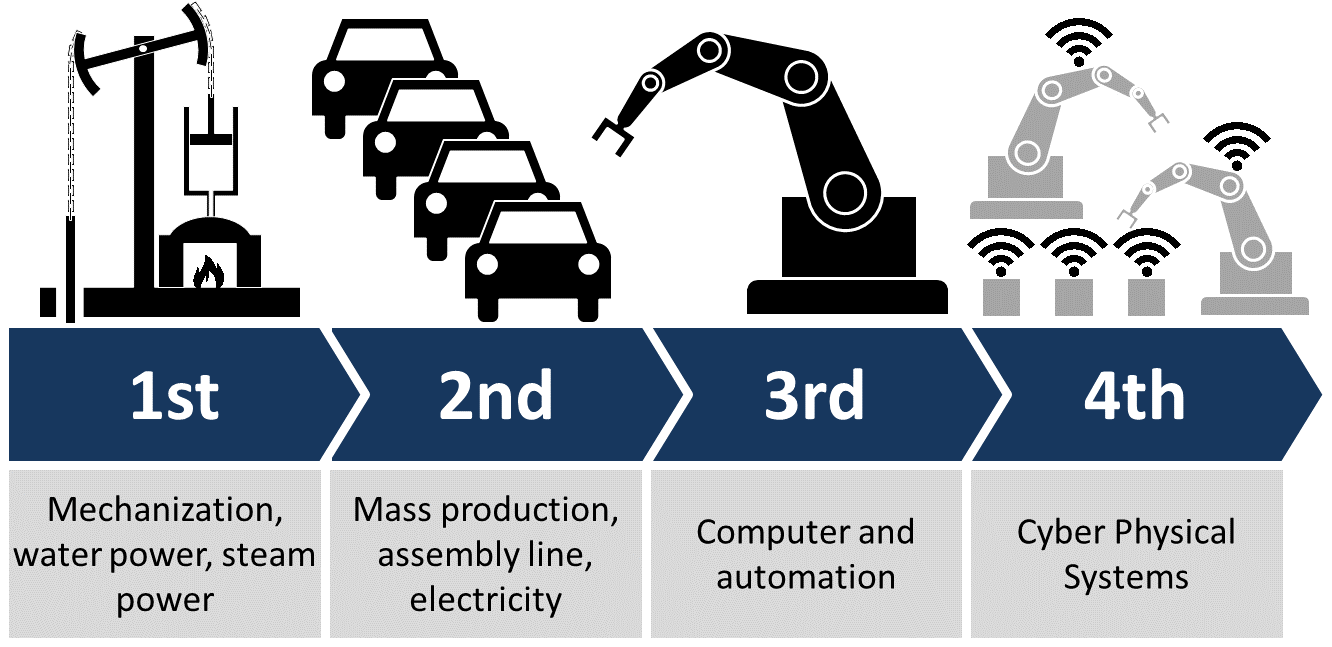
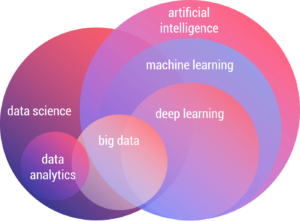
**DAY 1 - Introduction to Data Engineering & SQL REFRESHER**

**What is Data Engineering?**

Data Engineering ensures **businesses** can **efficiently collect**, **store** and **utilize data** to drive informed decision-making & innovation.

**🏭 Industrial Revolutions: 1.0 to 4.0**



* **Further Reading**: [Industry 1.0 Overview](https://www.coretigo.com/industrial-revolution-from-industry-1-0-to-industry-5-0/)
* **Further Reading**: [Industry 2.0 Insights](https://www.coretigo.com/industrial-revolution-from-industry-1-0-to-industry-5-0/)
* **Further Reading**: [Industry 3.0 Explained](https://www.coretigo.com/industrial-revolution-from-industry-1-0-to-industry-5-0/)
* **Further Reading**: [Industry 4.0 Details](https://www.coretigo.com/industrial-revolution-from-industry-1-0-to-industry-5-0/)

**🧪 Data Science -**

* **Definition**: It is a **core** branch of AI. An interdisciplinary field that uses scientific methods, processes, algorithms, and systems to extract knowledge and insights from data.

