# CASE STUDY

**Data Governance System
for One of Big Four**

softserve

## Client Background

Our client is a multinational professional services firm, one of the "Big Four" audit companies. It is one of the largest professional service companies in the world and in the top ten largest privately owned organizations in the US.

The organization operates as a network of member firms which are separate legal entities in individual countries. It has more than 270,000 employees in over 700 offices across 150 countries in the world. It provides financial audit and assurance, tax, consulting and advisory services to companies.

## The Problem

Our client's personnel offer proactive information governance advice aimed at reducing the risk and cost associated with electronically stored information. They can create an information governance strategy, test both governance and operational adherence or implement performance monitoring.

Additionally, the ETL (Extract, Transform and Load) process wasn't standardized, which led to a mismatch between requested and required data for analysis.Furthermore, it required weeks of time to do the ETL process one additional time.

Also, the same data from one customer could be stored in various structures (depending on a solution), which was not beneficial.

According to statistics:

- 80% of all enterprise data is unstructured

- Unstructured data in the enterprise is predicted to grow up to 80% annually

- 69% of all that unstructured enterprise data no longer has any legal or business value

- In 2020, 55% of your data storage is being spent on keeping and maintaining Redundant, Outdated, and Trivial (ROT) data
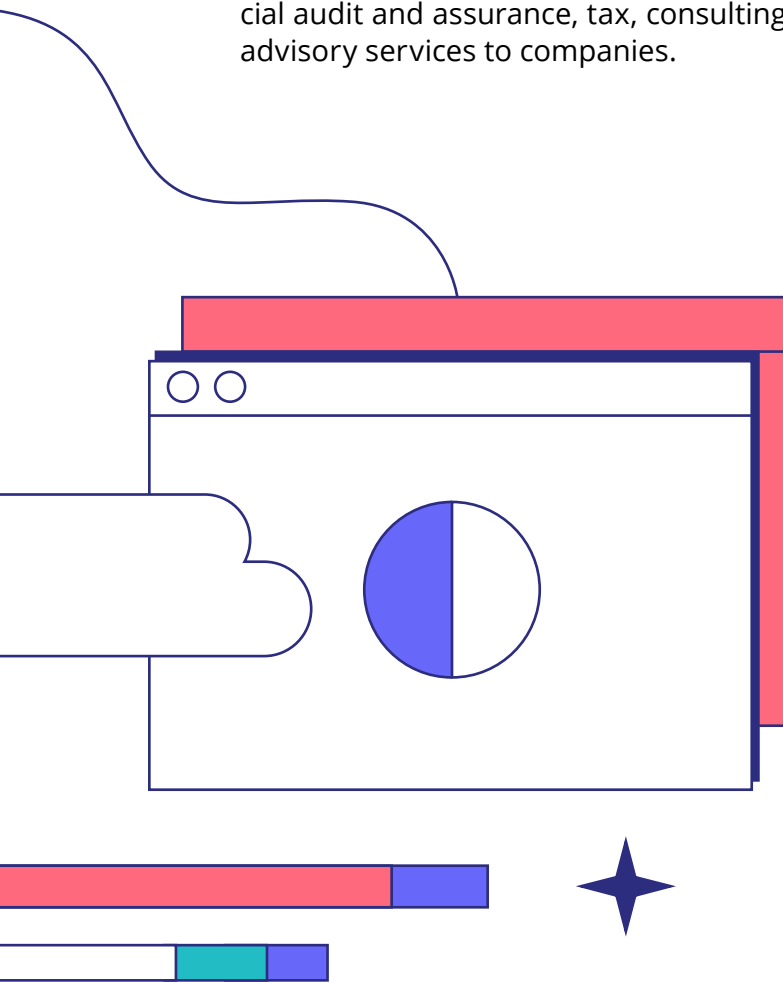
## Business Challenge

As our customers' data became larger and less manageable, cost avoidance became impossible, and the dangers from excessive data retention became unavoidable. Specific challenges included:

- Limited data visibility

- Inability to act

- Lack of context

- Lack of simple to use boxed solution

## Project Description

Our client partnered with SoftServe to design and implement a solution that meets even the most challenging demands:

