
Analytics on Azure
PLURALSIGHT

Contents

01

Big Data Analytics

Benefits of Big Data Analytics

Improved Decision Making

Improved decision-making means businesses can make data-driven decisions based on comprehensive analytics

Enhanced Operational Efficiency

Increases efficiency by analyzing operational data to identify and eliminate bottlenecks

Customer Insights

Provides deep insights into customer behaviors and preferences, improving engagement and satisfaction

Big Data Tools in Azure

Azure Data Factory

A service that supports build and operationalization of Big Data pipelines for ETL

Azure Synapse Analytics

A scalable analytics service to process big data jobs easily, using a variety of pool types

Azure HDInsight

A cloud service that makes it easy to process big data using popular open-source frameworks like Hadoop, Spark, etc.

Azure Databricks

An Apache Spark-based analytics platform that is optimized for Azure

Implementing Big Data Analytics

01

Set Up Infrastructure

Infrastructure setup involves choosing appropriate tools and services for your specific big data requirements (relative to your target platform)

02

Data Analysis Techniques

Discusses various techniques to analyze big data, including pattern recognition and machine learning

03

Visualization and Reporting

Effective visualization to present complex data in actionable insights

02

Why do you need a data warehouse?

Centralized Data Management

01

Unified Data Access

Centralizes data from different sources for better accessibility and management

02

Streamlined Reporting

Enables consistent and streamlined data reporting across departments

03

Enhanced Data Quality

Ensures high data quality through integrated and automated data cleansing and transformation processes

Business Intelligence

01

Informed Decision Making

Provides historical data analysis for informed business decisions

02

Competitive Advantage

Leveraging data insights to gain a competitive edge in the marketplace

03

Predictive Analysis

Uses historical data for forecasting future trends and behaviors

Cost Efficiency

Cost Reduction

Reduces costs by consolidating data infrastructure and optimizing data storage

Resource Optimization

Ensures optimal use of resources by streamlining data management processes

Data Archiving

Efficiently archives data reducing primary storage load, improving system performance

03

Creating and Managing Data Pipelines in Azure

Ingestion

Data Sources

Defines various data sources that can be ingested into Azure for analytics

Data Integration

Describes methods for integrating disparate data sources seamlessly

Real-time Ingestion

Explains the importance and methods of ingesting real-time data for timely analysis (e.g., streaming data)

Storage

Data Storage Options

Details storage options in Azure, including Azure Blob Storage, Azure Data Lake (Gen2 storage), relational, and NoSQL

