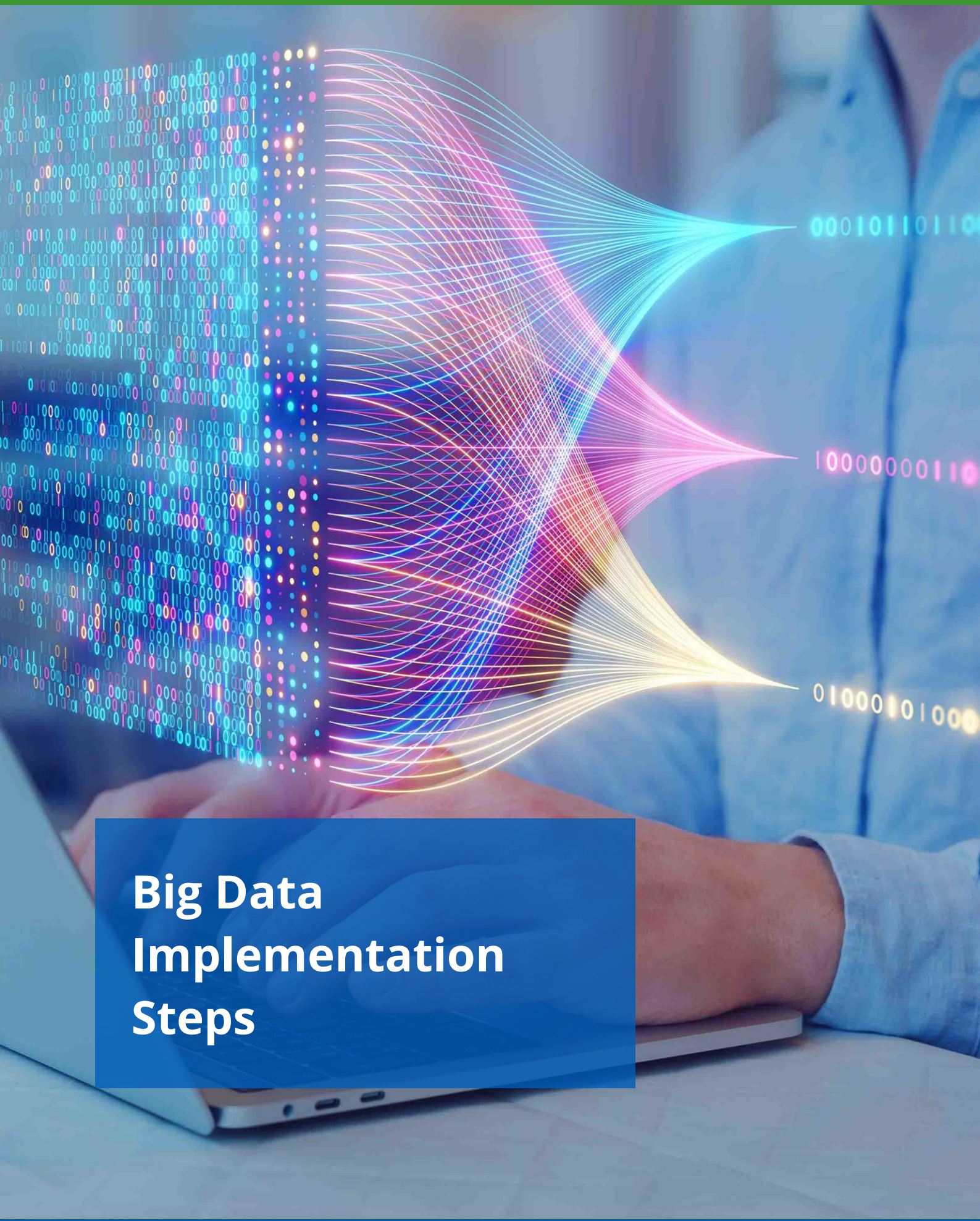
ScienceSoft
Professional Software Development

www.scnsoft.com



Big Data Implementation Steps

US, McKinney, TX +1 214 306 68 37 Finland, Vantaa +371 66 011 905 UAE, Fujairah +971 600 57 59 69 KSA, Riyadh +966 800 880 3035
www.scnsoft.com contact@scnsoft.com

Introduction

Real-life big data implementation steps may vary greatly depending on the business goals a solution is to meet, data processing specifics (e.g., real-time, batch processing, both), etc. However, from ScienceSoft's experience, there are six universal steps that are likely to be present in most projects.

