

Caterpillar IT Architect

Interview Preparation Handbook — Product Support & Logistics
Division

Morton, IL

\$128K-\$193K

On-site

Closes Aug 4, 2026

01 Role Overview

IT Architect — Product Support & Logistics Division (PSLD)

Core Mission

Design, administer, and optimize enterprise data platforms supporting analytics and business operations across PSLD. Lead data architecture and integration efforts ensuring data is securely connected, governed, and accessible across enterprise systems.

Key Responsibilities

- Technical architect and administrator for the PSLD Data Analytics Platform
- Design logical technical architectures for enterprise-wide data integration
- Administer and optimize Snowflake and enterprise database platforms
- Develop secure data integration strategies with governance and cybersecurity
- Partner with PSLD, ICD, and business functions for end-to-end analytics support
- Design integrations between SAP, PRISM, cloud solutions, and analytics platforms
- Administer cloud technologies: Azure, Kubernetes, Key Vault, IoT
- Support source control and deployment via Azure DevOps and GitHub
- Administer Linux and Windows server environments

Required Qualifications

Core Application Systems

Knowledge of major production application systems used for delivery of services to internal/external clients; ability to leverage them in diverse situations.

Application Modernization

Concepts, processes and technologies of application modernization; ability to design, develop and deploy modernized applications.

Architecture Modeling

Develop and modify enterprise architecture through conceptual, logical and physical approaches.

Business Assessment

Assess benefits, risks, and success factors of potential applications; translate business needs into technology solutions.

System Integration

Integrate and communicate among applications, databases and technology platforms.

Top Candidate Differentiators

- **SnowPro certification** or advanced Snowflake administration experience
- Experience with enterprise databases, cloud platforms, and secure data integrations
- Familiarity with **Caterpillar Logical Technical Architecture (LTA)**, SAP, or Dataiku
- Strong business acumen translating needs into scalable solutions
- Broad technical knowledge across analytics, data, and integration platforms
- Critical thinking and problem-solving in ambiguous environments
- Intellectual curiosity, adaptability, continuous learning mindset

Pay Range: \$128,470 – \$192,710 + bonus, 401(k), medical, tuition reimbursement, parental leave. **Must be on-site in Morton, IL — no remote/hybrid.**

02 Company Context

Caterpillar Inc. — Fortune 100, \$67B revenue, 113K+ employees

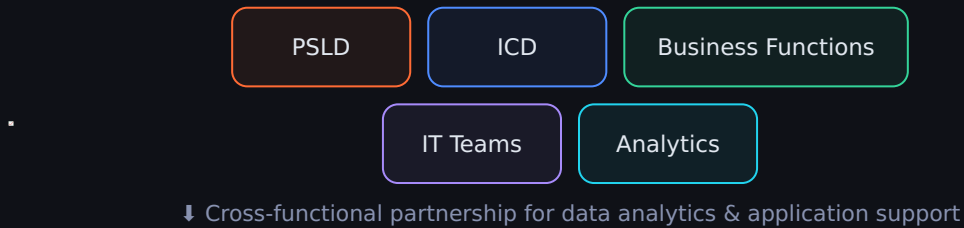
About Caterpillar

- World's leading manufacturer of construction & mining equipment, diesel/natural gas engines, industrial turbines, diesel-electric locomotives
- Nearly 100 years of operation — Fortune 100, S&P 500 component
- Dealer network: 180+ dealers across the globe
- Strong push toward **digital transformation, sustainability, and reduced-carbon future**

Product Support & Logistics Division (PSLD)

- One of the most innovative divisions within Caterpillar
- Responsible for **Aftermarket Parts Distribution, Integrated Logistics, and Global Services**
- Works with Cat dealers and customers to support the entire services value stream
- Leverages digital capabilities, technology, and operational excellence

Organizational Architecture



Why Caterpillar Now

- Major digital transformation initiative — modernizing legacy systems
- Investing heavily in **cloud (Azure)**, **data platforms (Snowflake)**, and **IoT**
- Sustainability push creates new data and analytics requirements
- PSLD is at the center of the aftermarket services digital evolution

03 Key Technologies

Technologies mentioned in the job description and what they mean for the role

Data Platform

Snowflake

Cloud-native data warehouse. Key focus: architecture, administration, optimization, performance tuning. SnowPro certification is a differentiator.