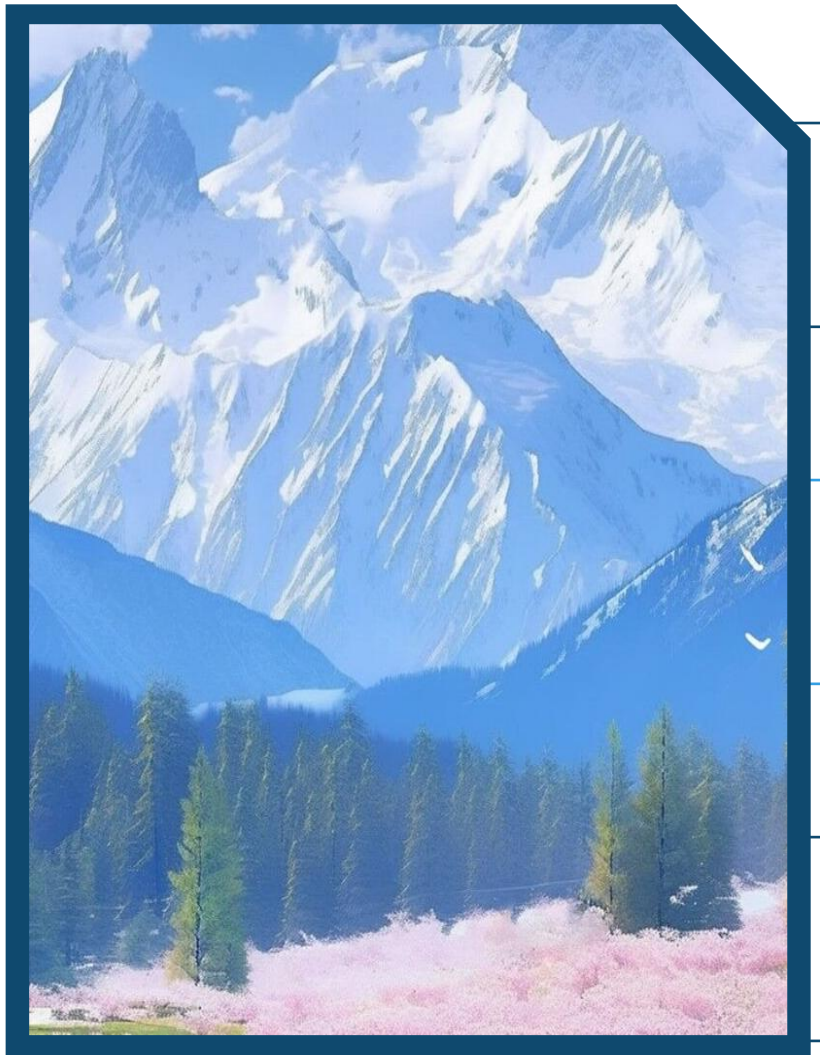
Backup and Recovery

Best practices for data backup and recovery to ensure data integrity and availability

Scalability

Scaling storage solutions to handle growing data volumes efficiently

Preparation



Data Cleaning

Processes and tools for cleaning data to ensure accuracy and reliability

Data Transformation

Transforming data into formats suitable for analysis

Data Enrichment

Enhancing raw data with additional information for deeper insights

04

Orchestration

Workflow Automation

Azure Data Factory

Azure Data Factory can be used for efficiently orchestrating and automating ETL workflows

Pipeline Scheduling

Pipelines can be executed on demand or can be scheduled for regular data processing tasks

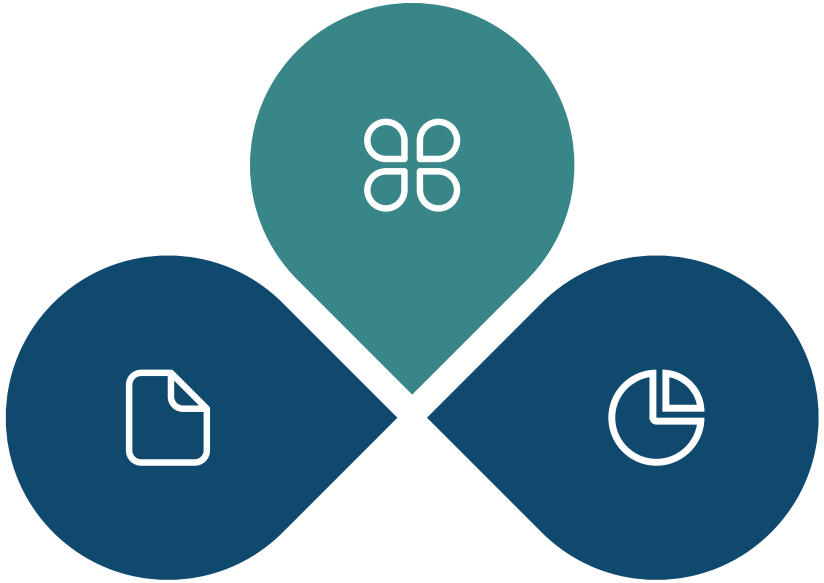
Monitoring Workflows

Important to ensure smooth operation of workflows through effective monitoring

Integration Services

Integration with Other Azure Services

Ensuring seamless integration with other Azure services for extended functionality



