MODULE 2

Database and Business Intelligence

1. Database Approach
2. Big Data
3. Data Warehouses and Data Marts
4. Managing data resources:
   - Establishing an information policy
   - ensuring data quality
5. Business intelligence (BI):
   - Decision Making Process
   - BI for Data analytics and Presenting Results

Managing Data

- All IT applications require data.
- These data should be of high quality, meaning that they should be accurate, complete, timely, consistent, accessible, relevant, and concise.
- Unfortunately, the process of acquiring, keeping, and managing data is becoming increasingly difficult.

- Difficulties of Managing Data
- Data Governance

Difficulties of Managing Data

- The amount of data increases exponentially over time
- Data are scattered throughout organizations
- Data are generated from multiple sources (internal, personal, external)
  - **Internal Data Sources** (e.g., corporate databases and company documents)
  - **Personal Data Sources** (e.g., personal thoughts, opinions, and experiences)
  - **External Data Sources** (e.g., commercial databases, government reports, and corporate Web sites).
- New sources of data (e.g., blogs, podcasts, videocasts, and RFID tags and other wireless sensors)

Difficulties of Managing Data(cntd)

- Data Degradation(e.g., customers move to new addresses, change their names, etc.)
- **Data Rot:** refers primarily to problems with the media on which the data are stored. Over time, temperature, humidity, and exposure to light can cause physical problems with storage media and thus make it difficult to access the data.
- Data security, quality, and integrity are critical
- Legal requirements change frequently and differ among countries & industries

Data Governance

- **Data governance** is the overall management of the **availability, usability, integrity, and security of data** used in an organization.

Key Components of Data Governance:

- **Data Quality** – Ensuring data is accurate, consistent, and up to date.
- **Data Ownership & Stewardship** – Defining who is responsible for specific data assets.
- **Data Policies & Standards** – Creating rules for data usage, access, sharing, retention, etc.
- **Data Security & Privacy** – Protecting data from unauthorized access and ensuring compliance with regulations .
- **Metadata Management** – Managing data about data (e.g., definitions, origins, structure).
- **Data Architecture & Integration** – Ensuring systems and processes work together and support governance goals.
- **Compliance & Risk Management** – Monitoring and ensuring data practices comply with laws and reduce business risk.

What is a Database?

- A **database** is a structured collection of data that allows easy access, management, and updating. Think of it like a smart digital filing cabinet.

Examples:

- A customer database for a business
- A student record system for a school
- A hospital patient management system

Database Approaches

There are different **approaches or models** to how databases are built and managed:

1. File-Based Approach (Old Method)

- Data is stored in separate files.
- Each application has its own data files.
- Problems: **Redundancy**, **inconsistency**, hard to **share** or **secure** data.

2. Database Approach (Modern Method)

- All data is stored in a centralized database.
- Managed by a **DBMS (Database Management System)**.
- Reduces redundancy, allows multi-user access, enforces integrity and security.

Types of Database Models

Model	Description
Hierarchical	Tree-like structure (like folders)
Network	More flexible connections; multiple parent-child links
Relational	Data in **tables** (rows and columns) — most common (e.g., MySQL, PostgreSQL
Object-Oriented	Data as objects (used in programming-heavy apps)
NoSQL	Non-tabular; for big, fast-changing data (e.g., MongoDB, Cassandra)

Common DBMS Tools

DBMS	Type
MySQL	Relational
PostgreSQL	Relational
Oracle DB	Relational
MongoDB	NoSQL
Microsoft SQL Server	Relational
Firebase	NoSQL (Cloud)

Databases in Information Systems

Roles in Information Systems

- **Data Storage:** Central repository for transactional and operational data.
- **Data Retrieval:** Enables fast querying using SQL (Structured Query Language).
- **Data Integrity & Security:** Enforces data accuracy, consistency, and access control.
- **Support for Applications:** Power business applications like ERP, CRM, HR systems, etc.

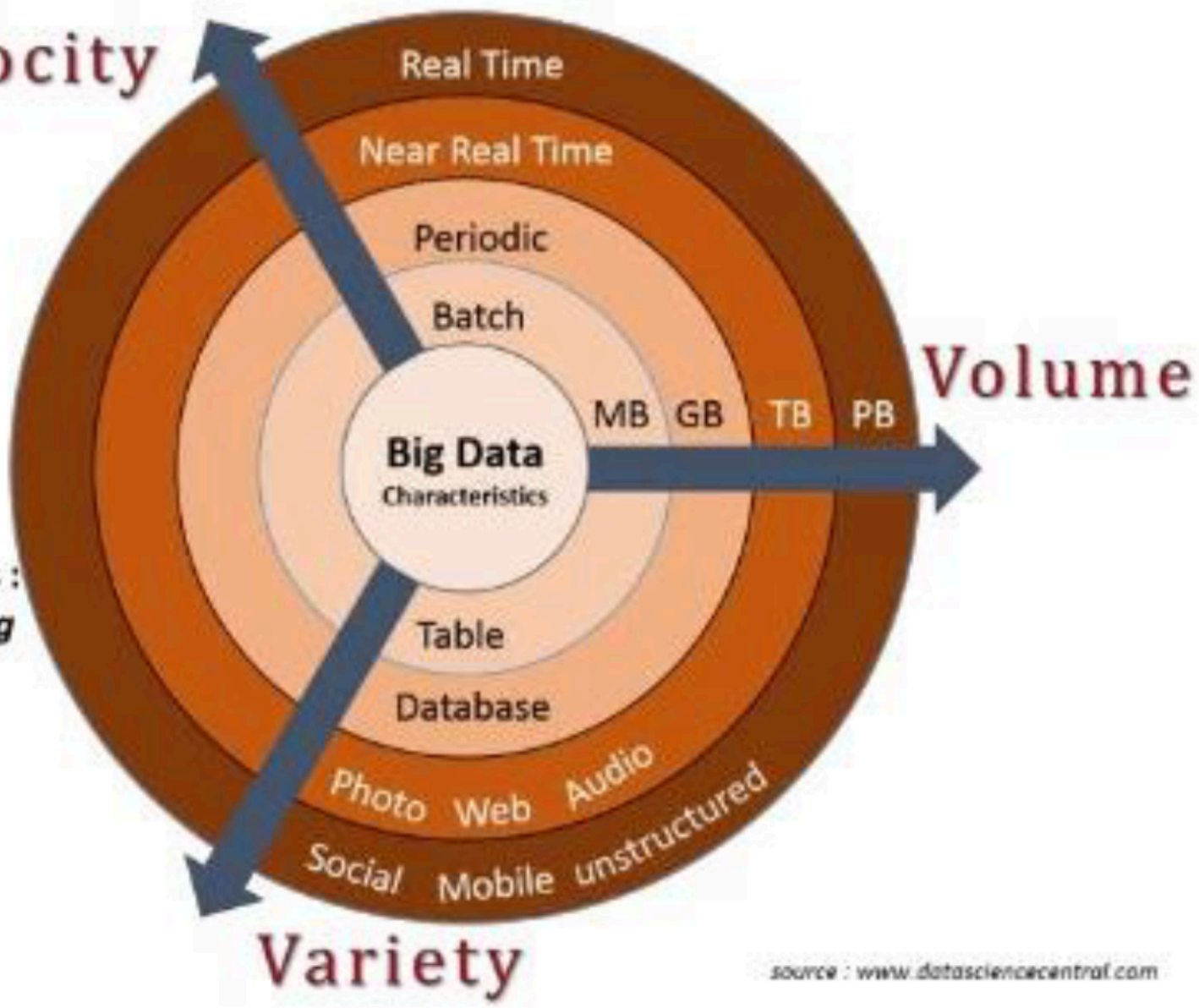
Big Data:

