



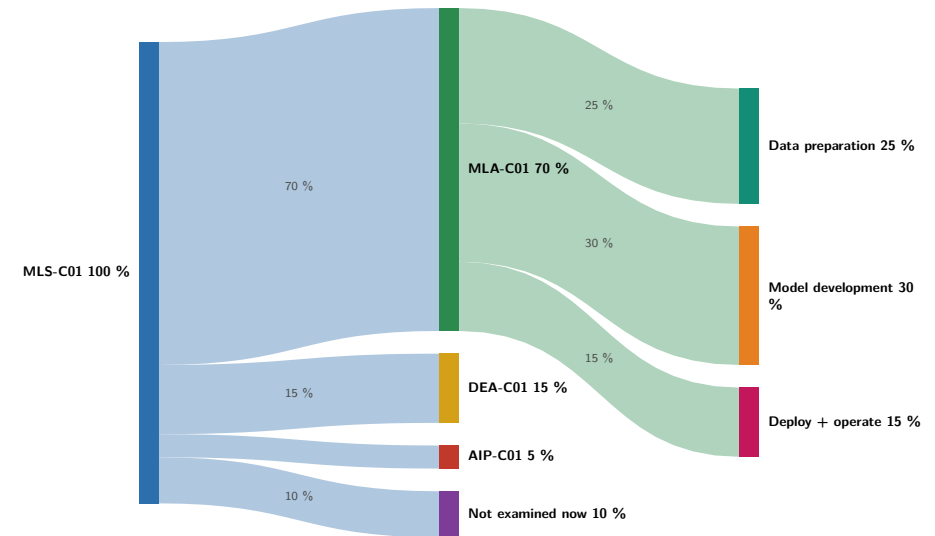
Machine Learning Specialty (MLS-C01): Legacy

AWS Certified Machine Learning - Specialty exam guide (MLS-C01) and its retirement announcement, as of 2026 | What it tested, the ML science still worth knowing, where each domain lives now (MLA-C01, AIP-C01, DEA-C01), reading MLS on a CV

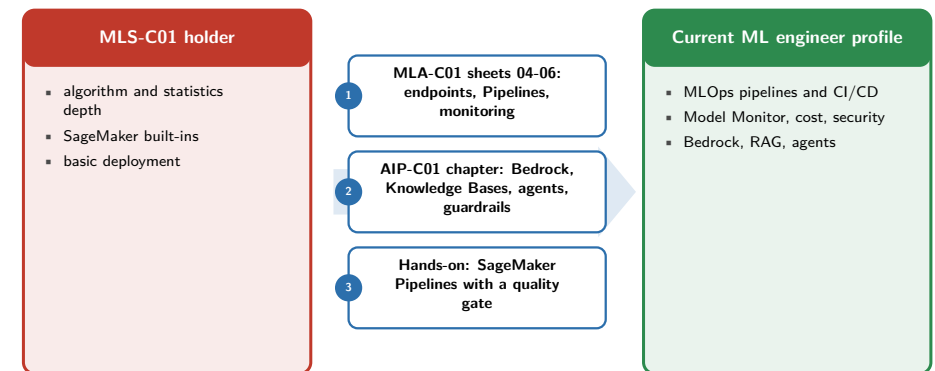
MLS-C01 Retired Machine learning SageMaker AI

What MLS-C01 tested and where it lives now

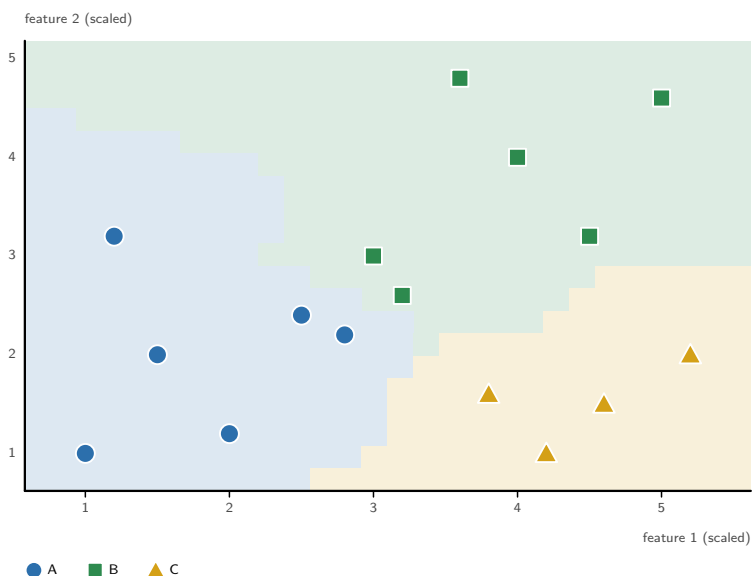
MLS-C01 domain (weight)	Typical content	Now taught in
1 Data engineering (20 %)	S3 data lakes, Kinesis (incl. Video Streams), Glue, EMR, batch vs streaming for ML	DEA-C01 01-03; MLA-C01 01
2 Exploratory data analysis (24 %)	cleaning, feature engineering, visualisation, statistics, labelling	MLA-C01 01
3 Modeling (36 %)	framing problems, algorithm choice, training, tuning, evaluation, built-in algorithms	MLA-C01 02-03 (less maths)
4 ML implementation and operations (20 %)	endpoints, scaling, security, AI services	MLA-C01 04-06; AIP-C01 for generative AI
Format (MLS-C01)	65 questions, 180 minutes, passing score 750 / 1,000	
Status	retired: last exam day 2026-03-31 (AWS exam page); earned certifications stay active 3 years	
Why	AWS split the role: ML engineering (MLA-C01, associate) and generative-AI development (AIP-C01, professional)	
Validity of earned MLS	3 years from the pass date, no renewal through MLS once retired	



