AN INDUSTRY INTERNSHIP REPORT ON

Data Validation and Authorization: Ensuring Quality Assurance

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

BATCHELOR OF TECHNOLOGY IN

COMPUTER ENGINEERING

BY

Aryan Shinde (12010823)

UNDER THE MENTORSHIP OF

Prof. Shailaja Uke

VISHWAKARMA
INSTITUTES

DEPARTMENT OF COMPUTER ENGINEERING

BANSILAL RAMNATH AGARWAL CHARITABLE TRUST'S

**VISHWAKARMA INSTITUTE OF TECHNOLOGY,
PUNE – 411037.**

(An Autonomous Institute affiliated to Savitribai Phule Pune University)

DEPARTMENT OF COMPUTER ENGINEERING

BANSILAL RAMNATH AGARWAL CHARITABLE TRUST'S

VISHWAKARMA INSTITUTE OF TECHNOLOGY, PUNE – 411037.

(An Autonomous Institute affiliated to Savitribai Phule Pune University)

<u>CERTIFICATE</u>

This is to certify that the industry internship report entitled **Data Validation and Authorization: Ensuring Quality Assurance** submitted by **Aryan Shinde (12010823)** is approved for partial fulfillment of the requirements for the award of degree of Batchelor of Technology in Computer Engineering of Vishwakarma Institute of Technology, Savitribai Phule Pune University. This report is a record of bonafide work carried out as a part of his/her internship in MSCI Inc. during the academic year 2023–24, Semester – 7 & 8.

Internal College Mentor

Shailja Uke

External Examiner   Head of Department

                                     Prof . (Dr.) Sandeep
Date: 17-05-2024   Shinde

ACKNOWLEDGEMENT

 I extend my sincere appreciation to Vishwakarma Institute of Technology, Pune for affording me the invaluable opportunity to embark on a transformative internship at MSCI Inc. This experience has been a cornerstone in my professional development, allowing me to bridge the gap between theoretical knowledge and practical application. The unwavering support and encouragement from Vishwakarma Institute of Technology, Pune have created an environment where learning transcended the confines of the classroom, and academic insights were seamlessly integrated into real-world scenarios.

I would also like to express profound gratitude to MSCI for providing a conducive and dynamic work environment that has been instrumental in my growth. The wealth of diverse projects and exposure to cutting-edge technologies has been an enriching journey, offering insights beyond the realms of conventional academic learning. The mentorship extended by seasoned professionals within the company has been a beacon, guiding me through the intricacies of real-world challenges and providing practical wisdom that goes beyond the scope of textbooks.

The trust and confidence vested in me by MSCI Inc. have been pivotal in fostering my professional growth. The company's commitment to nurturing talent and providing hands-on experiences has not only enhanced my technical skills but also instilled a profound understanding of the collaborative and dynamic nature of the industry.

This acknowledgment underscores the symbiotic relationship between academic institutions and industry partners. It highlights the seamless integration of theoretical knowledge and practical application, enriching the learning experiences for students. The opportunities provided by MSCI Inc. and the mentorship received at MSCI have collectively shaped my perspective, and I am eager to carry the acquired knowledge and skills into my future endeavours. I am deeply grateful for the collaborative efforts that have facilitated this profound learning journey, and I look forward to applying these lessons in my ongoing academic and professional pursuits.

CONTENTS

Abstract

During my internship at MSCI, I immersed myself in diverse projects, engaging with technologies such as Spring, Spring Boot, Spring Batch, SQL, and Linux commands. This experience proved transformative, enhancing my technical proficiency and providing insights into project intricacies.

My responsibilities ranged from orchestrating complex server deployments to executing seamless project migrations. Leading Proof of Concepts (PoCs) allowed me to explore the integration feasibility of emerging frameworks, contributing to a deeper understanding of cutting-edge technologies.

A pivotal contribution was the creation of comprehensive documentation, serving as a valuable knowledge repository for incoming team members. This documentation aimed to streamline onboarding by offering insights into project structures, employed technologies, and established best practices within the company.