- Big data
  - Massive sets of unstructured/semi-structured data from Web traffic, social media, sensors, and so on
  - Can reveal more patterns and anomalies

Defining Big Data: The Big Data Institute (TBDI)

- Vast Datasets that:
  - Exhibit variety
  - Include structured, unstructured, and semi-structured data
  - Generated at high velocity with an uncertain pattern
  - Do not fit neatly into traditional, structured, relational databases
  - Can be captured, processed, transformed, and analyzed in a reasonable amount of time only by sophisticated information systems.

Characteristics of Big Data



Big Data Characteristics : ***Ever Expanding horizon***

Characteristics of Big Data

- **Volume:** incredible volume of data. This refers to the sheer amount of data being generated and collected. Big data involves vast quantities of information, often measured in terabytes or petabytes
- **Velocity:** The rate at which data flow into an organization is rapidly increasing. This is the speed at which data is created, processed, and analyzed.
- **Variety:** Big Data formats change rapidly and can include satellite imagery, broadcast audio streams, digital music files, Web page content. Big data comes in many forms, including structured data (like databases), semi-structured data (like XML or JSON), and unstructured data (like text, images, and videos).

The diversity of data types and sources adds complexity to managing and analyzing big data

3Vs-5Vs-7Vs of Big data

4.Veracity: This aspect deals with the quality and accuracy of the data. Given the vast volume and variety of big data, ensuring that the data is reliable and accurate is a significant challenge.

5.Value: The ultimate goal of big data is to extract meaningful insights and actionable information that can drive decision-making and create value. This involves analyzing data to uncover patterns, trends, and correlations that can be beneficial.

In addition to these five Vs, some frameworks also include:

•**Variability**: Refers to the changing nature of data and its consistency. For example, social media data can have fluctuating patterns of use and behavior.

•**Visualization**: The capability to present data in a visually accessible manner, enabling easier interpretation of complex data sets.

These characteristics highlight the complexity and potential of big data and underscore the need for advanced technologies and methodologies to effectively manage and analyze it.

Issues with Big Data

- Big Data can come from **untrusted sources**.
- **Big Data is dirty:** Dirty data refers to inaccurate, incomplete, incorrect, duplicate, or erroneous data.

Big Data changes:

- Data quality in an analysis can change
- Data itself can change, because the conditions under which the data are captured can change.

Big Data in Information Systems

Roles in Information Systems

- **Advanced Analytics:** Enables predictive analytics, machine learning, and real-time insights.
- **Customer Insights:** Analyzes customer behavior across multiple channels.
- **Operational Efficiency:** Optimizes business processes using large-scale data analysis.
- **Decision Support:** Assists in data-driven strategic decisions.

Tools & Technologies

- **Storage/Processing:** Hadoop, Apache Spark, HDFS
- **Data Management:** Hive, Pig, HBase
- **Streaming Data:** Kafka, Flink
- **Cloud-Based Platforms:** AWS (Redshift, EMR), Azure (Data Lake), Google BigQuery

Integration in Information Systems

Modern **Information Systems** integrate both traditional databases and big data platforms to:

- Handle structured (databases) and unstructured data (big data).
- Enable **real-time decision-making** through **data pipelines and dashboards**.
- Support business intelligence, AI, and machine learning applications.

Role and Use of Data Warehousing and Data Marts in Information Systems

Introduction

- Data warehousing: Centralized repository for integrated data.
- Data marts: Smaller, department-specific repositories.
- Both essential for business intelligence and decision-making.

Example: Retail chain integrating sales data from all branches.

Role of Data Warehousing

- Integrates data from multiple sources.
- Stores historical data for trend analysis.
- Improves data consistency and quality.
- Supports advanced analytics.

Case: Amazon's data warehouse enables real-time recommendation engines using years of customer data.

Role of Data Marts

- Focused on specific business areas.
- Faster access for targeted analysis.
- Simplifies queries for non-technical users.
- More cost-effective for departmental needs.

Example: A bank's marketing department uses a data mart for campaign targeting based on customer demographics.

