

## BI – Unit 3 (Reporting Authoring) – END-SEM PYQ Answers

### MAY-JUNE 2023

#### Q1a) Explain Drill-Up and Drill-Down in detail.

[6 marks]

Drill-up and drill-down are OLAP (Online Analytical Processing) navigation techniques that allow users to change the level of detail in a report by moving up or down through a data hierarchy.

#### Drill-Down

Drill-down means moving from a high-level, summarized view to a more detailed, granular view by descending the dimension hierarchy.

- You start with summary data and click into it to expose finer details.
- Example — Sales hierarchy: Country → Region → City → Store. When you drill down from 'India' you see 'Maharashtra', then 'Pune', then individual stores.
- Useful for root-cause analysis — e.g., overall sales dropped, so you drill down to find which product category caused it.
- In BI tools like Power BI or Cognos, clicking a chart bar typically triggers drill-down automatically.

#### Drill-Up (Roll-Up)

Drill-up (also called roll-up) is the opposite — you move from detailed data back up to a summarized, aggregated view.

- Example — From individual store-level sales, you roll up to city-level totals, then region, then country.
- Aggregation functions (SUM, AVG, COUNT) are applied as you go up.
- Useful for executive dashboards that need high-level KPIs without clutter.
- Helps in trend analysis across periods — e.g., daily → monthly → quarterly → annual.

#### Comparison Table

Feature	Drill-Down	Drill-Up
Direction	High → Low detail	Low → High detail
Data scope	Narrower (more specific)	Broader (more aggregated)
Use case	Investigation, root cause	Summary, reporting to management
Hierarchy move	Descend the hierarchy	Ascend the hierarchy

**Note:** Drill-through is a related but distinct concept — it moves across dimensions rather than up/down within the same dimension (e.g., from a sales report to the underlying transaction records).

#### Q1b) Explain Drill-Through capability.

[6 marks]

Drill-through (sometimes called drill-across) allows a user to jump from one report or summary to a completely different, related report that provides detailed transactional data behind the number.

- It works across subjects or fact tables — not just within the same hierarchy.

- Example: A quarterly revenue report shows Rs. 5 crore for Q2. Drill-through takes you to the actual list of individual orders, invoices, and customer names that make up that figure.
- Unlike drill-down (which stays within the same dimension), drill-through typically opens a new report or screen.
- In enterprise BI tools (SAP BusinessObjects, IBM Cognos), drill-through is configured via 'report links' or 'go to' actions.
- Practical value: Finance managers use it to audit summary figures; sales managers use it to see which deals drove the quarter.

### **Q1c) Explain Data Grouping & Sorting, and Filtering in BI Reports with examples. [5 marks]**

#### **Data Grouping**

Grouping organizes report data into logical clusters based on a common attribute value, collapsing individual rows into a group header with aggregated metrics.

- Example: Group 'Sales' by 'Region' — all rows for 'West India' collapse under one header showing total West India sales.
- Nested grouping is also possible — Group by Region, then by Product Category within each Region.
- Grouping makes large datasets readable and reveals patterns within categories.

#### **Sorting**

Sorting arranges the rows or groups in a specific order — ascending or descending — based on one or more columns.

- Example: Sort a customer list alphabetically (A–Z) on 'Customer Name'.
- Multi-level sort: Sort first by 'Region' (A–Z), then by 'Revenue' (high to low) within each region.
- BI tools support both static (design-time) and dynamic (user-driven) sorting.

#### **Filtering**

Filtering restricts which data rows appear in a report based on conditions, reducing noise and focusing on relevant information.

- Example: Filter to show only transactions where 'Status = Completed' and 'Amount > 10000'.
- Filters can be applied at the data source level (pre-aggregation) or at the report display level (post-aggregation).
- Interactive filters (slicers in Power BI, prompt filters in Cognos) let the end user dynamically change what they see.
- Date range filters are extremely common — e.g., 'Show data only for the current financial year'.

### **Q2a) Explain Multidimensional Data Model with example. [6 marks]**

A multidimensional data model organizes data into a cube structure (often called an OLAP cube) where data is viewed along multiple dimensions simultaneously. This is in contrast to the flat, two-dimensional tabular structure of relational databases.

#### **Core Concepts**

- Dimensions — The axes of analysis. Common dimensions: Time, Geography, Product, Customer.