An achievement during this period was progressing through the preliminary round of the MSCI Hackathon, fostering collaboration with professionals from diverse offices. Complementing this, I authored a blog exploring Spring Batch 5 configuration intricacies, contributing to a culture of knowledge-sharing within the industry. [4]

In retrospect, my internship at MSCI was a valuable journey that extended beyond technical skills, providing a holistic understanding of project management and team dynamics. As I transition into the next phase of my career, the experience gained will shape my approach to challenges, emphasizing continuous learning, adaptability, and effective collaboration.

About Company

MSCI Inc. stands as a global trailblazer in the financial services sector, renowned for its commitment to providing cutting-edge analytics and innovative solutions. With a profound legacy, MSCI has become synonymous with empowering clients with actionable intelligence that navigates the complexities of global financial markets. At the heart of MSCI's comprehensive offerings is a suite of advanced tools spanning various sectors. From pioneering analytics and climate solutions to leading the charge in Environmental, Social, and Governance (ESG) investing, MSCI is at the forefront of shaping the future of sustainable finance. The company's dedication to Net-Zero initiatives underscores its commitment to addressing the pressing challenges of climate change through impactful financial strategies.

MSCI's influence extends to Factor Investing, where the company continues to innovate and redefine investment strategies. The creation of diverse indexes, including widely used benchmarks like MSCI World and MSCI Emerging Markets, has positioned MSCI as a global leader in the index provider space. These indices play a pivotal role in guiding investment decisions for professionals and institutions worldwide.

Cutting-Edge Analytics:
MSCI's suite of advanced analytics empowers clients with actionable intelligence, providing a comprehensive understanding of global financial markets.
The company's analytics solutions enable investors to make informed decisions in a rapidly evolving financial landscape.

Sustainable Finance Leadership:
MSCI leads the industry in climate considerations, Net-Zero initiatives, and Environmental, Social, and Governance (ESG) investing.
The company's focus on sustainable finance underscores its dedication to responsible investing practices.

Innovative Investment Strategies:
MSCI is a trailblazer in reshaping investment approaches through innovative strategies, notably in the domain of Factor Investing.

The company's forward-thinking investment methodologies contribute to the evolution of global investment practices.

Premier Index Provider:

Serving as a premier index provider, MSCI's benchmarks, including globally recognized ones like MSCI World and MSCI Emerging Markets, guide investment decisions worldwide.
MSCI's indices are instrumental in benchmarking and assessing the performance of investment portfolios.

Insights into Real Assets:

MSCI's influence extends to Real Assets, offering comprehensive solutions for navigating investments in physical assets such as real estate and infrastructure.
The company's expertise in Real Assets contributes to the diversification and resilience of investment portfolios.
Corporate Culture:

MSCI fosters a corporate culture marked by collaboration, innovation, and a relentless commitment to excellence.
The dynamic and inclusive work environment encourages professionals to thrive, fostering creativity and continuous improvement.

Introduction

Embarking on my internship journey at MSCI, I found myself stepping into a dynamic realm teeming with possibilities. This report encapsulates the intricate narrative of my immersive experience, unfolding a tapestry woven from enthusiastic engagement with diverse projects. Against the backdrop of key technologies—Spring, Spring Boot, Spring Batch, SQL, and Linux commands—each project became a canvas for exploration and growth.

From the outset, the dynamic environment at MSCI presented a myriad of opportunities to delve into the intricacies of cutting-edge technologies. The vibrant atmosphere, coupled with the complexity of projects, allowed for a profound exploration of the potentialities within the technological landscape. The narrative of my internship is a testament to the depth of my involvement, the challenges faced, and the invaluable lessons learned.

The projects became a canvas for hands-on learning, where Spring technologies emerged as the brushstrokes shaping my technical proficiency. Spring, with its versatile ecosystem, offered a robust framework for the development of scalable and efficient applications. Spring Boot streamlined the process, simplifying the configuration and deployment of applications. Concurrently, Spring Batch facilitated the seamless execution of batch processing tasks, adding another layer of versatility to my skill set.

I was tasked with the significant responsibility of platform unification between two teams, focusing on the removal of ETL (Extract, Transform, Load) processes and Spring Batch, replacing them with Spring Boot to streamline and modernize our workflow. The primary objective was to enhance the efficiency, maintainability, and scalability of our data processing operations. This project required an in-depth understanding of the existing ETL and Spring Batch frameworks, their integration points, and the potential challenges in transitioning to a new architecture. I began by conducting a thorough analysis of the current system, identifying the key processes and data flows that were handled by ETL and Spring Batch. This analysis involved reviewing extensive documentation, engaging in detailed discussions with team members, and examining the underlying SQL scripts and Linux-based deployment scripts.