Benefits in Information Systems

- Better, data-driven decision-making.
- Enables predictive analytics and forecasting.
- Improves operational efficiency.
- Helps meet compliance requirements.

Case: Healthcare providers use data warehouses to track patient outcomes and meet reporting standards.

Case Study 1: Walmart – Retail Inventory Optimization

Background:

Walmart operates thousands of stores worldwide and sells millions of products daily. Managing inventory levels across locations is complex.

Problem:

1. Stockouts and overstock situations were costing Walmart millions.
2. Data was scattered across regional systems, making it hard to see the full picture in real time.

Solution (Data Warehouse Role):

1. Built a massive enterprise data warehouse that integrates sales, supplier, and logistics data from all stores.
2. Uses real-time analytics and predictive algorithms to forecast demand.
3. Data marts were created for individual departments (e.g., apparel, electronics) for more focused analysis.

Results:

1. Reduced stockouts by **30%**.
2. Optimized inventory turnover, saving **$1 billion annually**.
3. Faster decision-making: managers can see demand trends in near real-time.

Case Study 2: Bank of America – Marketing Campaign Optimization

Background:

Bank of America serves millions of customers with diverse products — loans, credit cards, savings accounts, and investments.

Problem:

1. Customer data was spread across different systems (credit card division, mortgage division, etc.).
2. Marketing teams couldn't personalize offers effectively.

Solution (Data Mart Role):

1. Centralized customer profiles in a data warehouse.
2. Created a marketing data mart containing only relevant fields: demographics, transaction history, credit score, and product usage.
3. Used this mart to run targeted campaigns with predictive analytics.

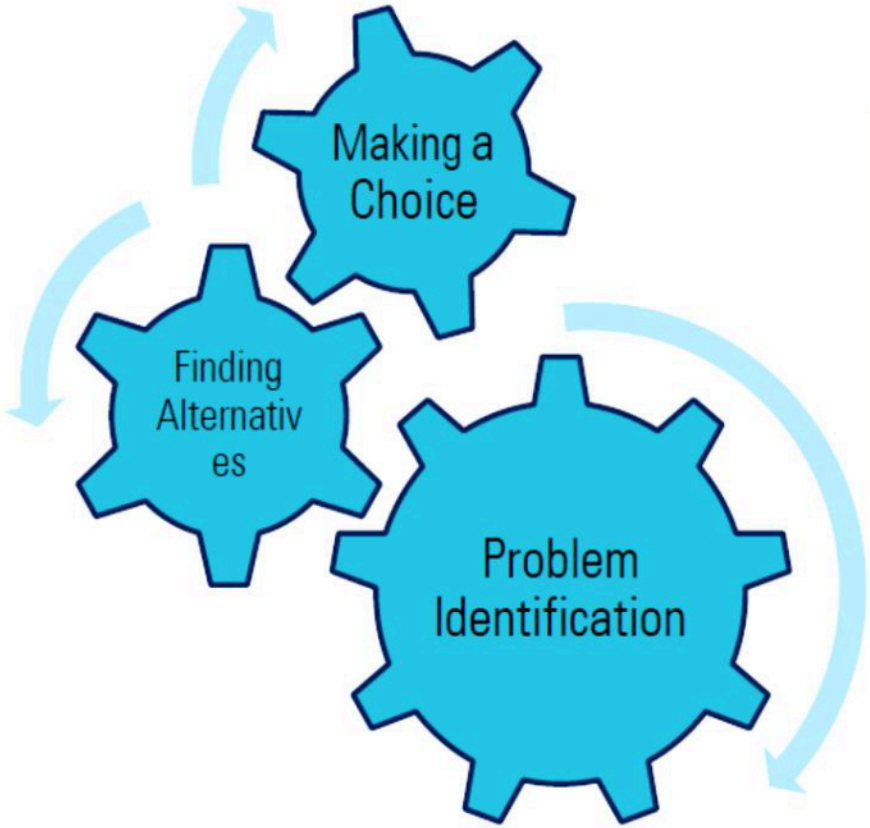
Results:

1. Response rates to campaigns increased by **25%**.
2. Reduced marketing costs by focusing on high-probability customers.
3. Strengthened customer loyalty through personalized offers.

CHAPTER

Business Intelligence

Decision Making Basics



"Information and knowledge form the backbone of the decision making process"

Who Makes Decisions?

Decision making at different levels:

– Operational

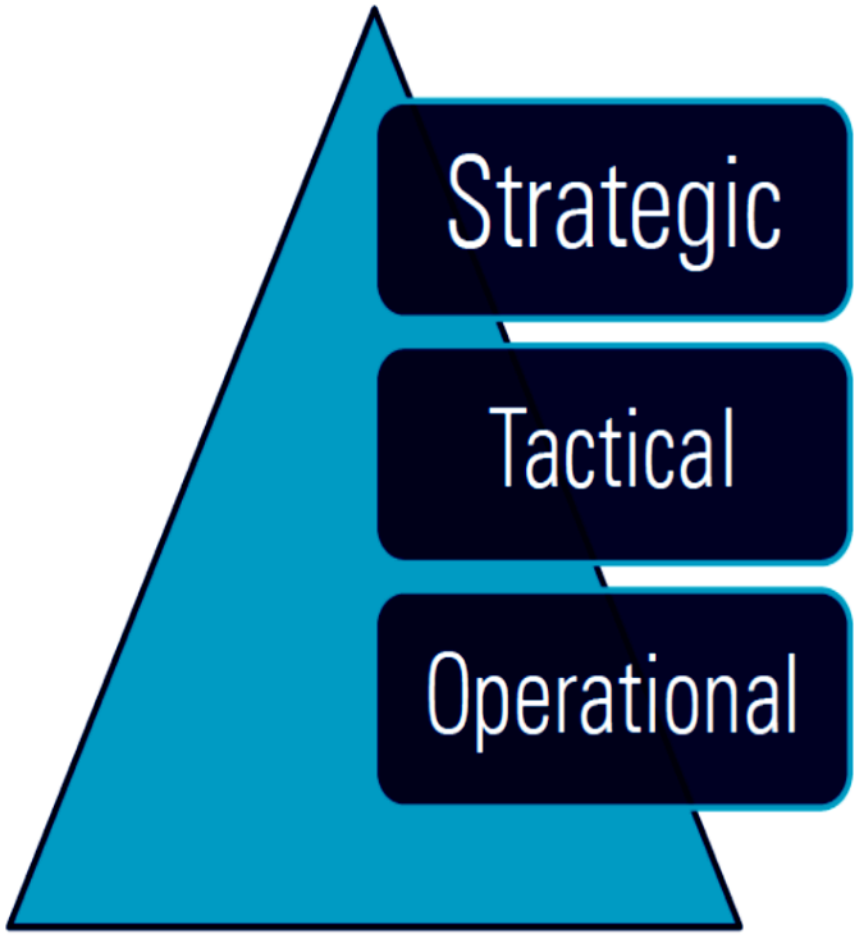
- Related to daily activities with short-term effect
- Structured decisions taken by lower management

