISTORY

[](https://en.wikipedia.org/wiki/File:NoSQL_early_schema_2009.jpg)

First visual representation of NoSQL initiatives [[15]](https://en.wikipedia.org/wiki/NoSQL#cite_note-msjavan-15)

The term *NoSQL* was used by [Carlo Strozzi](https://en.wikipedia.org/w/index.php?title=Carlo_Strozzi_(developer)&action=edit&redlink=1) in 1998 to name his lightweight, [Strozzi NoSQL open-source relational database](https://en.wikipedia.org/wiki/Strozzi_NoSQL_(RDBMS)" \o "Strozzi NoSQL (RDBMS)) that did not expose the standard [Structured Query Language](https://en.wikipedia.org/wiki/SQL) (SQL) interface, but was still relational. His NoSQL RDBMS is distinct from the circa-2009 general concept of NoSQL databases. Strozzi suggests that, because the current NoSQL movement "departs from the relational model altogether, it should therefore have been called more appropriately 'NoREL', referring to 'No Relational'.