I was involved in a critical project aimed at the platform unification of two separate teams, with a focus on removing ETL (Extract, Transform, Load) processes and Spring Batch, and transitioning to Spring Boot. This initiative aimed to modernize and streamline data processing workflows, enhancing the system's efficiency, maintainability, and scalability. The technologies I utilized included Spring Boot, SQL, and Linux.

The project began with an in-depth analysis of the existing ETL and Spring Batch frameworks. This analysis was crucial to understand the existing data flows and processing logic deeply. I examined extensive documentation, engaged in detailed discussions with team members, and reviewed the SQL scripts and Linux-based deployment procedures. This comprehensive understanding helped in identifying the key components and processes that required migration.

Based on this analysis, I designed a solution architecture using Spring Boot. The choice of Spring Boot was due to its capability to create stand-alone, production-grade Spring applications with minimal configuration. Its embedded server, extensive ecosystem, and ease of deployment made it an ideal fit for our needs. I initiated the implementation by setting up a new Spring Boot project, adhering to best practices for project structure and ensuring a clear separation of concerns.

One of the primary tasks was to rebuild the data ingestion pipelines that were previously managed by ETL processes. Utilizing JDBC, I created repository interfaces for database interactions, encapsulating the data ingestion logic within well-defined service layers. These service layers managed data transformation, validation, and interaction with SQL databases, ensuring modularity and testability.

This detailed narrative seeks to convey not just a chronicle of tasks undertaken but a journey of profound exploration and growth. The projects undertaken at MSCI, set against the canvas of key technologies, have sculpted a professional narrative rich in experiences and insights. As I reflect on this transformative journey, I am equipped with a nuanced understanding of the dynamic interplay between technology and innovation, paving the way for continued growth and contribution in the ever-evolving landscape of IT.

Literature Review

In the realm of software development, project documentation serves as a vital tool for developers to comprehend both the code and the intricate business logic embedded within a project. The literature emphasizes that structured documentation, coupled with visual aids like flow diagrams, significantly contributes to a comprehensive understanding of projects. Using The technology and reading its documentation helped in implementation and its rules. Below listed are the technologies used throughout my internship.

Spring Framework: Elevating Java Development

The Spring Framework stands as a cornerstone in modern Java development, offering a comprehensive solution to the challenges posed by traditional Java EE development. Spring introduces a unique approach to inversion of control (IoC) for dependency injection. This not only provides a modular and loosely coupled architecture but also allows developers to concentrate on business logic, promoting clean and efficient code. Over the years, Spring has evolved into a versatile and widely adopted framework, enriching the Java ecosystem with its robust feature set.
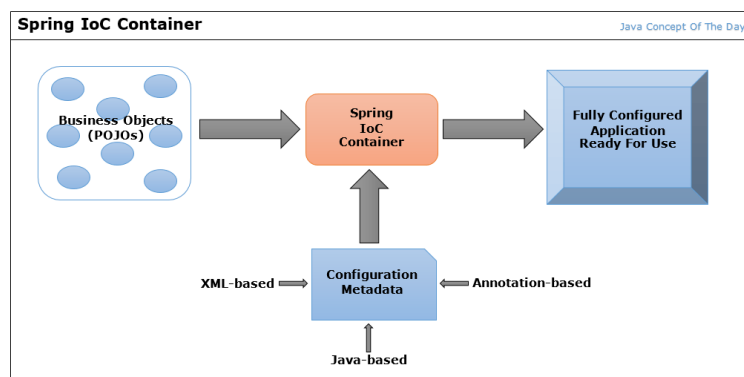

Figure 1: Spring IoC Container [5]

