# DataFlint white Fully SAAS paper

## 1. Executive Summary

Big data processing has become unnecessarily complex. Without deep system expertise, data teams are forced to rely on specialists and generic AI tools that lack critical production infrastructure context. DataFlint solves this with an infrastructure-aware AI Co-Pilot for Apache Spark that understands your production data environment, delivering code suggestions directly into your IDE and runtime.

Security is foundational to this vision. We embed robust security practices into every layer of our product and operations to ensure trust, reliability, and compliance. This white paper outlines our current security architecture, policies, and roadmap as part of our long-term commitment to secure and scalable data infrastructure.

**2. Security Philosophy**

- **Security by Design**: Integrated into every layer of our stack and SDLC.
- **Minimum Viable Compliance**: Practical alignment with SOC 2 and cloud best practices.
- **Developer Enablement**: Secure-by-default tooling, pre-commit checks, and automation to maintain speed.
- **Threat Modeling**: Built into every feature's design phase to proactively mitigate risk.

---

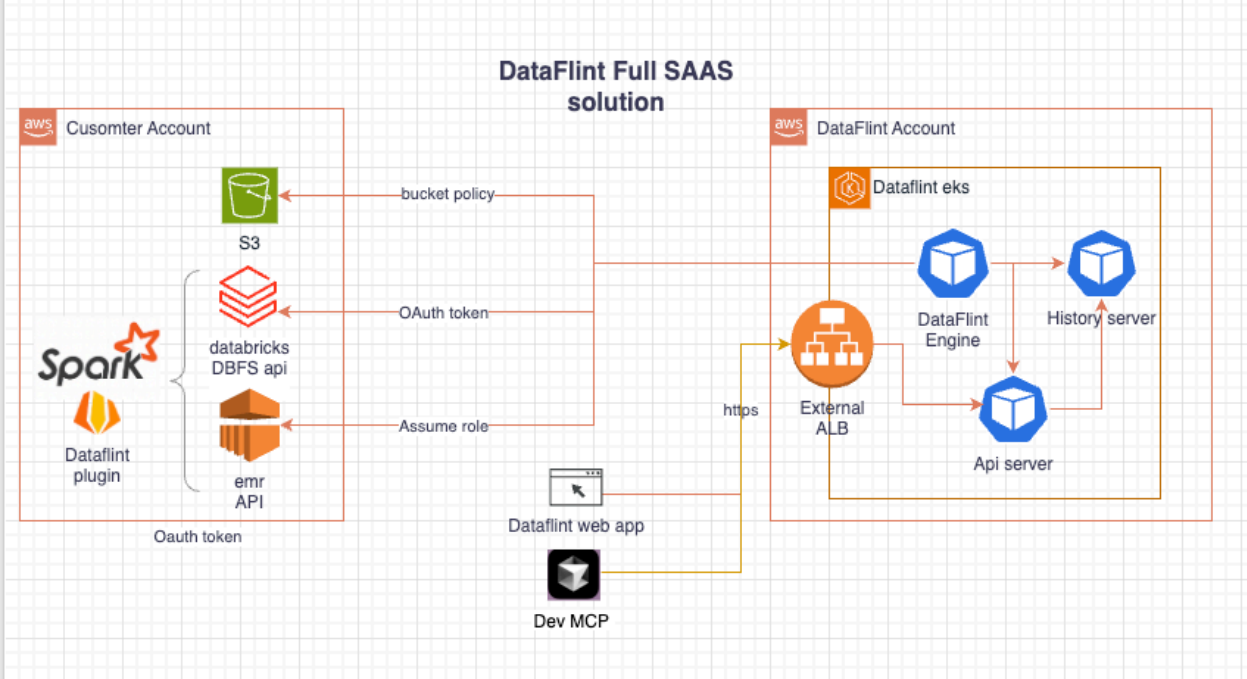
## 3. Infrastructure & Platform Overview

- **Cloud Provider**: Amazon Web Services (AWS)
- **Container Orchestration**: Amazon Elastic Kubernetes Service (EKS)
- **Infrastructure as Code**: Terraform (provisioning), Ansible (configuration)

