Lecture 01

Data

Data refers to facts, figures, or information collected for reference, analysis, or processing. It is the foundation of decision-making and computing.

Data represents raw elements or unprocessed facts, including numbers and symbols to text and images. When collected and observed without interpretation, these elements remain just data—simple and unorganized. When these pieces are analyzed and contextualized, they transform into something more meaningful.

What is Information?

we get information when data is processed, organized, interpreted, and structured. The comprehensible output derived from raw data helps inform decisions, strategies, and actions. Information is essentially data made valuable and accessible—an integral component of decision-making.

For instance, if data points include daily temperature readings over a year, information is recognizing the trend of temperatures, understanding seasonal changes, and predicting future weather conditions.

Difference Between Data and Information

Data is raw and often chaotic, lacking meaningful structure or context. On the other hand, information is the refined, analyzed, and structured output derived from this data, tailored to provide actionable insights and facilitate strategic decision-making.

The journey from data to information involves several key distinctions:

- **Data** is raw and unstructured, like individual customer interactions or transaction logs.
- **Information** provides context and insights, like a trend analysis that shows increasing customer satisfaction or sales figures over time.
- **Data** is often abundant and readily available but can be overwhelming without interpretation.
- **Information** is curated and actionable, offering strategic insights to guide business decisions.

Types of Data

Data can be classified broadly into **two categories**:

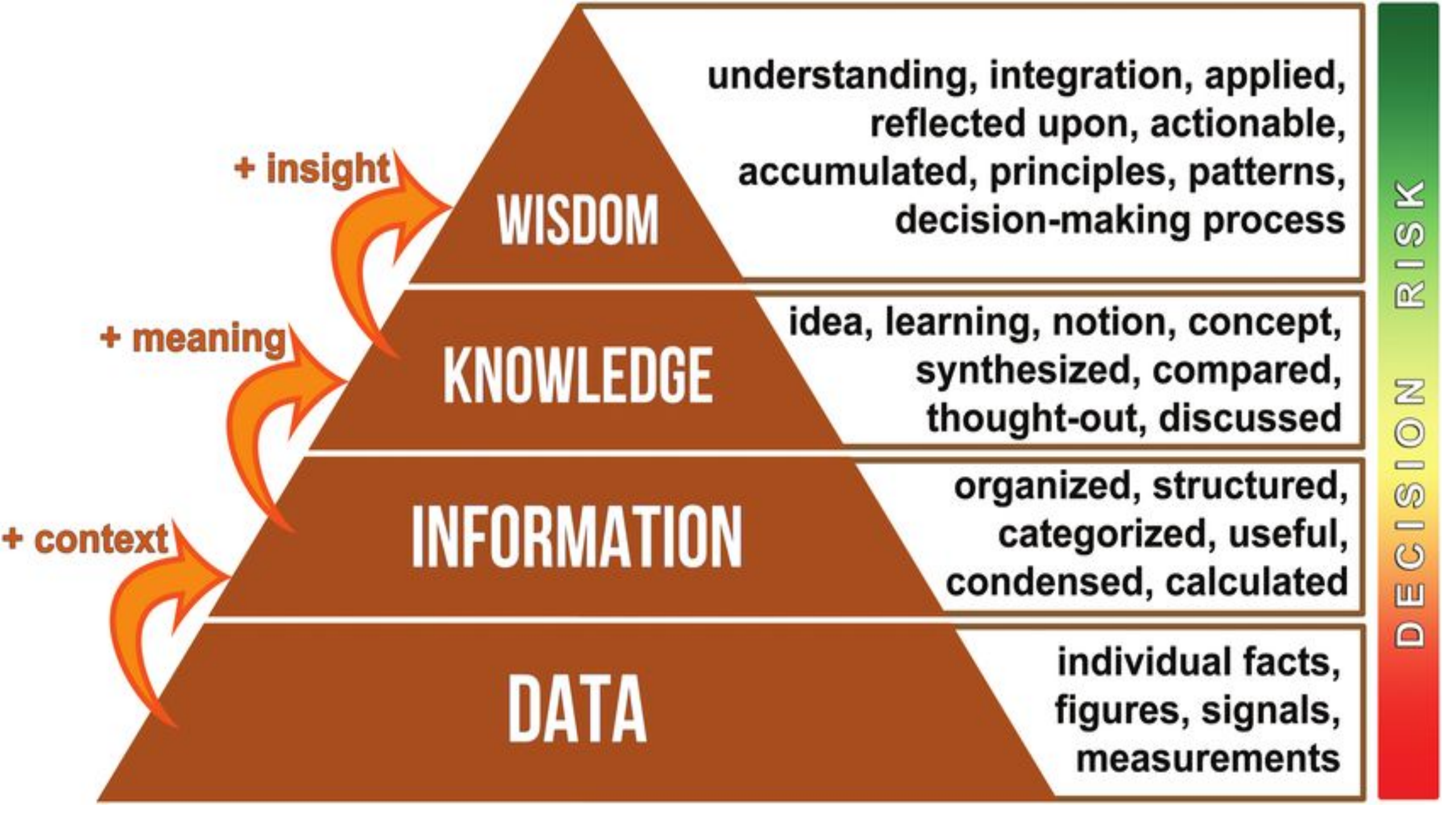
- **1. Quantitative Data (Numerical Data)**
- Quantitative data represents measurable quantities and is expressed in numbers.
- **Types of Quantitative Data**:
  - **Discrete Data**: Countable values (e.g., number of students, age in years).
  - **Continuous Data**: Measurable values that can take any value within a range (e.g., height, weight, temperature).