Spring Boot: Streamlining Application Development

Building upon the success of the Spring Framework, Spring Boot takes a revolutionary approach to Java application development. By embracing a convention-over-configuration paradigm, Spring Boot significantly reduces the boilerplate code traditionally associated with Java applications. It simplifies application configuration and setup, offering built-in features and an embedded server approach. This convention-driven framework expedites project initiation, enabling developers to create production-ready applications with minimal effort. The seamless integration of Spring Boot with the broader Spring ecosystem contributes to a more agile and efficient development process.
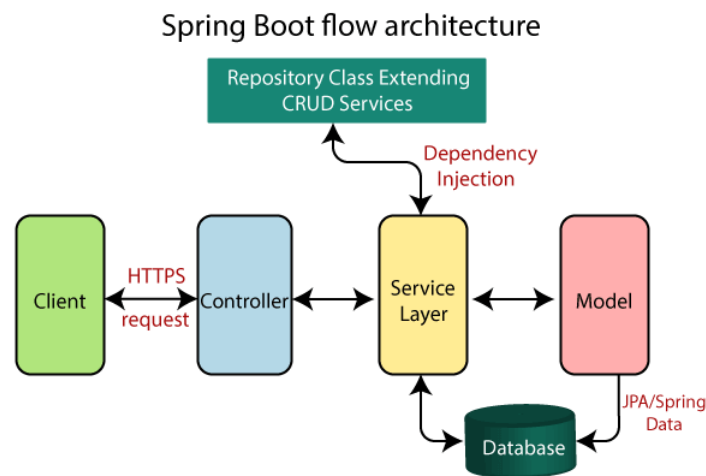

Figure 2: Spring Boot Architecture [6]

Spring Batch: Empowering Data Processing

In the domain of data processing, Spring Batch emerges as a powerhouse for batch applications. Addressing the unique challenges of large-scale data processing, Spring Batch simplifies development through reusable components and a declarative configuration approach. This framework excels in managing complex data processing tasks, offering features like parallel processing, transaction management, and robust error handling. With its modular architecture, Spring Batch facilitates the creation of scalable and maintainable batch applications, making it a preferred choice for enterprises dealing with data-intensive processes. Its integration with the

12

broader Spring family reinforces a cohesive and efficient approach to handling diverse data processing requirements.
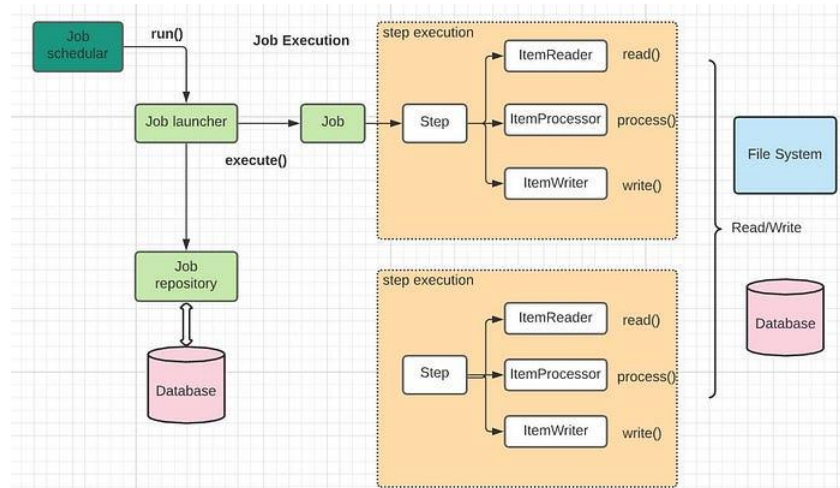

Figure 3: Spring Batch Flow [7]

Structured documentation provides a systematic and organized approach to presenting project information. The inclusion of flow diagrams is highlighted to enhance the clarity of project structures, offering a visual roadmap for developers to navigate through complex systems.

Moreover, effective project documentation should not solely focus on technical details but also encapsulate underlying business logic. This dual-layered approach, encompassing both technical and business processes, creates a holistic resource that empowers developers to understand the code implementation in the broader context of business objectives.

In our project documentation practices, integrating these insights is crucial. Providing structured documentation with flow diagrams facilitates a visual representation of project architecture, aiding developers in navigating and comprehending the codebase more efficiently. Simultaneously, maintaining documentation that encapsulates business logic aligns with best practices, ensuring that developers can make informed decisions based on a comprehensive understanding of both technical and business aspects.

13

The literature consistently underscores that well-documented projects enhance collaboration among team members, reduce onboarding time for new developers, and contribute to the overall maintainability of the codebase. Going forward, our project documentation practices will continue to align with these scholarly insights, ensuring that our documentation remains an asset for developers seeking to comprehend the intricate interplay between business requirements and code logic.

