JUPITER PRIME

Consulting • Training • Advisory

جوبيتر برايم للاستشارات والتدريب

COURSE BROCHURE

Big Data Management and Analysis Essentials Training Course

7 – 11 June 2027
Dar es Salaam

Big Data Management and Analysis Essentials Training Course

Ref: 2972_112678 **Date:** 7 – 11 June 2027 **Location:** Dar es Salaam **Fees:** 7900 GBP

Big Data Management and Analysis Essentials is a 1-week training course for managers and analysts handling very large data sets, ending with a big data use case and delivery plan. Sensor feeds and transaction logs now arrive in volumes single servers cannot hold, so leaders must judge where scale adds insight and where it only adds cost. Jupiter Prime Training offers this course on Big Data Management and Analysis Essentials.

Why Big Data Management and Analysis Essentials Matters Now

Three forces shape the subject in organisations today. Cheap cloud storage tempts organisations to keep everything, yet unmanaged collections become costly. Business sponsors expect answers from machine logs, clickstreams and equipment readings within days, not quarters. Distributed processing across many machines changes how analysts frame questions and check results. Careful use case choice brings earlier warnings and sharper insight; careless choice leaves idle archives. The course follows a staged analytics life cycle, moving each use case from business question to deployment.

Who Gains Most from Big Data Management and Analysis Essentials

- Those who sponsor initiatives built on high-volume records
- Those who query, sample and summarise large collections to answer business questions
- Those who decide storage tiers, retention and access for growing data holdings
- Those who link business sponsors with technical specialists

An operations or customer analytics unit gains most when a sponsor, two analysts and a technical liaison attend together, because scoping and feasibility judgements are then shared.

Big Data Management and Analysis Essentials Capabilities

Participants weigh volume, variety and velocity against a business question before committing effort.

- Assess whether a business question genuinely needs big data methods — sponsors choosing where to invest
- Specify storage tiers and partitioning for a large data holding — cost owners and infrastructure leads
- Apply sampling and aggregation to explore large collections — analysts framing findings
- Design preparation steps for duplicates and gaps at scale — report users
- Calculate the effort and running cost of a distributed analysis job — budget holders
- Build a big data use case with milestones and success measures — steering groups approving funding

Key Big Data Management and Analysis Essentials Terms

- Volume, variety and velocity — the traits of size, mixed formats and arrival speed that mark big data.
- Distributed processing — splitting a computation across many machines, then combining their results.
- Partitioning — dividing a large data set into segments, often by date, so queries read only what they need.
- Schema on read — applying structure to raw records only when they are queried.
- Data lake — a central store holding raw records in their original formats until a use case calls for them.
- Sampling — selecting a representative subset of records to explore patterns before full-volume analysis.

Big Data Management and Analysis Essentials Day by Day

Day 1: Scale, Sources and Business Questions

- Volume, Variety and Velocity Profile — Judging Whether a Problem Needs Scale
- Source Inventory of Logs, Sensors and Transactions — Locating Usable Raw Material
- Business Question Framing Canvas — Linking Analysis to a Sponsor Decision
- Sector Cases in Utilities and Retail — Comparing Payback From Large Data Holdings
- Use Case Value Scoring Sheet — Ranking Candidate Initiatives for Sponsorship

Day 2: Large-Scale Storage Choices

- Data Lake Zones for Raw and Refined Records — Separating Exploration From Trusted Use
- Storage Tiering by Access Frequency — Balancing Retrieval Speed Against Holding Cost
- Partitioning and File Format Selection — Cutting Query Time on Large Collections
- Metadata Catalogue and Lineage Notes — Helping Analysts Find and Trust Holdings
- Retention and Archiving Rules — Deciding What to Keep, Compress or Delete

Day 3: Distributed Processing Concepts

- Split, Process and Combine Pattern — Planning Work Across Many Machines
- Batch Versus Near Real-Time Processing — Matching Latency to the Business Need
- Cluster Sizing and Job Cost Estimates — Setting Budgets for Analysis Runs
- Schema on Read Exploration — Querying Raw Records Without Prior Modelling
- Access Controls for Shared Holdings — Protecting Sensitive Fields in Large Stores

Day 4: Preparing and Analysing at Scale

- Deduplication and Gap Handling Rules — Cleaning Millions of Records Consistently
- Sampling Strategies for Exploration — Finding Patterns Before Full-Volume Runs
- Aggregation and Windowed Summaries — Turning Event Streams Into Decision Measures
- Clustering and Anomaly Detection Basics — Spotting Unusual Customers or Equipment
- Result Validation Against Known Totals — Confirming Outputs Before Sponsors Act

Day 5: Practising the Big Data Use Case and Delivery Plan

- Use Case Selection Review — Applying Value Scores to a Live Candidate
- Storage and Partitioning Design — Fitting Day Two Choices to the Chosen Case
- Processing Cost and Effort Estimate — Using Sizing Methods From Day Three
- Success Measures and Validation Checks — Defining Evidence the Sponsor Will Accept
- Big Data Use Case and Delivery Plan — Sequencing Milestones, Roles and Funding Gates

Applying Big Data Management and Analysis Essentials After the Course

- Score three candidate initiatives with the value sheet and present the leader to a sponsor.
- Sample one large internal holding and record the partitioning changes it suggests.

Big Data Management and Analysis Essentials Questions FAQ

What is Big Data Management and Analysis Essentials and why does it matter now?

It is the practice of storing, preparing and analysing data sets too large or varied for ordinary tools. It matters now because organisations collect records faster than they can use them.

Which roles gain most from Big Data Management and Analysis Essentials?

Managers who sponsor analytics initiatives and analysts who work with high-volume records gain most. Technical liaisons also benefit directly.

How does Big Data Management and Analysis Essentials differ from an advanced data engineering course?

This course works at foundation level on use case choice, storage options and analysis. An advanced engineering course goes deeper into automated data flows and storage architecture.

What do you bring back from Big Data Management and Analysis Essentials?

You bring back a scored use case, a storage and partitioning outline, a processing cost estimate and a delivery plan with funding gates, each shaped around your own data holding.

Big Data Management and Analysis Essentials: The Way Forward

The big data use case and delivery plan gives a sponsor a clear funding decision: whether the chosen initiative proceeds to its first milestone. Analysts use it to sequence storage and processing work. After the first validated result, the value scoring behind it is worth revisiting with fresh cost evidence.