- Measures (Facts) — The numeric values being analyzed, e.g., Sales Amount, Quantity Sold, Profit.
- Cells — Each cell in the cube holds the measure value for a unique combination of dimension values.
- Hierarchies — Each dimension has a hierarchy, e.g., Time: Year → Quarter → Month → Day.

### Case Study: Retail Sales Cube

Consider a supermarket chain analyzing sales data across three dimensions: Time, Product, and Store Location.

- Dimension 1 — Time: Year (2024) → Quarter (Q1) → Month (Jan)
- Dimension 2 — Product: Category (Electronics) → Sub-category (Phones) → SKU (iPhone 15)
- Dimension 3 — Location: Country (India) → State (Maharashtra) → City (Pune) → Store
- Measure: Revenue (Rs.), Units Sold

A query like 'What were total phone sales in Pune in Q1 2024?' navigates to that exact cell in the cube instantly, without scanning millions of rows.

### Advantages of Multidimensional Data Model

- Fast query performance — pre-aggregated data means even complex cross-dimensional queries run in seconds.
- Intuitive navigation — business users can slice, dice, and pivot data without SQL knowledge.
- Supports OLAP operations — drill-down, roll-up, slice, dice, and pivot are all natively supported.
- Consistent calculations — measures like YTD sales or growth % are computed once and stored.
- Ideal for dashboards and executive reporting where summarized, multi-angle views are needed.

## Q2b) Importance of Conditional Formatting and Adding Calculations in Reports. [6 marks]

### Conditional Formatting

Conditional formatting applies visual styling (colors, icons, data bars) to cells or values automatically based on rules, making patterns and anomalies immediately visible without manual inspection.

- Traffic-light indicators: Sales cells colored red/yellow/green based on whether targets are met.
- Heatmaps: Revenue by region where darker blue = higher revenue — geographic patterns stand out instantly.
- Threshold highlighting: Any expense exceeding budget automatically appears in bold red.
- Icon sets: KPI dashboards use up/down arrows to show improvement or decline versus the previous period.
- Business value: Reduces the cognitive load on report readers — the eye is drawn to what needs attention.

### Adding Calculations

Calculations extend raw data into derived metrics that have direct business meaning. They can be in-report (calculated columns) or aggregated (summary measures).

- Simple arithmetic: Profit = Revenue – Cost, automatically computed for every row.
- Percentage calculations: Market share = (Brand Sales / Total Sales) × 100.
- Running totals and cumulative sums: Track year-to-date revenue as each month is added.

- Variance analysis: Actual vs. Budget = Actual – Budget; % Variance = (Actual – Budget) / Budget × 100.
- Ranking: Rank stores by revenue to instantly identify top and bottom performers.

**Note:** In Cognos, calculations are added via the 'Query Explorer' as expression items. In Power BI, they are written as DAX measures.

### Q2c) How does a Business Report help an organization?

[5 marks]

Business reports are structured documents that present analyzed data to support decision-making across all levels of an organization.

- Informed decision-making: Leaders make strategic choices based on data evidence rather than intuition.
- Performance monitoring: KPI reports track progress towards goals (sales targets, production efficiency, cost controls).
- Trend identification: Time-series reports expose growth or decline patterns before they become critical problems.
- Accountability: Departmental reports make teams responsible for their numbers — visible performance drives behavior.
- Communication: Reports standardize how information is shared across teams, ensuring everyone works from the same data.
- Compliance and auditing: Financial and operational reports provide audit trails for regulatory requirements.
- Resource optimization: Reports reveal inefficiencies (e.g., high inventory holding costs) enabling better allocation.

## NOV-DEC 2023

### Q1a) State different types of reports with their applications.

[6 marks]

BI reports come in several structural types, each suited to different analytical purposes.