Step 1. Feasibility study

Analyzing business specifics and needs, validating the feasibility of a big data solution, calculating the estimated cost and ROI for the implementation project, assessing the operating costs.

> *Analyzing business specifics and needs, validating the feasibility of a big data solution, calculating the estimated cost and ROI for the implementation project, assessing the operating* costs.
>
> ScienceSoft

Step 2. Requirements engineering and big data solution planning

- Defining the types of data (e.g., SaaS data, SCM records, operational data, images and video) to be collected and stored, the estimated data volume, and the required data quality metrics.
- Forming a high-level vision of the solution, outlining:
  - Data processing specifics (batch, real-time, or both).
  - Required storage capabilities (data availability, data retention period, etc.).
  - Integrations with the existing IT infrastructure components (if applicable).
  - The number of potential users.
  - Security and compliance (e.g., HIPAA, PCI DSS, GDPR) requirements.

  - Analytics processes to be introduced to the solution (e.g., data mining, ML-based predictive analytics).
- Choosing a deployment model: on-premises vs. cloud (public or private) vs. hybrid.
- Selecting an optimal technology stack.
- Preparing a comprehensive project plan with timeframes, required talents, and budget outlined.

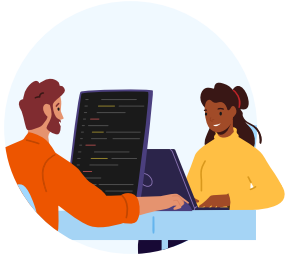
ScienceSoft can provide you with expert guidance on all aspects of big data planning.

Step 3. Architecture design

- Creating the data models that represent all data objects to be stored in big data databases, as well as associations between them, to get a clear picture of data flows, the ways data of certain formats will be collected, stored, and processed in the solution-to-be.
- Mapping out data quality management strategy and data security mechanisms (data encryption, user access control, redundancy, etc.).
- Designing the optimal big data architecture that enables data ingestion, processing, storage, and analytics.

> *As your business grows, the number of big data sources and the overall data volume will likely grow too. For instance, if we compare infographics for 2020 and 2021, we'll see that the volume of video streams on YouTube per minute grew from 500 to almost 700 hours in just a year. This makes scalable architecture the cornerstone of efficient big data implementation that can save you from costly redevelopments down the road.*
>
> ScienceSoft

Step 4. Big data solution development and testing



- Setting up the environments for development and delivery automation (CI/CD pipelines, container orchestration, etc.).
- Building the required big data components (e.g., ETL pipelines, a data lake, a DWH) or the entire solution using the selected techs.
- Implementing data security measures.
- Performing quality assurance in parallel with development. Conducting comprehensive testing of the big data solution, including functional, performance, security and compliance testing. If you're interested in the specifics of big data testing process, see expert guide by ScienceSoft.

With 10 years of experience in delivering end-to-end big data solutions, ScienceSoft is ready to help you develop and deploy your big data software.

Step 5. Big data solution deployment

- Preparing the target computing environment and moving the big data solution to production.
- Setting up the required security controls (audit logs, intrusion prevention system, etc.).
- Launching data ingestion from the data sources, verifying the data quality (consistency, accuracy, completeness, etc.) within the deployed solution.
- Running system testing to validate that the entire big data solution works as expected in the target IT infrastructure.
- Selecting and configuring big data solution monitoring tools, setting alerts for the issues that require immediate attention (e.g., server failures, data inconsistencies, overloaded message queue).

- Delivering user training materials (FAQs, user manuals, a knowledge base) and conducting Q&A sessions and trainings, if needed.

Step 6. Support and evolution (continuous)

- Establishing support and maintenance procedures to ensure trouble-free operation of the big data solution: resolving user issues, refining the software and network settings, optimizing computing and storage resources utilization, etc.
- Evolution may include developing new software modules and integrations, adding new data sources, expanding the big data analytics capabilities, introducing new security measures, etc.

About ScienceSoft

ScienceSoft is a global IT consulting and software development company headquartered in McKinney, TX. Since 2013, we have been delivering end-to-end big data services to businesses in 30+ industries. Being **ISO 9001** and **ISO 27001**-certified, we ensure robust quality management system and full security of our customers' data.