– Tactical

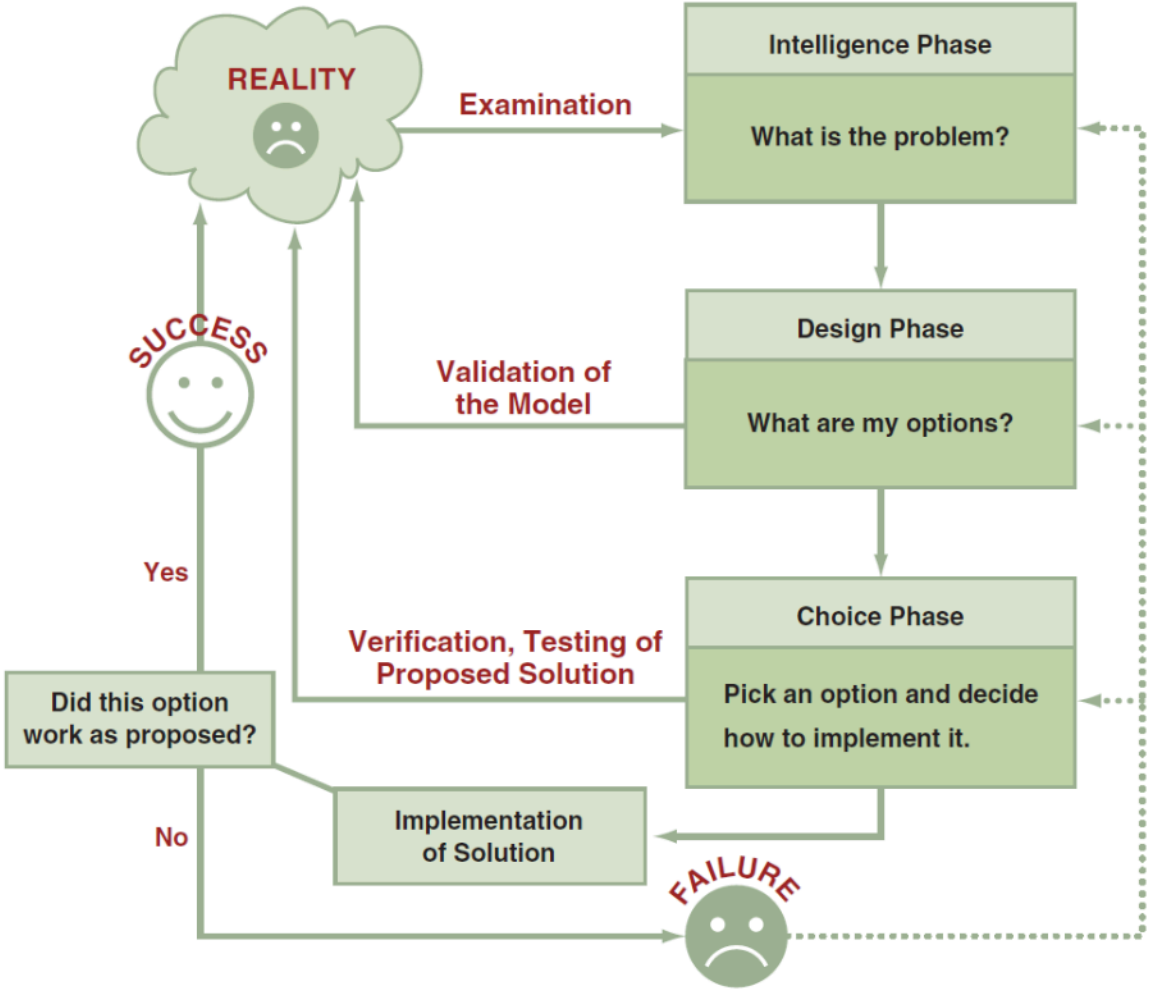
- Semi-structured decisions taken by middle management

– Strategic

- Long-term effect
- Unstructured decisions taken by top management



The Process and Phases in Decision Making



The Process and Phases in Decision Making

Decision: a choice among two or more alternatives that individuals and groups make. Decisions are diverse and are made continuously.

Phases of the Decision Making Process:

- **Intelligence Phase:** managers examine a situation and then identify and define the problem or opportunity.

- **Design Phase:** decision makers construct a model for addressing the situation. They perform this task by making assumptions that simplify reality and by expressing the relationships among all of the relevant variables. Managers then validate the model by using test data. Finally, decision makers set criteria for evaluating all of the potential solutions that are proposed.

- **Choice Phase:** involves selecting a solution or course of action that seems best suited to resolve the problem. This solution (the decision) is then implemented.

- **Implementation Phase:** is successful if the proposed solution solves the problem or seizes the opportunity. If the solution fails, then the process returns to the previous phases. Computer-based decision support assists managers in the decision-making process.