- List Reports — The most basic form; display raw or filtered tabular data row by row. Application: Employee roster, transaction log, inventory list.
- Crosstab (Pivot) Reports — Display data in a matrix with row and column dimensions and summary cells at intersections. Application: Sales by Product vs. Region, headcount by Department vs. Grade.
- Statistical Reports — Present descriptive statistics (mean, median, standard deviation) and distribution summaries. Application: Quality control in manufacturing, academic performance analysis.
- Chart / Graphical Reports — Visualize data using bar charts, line charts, pie charts, scatter plots. Application: Sales trend over time (line), market share (pie), regional comparison (bar).
- Map Reports — Overlay data on geographic maps using choropleth shading or bubble markers. Application: Store performance by city, disease spread visualization, logistics route analysis.
- Financial Reports — Structured around accounting conventions: P&L statements, balance sheets, cash flow statements. Application: Quarterly earnings reports, budget variance analysis.
- Dashboard Reports — Aggregate multiple KPIs and mini-charts on a single screen for at-a-glance status. Application: Executive dashboards, operations control centers.

**Q1b) What are the best practices in dashboard design?****[6 marks]**

A well-designed dashboard communicates critical information at a glance. Poor design leads to cognitive overload and misinterpretation.

- Define the audience first — an executive needs high-level KPIs; an operations analyst needs granular drill-down capability. Design for the specific user.
- Limit to 5–7 KPIs per screen — too many metrics dilute focus. Prioritize what truly drives decisions.
- Place the most important metric in the top-left position — users follow the Z or F reading pattern.
- Use consistent color semantics — red always means bad/alert, green always means good/on-target throughout the dashboard.
- Choose the right chart type — use line charts for trends over time, bar charts for comparisons, pie charts only for parts-of-a-whole (max 5 slices).
- Minimize chart junk — remove gridlines, 3D effects, and decorative elements that add no information value (Tufte's data-ink ratio principle).
- Include interactivity — filters, drill-downs, and date selectors let users self-serve without needing a new report.
- Ensure mobile responsiveness — dashboards must render well on tablets and phones for field users.
- Refresh cadence — clearly display the 'Last Updated' timestamp so users know data freshness.

**Q1c) Difference between Relational and Multidimensional Data Models.****[6 marks]**

Aspect	Relational Model	Multidimensional Model
Structure	2D tables (rows & columns)	N-dimensional cube (dimensions & measures)
Query language	SQL	MDX (Multi-Dimensional Expressions)
Optimized for	OLTP — transactional inserts/updates	OLAP — analytical queries & aggregations
Data storage	Normalized (3NF) — minimal redundancy	De-normalized or pre-aggregated — fast reads
Query speed	Slower for complex aggregations	Very fast for pre-aggregated queries
Data model	Star/Snowflake schema for analytics	Cube with dimension hierarchies
User skill	Requires SQL knowledge	Business users can use GUI navigation
Tools	MySQL, PostgreSQL, Oracle	Microsoft SSAS, IBM Cognos TM1, SAP BW

**Q2a) [REPEATED] Suggest the use of Data Grouping & Sorting, Filtering Reports. [6 marks]****Q2b) What is a File Extension? Explain the structure of a CSV file. [6 marks]**

A file extension is the suffix appended to a filename (after the dot) that indicates the file format and the type of application needed to open it. For example, '.pdf' indicates a Portable Document Format file, '.xlsx' indicates an Excel spreadsheet.

**Common BI Report Output Formats**

- .pdf — Portable Document Format: Fixed layout, ideal for sharing printed-style reports that must look identical on any device.
- .xlsx — Excel: Best when recipients need to perform further analysis, filtering, or pivot table operations on the data.
- .csv — Comma-Separated Values: Plain text, universally compatible, ideal for data interchange between systems.
- .xml — Extensible Markup Language: Structured text with tags, used for data exchange between applications.
- .html — Web format, used when reports are embedded in web portals or email.

**Structure of a CSV File**

A CSV (Comma-Separated Values) file is a plain text file where each line represents one data record, and within each line the individual field values are separated by commas.

- Header row (optional but standard): The first line contains column names — e.g., 'EmployeeID,Name,Department,Salary'
- Data rows: Each subsequent line contains the actual values in the same order — e.g., '101,Riya Sharma,Engineering,75000'
- Quoting: If a field itself contains a comma, it must be enclosed in double quotes — e.g., 'Pune, MH' becomes '"Pune, MH"'
- Line endings: Each record ends with a newline character (\n on Unix, \r\n on Windows).
- No data types: CSV stores everything as text; the consuming application interprets data types.

**Note:** CSVs are the most universally supported export format. Every BI tool, database, and spreadsheet application can import/export CSV.