- All-in-one

- Purpose built

- Hyper fast

- Big data scalable

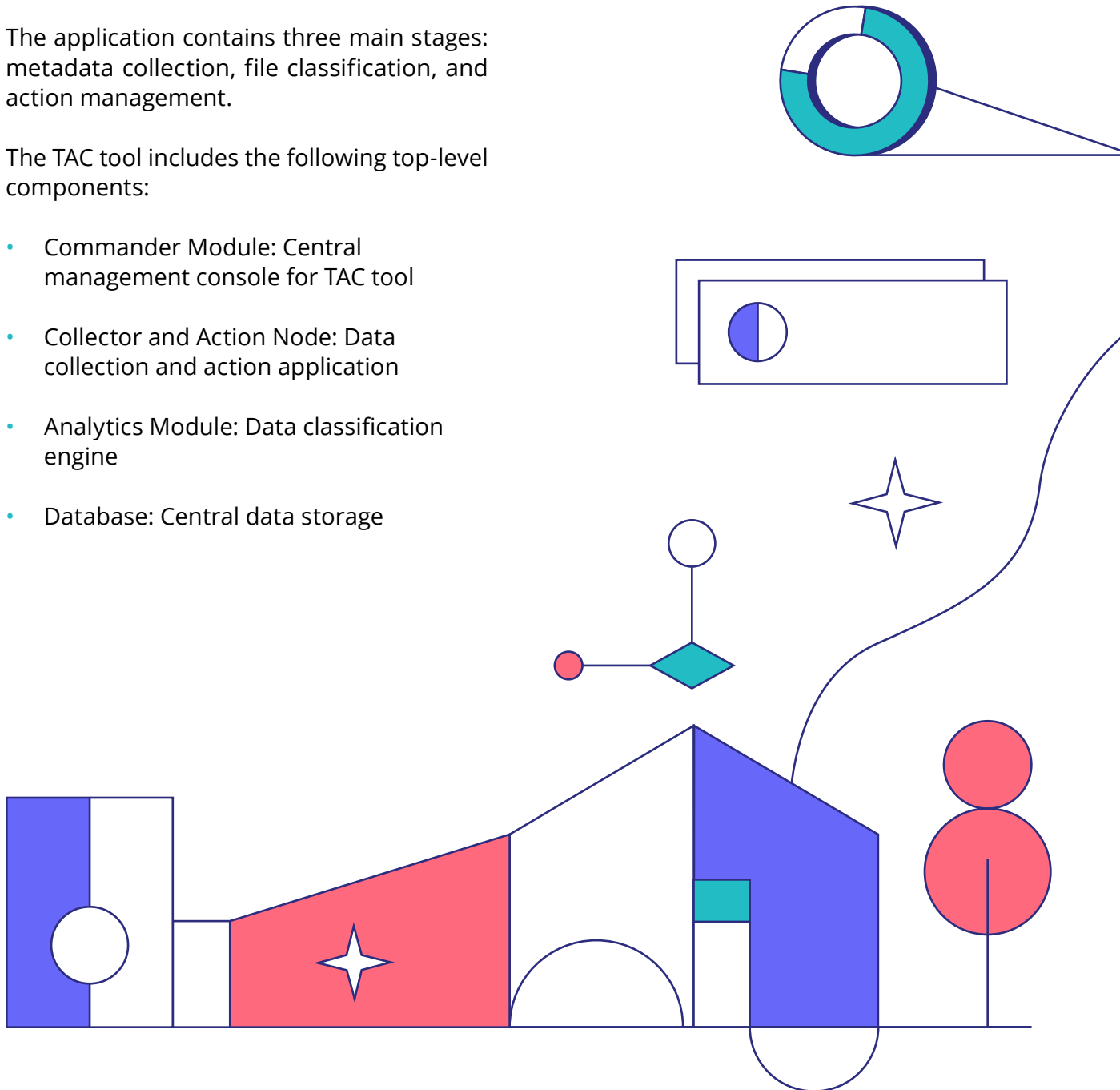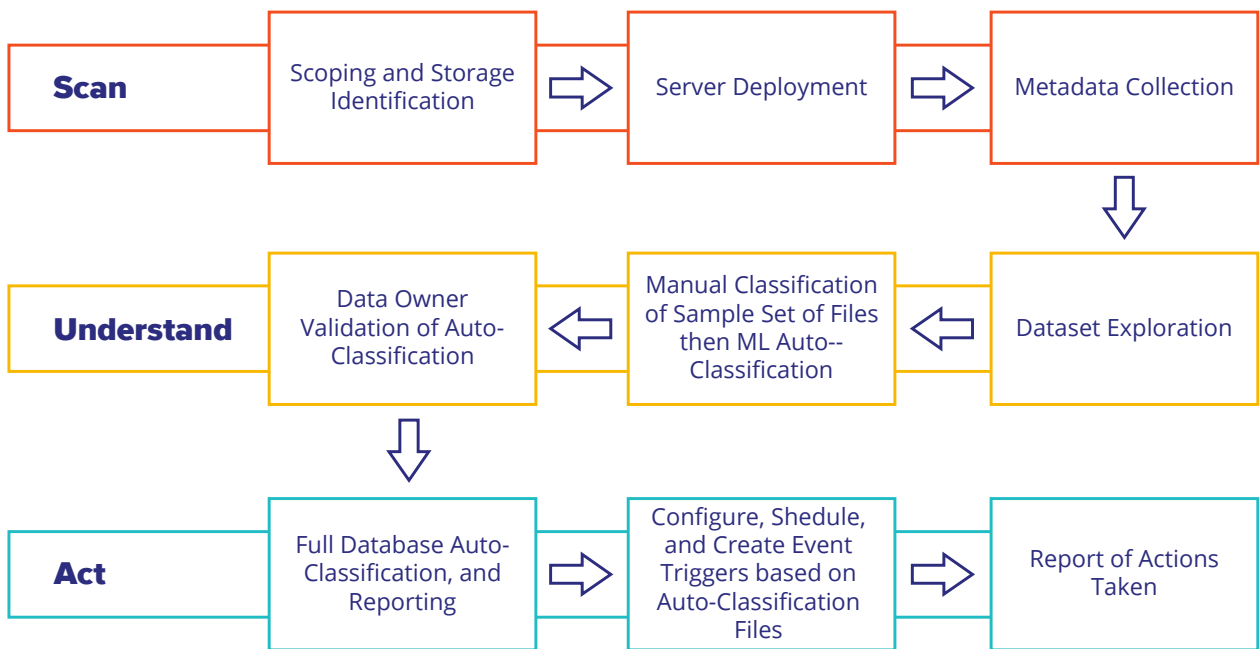- Analytic driven