- **2. Qualitative Data (Categorical Data)**
- Qualitative data represents categories or labels used to classify objects or information.
- **Types of Qualitative Data**:
  - **Nominal Data**: Categories without a specific order (e.g., gender, colors, countries).
  - **Ordinal Data**: Categories with a specific order or ranking (e.g., grades: A, B, C; satisfaction levels: low, medium, high).

Data Types in Programming

- In programming, data types define the kind of data a variable can store. Common data types include:
- **1. Primitive Data Types**
- **Integer (int)**: Whole numbers (e.g., 1, 42, -7).
- **Floating-point (float)**: Decimal numbers (e.g., 3.14, -0.001).
- **Character (char)**: Single characters (e.g., 'a', 'z').
- **String (str)**: Sequence of characters (e.g., "hello", "world").
- **Boolean (bool)**: Logical values (True or False).

- **Non-Primitive Data Types**
- **List/Array**: Collection of items (e.g., [1, 2, 3], ["apple", "banana"]).
- **Dictionary (dict)**: Key-value pairs (e.g., {"name": "John", "age": 30}).
- **Tuple**: Immutable collection of items (e.g., (1, 2, 3)).
- **Set**: Collection of unique items (e.g., {1, 2, 3, 4}).


+ insight
+ meaning
+ context
WISDOM
KNOWLEDGE
INFORMATION
DATA
understanding, integration, applied, reflected upon, actionable, accumulated, principles, patterns, decision-making process
idea, learning, notion, concept, synthesized, compared, thought-out, discussed
organized, structured, categorized, useful, condensed, calculated
individual facts, figures, signals, measurements
DECISION RISK

Data Sources

1. ***Relational Databases***

Examples: MySQL, PostgreSQL, Oracle Database, Microsoft SQL Server.

Features: Structured data stored in rows and columns; uses SQL for querying.

Usage: Customer data, financial records, inventory management.

2. ***NoSQL Databases***

Examples: MongoDB, Cassandra, Couchdb.

Features: Designed for unstructured or semi-structured data; includes document-based, key-value, wide-column, and graph databases.

Usage: Real-time analytics, IoT applications, social networks.

3. ***Flat Files***

Examples: CSV, Excel, JSON, XML.

Features: Simple text or tabular files; used for data exchange or lightweight storage.