**Q2c) [REPEATED] Explain Drill-Up and Drill-Down in detail. [6 marks]****MAY-JUNE 2024****Q1a) What are the important BI Reporting Practices? [6 marks]**

BI reporting practices are the guidelines and standards that ensure reports are accurate, useful, and efficiently consumed by their audience.

- Define clear objectives: Every report must answer a specific business question. Avoid creating reports 'just to have data'.
- Know your audience: Tailor language, level of detail, and visual style to the consumer — executive, analyst, or operational staff.
- Data quality first: Validate and clean source data before report publication. A report built on dirty data misleads rather than informs.

- Use standard templates: Consistent headers, footers, color schemes, and KPI definitions across all reports in the organization.
- Version control and scheduling: Reports should be scheduled to refresh automatically and versioned so historical snapshots can be retrieved.
- Appropriate visualization: Match the chart type to the data relationship being communicated. Do not use pie charts for more than 5 categories.
- Drill-down capability: Allow users to move from summary to detail rather than producing separate reports for every level.
- Security and access control: Sensitive reports (HR, Finance) must be access-restricted to authorized roles only.
- Documentation: Maintain a data dictionary that defines every metric, formula, and filter used in the report.

**Q1b) [REPEATED] Discuss the importance of Drill-Up, Drill-Down, and Drill-Through capabilities. [6 marks]**

**Q1c) [REPEATED] Explain Data Grouping, Sorting, and Filtering in BI Reports with examples. [5 marks]**

**Q2a) [REPEATED] Explain the Multi-Dimensional Data Model with a case study and its advantages. [6 marks]**

**Q2b) [REPEATED] Importance of Conditional Formatting and Adding Calculations in Reports. [6 marks]**

**Q2c) [REPEATED] How does a Business Report help an organization? [5 marks]**

## Additional Concepts & Quick Reference

### OLAP Operations Summary

Operation	Description	Example
Slice	Fix one dimension to a single value and view the rest	Sales for Q1 only (Time dimension fixed)
Dice	Fix multiple dimensions to ranges to get a sub-cube	Electronics sales in West India for Q1–Q2
Drill-down	Move to finer granularity within a hierarchy	Country → State → City → Store
Roll-up	Aggregate to coarser granularity	Daily → Monthly → Quarterly → Annual
Pivot	Rotate the cube to view a different face	Swap rows and columns in a crosstab
Drill-through	Jump to underlying transactional detail	Click revenue figure → see individual invoices

## Report Output Formats Comparison

Format	Best For	Limitation
PDF	Fixed-layout final reports for printing/sharing	Cannot do further analysis on data
Excel (XLSX)	Further analysis, pivot tables, charts by recipient	Formatting can break on large datasets
CSV	Data exchange between systems, import into other tools	No formatting, no formulas
XML	Machine-to-machine data exchange	Not human-readable without a parser
HTML	Web portal embedding, email delivery	Complex formatting is hard to maintain

## Types of Filters in BI

- Pre-query filters — applied before data is fetched from the database (most efficient, reduces data transfer).
- Post-query filters — applied after data is retrieved, filtering the result set in the report layer.
- Parameterized / Prompted filters — the user is asked to select a value when the report runs (e.g., 'Select a date range').
- Context filters (Cascading) — one filter's selection changes the available options in another filter (e.g., selecting State filters the City dropdown to only show cities in that state).
- Report-level vs. Visual-level filters (Power BI) — report-level filters affect all visuals; visual-level filters affect only one chart.

## NOV-DEC 2025

**Q1a) [REPEATED] State different types of reports with their applications. [6]**

**Q1b) [REPEATED] Explain the Multidimensional Data Model with example. [6]**

**Q1c) [REPEATED] Difference between Relational and Multidimensional Data Model. [6]**

**Q2a) [REPEATED] Suggest the use of Data Grouping & Sorting, Filtering Reports. [6]**

**Q2b) Explain the Relational Data Model with example. [6]**

The relational data model, introduced by E.F. Codd in 1970, organizes data into a collection of two-dimensional tables called relations. Each table has a fixed set of named columns (attributes) and stores data in rows (tuples). The model is based on set theory and predicate logic, giving it a rigorous mathematical foundation that makes data manipulation precise and predictable.

## Core Concepts

- Relation (Table): A named table with a fixed schema. Example: a table called EMPLOYEE has columns EmployeeID, Name, DepartmentID, and Salary.



