Here is a **detailed study guide** for **Task Statement 1.3** of the **AWS Certified AI Practitioner (AIF-C01)** exam.

✅ Task Statement 1.3: Describe the ML Development Lifecycle

🔷 1. Components of an ML Pipeline

The machine learning (ML) pipeline is a series of steps used to develop and deploy ML models.

Stage	Purpose	Example Tools/Services
1. Data Collection	Gather data relevant to the problem (structured/unstructured).	Amazon S3, AWS Glue
2. Exploratory Data Analysis (EDA)	Understand data distributions, outliers, trends.	Jupyter notebooks, Pandas, SageMaker Studio
3. Data Pre-processing	Clean and transform data (remove duplicates, handle missing values).	Amazon SageMaker Data Wrangler
4. Feature Engineering	Create new input features to improve model performance.	Amazon SageMaker Feature Store
5. Model Training	Train model on labeled data using an algorithm.	Amazon SageMaker Training Jobs
6. Hyperparameter Tuning	Optimize model settings to improve accuracy.	SageMaker Automatic Model Tuning
7. Model Evaluation	Test model with unseen data to assess performance.	Confusion matrix, AUC, F1 Score
8. Deployment	Make the model available for use (inference).	Amazon SageMaker Endpoints
9. Monitoring	Track model predictions, drift, and accuracy over time.	Amazon SageMaker Model Monitor

---

🔷 2. Sources of ML Models

Model Source	Description	When to Use
Open Source Pre-trained Models	Public models trained on large datasets (e.g., Hugging Face, TensorFlow Hub).	Save time, use for common tasks (e.g., sentiment analysis, image classification).
Custom Trained Models	Built and trained from scratch using your own data.	When business problem is domain-specific or proprietary.

◈ 3. Using a Model in Production

Deployment Method	Description	Pros/Cons
Managed API Service	Host model using cloud-managed infrastructure (e.g., SageMaker Endpoint).	Easy to scale, fully managed, low ops.
Self-hosted API	Package model and serve via a custom API (e.g., Flask app on EC2).	More control, but requires ops expertise.

◈ 4. Relevant AWS Services for Each Pipeline Stage

Stage	AWS Service	Functionality
Data Collection	AWS Glue, Amazon S3	ETL and data storage
EDA & Notebooks	SageMaker Studio, Jupyter	Interactive data analysis
Data Prep	SageMaker Data Wrangler	Clean and transform data
Feature Store	SageMaker Feature Store	Central place to store and retrieve features
Training	SageMaker	Train models on built-in or custom algorithms
Tuning	SageMaker Hyperparameter Tuning	Automatically optimize model settings
Deployment	SageMaker Inference Endpoints	Real-time or batch deployment
Monitoring	SageMaker Model Monitor	Monitor data quality, drift, and prediction outcomes

◈ 5. ML Operations (MLOps) Fundamentals

Concept	Definition	Why It Matters
Experimentation	Trying different models, features, and configurations.	Helps identify best performing model.
Repeatability	Re-running ML pipelines with the same results.	Essential for debugging and collaboration.
Scalable Systems	Handle large data volumes and traffic.	Required for production-level usage.
Managing Technical Debt	Avoid hardcoded logic, undocumented hacks, or manual steps.	Makes models easier to maintain and update.
Production Readiness	Includes monitoring, logging, version control.	Prevents model failures in real-world use.

Concept	Definition	Why It Matters
Model Monitoring	Track input data and predictions over time.	Detects data or concept drift.
Model Re-training	Updating model when performance drops.	Keeps predictions accurate as data evolves.

🔷 6. Model Performance Metrics

Metric Type	Metric	Definition	Use Case
Model Metrics	**Accuracy**	Correct predictions / total predictions	Good for balanced classification problems
	AUC (Area Under Curve)	Probability model ranks a positive example higher than a negative one	Evaluates classification over all thresholds
	F1 Score	Harmonic mean of precision and recall	Good for imbalanced data
Business Metrics	**Cost per user**	Cost to serve model per user	Helps measure efficiency
	Customer feedback	User satisfication, usability reports	Measures model usefulness
	ROI (Return on Investment)	Benefit gained vs. cost incurred	Business justification for ML investment
	Development Costs	Total cost to build and deploy the model	Used for cost-benefit analysis

📌 Summary Table: ML Lifecycle and AWS Tools

Lifecycle Stage	AWS Service
Data Ingestion	AWS Glue, Amazon S3
EDA & Prep	SageMaker Studio, Data Wrangler
Feature Engineering	SageMaker Feature Store
Model Training	SageMaker
Hyperparameter Tuning	SageMaker Automatic Tuning
Evaluation	Built-in metrics, notebooks
Deployment	SageMaker Endpoints
Monitoring	SageMaker Model Monitor

Lifecycle Stage	AWS Service
MLOps & Automation	SageMaker Pipelines, CI/CD Tools (e.g., CodePipeline)

Study Tips

- Use diagrams to visualize the ML pipeline from data to deployment.
- Try AWS SageMaker Studio Lab for free to explore a simplified ML pipeline.
- Understand *why* each stage matters, not just what tools are used.
- Practice matching AWS services to the lifecycle stages in multiple-choice format.
- Learn the **tradeoffs** between open source models vs. custom training.

Would you like a **one-page printable cheat sheet**, **flashcards**, or **slides** for this study guide?