Usage: Data transfer, small-scale data storage.

4. ***APIs (Application Programming Interfaces)***

Examples: REST APIs, GraphQL APIs.

Features: Enables access to live data from external sources over the web.

Usage: Weather data, financial markets, e-commerce integrations.

5. ***Data Warehouses***

Examples: Amazon Redshift, Google BigQuery, Snowflake.

Features: Centralized repository for structured data; optimized for analytics and reporting.

Usage: Business intelligence, data-driven decision-making.

6. ***Data Lakes***

Examples: Amazon S3, Azure Data Lake, Hadoop Distributed File System (HDFS).

Features: Stores large volumes of raw, unprocessed data; supports structured, semi-structured, and unstructured formats.

Usage: Big data analysis, machine learning, advanced analytics.

7. ***Streaming Data Sources***

Examples: Apache Kafka, Apache Flink, Amazon Kinesis.

features: Real-time data streams from devices or applications.

Usage: IoT telemetry, stock market monitoring, live event tracking.

8. ***Cloud Storage***

Examples: Google Drive, Dropbox, OneDrive, AWS S3.

Features: Scalable, secure, and accessible online storage.

Usage: Document management, file sharing, archival storage.

9. ***Web Scraping***

Tools: BeautifulSoup, Scrapy, Selenium.

Features: Extracts data directly from websites.

Usage: Market research, price comparison, sentiment analysis.

10**. *Social Media Platforms***

- **Examples**: Twitter, Facebook, Instagram, LinkedIn.
- **Features**: User-generated content in text, image, and video formats.
- **Usage**: Sentiment analysis, brand monitoring, trend analysis.

11. ***Logs and Machine-Generated Data***

- **Examples**: Application logs, server logs, network logs.
- **Features**: Sequential records of events or activities.
- **Usage**: Troubleshooting, security analysis, system monitoring.

14. ***Manual Data Collection***

- **Examples**: Surveys, forms, interviews.
- **Features**: Human-entered data; often qualitative or structured.
- **Usage**: User feedback, academic research, customer satisfaction analysis.

Business Intelligence

BI is a set of tools and techniques that enable organizations to collect, analyze, and report on data to make informed business decisions.

Business intelligence (**BI**) consists of strategies, methodologies, and technologies used by enterprises to analyse data and manage business information.

Common functions of BI technologies include reporting, online analytical processing, analytics, dashboard development, data mining, process mining, complex event processing, business performance management, benchmarking, text mining, predictive analytics, and prescriptive analytics.

Key Components of Business Intelligence:

1. Data Collection:
Gathering data from various sources such as databases, APIs, ERP systems, and external platforms.

2. Data Integration and Transformation:
Combining data from different sources, cleaning it, and transforming it into a unified format for analysis.

3. Data Warehousing:
Storing structured and organized data in centralized repositories, such as a data warehouse, for easy access and analysis.

4. Data Analysis and Reporting:
Using analytical tools to explore data and generate insights. Reports and dashboards are created to visualize metrics, trends, and performance.

5. Visualization Tools:
Tools like Power BI, Tableau, QlikView, or Looker provide interactive dashboards and charts for presenting insights in a user-friendly manner.

6. Advanced Analytics:
Incorporating predictive analytics, machine learning, and statistical modeling to uncover patterns and make forecasts.

7.Decision Support:
Enabling data-driven decision-making across all organizational levels by delivering actionable insights.

Benefits of BI

Improved Decision-Making: Provides accurate and real-time data to support strategic and operational decisions.

Increased Efficiency: Automates data collection, reporting, and analytics, saving time and resources.

Enhanced Customer Insights: Helps in understanding customer behaviors, preferences, and trends.

Competitive Advantage: identify market opportunities and optimize operations.

Performance Monitoring: Tracks key performance indicators (KPIs) to evaluate progress toward business goals.