Leveraging APIs

Using APIs to integrate external data sources and services with Azure pipelines

Continuous Integration/Continuous Deployment (CI/CD)

Practices for CI/CD in orchestrating data pipelines for faster and more reliable deployments

Security

Data Encryption

Measures taken to encrypt data at rest and in transit

Access Controls

Defining role-based and attribute-based access controls to secure data pipelines (zero trust architecture)

Compliance

Ensuring data handling practices comply with industry regulations and standards

05

Monitoring

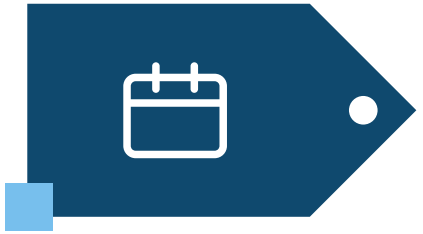
Tracking Data Pipeline Performance

Performance Metrics

Key metrics to track pipeline performance

Alerts and Notifications

Setting up alerts for any issues in data pipeline processes



Logs and Audit Trails

Maintaining logs and audit trails for troubleshooting and compliance

Performance Optimization

Identifying Bottlenecks

Techniques to identify and resolve performance bottlenecks in data pipelines

Resource Allocation

Optimizing resource allocation for peak efficiency in pipeline operations

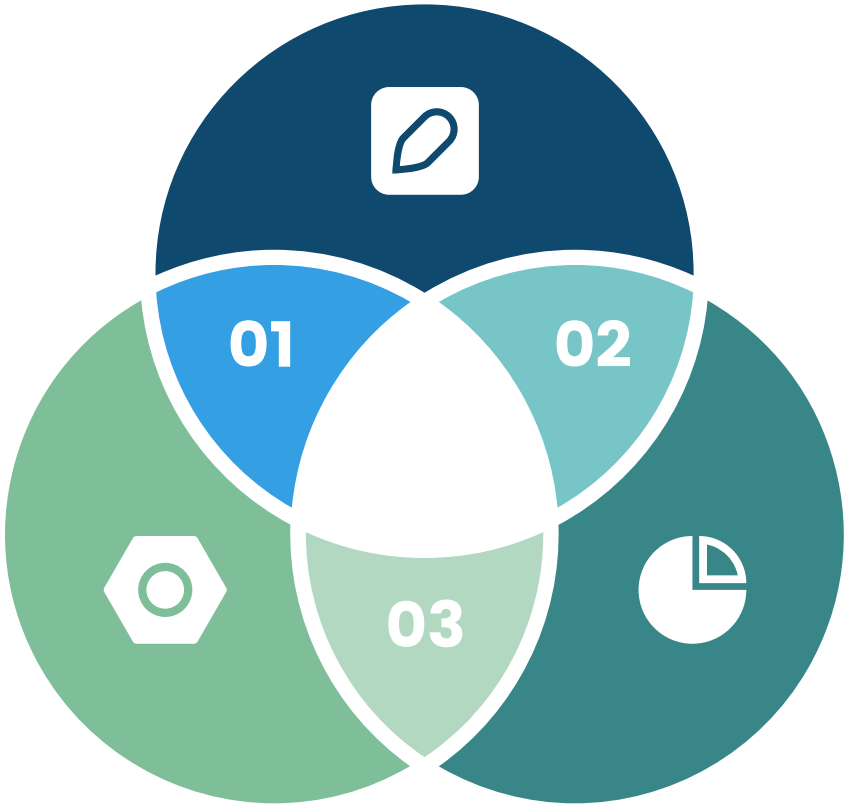
Load Balancing

Implementing load balancing to ensure reliable and efficient data processing

Reporting

Generating Reports

Generating detailed reports on pipeline performance and analytics – both operational and business-related



Data Visualization

Leveraging visualization tools to present data insights in order to tell a compelling story

