
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

ANT307

Improve search relevance with ML in Amazon OpenSearch Service

Jianwei Li

Senior Analytics Specialist TAM
AWS

Muthu Pitchaimani

Senior Specialist Solution Architect
AWS

Agenda

What is search relevance?

How to measure relevance?

Improving relevance

Text search versus semantic search

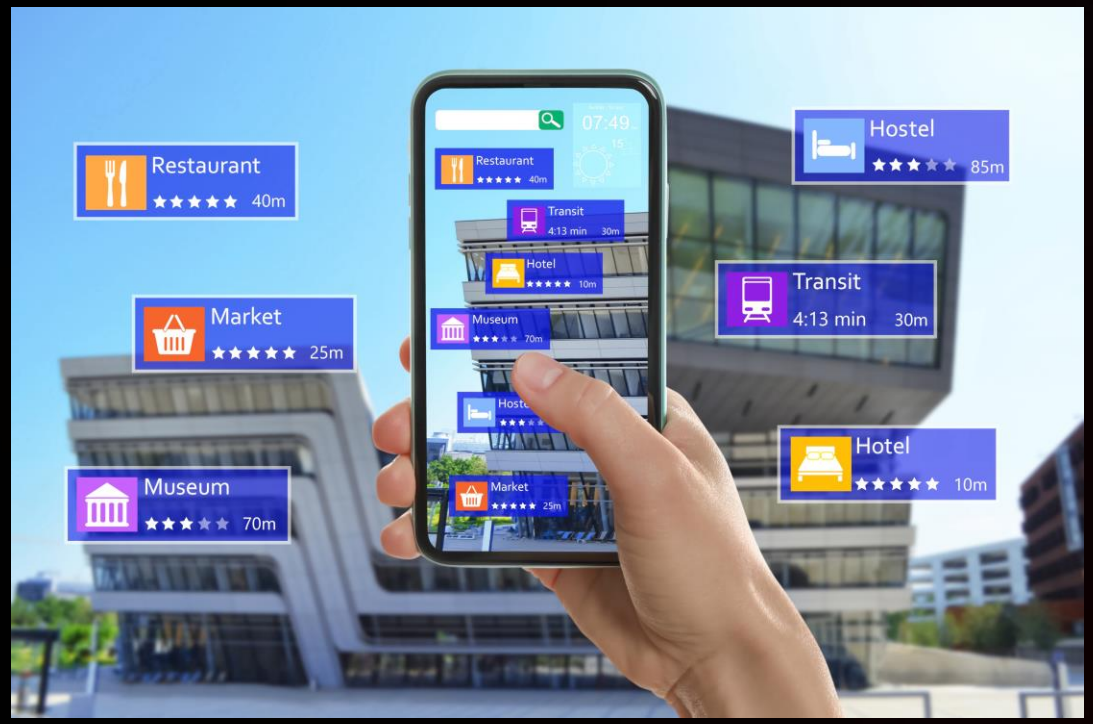
Semantic search architecture

Workshop introduction

Get started

Q&A

What is search relevance?



Web search, shopping, food delivery, watch a movie…

“Relevance is a measure of how good the results are in satisfying a particular information need”

How to measure relevance?

How do you tell whether something is relevant or not?