- **Attribute (Column):** A named property of the relation. Each attribute has a defined domain — the set of permissible values (e.g., Salary domain = positive integers).
- **Tuple (Row):** A single record in the table. Each tuple represents one real-world entity instance (e.g., one specific employee).
- **Primary Key:** One or more attributes that uniquely identify each tuple. In EMPLOYEE, EmployeeID is the primary key — no two employees share the same ID.
- **Foreign Key:** An attribute in one table that references the primary key of another, establishing a relationship between tables. DepartmentID in EMPLOYEE references the DEPARTMENT table's primary key.
- **Schema vs. Instance:** The schema is the structure (column names and types — static). The instance is the current content of the table (rows — changes over time).

### Case Study: A Company Database

Consider a small company database with three tables:

EMPLOYEE			
EmplID (PK)	Name	DeptID (FK)	Salary
E001	Ananya Kulkarni	D01	72000
E002	Rohan Mehta	D02	65000
E003	Priya Singh	D01	80000

DEPARTMENT	
DeptID (PK)	DeptName
D01	Engineering
D02	Marketing

A query like 'Find the names and salaries of all Engineering employees' is expressed in SQL as: `SELECT E.Name, E.Salary FROM EMPLOYEE E JOIN DEPARTMENT D ON E.DeptID = D.DeptID WHERE D.DeptName = 'Engineering'`. This returns Ananya Kulkarni (72000) and Priya Singh (80000).

### Key Properties (Codd's Rules — simplified)

- Data is represented exclusively in tables — no hidden pointers or links.
- Every value is accessible by specifying the table name, primary key value, and column name.
- NULL values are supported to represent missing or inapplicable information.
- Data manipulation is done through a high-level language (SQL) — the user specifies what to retrieve, not how.
- Data integrity is enforced through constraints: primary key uniqueness, foreign key referential integrity, and domain constraints.

### Advantages and Limitations

- **Advantages:** Simple, intuitive table structure; strong theoretical foundation; mature tooling (MySQL, PostgreSQL, Oracle); excellent for OLTP workloads with frequent inserts, updates, and deletes; ACID compliance ensures transactional reliability.

- Limitations: Joins across multiple tables become expensive for complex analytical queries spanning many tables; does not natively support the hierarchical, multi-dimensional navigation patterns needed for OLAP reporting; schema changes require ALTER TABLE operations which can be disruptive.

**Note:** *The relational model powers virtually all enterprise transaction systems (ERPs, CRMs, banking systems). For analytical workloads, data is often copied from relational OLTP systems into a dimensional (star schema) model in a data warehouse.*

**Q2c) [REPEATED] Explain Drill Up and Drill Down in detail.**

**[6]**

## MAY-JUN 2025

---

**Q1a) Explain Business Intelligence Architecture in detail with a suitable diagram.**

**[6]**

Business Intelligence architecture describes the layered technical framework through which raw data flows from source systems all the way to decision-makers as actionable insights. Understanding this architecture is critical because each layer has its own tools, responsibilities, and quality requirements — a failure at any layer cascades downstream.

### Layer 1: Data Sources

The foundation of BI is data originating from diverse internal and external sources. Internal sources include transactional databases (ERP systems like SAP, CRM systems like Salesforce, HR systems, POS terminals). External sources include market data feeds, government statistical datasets, social media APIs, and weather data. These systems use different formats, schemas, and update frequencies, making integration the first major challenge.

### Layer 2: ETL (Extract, Transform, Load) / Data Integration

ETL processes extract data from source systems on a scheduled or real-time basis, transform it by cleaning errors, resolving inconsistencies, standardizing formats, and applying business rules, then load it into the central data store. Modern architectures increasingly use ELT (Extract, Load, Transform) where raw data is loaded first and transformed on-demand using the processing power of cloud data warehouses.

- **Extract:** Connect to each source system and pull the relevant data (full extract or incremental delta since last run).
- **Transform:** Apply data quality rules — remove duplicates, fill missing values, convert date formats, unify naming conventions (e.g., 'Male'/'M'/'1' all mapped to 'Male'), derive calculated fields.
- **Load:** Write the cleaned, transformed data into the target data warehouse or data mart, maintaining historical snapshots.