Stakeholder Communication

Effective communication of pipeline performance and analytics to stakeholders

06

Data Security and Compliance

Regulatory Compliance

01 GDPR

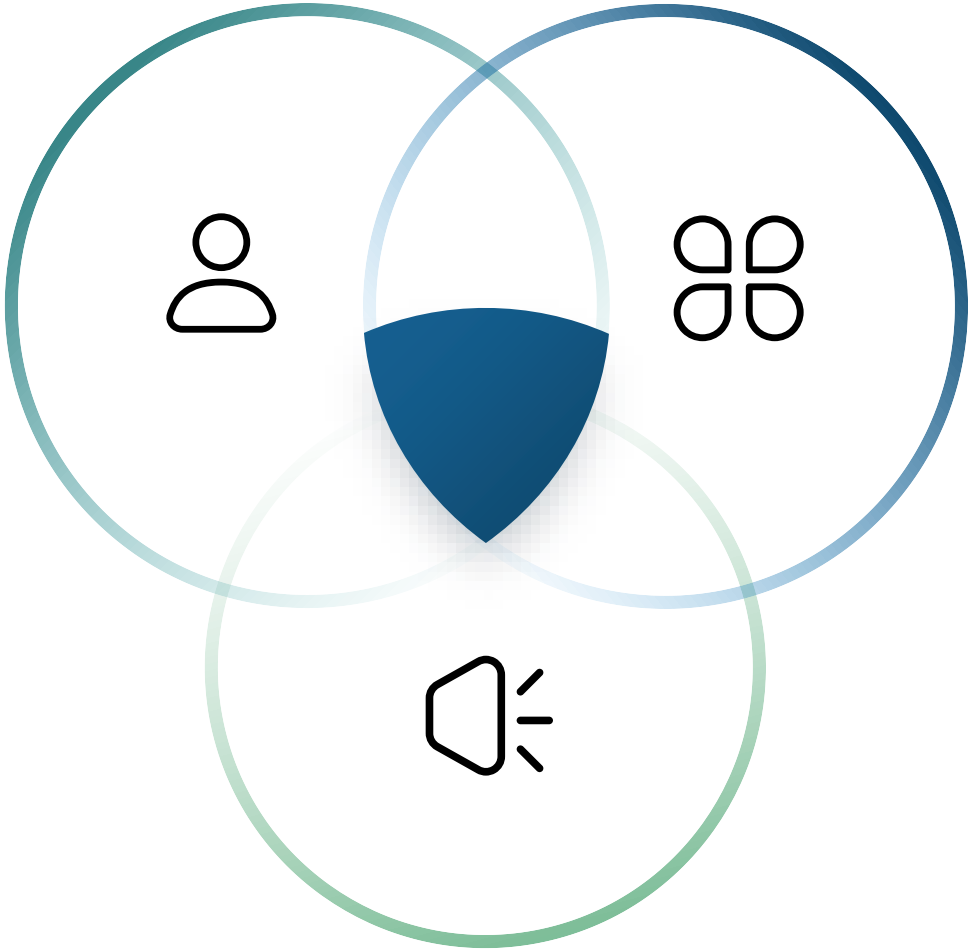
Ensuring compliance with General Data Protection Regulation (GDPR) for data handling

02 HIPAA

Complying with Health Insurance Portability and Accountability Act (HIPAA) standards

03 CCPA

Adhering to California Consumer Privacy Act (CCPA) requirements.



Data Security Best Practices

Data Encryption

Recommended best practices for encrypting data to ensure its security

Secure Access Controls

Implementing strict access controls to protect data from unauthorized access

Regular Security Audits

Conducting periodic security audits to identify and address vulnerabilities

Incident Response

Identifying Breaches

Processes for identifying data breaches quickly

Mitigating Impact

Steps to mitigate the impact of a data breach

Recovery Plan

Developing a comprehensive recovery plan to restore data and operations


Thanks