Tools and Platforms:

BI Tools:

- Microsoft Power BI
- Tableau
- QlikView
- Looker
- Sisense

Data Integration Tools:

- Talend
- Apache NiFi
- Informatica
- Microsoft SSIS

Cloud BI Solutions:

- AWS QuickSight
- Google Data Studio
- Snowflake (for data warehousing

Decisionsupport system
Changing Business Environment

Companies are moving aggressively to computerized support of their operations => Business Intelligence

Business Pressures–Responses–Support Model

- Business pressures result of today's competitive business.
- climate Responses to counter the pressures.
- Support to better facilitate the process.

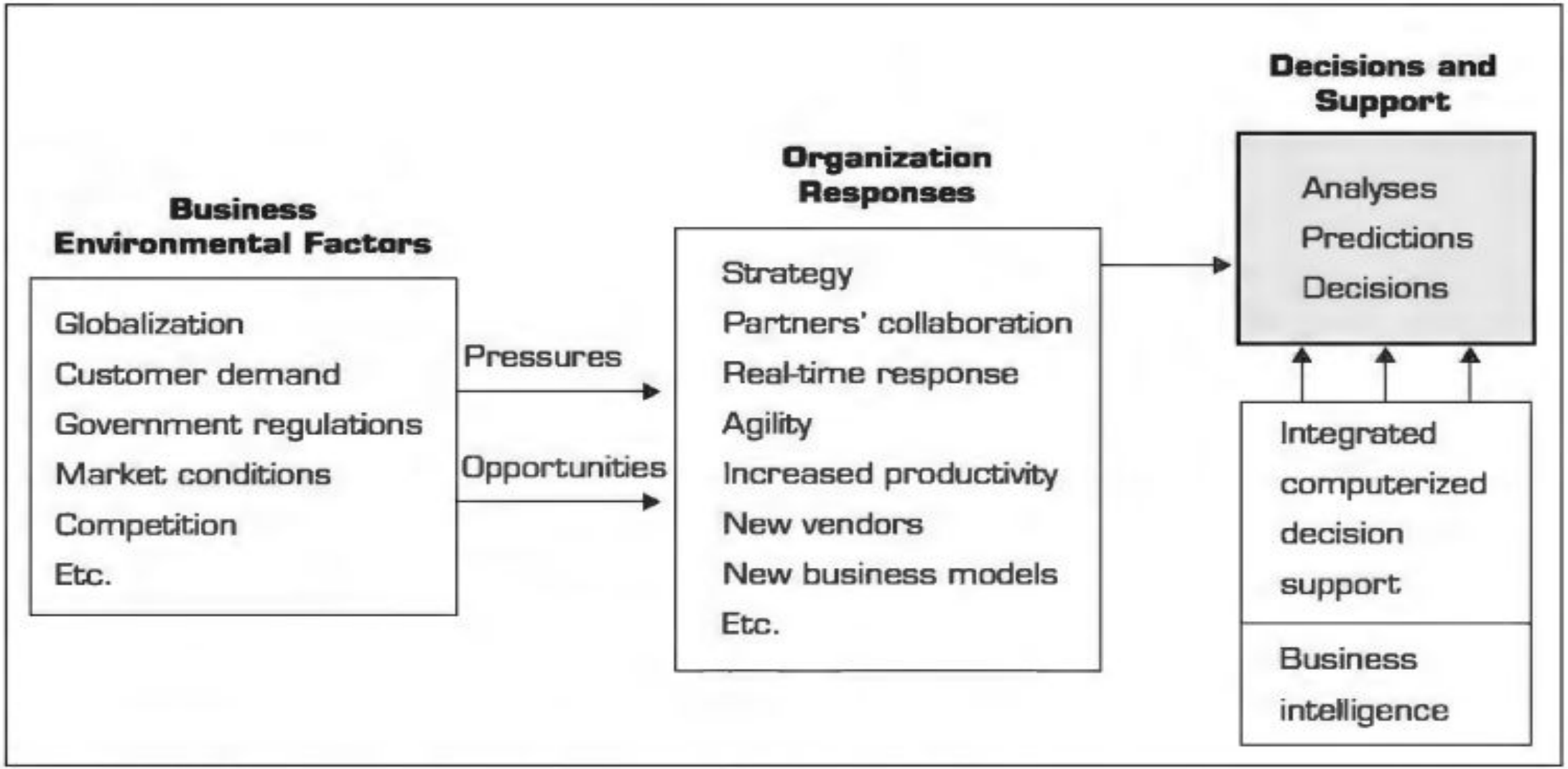


FIGURE 1.1 The Business Pressures–Responses–Support Model.