**🧠 Core Concepts of Data Science**

**1. Data Collection**

* **Definition**: The process of gathering raw data from various sources.
* **Sources**: APIs, web scraping, sensors, databases, surveys.
* **Example**: Scraping tweets from Twitter for sentiment analysis.
* 📚 [Web scraping with Python](https://realpython.com/python-web-scraping-practical-introduction/)

**2. Data Cleaning**

* **Definition**: Removing errors, handling missing values, and standardizing data.
* **Example**: Replacing null values with averages, fixing spelling in categorical variables.
* 🔧 Tools: Pandas, OpenRefine

**3. Exploratory Data Analysis (EDA)**

* **Definition**: Summarizing the main characteristics of data using visuals and statistics.
* **Example**: Using box plots and histograms to explore distribution.
* 🎥 [EDA with Python](https://www.youtube.com/watch?v=GcXcSZ0gQps)

**4. Feature Engineering**

* **Definition**: Creating new input features from existing data to improve model performance.
* **Example**: Extracting “day of week” from a timestamp for time series modeling.
* 🔍 [Kaggle Guide on Feature Engineering](https://www.kaggle.com/code/ryanholbrook/feature-engineering-for-machine-learning)

**5. Data Visualization**

* **Definition**: Using graphs and charts to represent data trends and insights.
* **Example**: Line charts for stock price trends, heatmaps for correlation.
* 📊 Tools: Power BI, Tableau, Plotly

**6. Machine Learning**

* **Definition**: Training algorithms to make predictions or decisions from data.
* **Types**:
  + **Supervised Learning**: Labeled data (e.g., predicting house prices)
  + **Unsupervised Learning**: No labels (e.g., customer segmentation)
  + **Reinforcement Learning**: Reward-based learning (e.g., robotics, game AI)
* 🎥 [Machine Learning](https://developers.google.com/machine-learning/crash-course)

**🤖 Automation**

* **Definition**: The use of technology to perform tasks without human intervention.
* **Real-World Example**: Automated guided vehicles (AGVs) in warehouses transporting goods.
* **Further Reading**: [Understanding Automation](https://www.isa.org/about-isa/what-is-automation)

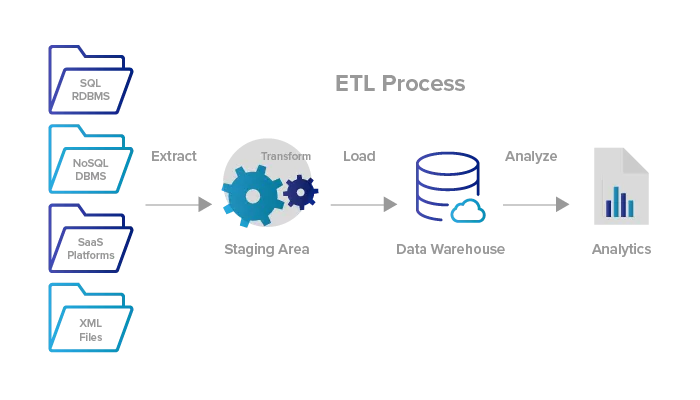
**🧠 Intelligence Age**

* **Definition**: A proposed era following the Information Age, characterized by the integration of artificial intelligence into all aspects of society.
* **Real-World Example**: AI-driven personal assistants managing daily tasks and schedules.
* **Further Reading**: [Exploring the Intelligence Age](https://www.forbes.com/sites/forbestechcouncil/2025/02/03/top-10-technology-trends-for-2025/)([Simplilearn.com](https://www.simplilearn.com/data-engineer-role-article?utm_source=chatgpt.com))