Why Managers Need IT Support?

- The number of alternatives is constantly increasing
- Most decisions are made under time constraints
- Uncertainty in the decision environment Group decision making required

What IT are Available to Support Managers

- **Business Intelligence (BI):** is a broad category of applications, technologies, and processes for gathering, storing, accessing, and analyzing data to help business users make better decisions.
- BI applications enable decision makers to quickly ascertain the status of a business enterprise by examining key information.

What is Business Intelligence

Technology that Allows:

- Gathering, storing, accessing & analyzing data to help business users make better decisions

Set of Applications that Allow:

- Decision support systems
- Query and reporting
- online analytical processing (OLAP)
- Statistical analysis, forecasting, and data mining

Help in analyzing business performance through data-driven insight:

- Understand the past & predict the future

Case Study : Netflix – Content Recommendation

Background:

Netflix streams thousands of shows and movies to over 200 million subscribers.

Problem:

- Needed to keep users engaged to reduce churn.
- Enormous content library made it hard for users to find relevant content.

BI Solution in Information Systems:

- Built a BI system that analyzes viewing history, ratings, and user behavior in real-time.
- Implemented machine learning algorithms to recommend personalized content.
- Used A/B testing dashboards to measure recommendation effectiveness.

Results:

- Increased viewing time by **80%** in some user segments.
- Reduced churn by **9%** year-over-year.

A Framework for Computerized Decision Analysis

Problem Structure: where decision-making processes fall along a range from highly structured to highly unstructured.

Nature of Decisions:

- All managerial decisions fall into one of three broad categories:
  - **Operational Control:** executing specific tasks efficiently and effectively.
  - **Management Control:** acquiring and using resources efficiently in accomplishing organizational goals.
  - **Strategic Planning:** the long-range goals and policies for growth and resource allocation.

Apollo Hospitals (Healthcare)

- **Scenario**: A 350-bed Apollo hospital faced long emergency room wait times (>90 min), leading to patient dissatisfaction and staff burnout.
  BI Solution: Analyzed two years of ER data to forecast patient inflow and deployed a real-time BI dashboard (e.g., Power BI) to monitor patient flow, bed availability, and queue lengths.

Outcomes:

- ER wait times dropped by **40%**
- Patient satisfaction improved by **18%**
- Staff efficiency increased by **22%**

Case Study : Coca-Cola – Customer Retention

Background:

Coca-Cola serves millions of customers through its distribution networks.

Problem:

- Declining customer loyalty in certain regions.
- Lack of unified visibility into customer purchasing patterns.

BI Solution in Information Systems:

- Integrated sales and marketing data into a **BI dashboard**.
- Used predictive analytics to identify customers at risk of leaving.
- Implemented personalized promotions for at-risk customers.

Results:

- Customer retention improved by **15%**.
- Revenue increased in targeted regions by **$10 million** in the first year.

Decision Support Framework

	Operational Control	Management Control	Strategic Planning	IS Support
Structured	Accounts receivable, order entry 1	Budget analysis, short-term forecasting, personnel reports, make-or-buy analysis 2	3	MIS, statistical models (management science, financial, etc.)
Semistructured	Production scheduling, inventory control 4	Credit evaluation, budget preparation, plant layout, project scheduling, reward systems design 5	Building a new plant, mergers and acquisitions, planning (product, quality assurance, compensation, etc.) 6	Decision support systems, business intelligence
Unstructured	7	Negotiating, recruiting an executive, buying hardware, lobbying 8	New technology development, product R&D, social responsibility planning 9	Decision support systems, expert systems, enterprise resource planning, neural networks, business intelligence, big data

The Scope of Business Intelligence

- The Development of One or a Few Related BI Applications
- The Development of Infrastructure to Support Enterprise wide BI
- Support for the Organizational Transformation

Business Intelligence Applications for Data Analysis

- Multidimensional Analysis or Online Analytical Processing
- Data Mining
- Decision Support Systems

Multidimensional Analysis or Online Analytical Processing

- **Online Analytical Processing** (OLAP - also referred to as **multidimensional analysis)** capabilities. OLAP involves "slicing and dicing" data stored in a dimensional format, drilling down in the data to greater detail, and aggregating the data.

Data Mining

- **Data Mining:** the process of searching for valuable business information in a large database, data warehouse, or data mart.
- **Data Mining Can Perform Two Basic Operations:**
- (1) predicting trends and behaviors
- (2) identifying previously unknown patterns.

Data Mining

The types of information obtainable from data mining include:

- Associations
- Sequences
- Classifications
- Clusters
- Forecasts

Data Mining

The types of information obtainable from data mining include:

- Associations
- Sequences
- Classifications
- Clusters
- Forecasts

Data Mining

- *Associations* are occurrences linked to a single event.
- In *sequences*, events are linked over time.
- *Classification* recognizes patterns that describe the group to which an item belongs by examining existing items that have been classified and by inferring a set of rules.
- *Clustering* works in a manner similar to classification when no groups have yet been defined.
- Although these applications involve predictions, *forecasting* uses predictions in a different way.
- It uses a series of existing values to forecast what other values will be.

Data Mining

- *Associations* are occurrences linked to a single event.
- For instance, a study of supermarket purchasing patterns might reveal that, when corn chips are purchased, a cola drink is purchased 65 percent of the time, but when there is a promotion, cola is purchased 85 percent of the time.
- This information helps managers make better decisions because they have learned the profitability of a promotion.

- In *sequences*, events are linked over time.
- We might find, for example, that if a house is purchased, a new refrigerator will be purchased within two weeks 65 percent of the time, and an oven will be bought within one month of the home purchase 45 percent of the time.

- *Classification* recognizes patterns that describe the group to which an item belongs by examining existing items that have been classified and by inferring a set of rules.
- For example, businesses such as credit card or telephone companies worry about the loss of steady customers.
- Classification helps discover the characteristics of customers who are likely to leave and can provide a model to help managers predict who those customers are so that the managers can plan special campaigns to retain such customers.

