

# CS951-30 Computational Data Analytics

**26/27**

**Department**

Computer Science

**Level**

Taught Postgraduate Level

**Module leader**

Weiren Yu

**Credit value**

30

**Module duration**

10 weeks

**Assessment**

Multiple

**Study location**

University of Warwick main campus, Coventry

---

## Description

### Introductory description

This advanced data analytics module explores state-of-the-art techniques for analysing complex, high-dimensional, and multimodal data at scale. Students will apply modern methods in data integration, embedding, visualisation, and privacy-preserving analytics using scalable tools. The module aims to deepen students' understanding of applied machine learning and data infrastructure in practical, data-intensive environments.

### Module aims

The module builds on the module Foundations of Computational Data Analytics , equipping students with a deep understanding of state-of-the-art methods for analysing complex, high-dimensional, and multimodal data at scale. Data Analytics is a core discipline within computer science, with increasing importance in the age of digital transformation and emerging technologies, with significant economic impact. Because of the highly interdisciplinary nature of Data Analytics students will benefit from being able to pursue working in a wide range of application domains.

## Outline syllabus

This is an indicative module outline only to give an indication of the sort of topics that may be covered. Actual sessions held may differ.

Introduction to Data Representations: Structured, semi-structured, and unstructured data; metadata, provenance, and data lineage standards.

Multimodal Data and Fusion: Hybrid data structures (e.g. GeoJSON + time series), basics of multimodal fusion and alignment.

Scalable Data Cleaning: Outlier detection in high dimensions; standardisation, encoding, and processing with Dask, Spark, and Polars.

Exploratory Data Analysis (EDA): Techniques for text, time series, and geospatial data; correlation analysis and common EDA pitfalls.

Entity Resolution & Data Integration: Blocking, hashing, probabilistic matching, and schema alignment.

Dimensionality Reduction & Embedding: Linear and nonlinear projections (PCA, UMAP, autoencoders); manifold learning and interpretability.

Advanced Data Visualisation: Interactive, geospatial, network, and high-dimensional visualisation using tools (e.g. Plotly, Dash).

Privacy-Aware Analytics: Differential privacy, federated learning, homomorphic encryption, and legal frameworks (GDPR, CCPA).

Modern Data Infrastructure: Data warehousing (BigQuery, Snowflake), ETL/ELT pipelines, and lakehouse architectures for large-scale analytics.

Applied ML for Complex Data: Contrastive and masked modelling, pretraining across modalities, and trends in large-scale ML for complex data.

## Learning outcomes

By the end of the module, students should be able to:

- Analyse and transform complex, high-dimensional, and multimodal datasets using advanced data cleaning, integration, and embedding techniques.
- Apply scalable tools and frameworks to perform privacy-aware analytics and visualisations on large, heterogeneous data sources.
- Evaluate and implement modern data infrastructure and machine learning approaches for real-world, data-intensive applications.

## Indicative reading list

[Reading lists can be found in Talis](#)

## Research element

Coursework will contain a research element.

## Subject specific skills

Ability to analyse and transform complex, high-dimensional, and multimodal datasets using advanced data cleaning, integration, and embedding techniques.

Ability to apply scalable tools and frameworks to perform privacy-aware analytics and visualisations on large, heterogeneous data sources.

Skills to evaluate and implement modern data infrastructure and machine learning approaches for real-world, data-intensive applications.

## Transferable skills

Being able to apply Data Analytics knowledge and understanding of specialist theoretical and methodological approaches, suggesting and incorporating interrelationships with other relevant disciplines in abstract and unpredictably complex contexts.

Students will obtain the cognitive skills to critically contribute to existing discourses and methodologies in Data Analytics, suggesting new ideas, and designing systematic studies in Data Analytics based on critical analysis and evaluation.

Students will obtain practical skills in organising and communicating information, improving interpersonal, team and networking skills through engaging in classes and computer laboratories. Formative assessment will allow students to strategically enhance their own learning.

Data Analytics is an area with immediate relevance for increasing ethical awareness and its practical application regarding privacy concerns. The associated values will help understanding the importance of personal responsibility and ethical leadership.

---

## Study

### Study time

| Type                         | Required                    |
|------------------------------|-----------------------------|
| Lectures                     | 30 sessions of 1 hour (10%) |
| Seminars                     | 5 sessions of 2 hours (3%)  |
| Supervised practical classes | 9 sessions of 2 hours (6%)  |
| Private study                | 116 hours (39%)             |
| Total                        | 300 hours                   |

| Type       | Required        |
|------------|-----------------|
| Assessment | 126 hours (42%) |
| Total      | 300 hours       |

## Private study description

Private study, background reading and revision.

## Costs

No further costs have been identified for this module.

---

## Assessment

You do not need to pass all assessment components to pass the module.

### Assessment group D

|  | Weighting | Study time | Eligible for self-certification |
|--|-----------|------------|---------------------------------|
| Computational Data Analytics Coursework  | 30%       | 36 hours   | No                              |
| The coursework will test competency in applying tools and techniques to perform tasks in advanced data analytics and demonstrate in-depth understanding of methods and concepts. |           |            |                                 |
| Computational Data Analytics Exam  | 70%       | 90 hours   | No                              |

- Answerbook Gold (24 page)
- Students may use a calculator

### Assessment group R

|   | Weighting | Study time | Eligible for self-certification |
|---|-----------|------------|---------------------------------|
| Computational Data Analytics Resit Exam | 100%      |            | No                              |

- Answerbook Gold (24 page)
- Students may use a calculator

## Feedback on assessment

For coursework individual feedback will be provided. For the exam collective feedback will be

provided.

[Past exam papers for CS951](#)

---

## **Availability**

### **Courses**

This module is Core optional for:

- TCSA-G5PD Postgraduate Taught Computer Science
  - Year 1 of G5PD Computer Science
  - Year 1 of G5PG Computer Science with specialism in Artificial Intelligence and Machine Learning
  - Year 1 of G5PH Computer Science with specialism in Cyber Security
  - Year 1 of G5PI Computer Science with specialism in Data Analytics