**🛠️ Data Engineer Responsibilities**

* **Definition**: Data engineers design, build, and maintain systems for collecting, storing, and analyzing data.
* **Key Responsibilities**:
  + Developing data pipelines
  + Ensuring data quality and integrity
  + Collaborating with data scientists and analysts
* **Real-World Example**: Building a data pipeline to process user activity logs for a web application.
* **Further Reading**: [Data Engineer Role](https://www.simplilearn.com/data-engineer-role-article)([IBM](https://www.ibm.com/think/topics/data-engineering?utm_source=chatgpt.com), [odga.virginia.gov](https://www.odga.virginia.gov/media/governorvirginiagov/chief-data-officer/css/Job-Description---Data-Engineer-Sample.pdf?utm_source=chatgpt.com))

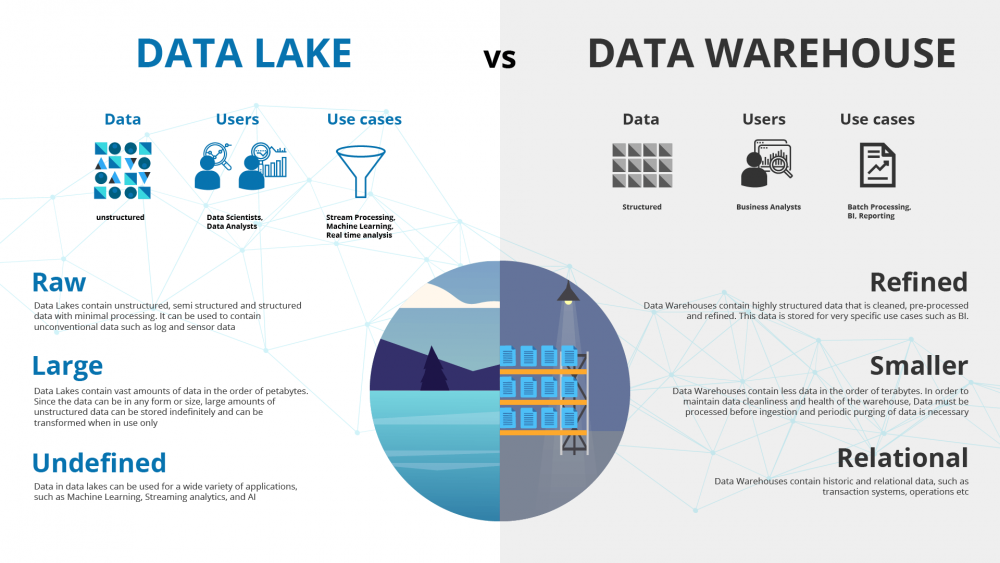
**🔄 ETL Process**

* **Definition**: ETL stands for Extract, Transform, Load – a process in data warehousing to extract data from sources, transform it into a suitable format, and load it into a destination system.
* **Real-World Example**: Extracting sales data from multiple stores, transforming it to a unified format, and loading it into a central database.
* **Further Reading**: [ETL Explained](https://aws.amazon.com/what-is/etl/)([Wikipedia](https://en.wikipedia.org/wiki/Extract%2C_transform%2C_load?utm_source=chatgpt.com))

**📊 Big Data**

* **Definition**: Large and complex data sets that traditional data processing tools cannot handle efficiently.
* **Characteristics**:
  + **Volume**: Massive amounts of data
  + **Velocity**: Rapid data generation and processing
  + **Variety**: Diverse data types and sources
* **Real-World Example**: Analyzing social media trends across millions of users in real-time.
* **Further Reading**: [What is Big Data?](https://www.oracle.com/big-data/what-is-big-data/)([Oracle](https://www.oracle.com/big-data/what-is-big-data/?utm_source=chatgpt.com))