Warehouse

ELT

Admin

Dataiku

Enterprise AI/ML platform. Familiarity is a plus — used for collaborative data science and MLOps workflows.

MLOps

Data Science

Enterprise Databases

Broader database ecosystem: relational, NoSQL, data lakes. Must understand performance, security, and governance.

RDBMS

NoSQL

Governance

Cloud & Infrastructure

Microsoft Azure

Primary cloud provider. Services mentioned: Azure DevOps, Key Vault, Kubernetes (AKS), IoT Hub, general PaaS/IaaS.

Azure

AKS

DevOps

Kubernetes

Container orchestration. Likely AKS (Azure Kubernetes Service). Need to understand deployments, scaling, secrets management.

K8s

Containers

AKS

Azure Key Vault

Cloud secrets management. Used for encryption keys, connection strings, certificates. IAM and RBAC integration.

Security

Secrets

Enterprise Systems

SAP

ERP backbone. Integration with Snowflake and analytics platforms via ETL/ELT pipelines. Familiarity is a plus.

ERP

Integration

PRISM

Caterpillar-specific logistics/parts distribution system. Data integration point for the PSLD analytics platform.

Internal

Logistics

GitHub / Azure DevOps

Source control, CI/CD, and deployment management. Git workflows, pipeline management, infrastructure-as-code.

CI/CD

Git

Operating Systems

Linux Servers

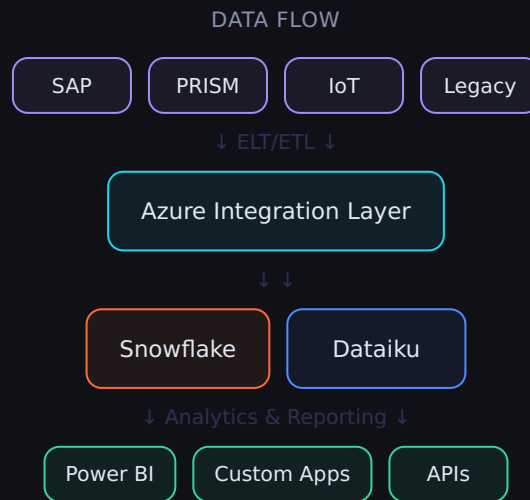
Windows Servers

Administration of both platforms supporting analytics and data management

04 Data Architecture

Architecture patterns relevant to the PSLD analytics platform

Modern Data Stack (PSLD Context)



Key Architecture Decisions

Decision	Approach	Rationale
Warehouse Platform	Snowflake	Cloud-native, separation of compute/storage, auto-scaling, zero-copy cloning
Cloud Provider	Azure	Caterpillar standard; integrates with Active Directory, Key Vault, AKS
Integration Pattern	ELT over ETL	Snowflake's compute power makes ELT viable; reduces pipeline complexity

Data Governance	RBAC + Row-level Security	Required for Caterpillar's compliance and cybersecurity standards
Container Orchestration	Kubernetes (AKS)	Standard for microservices; managed by Azure

Interview Tip: Know Snowflake's architecture (virtual warehouses, micro-partitions, time travel, cloning, data sharing). Be ready to compare Snowflake vs traditional warehouses vs Databricks vs BigQuery.

05 Integration Patterns

Common integration patterns relevant to enterprise data platforms

Batch ETL/ELT

Traditional scheduled data movement from SAP/PRISM to Snowflake. Use Snowpipe or custom pipelines. Idempotent, monitored, alertable.

Snowpipe

Airflow

Event-Driven

Real-time data ingestion via Event Hub / Kafka → Azure Functions → Snowflake. Used for IoT sensor data and logistics events.

Kafka

Event Hub

API Gateway

REST/gRPC APIs for application-to-database access. Secured via Azure API Management, OAuth2, and Key Vault for credentials.

APIM

OAuth

CDC (Change Data Capture)

Streaming database changes from SAP/PRISM to Snowflake for near-real-time sync.
Tools: Debezium, Azure Data Factory.