- Decisive

- Defensible
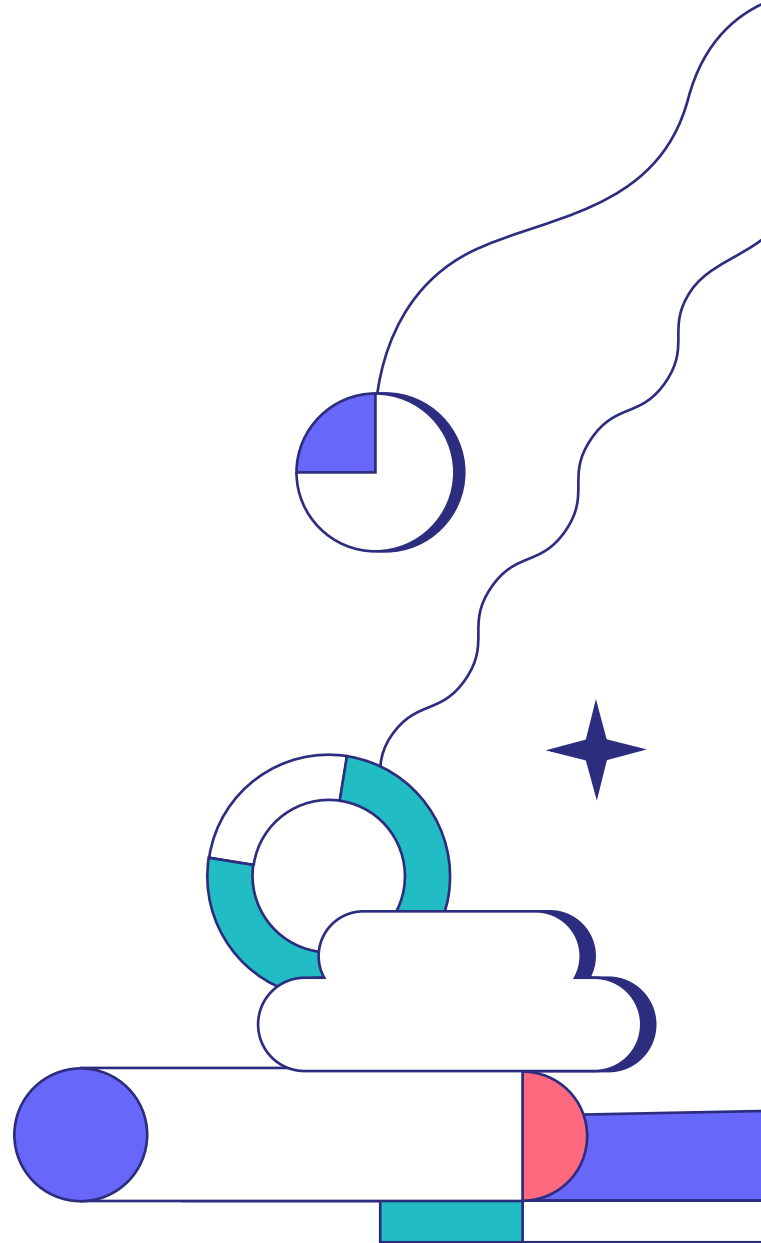
The solution is a Technically Assisted Classification (TAC) tool. It is a web application that helps users effectively govern the storage space by collecting files, analyzing them, and performing actions on them in the file system. This tool significantly reduces the cost of storage maintenance and helps optimize the business.

