

PGDBD-Core course I

IT-601 Data Collection and database management

Data collection through Questionnaires and surveys. Questionnaire design steps. Procedural and contextual factors for creating a questionnaire. Factual and retrospective questions. Survey sampling. Random, stratified, Cluster and quota sampling methods. Response rates and communication strategies for improving response rates. Observational data collection: Relationships and toles. More-structured and less-structured observations. Validity of observation.

Data collection through tests. Norm-references and criterion-referenced tests. Performance tests. Data collection from documents. Text mining. Data collection through interviews. Focus groups. Action planning. Measuring reaction and planned action. Customer satisfaction. Immediate adjustments. Areas of feedback. Measuring learning and confidence. Measuring application and implementation. Measuring impact and consequences.

Fundamental concepts of the main architecture of web bases access to databases: HTTP, HTML, XML, JSON, SQL.

Basic techniques to query web documents and data sets. XPath and regular expressions.

Data scraping and text management.

Data management. Data quality. Planning for data management. Data policies. Research notes and lab notebooks. Metadata. Standards. Organization. Naming conventions. Databases. Raw versus analyzed data. Sourced and anonymous data. Storage and backup. Long term storage and archiving. Data sharing. Intellectual property and data reuse.

Managing data. Data modeling and SQL. The single entity. The one-to-many, many-to-many and one-to-one and recursive relationships. The relational model and relational algebra.

Spatial and temporal data management. HDFS and MapReduce. SQL and Java. Data integrity and data administration. NoSQL and other database architectures.

Management for health data. National Health Information technology landscape. Healthcare information systems. Population health management systems. System acquisition process. Healthcare information system standars (HCIS).

Legal and ethical aspects of data collection and sharing.

Data collection from Internet of Things: Transform data to improve reporting and reduce storage volume. Store and aggregate your IoT data across multiple database servers. Building localized, low-cost MySQL database servers using small and inexpensive computers. Connect Arduino boards and other devices directly to MySQL database servers. Build high availability MySQL solutions among low-power computing devices.

Practicals:

-
1. Setting up an HDFS platform from scratch and evaluating its performance in comparison to single site database using benchmark data sets.
 2. Writing scripts to run on a cloud platform for selected tasks.
-