Tuesday, February 10

1:40 PM – 2:00 PM

Autonomous AI/ML Generated Software Development Test, Valdation and Reporting, Automation script updates & Evaluation Scripts

Mark Wells

Partner, DevSecOps & IT Automation Practice Lead

IBM

Michael Stack

Senior Software Engineer

IBM

<u>Abstract:</u>

To help the Department of Navy with its challenge to provide Artificial Intelligence Augmented Testing (AIAT), IBM's Platform and Product Design and Engineering (PDE) Factory brings AI-driven automation to software testing environments, enabling faster, safer, and more reliable development cycles. PDE Factory is a fully composed DevSecOps platform you can deploy in hours, not months, with a software bill of materials (SBOM) powered supply chain security baked into every layer. At a previous implementation at GSA FAS EOSS, the customer saw a 95% reduction in manual testing infrastructure management, a reduction in test script creation by 85–90%, and a 96-99% compliance with security standards out-of-the-box. Built on proven DevSecOps successes across federal agencies, PDE Factory ensures IL5 security compliance and empowering PEO MLB to modernize Navy software assurance at mission speed.

Architecture Overview:

Our proposed AI environment leverages a comprehensive, cloud-native architecture built upon IBM's PDE Factory framework (TRL 8) designed to seamlessly integrate advanced AI capabilities throughout the entire DevSecOps pipeline. The architecture employs modular, reusable continuous integration (CI) components that ensure scalability, maintainability, and accessibility compliance. The core foundation consists of architectural components that work in concert to deliver intelligent automation across all development and testing activities.

AI Technologies and Models:

The integration of advanced AI technologies forms the cornerstone of our technical approach. IBM utilizes Amazon Bedrock, a comprehensive, secure, and flexible platform for building generative AI applications and agents using IL4/5 compliant language models (such as Anthropic Claude and Meta Llama). The models work in conjunction with our knowledge graph integration system, which provides advanced mapping of components, parameters, and relationships to facilitate complex simulation understanding and management while enhancing decision-making through comprehensive relationship analysis. Department of Navy (DoN) tailored LLMs can also be leveraged that are specifically designed for problem statement formulation and understanding.

Our AI agents generate test scripts by analyzing code patterns, user stories from Jira, manual tests, system models and design requirements (see Figure 2). This engine seamlessly integrates with GitLab CI/CD pipelines enabling continuous test generation and execution throughout the development lifecycle. Other agent functionality includes intelligent patch management and automated test script updates based on code updates, so that developers can make changes while ensuring test quality. This engine also includes a comprehensive technical debt identification that systematically addresses code complexity, outdated dependencies, security vulnerabilities, performance bottlenecks, and architectural inconsistencies. The result is that systems are reliable and can be maintained in the long-term.

Our intelligent defect detection agent leverages AI Code Assistants, SonarQube integration, and specialized scanning tools to automatically detect code smells, logic errors, memory leaks, and compliance violations in real-time as developers commit code to GitLab repositories. The agent then reports on impact, probability and confidence scores so that developers can prioritize major fixes.

Finally, IBM PDE Factory supports human-in-the-loop and human-on-the-loop techniques to support seamless integration between AI-driven recommendations and human review. AI build optimizers or dependency updaters can be configured to approve changes or optimizations automatically, but humans typically review change summaries before explicit approval. This kind of integrated quality gating can occur at all steps of the SDLC to include code building, linting, testing, and scanning.

Tools and Technology Stack:

Our comprehensive technology stack integrates industry-leading tools with IBM's cutting-edge AI capabilities to deliver seamless DevSecOps automation. PDE Factory leverages existing DoD DevSecOps investments, requiring no proprietary infrastructure. The system is designed for modular deployment, enabling incremental adoption without full-stack replacement. GitLab CI/CD serves as the primary pipeline orchestration and version control platform, working in conjunction, for example, with Maven and npm/nx for build automation and dependency management. The testing framework utilizes JUnit for unit testing while supporting multi-framework automated testing through Cypress, Tricentis, and Insomnia to accommodate diverse testing requirements across different application types and technologies. IBM Accessibility Checker ensures Section 508 compliance validation throughout the development process.

Usability Features:

User experience is enhanced through plain language processing capabilities that enable natural language requirement interpretation, allowing business stakeholders to communicate requirements in everyday language while our platform automatically transforms these inputs into actionable development tasks, test scenarios, and validation criteria. The human-centric design approach provides intuitive interfaces for business stakeholders, while Behavior Driven Development practices ensure seamless translation of business requirements into technical specifications.

The PDE Factory integrates visualization tools and analytics platforms such as PowerBI, Qlik, and Tableau, as a part of its self-service. IBM's Historical Logging and Reporting capability provides comprehensive audit trails and analytics through our integrated observability platform within the PDE Factory, automatically capturing and storing detailed logs across the entire DevSecOps pipeline.

Security Measures:

PDE Factory is FedRAMP Ready and has standard policy enforcement compliant with IL5. The IBM Product Engineering Secure Software Supply Chain implements security protocols specifically designed for DoD environments with full network compatibility. The PDE Factory operates natively in AWS GovCloud ensuring compatibility with DoD network requirements and supporting air-gapped configurations using DoD SAFE. Our solution also supports Risk Management Framework compliance and authorization requirements through automated security controls and continuous monitoring capabilities. The security implementation follows CISA's Securing the Software Supply Chain requirements and NIST's Secure Software Development Framework (SP800-218), with security scanning embedded throughout the pipeline.

Integration:

PDE Factory can integrate seamlessly with Navy DevSecOps environments such as DSO or Platform One pipelines through standard CI/CD interfaces and IL5-approved cloud services, ensuring minimal disruption to existing workflows. PDE Factory is designed to handle large-scale problems. As the complexity of deployment increases, the tool can scale to accommodate more variables, higher dimensionality, and larger computational domains. This scalability is crucial for tackling real-world challenges that often involve numerous interacting components and extensive spatial and temporal scales. PDE Factory supports multiple platforms, enabling flexibility.

PDE Factory's AI-Generated Test Script recommendations are automatically surfaced through our Single Pane of Glass dashboards and integrated directly into JIRA user stories, enabling development teams to prioritize test improvements that deliver the highest impact on code quality while reducing the overall testing burden. The solution integrates with test management systems to streamline the testing process. This includes managing different testing scenarios, tracking the results, and comparing them against expected outcomes. Automated testing procedures can also be set up to save time and reduce manual errors.