The application contains three main stages: metadata collection, file classification, and action management.

The TAC tool includes the following top-level components:

- Commander Module: Central management console for TAC tool

- Collector and Action Node: Data collection and action application

- Analytics Module: Data classification engine

- Database: Central data storage

| Scan | Scoping and Storage Identification | → | Server Deployment | → | Metadata Collection |
|---|---|---|---|---|---|

| Understand | Data Owner Validation of Auto-Classification | ← | Manual Classification of Sample Set of Files then ML Auto--Classification | ← | Dataset Exploration |
|---|---|---|---|---|---|

| Act | Full Database Auto-Classification, and Reporting | → | Configure, Shedule, and Create Event Triggers based on Auto-Classification Files | → | Report of Actions Taken |
|---|---|---|---|---|---|

TAC tool features:

- Scanning engine

- Scalable architecture

- Scalable database

- Two-pass data analytics engine

- Auto-classification engineModularly expandable data analytics platform

- Modularly expandable visual analytics platform

- Modularly expandable data exploration platform

- Unified interface for above items

SoftServe worked independently and cooperated with the client on product requirements.

There were three main releases:

1. Data scan

2. Data classification

3. Operations on data

SoftServe's project ended with an extensive training of product stakeholders on deployment and testing.

## Technology Stack

- Python
- Tornado
- Go
- SQLAlchemy
- PostgreSQL
- REST API

- NFS
- Django
- Elasticsearch
- Pandas
- LDAP
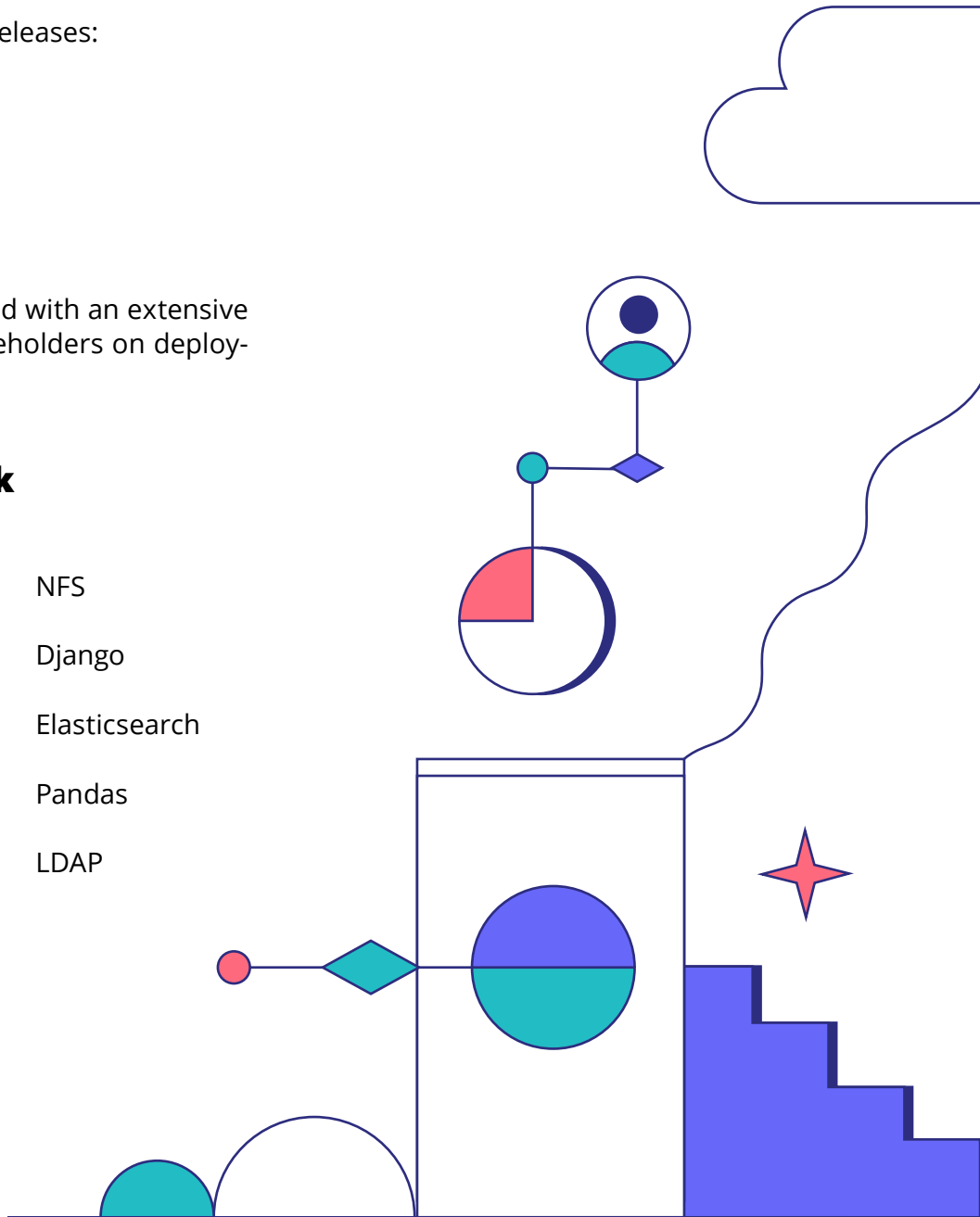
# Value Delivered

*S*oftServe provided a smooth delivery of the project implementation phase, despite the absence of product owner on the client's side during the last stage.

As a result, the Technically Assisted Classification tool provides the following values for each user type:

## General Counsel

• Reduces risk of privacy and data breaches

• Improves eDiscovery search performance and reduce cost

• Complies with regulatory data disposal mandates

## Corps Information Systems Control Officer

• Transitions security strategy from threat- centric to asset centric (truly risk-based)