- Debezium
- ADF

Data Governance Pipeline

Integration of data catalog, lineage tracking (Dataiku), and access controls. Ensure every pipeline logs source, transform, and destination.

- Lineage
- Catalog

Security-First Integration

All integrations route through Azure Key Vault for secrets, enforce TLS 1.2+, and log to Azure Monitor/Sentinel for audit.

- Key Vault
- Sentinel

06 Decision Matrices

Common architecture tradeoffs the role would encounter

Warehouse: Snowflake vs Databricks vs Traditional RDBMS

Factor	Snowflake	Databricks	Traditional RDBMS
Separation of compute/storage	Native	Yes	No
Auto-scaling	Virtual warehouses	Auto-scaling clusters	Manual
SQL-first	Native SQL	Spark SQL	Yes
ML/AI integration	Snowpark ML	Native MLflow, SparkML	No
Data sharing	Snowflake Data Cloud	Delta Sharing	Not native

Cost model	Credit-based	DBU-based	Fixed licensing
Governance	RBAC, row-level security	Unity Catalog	Mature

Integration: Batch vs Event-Driven vs CDC

Factor	Batch (ELT)	Event-Driven	CDC
Latency	Minutes-hours	Sub-second	Seconds
Complexity	Low	High	Medium
Data consistency	Strong	Eventual	Strong
Best for	Daily reporting	Real-time alerts	Database sync

Cloud: Azure vs AWS for Caterpillar Context

Service Category	Azure	AWS
Kubernetes	AKS	EKS
Secrets Management	Key Vault	Secrets Manager
Data Warehouse	Synapse (but Cat uses Snowflake)	Redshift
CI/CD	Azure DevOps	CodePipeline
Identity	Active Directory integration	IAM + Cognito
IoT	IoT Hub	IoT Core

07 Architecture Principles

Guiding principles for enterprise architecture decisions

Security by Design

Every integration, data store, and API must implement least-privilege access, encryption at rest and in transit, and audit logging.

Separation of Concerns

Compute separated from storage (Snowflake model). Decoupled systems with clear interface contracts.

Data as a Product

Treat datasets as products with owners, SLAs, quality metrics, and discoverable catalogs (Dataiku).

Resilience & Self-Healing

Design for failure. Retry logic, dead-letter queues, idempotent pipelines, automated failover.

Scalability First

Platforms must scale horizontally. Snowflake auto-scaling warehouses, AKS cluster auto-scaling, partitioned data pipelines.

Governance & Compliance

Every data asset has lineage, quality checks, access controls, and retention policies. Automate compliance where possible.

08 Q&A — Technical

Common technical interview questions for an IT Architect role

How would you design the PSLD data analytics platform?

Framework: Start with business requirements (parts availability, logistics KPIs, dealer reporting), then layer:

- **Ingestion layer:** Azure Data Factory / Event Hub for SAP and PRISM data

- **Storage layer:** Snowflake with separate virtual warehouses for ETL, analytics, and reporting
- **Compute layer:** AKS for custom microservices, Azure Functions for serverless processing
- **Governance layer:** Dataiku for catalog/lineage, Azure Purview for data map
- **Security layer:** Key Vault for secrets, RBAC, row-level security in Snowflake
- **Presentation layer:** Power BI, custom APIs, Dataiku dashboards

Key considerations: Scalability for global dealer network, 24/7 uptime for parts logistics, compliance with Caterpillar security standards.

Compare Snowflake and Databricks. When would you choose one over the other?

Snowflake (choose when): SQL-first teams, BI/reporting workloads, data warehouse replacement, need for data sharing across organizations, structured/semi-structured data.

Databricks (choose when): Heavy ML/AI workloads, data science collaboration, unstructured data processing (Spark), real-time streaming, MLflow for MLOps.

Hybrid approach: Many enterprises use both — Snowflake for warehousing/BI, Databricks for data science/ML. Caterpillar PSLD appears Snowflake-first with Dataiku for ML.

How do you approach data integration between SAP and Snowflake?