- *Clustering* works in a manner similar to classification when no groups have yet been defined.
- A data mining tool can discover different groupings within data, such as finding affinity groups for bank cards or partitioning a database into groups of customers based on demographics and types of personal investments.

- Although these applications involve predictions, *forecasting* uses predictions in a different way.
- It uses a series of existing values to forecast what other values will be.
- For example, forecasting might find patterns in data to help managers estimate the future value of continuous variables, such as sales figures.

Decision Support Systems (DSS)

- Decision support systems (DSS) are interactive software-based systems that help managers in decision-making by accessing large volumes of information generated from various related information systems
- DSS uses the summary information, exceptions, patterns, and trends using the analytical models.
- A decision support system helps in decision-making but does not necessarily give a decision itself.
- The decision makers compile useful information from raw data, documents, personal knowledge, and/or business models to identify and solve problems and make decisions.

DSS Capabilities

- **Sensitivity Analysis:** Sensitivity analysis is the study of the impact that changes in one or more parts of a decision-making model have on other parts.
- **What–If Analysis:** This analysis attempts to predict impact of assumptions (input data) on the proposed solution. The results depend on the accuracy of these assumptions, which can be highly subjective.
- **Goal-Seeking Analysis:** represents a "backward" solution approach. It attempts to calculate the value of the inputs necessary to achieve a desired level of output.

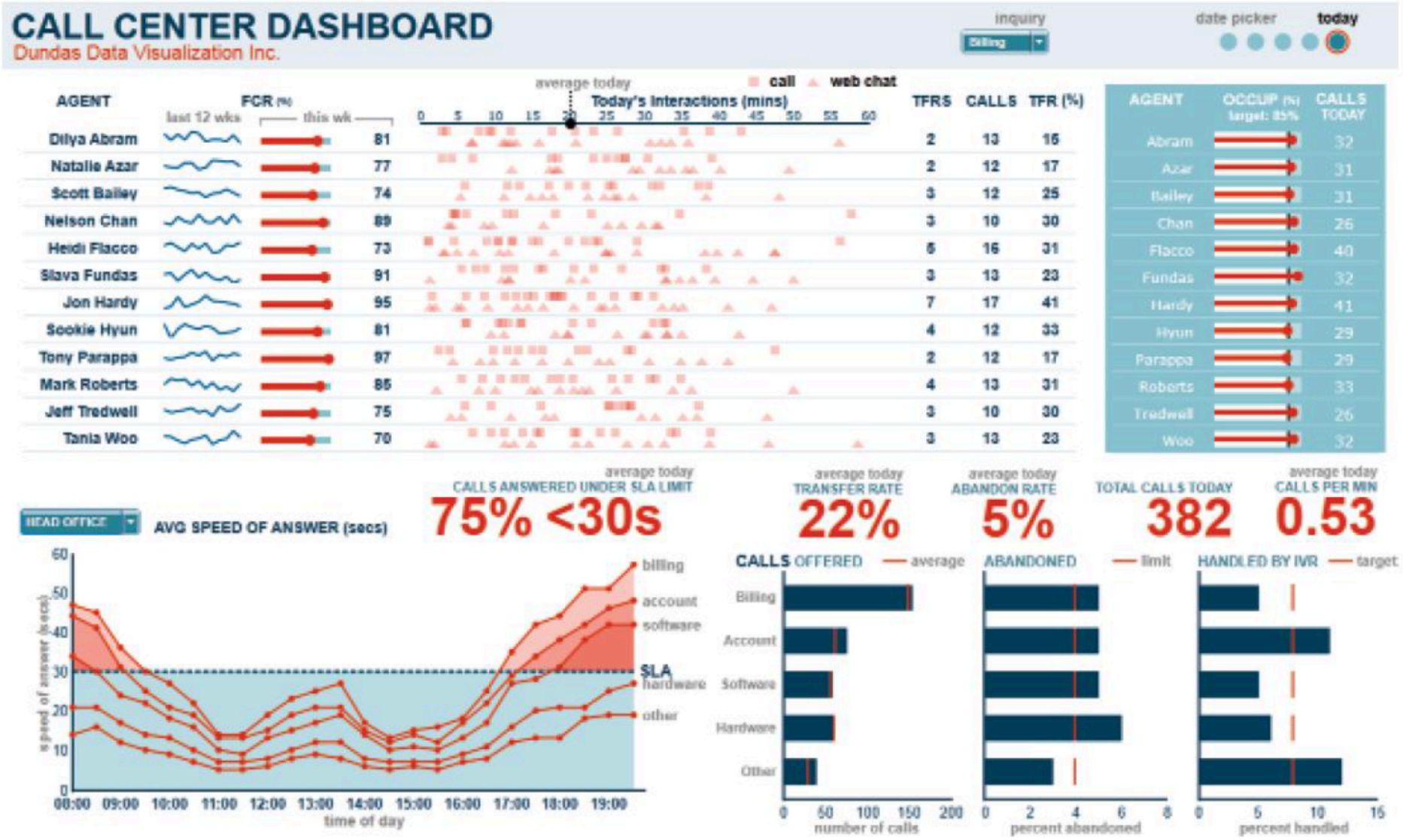
Business Intelligence Applications for Presenting Results

- **Dashboard:** provides easy access to timely information and direct access to management reports. They evolved from executive information systems, which were information systems designed specifically for the information needs of top executives
- **Data Visualization:** data presented to users in visual formats such as text, graphics, and tables following data processing. Data Visualization makes IT applications more attractive and understandable to users.
- **Real-Time Business Intelligence:** includes the use of real time data for analysis as it is created rather than using historical data for analysis.