• Enables big data protection capabilities

• Improves data loss prevention capabilities / reduces risk of breach

• Segments data by business unit
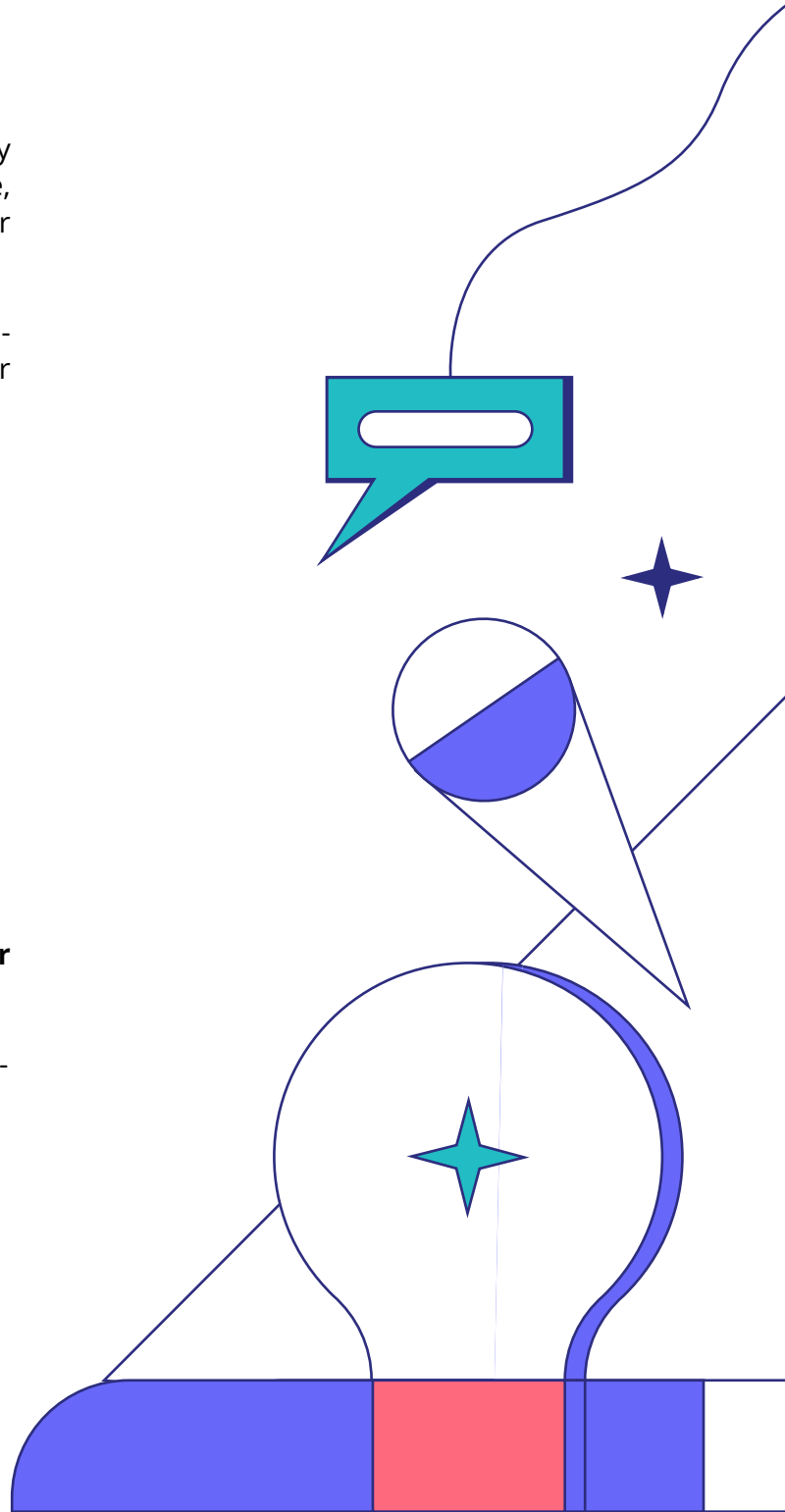
## Chief Financial Officer / Chief Operation Officer