**📥 Data Sources & Ingestion**

* **Definition**: Data ingestion is the process of collecting and importing data for immediate use or storage.
* **Common Data Sources**:
  + Databases
  + APIs
  + Sensor data
  + User-generated content
* **Real-World Example**: Ingesting real-time temperature data from IoT sensors in a smart building.
* **Further Reading**: [Data Ingestion Explained](https://www.geeksforgeeks.org/etl-process-in-data-warehouse/)([Wikipedia](https://en.wikipedia.org/wiki/Extract%2C_transform%2C_load?utm_source=chatgpt.com), [Tech-Labs](https://tech-labs.com/blog/evolution-industry-10-40-and-beyond?utm_source=chatgpt.com), [CENGN](https://www.cengn.ca/information-centre/innovation/industry-1-0-to-industry-4-0-and-whats-next/?utm_source=chatgpt.com))

**🗄️ Data Lake vs. Data Warehouse**

* **Data Lake**:
* **Definition**: A storage repository that holds vast amounts of raw data in its native format.
  + **Use Case**: Storing unstructured data like logs, images, and videos.
* **Data Warehouse**:
  + **Definition**: A system used for reporting and data analysis, storing structured data.
  + **Use Case**: Business intelligence and analytics.
* **Further Reading**: [Data Lake vs. Data Warehouse](https://www.oracle.com/big-data/what-is-big-data/)([Coursera](https://www.coursera.org/articles/what-does-a-data-engineer-do-and-how-do-i-become-one?utm_source=chatgpt.com), [Simplilearn.com](https://www.simplilearn.com/top-technology-trends-and-jobs-article?utm_source=chatgpt.com))

**📈 Business Intelligence (BI)**

* **Definition**: Technologies and strategies used by enterprises for data analysis and management of business information.
* **Real-World Example**: Using BI tools to visualize sales performance across different regions.
* **Popular BI Tools**:
  + Tableau
  + Power BI
  + Looker
* **Further Reading**: [Understanding Business Intelligence](https://www.oracle.com/business-analytics/business-intelligence/)([Encyclopedia Britannica](https://www.britannica.com/science/IQ?utm_source=chatgpt.com" \o "IQ | Intelligence Testing, Cognitive Ability & Mental Age - Britannica), [odga.virginia.gov](https://www.odga.virginia.gov/media/governorvirginiagov/chief-data-officer/css/Job-Description---Data-Engineer-Sample.pdf?utm_source=chatgpt.com))

**🌟 Trending Techn ologies Beyond Data Science (2025)**

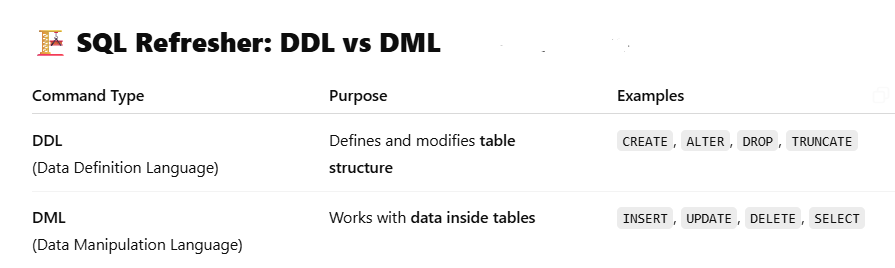
**1. Quantum Computing -**

* **Definition**: Quantum computing utilizes the principles of quantum mechanics to process information, enabling computations far beyond the capabilities of classical computers.
* **Real-World Example**: IBM's quantum systems are being explored for applications in cryptography, material science, and complex optimization problems.
* **Video**: [Quantum Computing Explained](https://www.youtube.com/watch?v=JhHMJCUmq28)
* **Further Reading**: [IBM Quantum Computing](https://www.ibm.com/quantum-computing/)