### Layer 3: Data Warehouse / Data Mart

The central repository that stores integrated, historical, subject-oriented data optimized for analytical queries. A data warehouse serves the entire enterprise. A data mart is a subject-specific subset (e.g., the Sales data mart, the Finance data mart) that serves a specific business unit. The schema used is typically a star schema or snowflake schema — not the normalized 3NF of OLTP — to maximize query performance.

### Layer 4: OLAP Server (Analytical Processing)

The OLAP layer pre-aggregates the warehouse data into multidimensional cubes, enabling business users to perform drill-down, roll-up, slice, dice, and pivot operations in sub-second response times — even on billions of records. ROLAP (Relational OLAP) keeps data in relational tables and generates dynamic SQL. MOLAP (Multidimensional OLAP) stores pre-aggregated data in a proprietary multidimensional cube structure for maximum query speed.

### Layer 5: BI Presentation / Front End

The top layer where analytical results are delivered to end users through dashboards, reports, charts, maps, and alerts. Tools in this layer include Tableau, Power BI, IBM Cognos, and SAP BusinessObjects. Self-service BI tools in this layer allow business analysts to build their own reports without IT involvement.

### Architecture Diagram (Text Representation)

Layer	Components	Purpose
1. Data Sources	ERP, CRM, HRMS, External APIs, Flat files	Raw business data origin
2. ETL / Integration	Informatica, SSIS, Talend, Apache NiFi	Extract, clean, transform, and load data
3. Data Warehouse / Mart	Amazon Redshift, Snowflake, Google BigQuery	Centralized analytical data store
4. OLAP Server	Microsoft SSAS, IBM Cognos TM1, SAP BW	Pre-aggregated cubes for fast queries
5. BI Presentation	Tableau, Power BI, IBM Cognos Analytics	Dashboards, reports, alerts for users

**Note:** Modern cloud BI architectures often collapse Layers 3 and 4 using lakehouse platforms (Databricks, Snowflake) that perform on-demand OLAP-style queries directly on the data lake without pre-building cubes.

### Q1b) Explain the Role of Mathematical Models in Business Intelligence.

[6]

Mathematical models are the analytical engine inside BI — they transform raw data into quantified predictions, optimized decisions, and scientifically defensible insights. Without mathematical models, BI would be limited to descriptive summaries; with them, it becomes predictive and prescriptive.

#### 1. Statistical Models

Statistical models quantify relationships between variables and provide measures of uncertainty around conclusions. They form the foundation of all analytical BI.

- Descriptive statistics (mean, median, standard deviation, correlation coefficients) summarize datasets and identify patterns.
- Regression models (linear, logistic, polynomial) quantify how changes in input variables predict changes in an output variable — e.g., 'For every additional Rs. 1 lakh in advertising spend, revenue increases by Rs. 3.2 lakh.'

- Hypothesis testing (t-test, chi-squared test, ANOVA) provides rigorous statistical evidence for whether observed differences are genuine or due to chance — critical for A/B testing of business decisions.

## 2. Forecasting and Time-Series Models

These models analyze historical temporal patterns to predict future values, which is essential for demand planning, financial budgeting, and capacity management.

- ARIMA (Autoregressive Integrated Moving Average): Models time series based on its own past values and past forecast errors. Used for short-to-medium term sales and inventory forecasting.
- Exponential Smoothing: Assigns exponentially decreasing weights to older observations, making the model more responsive to recent trends. Used in supply chain inventory replenishment.
- Facebook Prophet: Handles seasonality, holidays, and trend shifts automatically — widely adopted for business time-series forecasting.

## 3. Optimization Models

Optimization models find the best possible decision from all feasible options, subject to constraints. They represent the prescriptive analytics tier.

- Linear Programming (LP): Finds the optimal allocation of resources (raw materials, workforce, machine time) to maximize profit or minimize cost, subject to linear constraints. Used in production planning and logistics routing.
- Integer Programming: Same as LP but decisions must be whole numbers — used for workforce scheduling ('How many nurses do we need on each shift?').

## 4. Machine Learning Models

ML models discover complex, non-linear patterns in data that traditional statistical models cannot capture. Decision trees, neural networks, random forests, and support vector machines all provide predictive capabilities that power modern BI applications from churn prediction to fraud detection.