Problem Statement

In the dynamic landscape of our project, a significant challenge revolves around the effective handling of data, encompassing both ensuring its quality and managing the substantial load during database uploads. The first facet of this challenge lies in the need for a robust solution to acquire data while concurrently applying stringent quality checks. The objective is to guarantee that the data obtained is reliable, accurate, and well-structured, laying a solid foundation for downstream processes. This necessitates the implementation of systematic quality assurance measures to identify and rectify anomalies, ensuring that only high-quality data proceeds through the pipeline.

The second dimension of the problem centers on the efficient handling of substantial data uploads into the database. Current practices involve encountering bottlenecks when dealing with large volumes of data, hindering the seamless flow of operations. To address this, a solution is sought to optimize the process, enabling the system to gracefully manage and process a significant load of data at a single time.

In response to these challenges, our approach involves a comprehensive strategy that encompasses both unit and integration testing. This ensures that not only individual components but the entire system functions seamlessly, providing a holistic solution to the issues at hand. Through meticulous testing practices, we aim to fortify our data acquisition and processing mechanisms, fostering a more resilient and efficient data management system.

14

Scope and Objectives

My internship unfolded as an immersive journey into the heart of Spring technologies, revealing an expansive scope where my objectives intricately intertwined with ongoing projects. At its core, my role aimed to fortify these projects through the implementation of a meticulous testing regime, optimization of deployment processes, and the development of exhaustive documentation.

A pivotal facet of my responsibilities was contributing to the seamless assimilation of new team members. I played a crucial role in providing comprehensive insights into ongoing projects, acting as a bridge for newcomers to navigate the project landscape. This involved creating detailed documentation that not only served as a reference guide but also facilitated a smoother onboarding process. By ensuring that new team members had a profound understanding of project structures, technologies in use, and established best practices, I played a key role in fostering a collaborative and efficient work environment.

In essence, my internship experience was not merely about fulfilling tasks but about actively contributing to the cohesion and success of ongoing projects. It was a journey into the intricate dynamics of project management, testing methodologies, and knowledge transfer within the team.

Proposed Methodology

The proposed methodology for my internship was a carefully crafted approach aimed at navigating the intricate landscape of projects, with a particular emphasis on Spring technologies. This methodology, serving as the guiding framework for my tasks, was designed to ensure project robustness, efficient deployment, and seamless onboarding for new team members.

At the outset, a comprehensive understanding of the business requirements formed the bedrock of my approach. This involved close collaboration with project stakeholders to grasp the nuances of their needs and expectations. By immersing

myself in the intricacies of these requirements, I could align technological solutions with the broader objectives of the projects.

Understanding Business Requirements:
- Initiated collaborative sessions with project stakeholders to foster a deep understanding of business needs and expectations. This involved conducting detailed interviews and workshops to extract nuanced requirements.
- Employed advanced analytical tools to dissect intricate requirements, ensuring a comprehensive understanding. This process involved iterative feedback loops to align technological solutions seamlessly with overarching project objectives.

Utilization of Spring Technologies:
- Capitalized on the versatility of Spring, a renowned framework for building enterprise-level applications, leveraging its robust features to fortify the core of numerous projects. This included harnessing Spring's dependency injection for enhanced modularity.
- Integrated Spring Boot, a powerful extension of the Spring framework, to streamline configuration and enhance development efficiency. This involved customizing auto-configurations and utilizing Spring Boot Starters for rapid project setup, simplifying and expediting the deployment process.