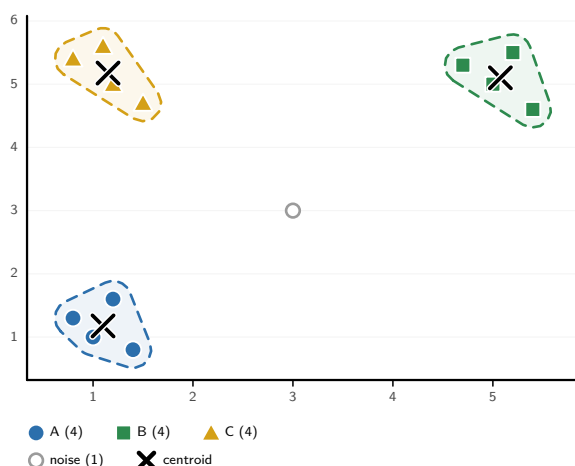
Where MLS-C01 content went (share of the MLS blueprint, %, approximate mapping)



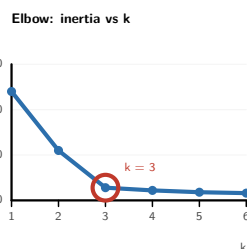
ML science MLS emphasised (still useful)



k-NN ($k = 3$) decision regions on two features: misclassified points ringed; scale features first, distances depend on units



k-means with an elbow chart: inertia drops sharply until $k = 3$, then flattens



Concept	What MLS asked	Rule of thumb
Missing data	impute vs drop	median for skewed numerics, mode for categories, model-based (k-NN, MICE) for important features
Correlated features	multicollinearity	PCA or drop one; trees are robust, linear models are not
Dimensionality	curse of dimensionality	PCA, feature selection, embeddings
Overfitting	train $\gg$ validation score	regularise (L1/L2, dropout), early stopping, more data, simpler model
Learning rate	loss diverges or crawls	too high diverges, too low is slow; schedules and Adam
Batch size	noisy or slow training	larger batches need a higher learning rate; smaller generalise better sometimes
Imbalance	accuracy misleading	SMOTE or weights; precision, recall, F1, PR-AUC
Text features	bag of words, TF-IDF, n-grams, Word2Vec	now mostly transformer embeddings
Image models	CNNs, transfer learning, augmentation	fine-tune a pre-trained backbone

PROBLEM

TF-IDF of the word 'refund' in one ticket: it appears 3 times in a 100-word ticket; 10,000 tickets in the corpus, 100 contain 'refund' (log base 10)

1. Term frequency

$$\Rightarrow TF = 3 / 100 = 0.03$$

2. Inverse document frequency

$$\Rightarrow IDF = \log_{10}(10,000 / 100) = \log_{10}(100) = 2$$

3. Weight

$$\Rightarrow TF-IDF = 0.03 \times 2 = 0.06$$

why: frequent in this ticket, rare in the corpus $\rightarrow$ informative

GIVEN

- TF = term count / words in the document
- IDF = $\log_{10}(N / n)$, N = documents, n = documents containing the term
- TF-IDF = TF $\times$ IDF

ANSWER

0.06; a word present in every ticket would get IDF = $\log_{10}(1) = 0$ and weight 0

Check: Stop words such as 'the' score near 0 because $n \approx N$

✓ Myths, recap and self-test

YOU MAY HEAR

"MLS-C01 holders must start from zero"

BUSTED

✓ ACTUALLY

Most modelling knowledge transfers to MLA-C01

Add MLOps, monitoring, cost and security; AIP-C01 adds generative AI.

YOU MAY HEAR

"The retirement voids existing MLS certifications"

BUSTED

✓ ACTUALLY

They stay valid until their 3-year expiry

Only new attempts and renewals stop.

YOU MAY HEAR

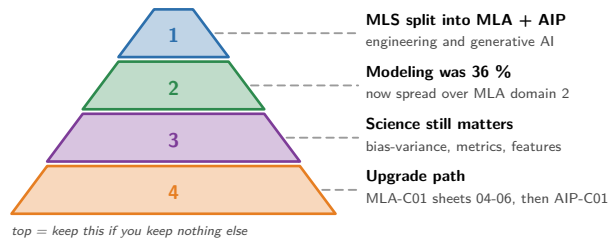
"MLA-C01 is the same exam renamed"

BUSTED

✓ ACTUALLY

MLA-C01 is an associate engineering exam

Less statistics and algorithm theory, more deployment and operations.



Q1 MLS-C01 domains and weights?

A Data engineering 20 %, exploratory data analysis 24 %, modeling 36 %, ML implementation and operations 20 %.

Q2 MLS-C01 format?

A 65 questions, 180 minutes, pass at 750.

Q3 Which exams succeed MLS-C01 for an ML engineer and a GenAI developer?

A MLA (MLA-C02 since late 2026) and AIP-C01.

Q4 L1 vs L2 regularisation?

A L1 drives some weights to zero (feature selection); L2 shrinks all weights.

Q5 Choose k for k-means?

A Elbow method on inertia (or silhouette score).

Q6 Reduce 200 correlated numeric features before k-means?

A PCA.

Q7 What does IDF penalise?

A Words that appear in many documents.

Q8 Algorithm that needs feature scaling: XGBoost or k-NN?

A k-NN (distance-based); trees do not need scaling.