THREE CORE AREAS IN DATA

Data Science – Data Engineering – Data Analytics

PhD. Ha Minh Tan

Vietnam National University, Ho Chi Minh City

✉ tanhm@uit.edu.vn

0879 226 448

BACKGROUND & SIGNIFICANCE

Data Revolution: Exponential growth → require processing, analysis, and inference systems at scale.

Data-driven business: Strategic decisions based on quantitative evidence.

Career ecosystem: The three fields complement each other, forming a complete data value chain.

OVERVIEW OF THE THREE-SECTOR RELATIONSHIP

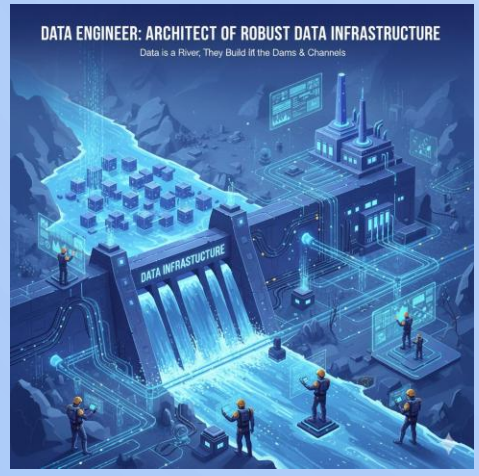
Data Engineering → build a data infrastructure foundation.

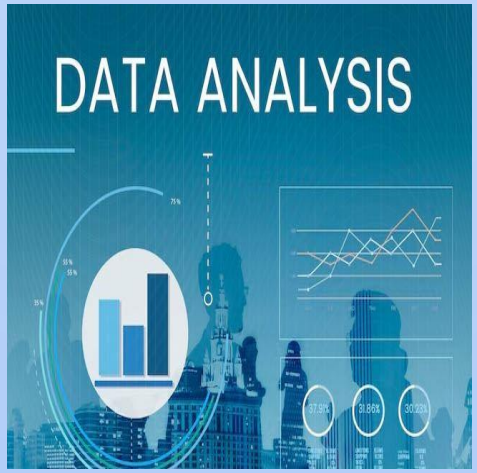
Data Analytics → harness descriptive and diagnostic knowledge from data**.**

Data Science → advanced algorithm-based forecasting and optimization model.

COLLECTION → PROCESSING → ANALYSIS → MODELING → DEPLOYMENT

SEQUENTIAL PIPELINE





I. DATA ENGINEERING

Focus: Design, architecture, and operation of large-scale data systems.

Core values: Ensure the right data – clean – stable – ready for the whole organization.

Features: Systems thinking, performance optimization, distributed data processing.

GOALS & PRODUCTS(DATA ENGINEERING)

- *Building sustainable ETL/ELT pipelines*.
- *Data Architecture Design: Data Lake, Data Warehouse, Lakehouse.*
- *Data Quality Assurance*(Data Quality, Observability).
- *Multi-source data integration*: streaming + batch.
- *Support Data Analyst and Data Scientist with trusted data.*

CORE TECHNOLOGY(DATA ENGINEERING)

- **Distributed Processing System**: Hadoop, Spark.
- **Data warehouse**: BigQuery, Snowflake, Redshift.
- **Pipelines**: Airflow, Dagster, Prefect.
- **Streaming**: Kafka, Flink.
- **Language**: Python, SQL, Scala.

II. DATA ANALYTICS

Focus: Explain, describe, and diagnose data-driven operational status.

Core Values: Turn data into insights for decision-making.

Features: Visualization, descriptive statistics, behavior analysis.

GOALS & PRODUCTS (DATA ANALYTICS)

- *Report Description.*
- *Cause Analysis* (Diagnostic).
- *Performance Evaluation*(KPIs, dashboards).
- *Strategy recommendations based on historical data.*

CORE TECHNOLOGY(DATA ANALYTICS)

- **Dashboarding**: Power BI, Tableau, Looker Studio.
- **Language:**SQL, Python (pandas), R.
- **Technique**: A/B Testing, descriptive statistics, correlation analysis.
- **Data warehouse**: use directly from the DE system.

III. DATA SCIENCE

Focus: Building predictive models, optimizations, and automated decision-making.

Core values: Creating forecasting capabilities and applied artificial intelligence.

Features: Mathematics – algorithms – modeling – statistical inference.

GOALS & PRODUCTS(DATA SCIENCE)

- *Forecast model*: Regression, Classification, Time-series.
- *Optimize operations*: Recommendation, Optimization, Clustering.
- *Model Implementation*: MLOps, monitoring.
- *Algorithm Research*: ML, DL, NLP, CV,…

CORE TECHNOLOGY(DATA SCIENCE)

- **LibraryML/DL**: scikit-learn, TensorFlow, PyTorch.
- **Notebook & Modeling**: Jupyter, Colab.
- **MLOps**: MLflow, Kubeflow, Vertex AI.
- **Language**: Python, R.
- **Computing infrastructure**: GPU/TPU, cloud AI.

IV. ACADEMIC COMPARISON BETWEEN THE THREE FIELDS

Data Engineering

- Focus on infrastructure, architecture, and system reliability.
- Suitable for technical – system – performance optimization.
- Focus on infrastructure, architecture, and system reliability.

Data Analytics

- Focus on interpretation & visualization.
- Suitable for business thinking – analysis – storytelling.
- Focus on interpretation & visualization.

Data Science

- Focus on modeling & algorithms.
- Suitable for mathematical thinking – forecasting model – research.
- Focus on modeling & algorithms.
- Suitable for mathematical thinking – forecasting model – research.

MỨC ĐỘ GIAO THOA (INTERDISCIPLINARY OVERLAP)

DE + DA: Access and process data for analysis.

DA + DS: Input data analysis & model evaluation.

DE + DS: Deploy the model into the production system.

→ **All three: Together form a data lifecycle.**

Keywords:

- **DE:** Data Analytics
- **DA:** Data Engineering
- **DS:** Data Science

V. KẾT LUẬN

- The three fields are not opposites, but complement each other.
- Market demand: increased sharply in the direction of:
  - **DE (infrastructure) → DA (decision-making) → DS (forecasting/AI).**
- The choice of direction depends on: basic capacity, technical thinking, interests and research orientation.

Từ khóa:

- **DE:** Data Analytics
- **DA:** Data Engineering
- **DS:** Data Science

THANK FOR WATCHING!

PhD. Ha Minh Tan

Vietnam National University, Ho Chi Minh City

tanhm@uit.edu.vn

0879 226 448