The Capabilities of Dashboards

Capability	Description
Drill down	The ability to go to details, at several levels; it can be done by a series of menus or by clicking on a drillable portion of the screen.
Critical success factors (CSFs)	The factors most critical for the success of business. These can be organizational, industry, departmental, or for individual workers.
Key performance indicators	The specific measures of CSFs.
Status access	The latest data available on KPI or some other metric, often in real time.
Trend analysis	Short-, medium-, and long-term trend of KPIs or metrics, which are projected using forecasting methods.
Exception reporting	Reports that highlight deviations larger than certain thresholds. Reports may include only deviations.

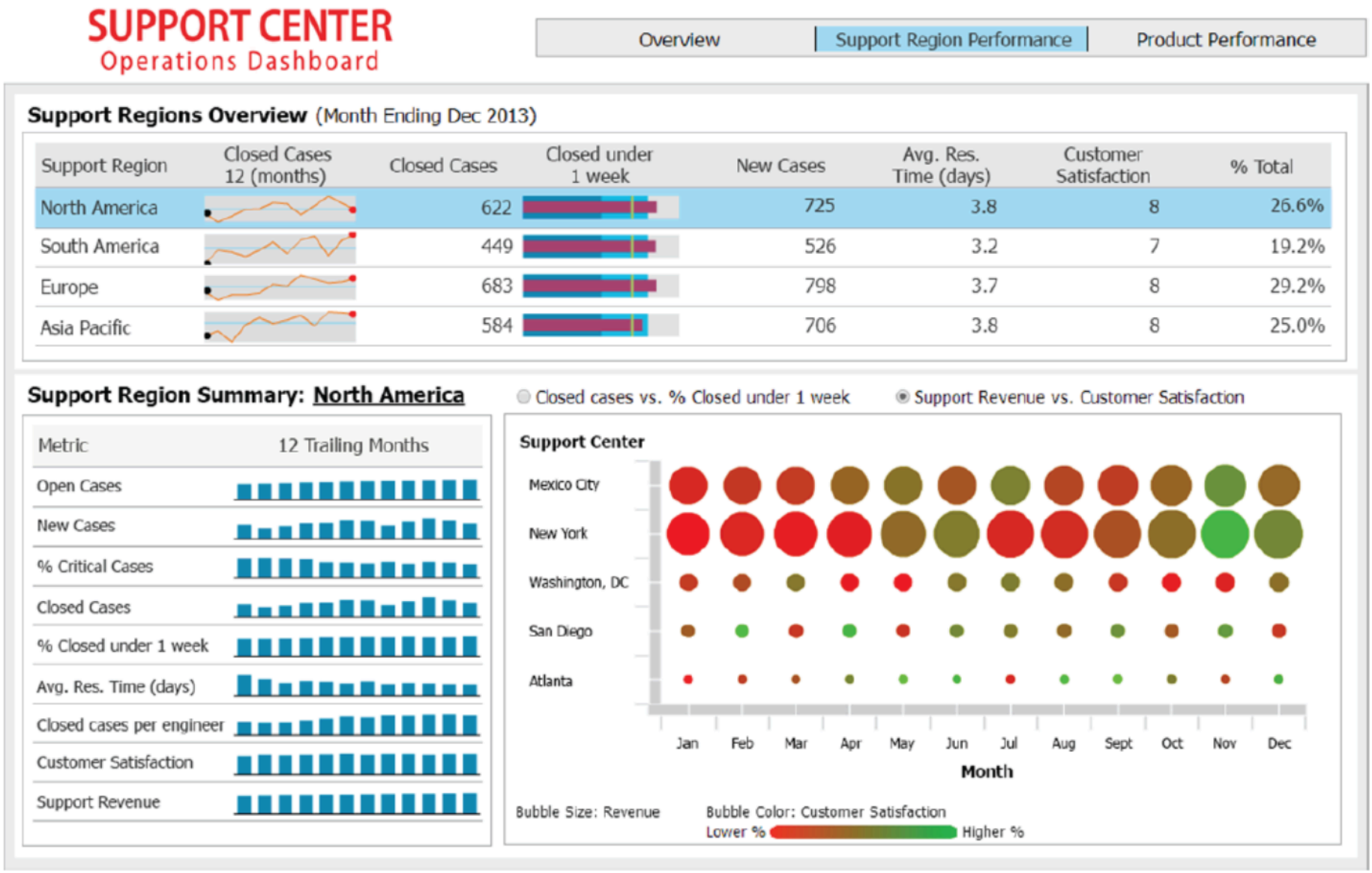