Approach:

1. **CDC first:** Use SAP SLT or third-party CDC tools (Debezium, HVR) to stream changes to Azure Event Hub
2. **Batch fallback:** Scheduled SAP BO Data Services or Azure Data Factory pipelines for nightly bulk loads
3. **Landing zone:** Raw data lands in Snowflake staging schemas with minimal transformation
4. **ELT transforms:** dbt or Snowflake SQL for cleansing, deduplication, and business logic
5. **Data quality:** Automated checks at each stage; failed records routed to dead-letter queues

Key considerations: SAP extraction impact on transactional systems, data volume management, historical vs incremental loading.

How do you ensure data security and governance in a multi-system enterprise?

Multi-layer approach:

- **Authentication:** Azure AD / Active Directory integration for SSO across all platforms
- **Authorization:** RBAC at Snowflake, AKS, and Azure levels; row-level security in Snowflake
- **Encryption:** TLS 1.2+ in transit, AES-256 at rest (Snowflake, Azure Storage)
- **Secrets:** All credentials in Azure Key Vault, rotated automatically

- **Audit:** Snowflake query history, Azure Monitor / Sentinel for SIEM integration
- **Data catalog:** Dataiku for lineage, Data classification, and PII tagging
- **Compliance:** Caterpillar-specific cybersecurity standards, SOX compliance for financial data

Describe your experience with Kubernetes in production.

Key talking points:

- Cluster architecture: control plane, node pools, namespaces for environment isolation
- Workload types: Deployments, StatefulSets (for databases), CronJobs for scheduled tasks
- Networking: Ingress controllers, service mesh (Istio/Linkerd), network policies
- Secrets: External Secrets Operator with Azure Key Vault integration
- Monitoring: Prometheus + Grafana, Azure Monitor for containers
- CI/CD: GitOps with Flux/ArgoCD, Helm charts for templating
- Security: Pod Security Standards, container image scanning, RBAC

Tip: If you've managed AKS specifically, emphasize that. Caterpillar uses Azure.

How do you handle performance tuning in Snowflake?

Key strategies:

- **Warehouse sizing:** Right-size virtual warehouses; use multi-cluster warehouses for concurrent workloads
- **Clustering:** Define cluster keys on large tables to reduce micro-partition scan
- **Materialized views:** Pre-aggregate common query patterns
- **Query profiling:** Use EXPLAIN and QUERY_HISTORY to identify full scans, spills, and joins
- **Data modeling:** Star schemas over normalized; appropriate use of Snowflake search optimization
- **Caching:** Result caching for repeated queries; warehouse caching for hot data
- **Resource monitors:** Set credit alerts to prevent runaway costs

09 Q&A — Behavioral

STAR-format answers tailored to Caterpillar PSLD context

Tell me about a time you led a complex data platform migration.

Situation: Legacy on-premise data warehouse reaching capacity limits, unable to scale for growing analytics demand.

Task: Lead migration to cloud-based data platform with zero downtime for critical reports.

Action: Designed parallel-run architecture — old and new systems active simultaneously. Built CDC pipelines for continuous sync. Orchestrated cutover in phases: reporting first, then operational dashboards, then ad-hoc analytics.

Result: Successful migration with 0 downtime. Query performance improved 5x, storage costs reduced 40%, and the platform auto-scales to handle peak demand.

How do you handle conflicting priorities between business stakeholders and IT constraints?

Framework: I use a structured approach — understand the business need first, then map it to technical options with clear tradeoffs.

Example: A business team wanted real-time inventory dashboards, but the source SAP system couldn't support sub-minute queries without impacting transactional performance. We compromised: implemented CDC-based near-real-time sync (30-second latency) to Snowflake, which met the business need without risking the ERP system. Documented the architecture decision with rationale.

Key principle: Never say "no" — say "here's what it would take, and here's an alternative that gets you 90% of the value."

Describe a situation where you had to learn a new technology quickly.

Approach: Systematic learning — official docs first (Snowflake SnowPro certification path), then hands-on labs, then production pilot with a non-critical workload.

Result: Achieved SnowPro certification within 3 weeks of starting. Designed and deployed a proof-of-concept data pipeline that became the template for the full production deployment.