Thorough Testing Approach:
- Implemented a comprehensive and systematic unit testing strategy using JUnit, ensuring the thorough validation of individual components in isolation. This included developing parameterized tests and mock objects for comprehensive coverage.
- Executed rigorous integration testing protocols to scrutinize the seamless collaboration between diverse modules. Utilized tools like Mockito for mock testing and designed test suites for end-to-end validation, validating the system's overall harmony and reliability.
- Conducted formal black box testing by assessing projects using different libraries, covering unit tests as well as integration tests. This involved treating the system as a "black box" and systematically evaluating its outputs against expected results, ensuring a robust and unbiased assessment.
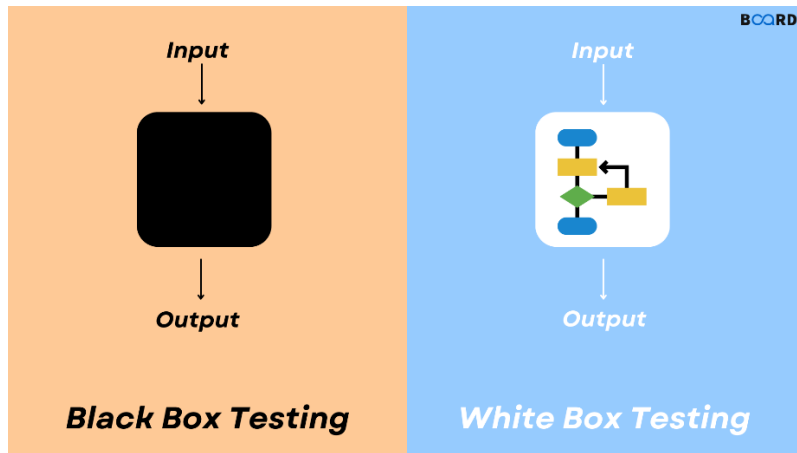
Figure 4: Black Box & White Box Testing [8]

Proof of Concepts (PoCs):
- Orchestrated proactive exploration of emerging frameworks through meticulously designed PoCs, systematically evaluating their integration potential into existing projects. This process included benchmarking performance metrics and assessing scalability.
- Assessed the benefits and challenges posed by new frameworks, providing a robust foundation for informed decision-making regarding their adoption. PoCs were designed to simulate real-world scenarios, ensuring a thorough evaluation of their viability.

Comprehensive Documentation:
- Crafted strategic and exhaustive documentation that encapsulated project architectures, design decisions, and coding standards. This was not a one-time task but an ongoing process, involving regular updates and version control.
- Positioned documentation as a proactive tool for knowledge transfer, expediting the onboarding process for new team members. Documentation extended beyond technical details to include project narratives, providing a holistic understanding.

Efficient Deployment Processes:
- Pioneered efficient deployment processes by leveraging cutting-edge tools and best practices. This involved utilizing containerization with Docker for consistency across development and production environments.
- Ensured a seamless transition from development to production, incorporating a holistic approach that considered not only the technical aspects but also

scalability, performance, and monitoring considerations. Automated deployment pipelines were implemented for efficiency.

Comprehensive Project Sustainability:
- Embraced a holistic approach that extended beyond immediate project requirements, contributing to the enduring success and sustainability of projects. This involved conducting regular project retrospectives for continuous improvement.
- Ensured that the methodology reflected a commitment to collaboration, strategic use of cutting-edge technologies, rigorous testing practices, and a dedication to knowledge transfer through comprehensive documentation. Continuous monitoring and feedback loops were integrated for adaptive project management.

Dedication to Continuous Improvement:
- Ensured that the proposed methodology fostered an environment conducive to continuous improvement and innovation within the project lifecycle. This included implementing feedback loops from end-users and stakeholders.
- Positioned the methodology as a dynamic framework, reflecting a commitment to delivering high-quality solutions and nurturing an atmosphere of continuous improvement. Regular skill enhancement sessions and participation in industry forums were integrated into the methodology for ongoing professional development.

Outcome

The culmination of my internship manifested in a rich tapestry of outcomes, reflecting a multifaceted growth across various dimensions. The practical application of technologies such as Spring and SQL significantly bolstered my problem-solving acumen. Engaging in complex server deployment and migration projects provided profound insights into the intricate lifecycles of projects, enhancing my project management skills and strategic thinking.

The emphasis on robust testing practices, spanning both unit and integration levels, significantly augmented my proficiency in quality assurance methodologies. Rigorous testing not only ensured the reliability of the solutions implemented but also honed my ability to identify and rectify potential issues, fostering a culture of precision and excellence.

Exploring new frameworks through Proof of Concepts (PoCs) expanded my technological prowess, providing me with a nuanced understanding of emerging technologies and their practical applications. This exploration not only enriched my skill set but also positioned me as an adaptive and forward-thinking contributor to the technological landscape.