Sample Performance Dashboard



Data Visualization Technologies

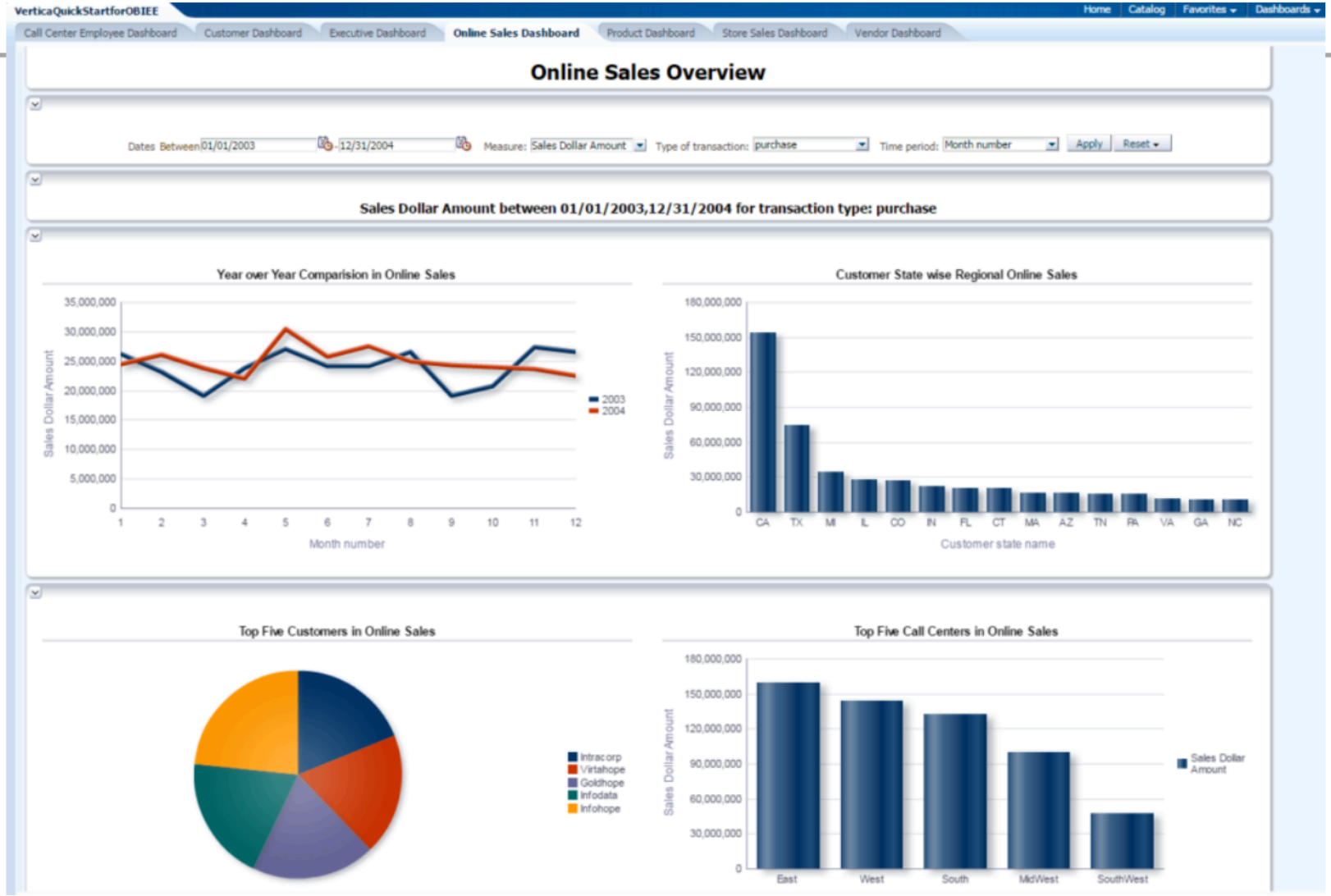
- **Geographic Information System (GIS):** a computer-based system for capturing, integrating, manipulating, and displaying data using digitized maps. Its most distinguishing characteristic is that every record or digital object has an identified geographical location.
- **Reality Mining:** Graphical Information Systems (GIS) and Geographic Positioning Systems (GPS) together to produce an interesting new type of technology which allows analysts to extract information from the usage patterns of mobile phones and other wireless devices.

Support Center Operations Dashboard

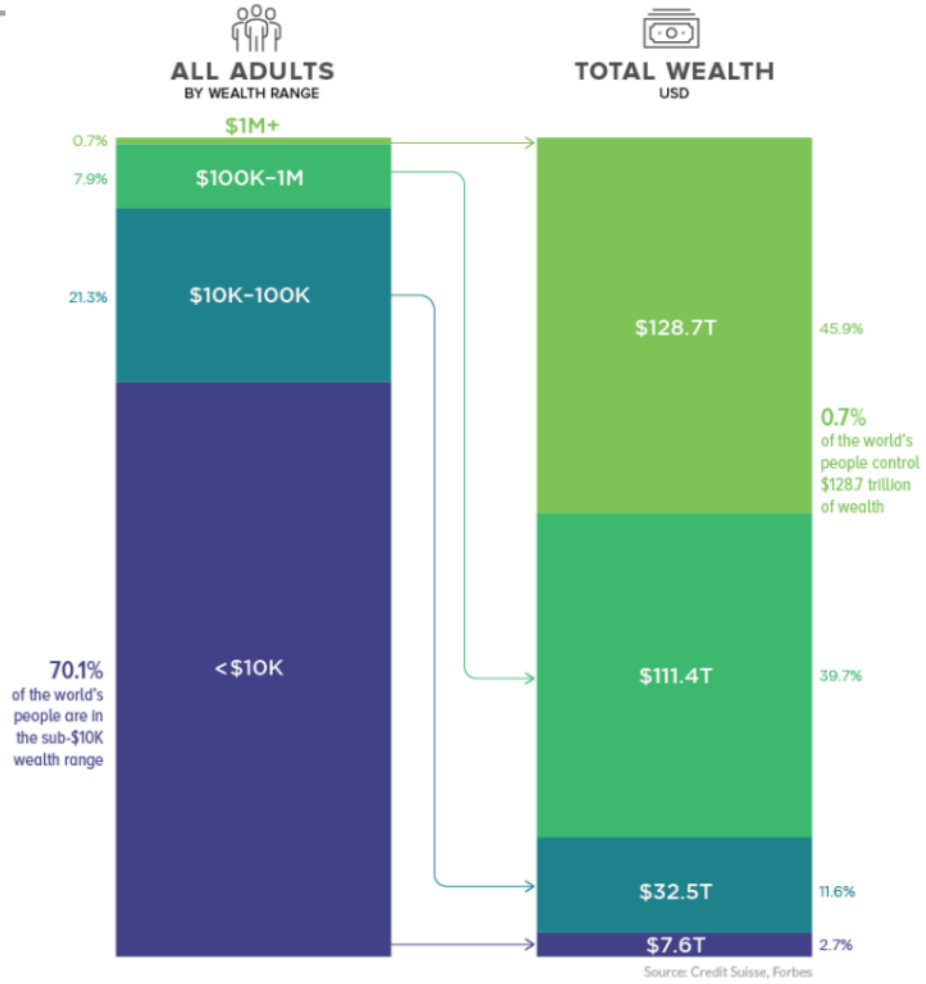


Applications of BI

- Sales Intelligence



Visualization



Report generation