**2. Green Technologies in Industry**

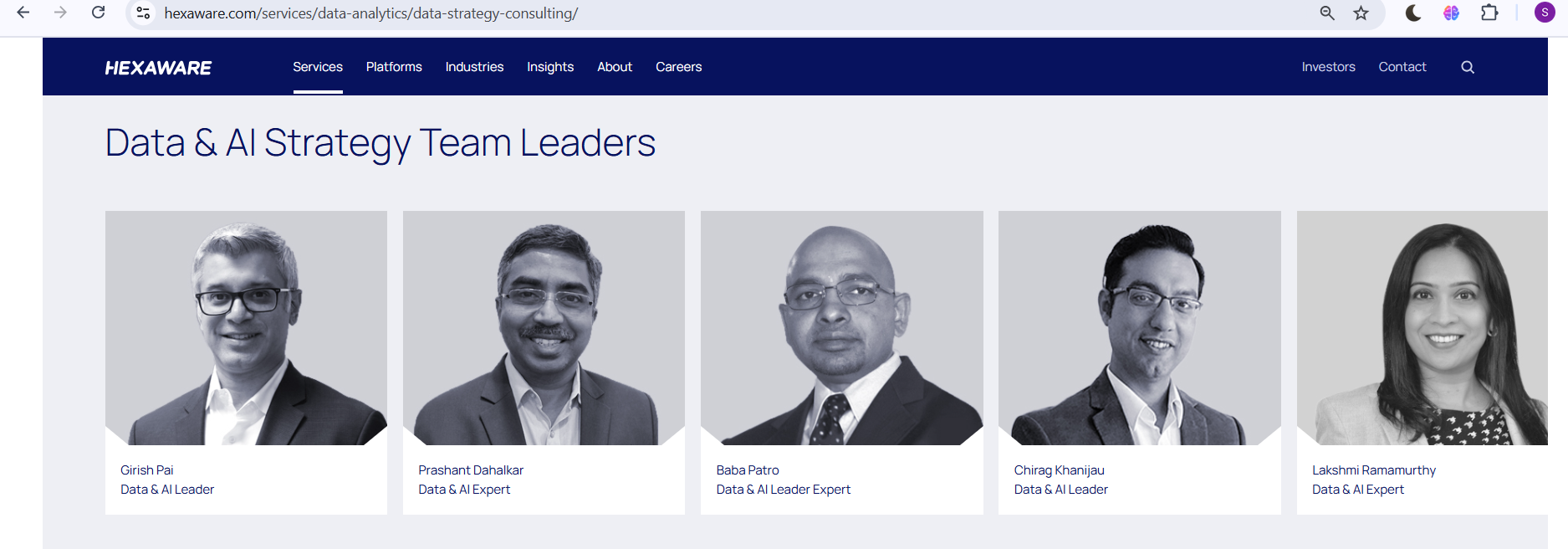
* **Definition**: Innovations aimed at reducing environmental impact, such as using hydrogen-based processes in steel manufacturing.
* **Real-World Example**: Sweden's Hybrit project is pioneering fossil-free steel production using hydrogen instead of coal.
* **Article**: [Green Technologies Transforming Steelmaking](https://www.reuters.com/sustainability/decarbonizing-industries/meet-green-technologies-set-transform-geopolitics-steelmaking-2025-05-28/)([Reuters](https://www.reuters.com/sustainability/decarbonizing-industries/meet-green-technologies-set-transform-geopolitics-steelmaking-2025-05-28/?utm_source=chatgpt.com))

**3. Humanoid Robotics**

* **Definition**: Robots designed to resemble and mimic human movements and interactions.
* **Real-World Example**: Agility Robotics' Digit robot is being tested for warehouse automation tasks.
* **Video**: [Agility Robotics' Digit](https://www.youtube.com/watch?v=5iV_hB08Uns)
* **Article**: [Humanoid Robots in Industry](https://www.ft.com/content/02f72125-dbc9-451d-84f8-1dc9e8bfb8ee)([Financial Times](https://www.ft.com/content/02f72125-dbc9-451d-84f8-1dc9e8bfb8ee?utm_source=chatgpt.com))

A screenshot of a chat

AI-generated content may be incorrect.

****HEXAWARE’S DATA & AI STRATEGY LEADERS