Furthermore, the documentation efforts undertaken during the internship not only served the immediate purpose of aiding project understanding but also contributed to the long-term knowledge repository of the team. This legacy of comprehensive documentation ensures the sustainability and continued efficiency of ongoing projects, laying the groundwork for seamless collaboration among team members both present and future. In essence, the outcomes of my internship extend beyond immediate project successes to encompass a lasting impact on the team's collective knowledge base and my individual professional development.

Conclusion and Future Scope

In conclusion, my internship at MSCI has been a transformative experience, transcending the realms of theoretical knowledge into practical application and professional growth. As I reflect on the multifaceted challenges and achievements, it's evident that this period has been instrumental in shaping my perspective as an emerging professional in the field of information technology.

The immersion into Spring technologies, alongside other key components like SQL and Linux commands, has not only deepened my technical acumen but has provided

a comprehensive understanding of their synergies within complex projects. Navigating through intricate server deployments and executing seamless project migrations underscored the practical intricacies of project lifecycles, enhancing my problem-solving skills and strategic thinking.

Beyond technical proficiency, the role of contributing to a meticulous testing regime, optimizing deployment processes, and creating exhaustive documentation has underscored the holistic nature of effective project management. The responsibility of ensuring a seamless onboarding process for new team members involved more than just knowledge transfer; it encapsulated the essence of creating a collaborative work culture where insights were shared transparently, and best practices were ingrained from the beginning.

Looking forward, the future scope expands into a dynamic horizon of continuous learning and evolving expertise. The experiences garnered during this internship serve as a solid foundation for navigating the ever-changing landscape of technology. The emphasis on implementing cutting-edge technologies, coupled with robust testing methodologies, positions me favourably for taking on more complex challenges in the realm of software development.

The journey doesn't end with the conclusion of the internship; instead, it extends into a proactive pursuit of emerging technologies and best practices. The commitment to knowledge-sharing, as exemplified through the participation in the MSCI Hackathon and the creation of a comprehensive blog,[4] is a testament to my dedication to contribute to the broader tech community. This ethos is integral to my future professional endeavours, where I envision not just keeping pace with industry trends but actively shaping them through innovation and collaboration.

In essence, my internship at MSCI has been a stepping stone, propelling me towards a future marked by continuous growth, adaptability, and a commitment to contribute meaningfully to the dynamic field of information technology. The lessons learned, the challenges overcome, and the achievements unlocked during this period have set the stage for a promising trajectory as I embark on the next chapter of my professional journey.

References

[1] - Spring Documentation - https://docs.spring.io/spring-framework/reference/index.html

[2] - Spring Boot Documentation - https://docs.spring.io/spring-boot/docs/current/reference/htmlsingle/

[3] - Spring Batch Documentation - https://spring.io/projects/spring-batch/

[4] - Configuration of Spring Batch with Spring boot - https://medium.com/@aryan.shinde.29/configuring-dual-data-sources-in-spring-batch-5-and-spring-boot-3-8a72bc00555c

[5] - Java Concept of the Day - https://javaconceptoftheday.com/spring-ioc-container/

[6] - JavaTpoint - https://www.javatpoint.com/spring-boot-architecture

[7] - Spring Batch Processing - https://shradhayewale.medium.com/spring-batch-processing-overview-371f1d9ee536

[8] - Board Infinity - https://www.boardinfinity.com/blog/white-box-vs-black-box/

Internship Achievements

Within the realm of achievements during my internship, a standout milestone was navigating through the preliminary round of the MSCI Hackathon—an intensely competitive event bringing together talented participants from diverse offices. This experience not only validated my problem-solving capabilities but also served as a testament to my adaptability in a fast-paced, collaborative setting. The Hackathon was more than a competition; it was an opportunity to engage with professionals from different locations, fostering meaningful connections and expanding my network within the company.

Additionally, I contributed to the knowledge-sharing ethos of the tech community by authoring a comprehensive blog focused on the intricacies of Spring Batch 5 configuration. The blog was more than a personal accomplishment;[4] it was a

commitment to effective communication and the dissemination of insights within the broader industry. By sharing my experiences and solutions, I aimed to contribute to the collective learning of the community, fostering a culture of collaboration and continuous improvement.

These achievements, whether in the competitive arena of the MSCI Hackathon or the knowledge-sharing realm of blogging, underscored not only my technical proficiency but also my commitment to growth, collaboration, and effective communication within the dynamic landscape of the tech industry.