- **Endpoint Security**: All company devices are macOS-based and managed with Kandji (MDM/EDR)

## 3a. Deployment Model – Fully Managed SaaS

- **Infrastructure**: Hosted entirely within DataFlint's AWS account
- **Data Access**: Via Databricks API, EMR interfaces, or S3
- **DataFlint engine Component**: Runs in DataFlint EKS; ingests and parses customer logs
- **Data Residency**: All processing and storage are confined to DataFlint-managed systems




Deployment model - Full SAAS, DataFlint engine hosted internaly

## 4. Access Control & Identity Management

- **SSO** via Google Workspace for internal tools
- **IAM & RBAC**: Least-privilege roles for all services and personnel

- **Access Reviews**: Conducted quarterly for AWS, GitHub, and Kubernetes
- **MFA** enforced for all critical systems
- **JIT Access (Planned)**: Exploring ephemeral privilege escalation mechanisms

---

## 5. Endpoint Security

- **MDM Enforcement**: Kandji used across all company devices
- **Disk Encryption**: FileVault on all endpoints
- **Patching & Firewall**: OS patching, firewall, and app whitelisting enforced
- **Zero Local Admin Rights**: Enforced for non-DevOps personnel
- **Device Lock Policies** and **Remote Wipe** capabilities

---

## 6. Application Security

- **Code Reviews** required for all pull requests
- **CI/CD Gatekeeping**: Reviews, automated tests, and deploy approvals via GitHub Actions
- **Static & Dependency Scanning**: Automated via CI pipelines
- **Secrets Management**: AWS Secrets Manager, no plaintext secrets in code
- **SBOM Tracking** and **PR-based deployments** with scoped deploy tokens
- **DAST Tools** under evaluation

---

## 7. Kubernetes & Cloud Security

- **CIS Benchmarks** applied to EKS node groups
- **GuardDuty, AWS Config, and CloudTrail** active across all regions
- **Pod Security & Network Policies** enforced
- **VPC Isolation and IAM Segmentation** in both deployment models
- **Customer EKS Clusters** receive the same hardening and audit policies as internal infrastructure

- **Admission Controllers** and signed image policies enforced

---

## 8. Data Protection

- **Encryption**: TLS 1.2+ for transit, AES-256 for data at rest
- **Data Minimization**: Only required metadata is ingested; raw logs are not persisted
- **Customer Data Isolation**: Separate ingestion pods and DB schemas
- **Backups**: Daily backups, monthly restore tests
- **Audit Logging**: CloudTrail logging across all access points
- **Customer-Controlled Keys (Planned)** for enterprise customers

---

## 9. Monitoring & Incident Response

- **Observability Stack**: CloudWatch, Prometheus, Grafana
- **Alerting**: Slack-integrated on-call system
- **Incident Playbooks**: Maintained and regularly drilled
- **MTTD/MTTR Tracking** for all incidents
- **Penetration Testing** performed as part of SOC 2 compliance
- **SIEM Evaluation** in progress

---

## 10. Compliance & Certifications

- **SOC 2 Type I Certified**
- **SOC 2 Type II Targeted for Q4 2025**
- **Vendor Risk Assessments** for all critical services
- **GDPR-Aligned Architecture**: Minimal PII collection
- **Policy Automation** using compliance tooling

---

## 11. Security Training and Awareness

- **Employee Training**: Security onboarding and awareness modules
- **Phishing Simulations**: Quarterly campaigns planned
- **Secure Coding Practices**: Training focused on OWASP Top 10 and supply chain hygiene
- **Customer Education**: Integrated into onboarding and documentation

---

## 12. Conclusion & Contact

At DataFlint, we view security as a continuous investment—foundational to our vision of enabling safe, scalable data infrastructure. Our commitment extends from internal processes to customer education, and we look forward to partnering with our users in creating a secure future.


- Security Lead: **Leon Shklyar**
- Email: **security@dataflint.io**