## 5. Simulation Models

Monte Carlo simulation uses random sampling to model the probability distribution of outcomes in complex systems with uncertainty. A financial BI tool might run 10,000 simulations of possible economic scenarios to estimate the probability distribution of next year's revenue — providing risk quantification rather than a single point forecast.

### Q1c) What is Data, Information, and Knowledge? Explain with suitable examples. [5]

These three terms represent a progression in the value and actionability of organized facts, forming the foundation of the Knowledge Hierarchy (also called the DIKW pyramid — Data, Information, Knowledge, Wisdom).

#### Data

Data is raw, unprocessed facts and observations, lacking context or meaning on their own. Data is the input to all analytical processes. It can be numerical, textual, image-based, or any other recorded observation.

- Example: '42', '2025-06-01', 'Pune', '15000', '72' are individual data points. They have no meaning in isolation.

- In business: transaction logs from a POS system record item codes, prices, and timestamps — raw data about purchases.

### Information

Information is data that has been processed, organized, structured, and given context so that it becomes meaningful and useful. The transformation from data to information answers the question 'Who? What? When? Where?'

- Example: 'Customer ID 42 from Pune purchased Product Code P15000 for Rs. 72 on June 1, 2025.' — the raw data points are now organized into a meaningful statement about a real-world event.
- In business: a monthly sales report showing total revenue by product category is information — the raw transactions have been aggregated and structured.

### Knowledge

Knowledge is information that has been synthesized, analyzed, and understood in context — combined with experience and judgment to enable understanding of patterns, cause-and-effect relationships, and principles. Knowledge answers 'How?' and 'Why?' and enables prediction.

- Example: 'Customers from Pune who purchase Product P15000 in June are 3x more likely to return and buy the extended warranty within 30 days.' — this is derived from analyzing patterns across thousands of information records and represents actionable understanding.
- In business: a data scientist's model that predicts customer churn based on behavioral patterns is operational knowledge embedded in a system.

Level	Nature	Example	Question Answered
Data	Raw facts, no context	'42', 'Pune', '15000'	None — just records
Information	Processed, structured, contextual	'Customer 42 from Pune bought Rs.72 product'	Who, What, When, Where
Knowledge	Synthesized patterns + understanding	'Pune customers re-buy warranty within 30 days'	How, Why — enables prediction
Wisdom	Applied knowledge with judgment	Decision to run a post-sale warranty upsell campaign	What should be done?

**Note:** BI operates primarily at the Information and Knowledge levels. Raw data enters the BI pipeline and is processed into information (reports, dashboards) and then into knowledge (predictive models, strategic insights).

**Q2a) [REPEATED] Explain different types of reports in detail.**

**[6]**

**Q2b) Explain Decision Support System (DSS) in detail.**

**[6]**

A Decision Support System (DSS) is an interactive, computer-based information system designed to help managers and analysts make semi-structured or unstructured decisions by providing data access, analytical models, and intuitive interfaces. Unlike transaction processing systems (which automate routine, structured operations), DSS focuses on the judgement-intensive decisions where human expertise must be combined with data-driven analysis.

## Characteristics of DSS

- Semi-structured problem support: DSS handles decisions that are too complex for fully automated rules but structured enough to benefit from computational analysis — e.g., 'Which new market should we enter?' or 'How should we price this new product?'
- Interactivity and flexibility: Decision-makers can change assumptions, run 'what-if' scenarios, and explore data from multiple angles in real time, rather than waiting for a pre-defined report.
- Model-driven or data-driven: Some DSS are built around mathematical models (optimization, simulation); others are primarily data-driven (warehoused data + OLAP tools).
- User-friendly interface: DSS must be accessible to managers who are domain experts, not IT specialists, so graphical dashboards and natural language interfaces are common.

## Components of a DSS

- Database Management System (DBMS): Stores the historical and current data needed for analysis — often sourced from an enterprise data warehouse.
- Model Management System: Contains a library of analytical models — financial models, forecasting models, optimization models — that can be invoked and parameterized by the user.
- User Interface (Dialogue Management): The presentation layer through which users interact with the system — dashboards, report generators, drill-down tools, and what-if scenario controls.
- Knowledge Base (in intelligent DSS): Expert system rules or machine learning models that encode domain knowledge and provide recommendations.