The Business Environment

The environment in which organizations operate today is becoming more and more complex, creating: opportunities, and problems

Example: globalization

Business environment factors:

markets, consumer demands, technology, and societal…

Business Environment Factors

Markets Strong competition

Expanding global markets

Blooming electronic markets on the Internet

Innovative marketing methods

Need for real-time, on-demand transactions

Consumer Desire for customization

demand Desire for quality, diversity of products, and speed of delivery

Technology More innovations,

new products, and new services

Increasing obsolescence

rate Increasing

information overload

Organizational Responses

Be Reactive, Anticipative, Adaptive, and Proactive

Managers may take actions, such as Employ strategic planning

Use new and innovative business models

Restructure business processes

Participate in business alliances

Improve corporate information systems

Improve partnership

Relationships

Encourage innovation and creativity

Managers actions

Improve customer service and relationships

Move to electronic commerce (e-commerce)

Move to make-to-order production and on-demand manufacturing and services.

Use new IT to improve communication, data access (discovery of information), and collaboration

Respond quickly to competitors' actions (e.g., in pricing, promotions, new products and services)

Automate many tasks of white-collar employees

Automate certain decision processes

Improve decision-making by employing analytics

Closing the Strategy Gap

One of the major objectives of computerized decision support is to facilitate closing the gap between the current performance of an organization and its desired performance, as expressed in its mission, objectives, and goals, and the strategy to achieve them.

Managerial Decision Making

Management is a process by which organizational goals are achieved by using resources

Inputs: resources

Output: attainment of goals

The measure of success: outputs/inputs

Management Decision Making

Decision making: selecting the best solution from two or more alternatives

Decision Making Process

Managers usually make decisions by following a four-step process (a.k.a. the scientific approach)

Define the problem (or opportunity)

Construct a model that describes the real-world problem

Identify possible solutions to the modeled problem and evaluate the solutions

Compare, choose, and recommend a potential solution to the problem

Decision making is difficult, because

Technology, information systems, advanced search engines, and globalization result in more and more alternatives from which to choose

Government regulations and the need for compliance, political instability and terrorism, competition, and changing consumer demands produce more uncertainty, making it more difficult to predict consequences and the future

Other factors are the need to make rapid decisions, the frequent and unpredictable changes that make trial-and-error learning difficult, and the potential costs of making mistakes

Why Use Computerized DSS

Computerized DSS can facilitate decision via:

Speedy computations

Improved communication and collaboration

Increased productivity of group members

Improved data management

Overcoming cognitive limits

Quality support; agility support

Using Web; anywhere, anytime support

A Decision Support Framework (by Gory and Scott-Morten, 1971)

Type of Decision	Type of Control: Operational Control	Type of Control: Managerial Control	Type of Control: Strategic Planning
Structured	1 Accounts receivable Accounts payable Order entry	2 Budget analysis Short-term forecasting Personnel reports Make-or-buy	3 Financial management Investment portfolio Warehouse location Distribution systems
Semistructured	4 Production scheduling Inventory control	5 Credit evaluation Budget preparation Plant layout Project scheduling Reward system design Inventory categorization	6 Building a new plant Mergers & acquisitions New product planning Compensation planning Quality assurance HR policies Inventory planning
Unstructured	7 Buying software Approving loans Operating a help desk Selecting a cover for a magazine	8 Negotiating Recruiting an executive Buying hardware Lobbying	9 R & D planning New tech development Social responsibility planning