COURSE BROCHURE

Big Data Analytics

An advanced, vendor-neutral course on distributed storage and processing, data lakes and real-time streaming, and on when big data really needs tools of its own.

Code	HT-DAT-109
Category	Data & Analytics
Duration	1 week
Level	Advanced
City	Istanbul

Introduction

The big data label is sometimes stuck on tables that a well-designed database would handle easily, and sometimes ignored until systems buckle under the load. The right call starts with understanding the real limits of each option.

Objectives

- Determine when a problem calls for big data tools and when conventional solutions suffice
- Explain the principles of distributed storage and processing and their effect on performance and cost
- Compare warehouses, data lakes and lakehouse platforms
- Assess batch and real-time processing needs for a given use case
- Design an analytics architecture that accounts for governance, cost and scalability

Who should attend

- Analytics and data team managers and leads
- Data engineers and architects
- Senior analysts dealing with growing data volumes
- IT managers involved in choosing data platforms

Skills you will develop

- Assessing big data use cases
- Distributed processing principles
- Data architecture design
- Streaming analytics

Learning Outcomes

- Prepare a criteria list that shows whether a case calls for distributed tools
- Draw a data lake architecture with its layers from ingestion to consumption
- Compare the cost and performance of two architectural options for one use case
- Prepare a documented architecture proposal for a case combining batch and real-time data

Course Outline

Day 1: What Makes Data Big

- Volume, velocity and variety and what each means in practice
- The limits of traditional databases, and when volume is not the real problem
- Typical use cases: logs, sensors, clickstreams and text

Day 2: Distributed Storage and Processing

- Distributed file systems and object storage
- Partitioning, parallelism and moving computation to the data
- Distributed processing engines and the distributed DataFrame model, with a live demo

Day 3: Warehouses, Lakes and Lakehouses

- The data warehouse versus the data lake, and the lakehouse concept
- Columnar formats such as Parquet and open table formats
- Data layers from raw to refined to consumption-ready

Day 4: Real-Time Streaming

- Events, message queues and stream processing
- Time windows, late-arriving events and delivery guarantees
- Trading off low latency against cost and complexity

Day 5: The Architecture Decision

- Cost in cloud and on-premises environments and keeping spend under control
- Governing the lake so it does not turn into a data swamp
- Applied project: an architecture proposal for an analytics platform combining transaction data with real-time sensor data

Training methodology

Live trainer demonstrations of distributed processing and streaming
Whiteboard architecture sketches for use cases from different sectors
A cost comparison exercise across architectural options using participants' own assumptions
Critical review of participants' proposals from the viewpoint of a technical committee

Certification

Participants receive a Harbor Training Center certificate of attendance on completing the course.

Upcoming dates

City	Dates	Language	Fee
Istanbul	15 – 19 February 2027	AR	EUR 5,400
Istanbul	16 – 20 August 2027	EN	EUR 5,400

Fees are per participant.

Contact us

Email info@harbor4training.com

Address istanbul Turkey

To register and for details, scan the code or visit:

https://harbor4training.com/en/courses/big-data-analytics