## Relationship between DSS and BI

Modern BI platforms are essentially sophisticated, enterprise-scale DSS. The distinction is primarily historical — DSS was the 1970s-1980s concept; BI evolved from it in the 1990s with the addition of data warehousing and OLAP. Today, a Power BI or Tableau dashboard with drill-down capabilities and what-if sliders is effectively a DSS for business users.

Example: A supply chain DSS allows a logistics manager to input projected demand figures, run an optimization model to determine the optimal warehouse stock levels across 15 locations, and compare three supplier contract scenarios side-by-side before committing to a sourcing decision.

## Q2c) Short note on Ethics and Business Intelligence.

[5]

As BI systems grow more powerful and pervasive, collecting and analyzing data about customers, employees, and partners, the ethical dimensions of their use have become critically important. Unethical BI practices damage trust, invite regulatory action, and cause real harm to individuals.

## Key Ethical Issues in BI

- Privacy and Data Collection: Organizations often collect far more personal data than is necessary for their stated purpose — a practice called data maximalism. Ethical BI limits collection to what is genuinely needed and obtains meaningful consent from individuals whose data is used.
- Data Security and Breach Responsibility: Storing large analytical datasets creates attractive targets for attackers. Organizations have an ethical obligation to implement robust security controls, disclose breaches promptly, and not collect data they cannot adequately protect.
- Algorithmic Bias and Fairness: Predictive models trained on historical data can perpetuate and amplify existing societal biases. A credit scoring model trained on historical loan data may systematically disadvantage certain demographic groups if the training data reflects past

discriminatory practices. Ethical BI requires bias auditing, fairness constraints, and diverse training data.

- **Transparency and Explainability:** Individuals affected by BI-driven decisions (e.g., denied a loan, flagged as a fraud risk) have a right to understand why. 'Black box' models that cannot be explained violate this principle and increasingly violate regulations like GDPR's 'right to explanation'.
- **Data Ownership and Consent:** Whose data is it? Ethical BI frameworks hold that individuals own their personal data. Organizations act as stewards, not owners, and must use data only within the scope of the consent given.
- **Misuse of Insights:** BI insights can be weaponized — e.g., using customer vulnerability data to target financially distressed individuals with predatory loan offers, or using employee monitoring data to unfairly penalize workers. Ethical BI requires governance frameworks defining permissible uses of analytical findings.

### Regulatory Frameworks

GDPR (EU General Data Protection Regulation) and India's DPDP Act (Digital Personal Data Protection Act, 2023) establish legal frameworks for data privacy that partially codify ethical requirements — consent, purpose limitation, data minimization, and rights of access/erasure. Ethical BI practice means going beyond mere legal compliance to genuinely respect the interests of data subjects.

**Note:** *The ACM and IEEE both have codes of ethics covering computing professionals' responsibilities around data privacy, algorithmic fairness, and transparency — relevant to any BI practitioner.*

**Cross-Reference: All Unit 3 Questions Across All 5 Years**

Question Topic	MJ-23	ND-23	MJ-24	ND-25	MJ-25
Types of Reports	-	Q1a	Q1a*	Q1a	Q2a
Multidimensional Data Model	Q2a	Q1c*	Q2a*	Q1b	-
Relational vs Multidimensional (diff)	-	Q1c	Q1c*	Q1c	-
Relational Data Model (with example)	-	-	-	Q2b	-
Data Grouping, Sorting, Filtering	Q1c	Q2a	Q1c*	Q2a	-
Conditional Formatting + Calculations	Q2b	-	Q1b*	-	-
Drill-Up and Drill-Down	Q1a	Q2c*	Q1b*	Q2c	-
Drill-Through	Q1b	-	-	-	-
CSV File / File Extensions	-	Q2b	-	-	-
BI Architecture (Unit 1/2)	-	-	-	-	Q1a
Mathematical Models in BI (Unit 1/2)	-	-	-	-	Q1b
Data, Information, Knowledge (Unit 1/2)	-	-	-	-	Q1c
Decision Support System (Unit 1/2)	-	-	-	-	Q2b
Ethics and BI (Unit 1/2)	-	-	-	-	Q2c

'\*' indicates the question was repeated from an earlier year. '-' means the topic was not asked that year. Questions appearing 3+ times are very high-priority topics.