CHRIST
(DEEMED TO BE UNIVERSITY)
BANGALORE | DELHI NCR | PUNE

Notice for the PhD Viva Voce Examination

Mr Mohammed Irfan, Registration Number: 2170231, PhD Scholar at the Department of Computer Science, School of Sciences, CHRIST (Deemed to be University) will defend his PhD thesis at the public viva-voce examination on Monday, 09 March 2026 at 11.30 am in Seminar Hall (Room No. 2208), Management Block, CHRIST (Deemed to be University), Pune Lavasa Off-Campus, Pune - 412112, Maharashtra, India.

Title of the Thesis	:	**An Efficient Generic Ingestion and Analytics Architecture for Big Data**
Discipline	:	**Computer Science**
External Examiner - I	:	**Dr Nilanjan Dey** Professor Department of Computer Science and Engineering Techno International New Town Kolkata - 700156 West Bengal
External Examiner - II	:	**Dr Atul Prabhakar Kulkarni** Professor and Dean Department of Mechanical Engineering Vishwakarma Institute of Information Technology Pune - 411048 Maharashtra
Supervisor	:	**Dr Jossy P George** Professor Department of Computer Science School of Sciences CHRIST (Deemed to be University) Bengaluru - 560029 Karnataka

The members of the Research Advisory Committee of the Scholar, the faculty members of the Department and the School, interested experts and research scholars of all the branches of research are cordially invited to attend this open viva–voce examination.

Registrar (Academics)

Place: Bengaluru
Date: 24 February 2026

ABSTRACT

The exponential proliferation of data across heterogeneous digital ecosystems has underscored the necessity for robust, adaptive, and validation-aware ingestion architectures. Data ingestion plays a pivotal role in data analysis, serving as the foundation for unifying data from diverse, heterogeneous sources into a unified system. This process ensures that data is collected, transformed, and stored in the desired format, creating data lakes. Apache NiFi, an open-source data ingestion framework, automates data flow between disparate systems, ensuring efficiency, scalability, and fault tolerance. This research addresses a critical gap in big-data engineering, that is, the absence of a generic, scalable ingestion framework capable of sustaining operational consistency across local and cloud environments. The study employs a comparative methodology across on-premises, Google Cloud Platform (GCP), and Amazon Web Services (AWS) environments to analyse ingestion throughput, latency, and elasticity under varying workloads. Two Custom validation processors are developed, namely, File Size Validation and Ingestion Frequency Validation, to introduce self-regulating intelligence within the ingestion pipeline, ensuring data correctness without performance degradation. Experimental outcomes demonstrate that NiFi achieves near-linear scalability, with minimal latency differentials (<10%) between local and cloud executions, thereby disproving long-held beliefs that validation or cloud orchestration inherently reduce performance. The study establishes that efficiency in data ingestion extends beyond raw throughput to encompass data quality, resilience, and deployment neutrality. The findings position NiFi as an adaptable ingestion solution suited to hybrid data ecosystems. This work contributes to both theory and practice, providing a reproducible blueprint for future research in scalable, validation-driven ingestion frameworks that bridge local–cloud interoperability and support sustainable data-driven innovation.

Keywords: *Data lakes, Data ingestion, Apache NiFi, Cloud Computing*

Publications:

1. **Irfan, M.**, George, J.P. (2022). A Systematic Review of Challenges, Tools, and Myths of Big Data Ingestion. In: Data Science and Security. Lecture Notes in Networks and Systems, vol 462. Springer, Singapore. https://doi.org/10.1007/978-981-19-2211-4_43
2. **M. Irfan**, Reena and J. George, Data Ingestion - Cloud based Ingestion Analysis using NiFi, 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India, 2023, pp. 1-9, doi: 10.1109/ICSSAS57918.2023.10331884.
3. **Irfan, M.**, Gangadhar, A., George, J. (2024). File Validation in the Data Ingestion Process Using Apache NiFi. In: Data Science and Security. IDSCS 2023. Lecture Notes in Networks and Systems, vol 922. Springer, Singapore. https://doi.org/10.1007/978-981-97-0975-5_27