10 Whiteboard Exercises

Practice prompts with architecture sketches

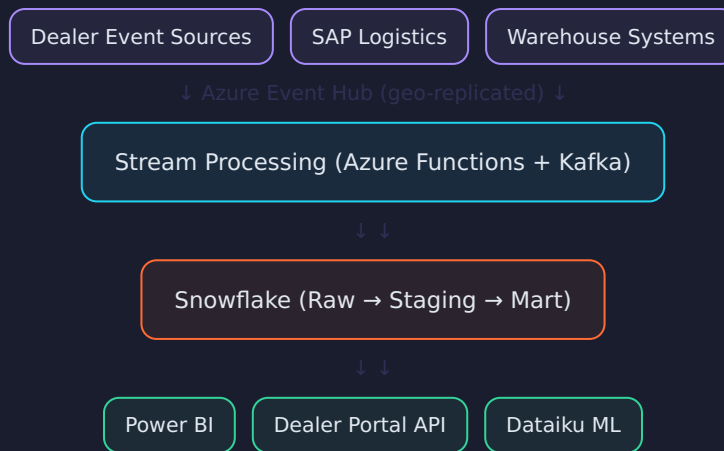
Exercise 1: Global Parts Logistics Data Pipeline

Prompt: Caterpillar's parts logistics network generates millions of events daily — order placements, warehouse movements, shipping updates, dealer inventory changes. Design a data pipeline that ingests, processes, and makes this data available for analytics and real-time dashboards across 180+ dealers.

Key considerations: Global timezones, variable dealer connectivity, data sovereignty, SAP integration, SLAs for reporting.

Suggested Approach:

HIGH-LEVEL ARCHITECTURE



Exercise 2: Enterprise Data Catalog & Governance

Prompt: PSLD has data spread across Snowflake, SAP, PRISM, and legacy systems with no centralized catalog. Business users can't find data; compliance can't verify data lineage. Design a governance framework.

Key components: Azure Purview or Dataiku for catalog, automated lineage tracking, data quality rules, certification workflows.

11 System Design Deep Dives

Full-stack architecture scenarios

Multi-Environment Data Platform (Caterpillar Context)

Design a data platform with development, QA, staging, and production environments supporting PSLD analytics.

Environment Strategy:

Environment	Purpose	Data	Snowflake Config
DEV	Development & testing	Anonymized subset	XS warehouse, cloned from PROD
QA	Integration testing	Full fidelity, 1-day lag	XS warehouse, cloned from PROD
STAGE	UAT / pre-prod	PROD mirror, 1-hour lag	S warehouse, cloned from PROD
PROD	Production	Live data	M-L warehouse, multi-cluster

Key design decisions: Snowflake zero-copy cloning for instant environment creation. AKS namespaces per environment. Azure DevOps pipelines with environment-specific variable groups.

Secure API Gateway for Analytics Data

Design an API layer that exposes Snowflake data to internal applications and dealer portals securely.

Components: Azure API Management → Azure Functions or AKS microservices → Snowflake. Authentication via Azure AD + OAuth2. Rate limiting, caching, and usage analytics at the APIM layer. Row-level security in Snowflake ensures dealers only see their own data.

12 Architecture Glossary

Key terms to know for the interview

Snowflake

Cloud-native data warehouse with separate compute and storage, virtual warehouses, micro-partitions, and zero-copy cloning.

Dataiku

Enterprise AI/ML platform for collaborative data science, MLOps, and data catalog.

AKS

Azure Kubernetes Service — managed Kubernetes for container orchestration.

ELT vs ETL

ELT loads raw data first, transforms in warehouse (Snowflake paradigm). ETL transforms before loading.

CDC

Change Data Capture — streaming database changes in real-time for sync.

LTA

Caterpillar Logical Technical Architecture — enterprise architecture framework.

PRISM

Caterpillar's parts/logistics management system.

SAP

Enterprise ERP system — Caterpillar's transactional backbone.

dbt

Data build tool — SQL-first transformation framework for ELT workflows.

RBAC

Role-Based Access Control — access management model used in Snowflake and Azure.

Snowpipe

Snowflake's auto-ingestion service for continuous data loading.

IaC

Infrastructure as Code — Terraform, ARM templates, Pulumi for reproducible cloud infrastructure.