- Reduces risk of privacy and data breaches

- Reduces total cost of ownership for enterprise storage management

- Improves business resiliency and disaster recovery capabilities

## Chief Information Officer / Chief Technology Officerr

- Reduces total cost of ownership for enterprise storage management

- Improves enterprise search efficiency and data quality

- Improves business resiliency and disaster recovery

- Improves data backup operations

- Improves data allocation enforcement

## Business Owners

- Improves data management and data sharing capabilities

- Improves enterprise search efficiency and data quality

- Improves business resiliency and disaster recovery capabilities
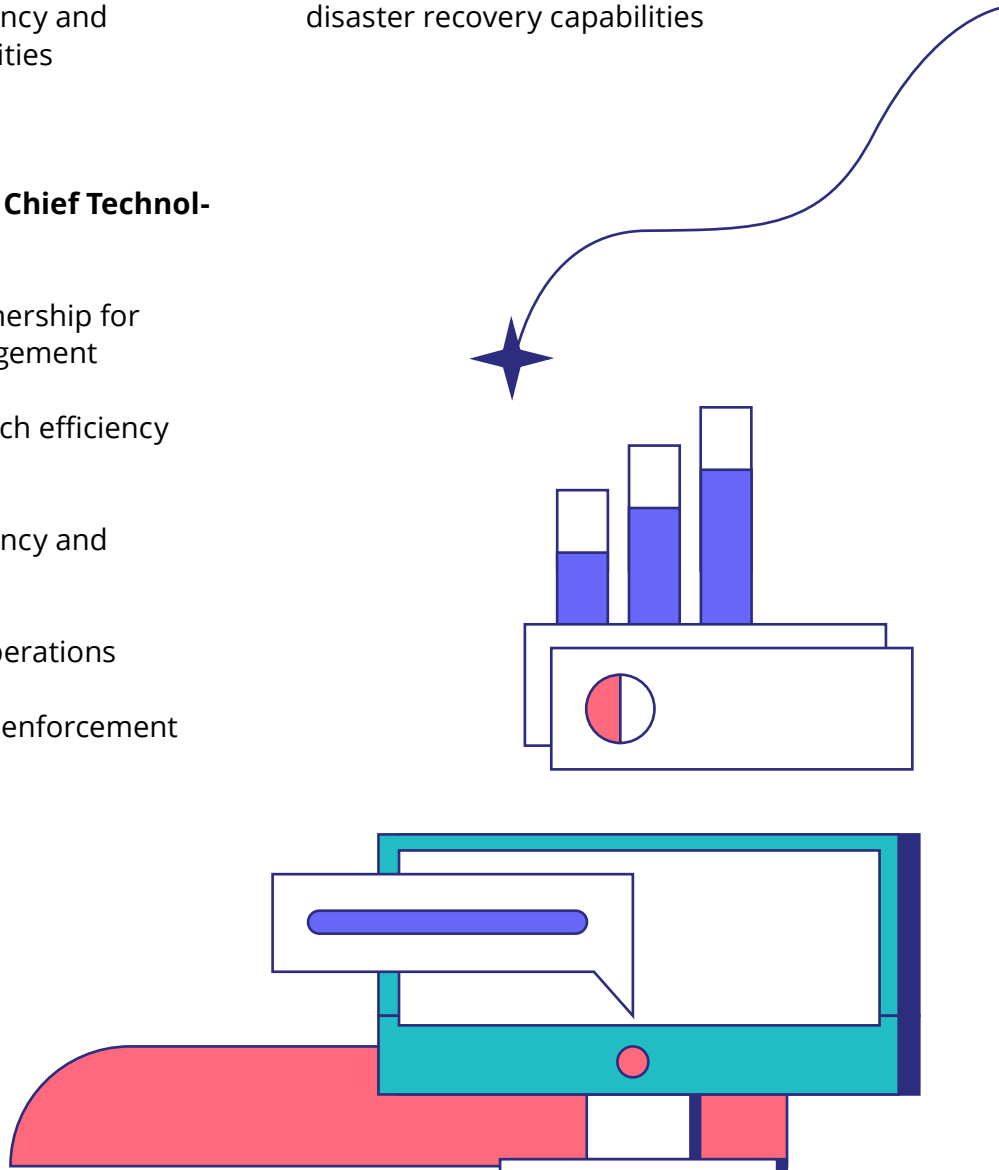
# ABOUT US

SoftServe is a digital authority that advises and provides at the cutting-edge of technology. We reveal, transform, accelerate, and optimize the way enterprises and software companies do business. With expertise across healthcare, retail, energy, financial services, and more, we implement end-to-end solutions to deliver the innovation, quality, and speed that our clients' users expect.

SoftServe delivers open innovation, from generating compelling new ideas, to developing and implementing transformational products and services.

Our work and client experience is built on a foundation of empathetic, human-focused experience design that ensures continuity from concept to release.

We empower enterprises and software companies to (re)identify differentiation, accelerate solution development, and vigorously compete in today's digital economy. No matter where you are in your journey.

Visit our **website**, **blog**, **LinkedIn**, **Facebook**, and **Twitter** pages.

### NORTH AMERICAN HQ

201 W 5th Street, Suite 1550
Austin, TX 75703
USA +1 866 687 3588 (USA)
+1 647 948 7638 (Canada)

### EUROPEAN HQ

14 New Street
London EC2M 4HE
United Kingdom
Level 39, One Canada Square

Canary Wharf, London E14 5AB
United Kingdom
+44 (0) 800 302 9436

info@softserveinc.com
www.softserveinc.com

softserve