

## OPTIMIZING DATA PIPELINES IN AZURE SYNAPSE: BEST PRACTICES FOR PERFORMANCE AND SCALABILITY

Afroz Shaik<sup>1</sup>, Ashvini Byri<sup>2</sup>, Sivaprasad Nadukuru<sup>3</sup>, Om Goel<sup>4</sup>, Niharika Singh<sup>5</sup> & Prof.(Dr.) Arpit Jain<sup>6</sup>

<sup>1</sup>Cleveland State University, Cleveland OH, USA

<sup>2</sup>Scholar, University of Southern California, Parel, Mumbai, India

<sup>3</sup>Andhra University, Muniswara Layout, Attur, Yelahanka, Bangalore, India

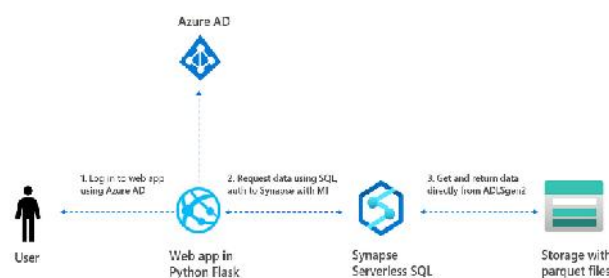
<sup>4</sup>ABES Engineering College Ghaziabad, India

<sup>5</sup>ABES Engineering College Ghaziabad, India

<sup>6</sup>KL University, Vijaywada, Andhra Pradesh, India

### ABSTRACT

The rapid evolution of big data analytics necessitates efficient and scalable data pipelines to handle large volumes of structured and unstructured data. Azure Synapse Analytics, Microsoft's unified data integration platform, offers robust capabilities for building, managing, and optimizing data pipelines. This paper explores the best practices for optimizing data pipelines in Azure Synapse to achieve high performance and scalability. It focuses on key areas such as workload management, data partitioning, parallelism, and query optimization techniques. Additionally, it discusses the integration of external data sources, the use of serverless SQL pools, and the role of Azure Data Lake for cost-effective data storage. Emphasis is placed on monitoring and troubleshooting pipelines using Azure Synapse Studio and applying automation through pipelines to reduce latency. By adopting these best practices, organizations can improve query response times, enhance resource utilization, and ensure seamless data processing at scale, ultimately driving better business insights and operational efficiency.



**KEYWORDS:** Azure Synapse Analytics, Data Pipelines, Performance Optimization, Scalability, Workload Management, Data Partitioning, Query Optimization, Parallel Processing, Serverless SQL Pools, Azure Data Lake, Automation, Monitoring, Troubleshooting, Resource Utilization, Big Data Analytics

### Article History

Received: 23 Jul 2021 | Revised: 26 Jul 2021 | Accepted: 29 Jul 2021