Relevance is inherently subjective

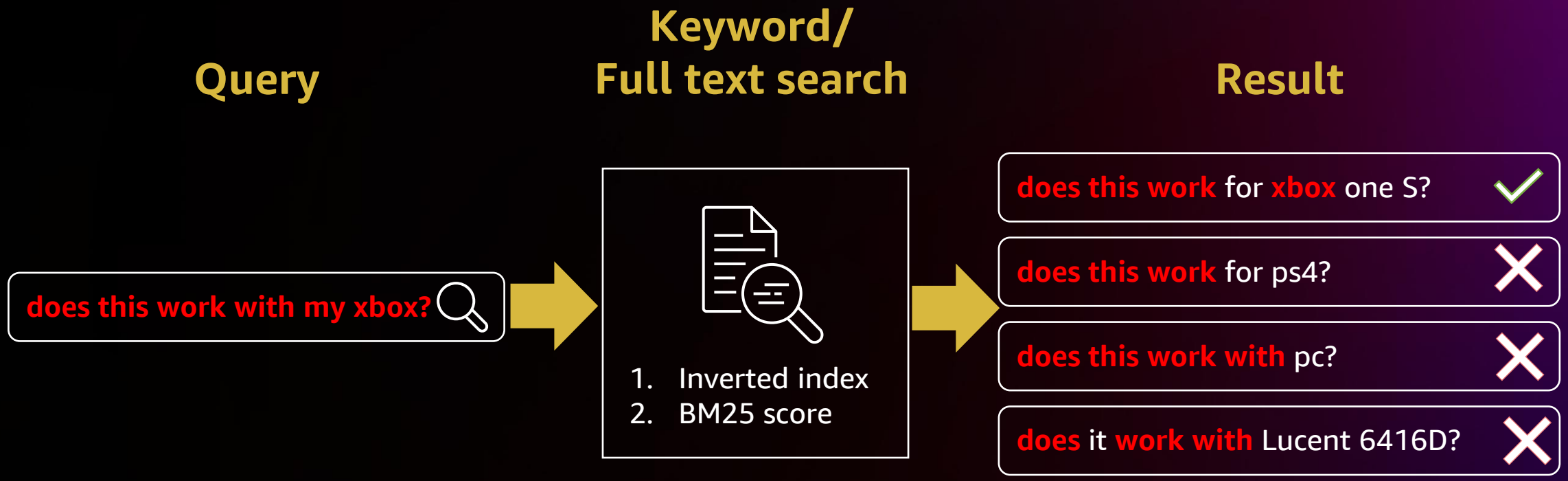
Search for ”big bang theory”

Fraction of relevant results
vs.
Fraction of relevant results retrieved

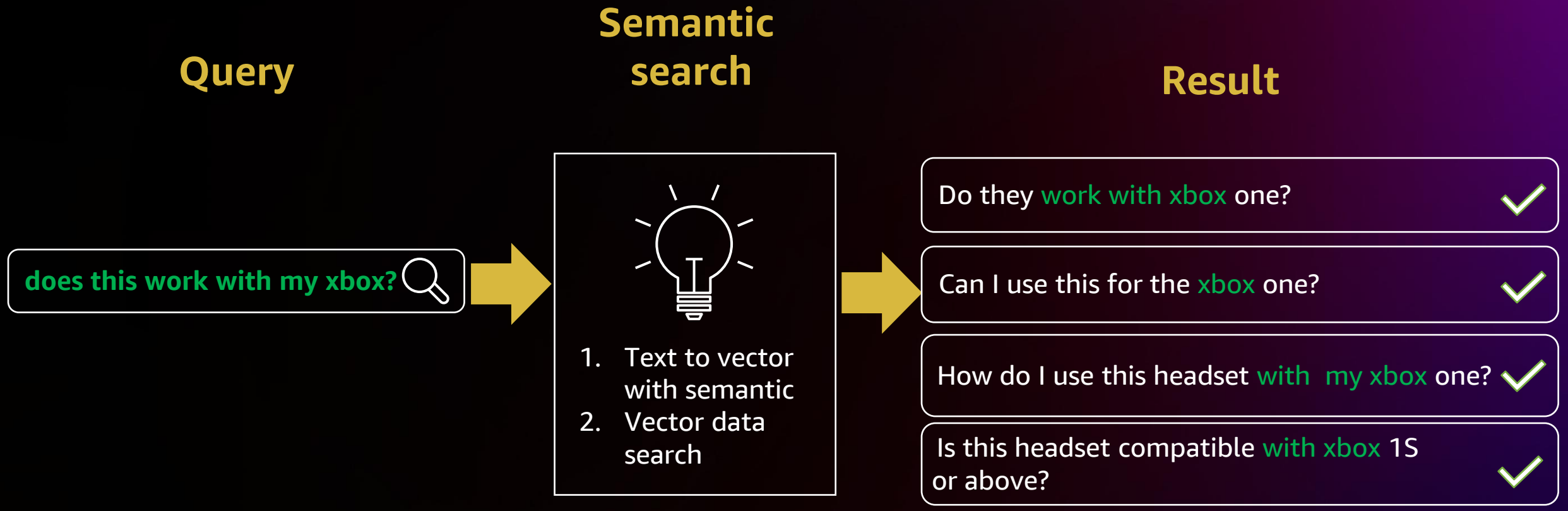
Improving relevance with ML

- Is ranking and scoring enough?
  - Matching, advanced statistics, synonyms (text search) can take you only so far
- Various techniques: learning to boost, learning to rank, vector similarity with ML
- Understanding meaning and context behind queries become important
- Vector similarity with machine learning (ML)
  - Not mutually exclusive with text search
  - Hybrid search expectedly higher relevance than simple matching

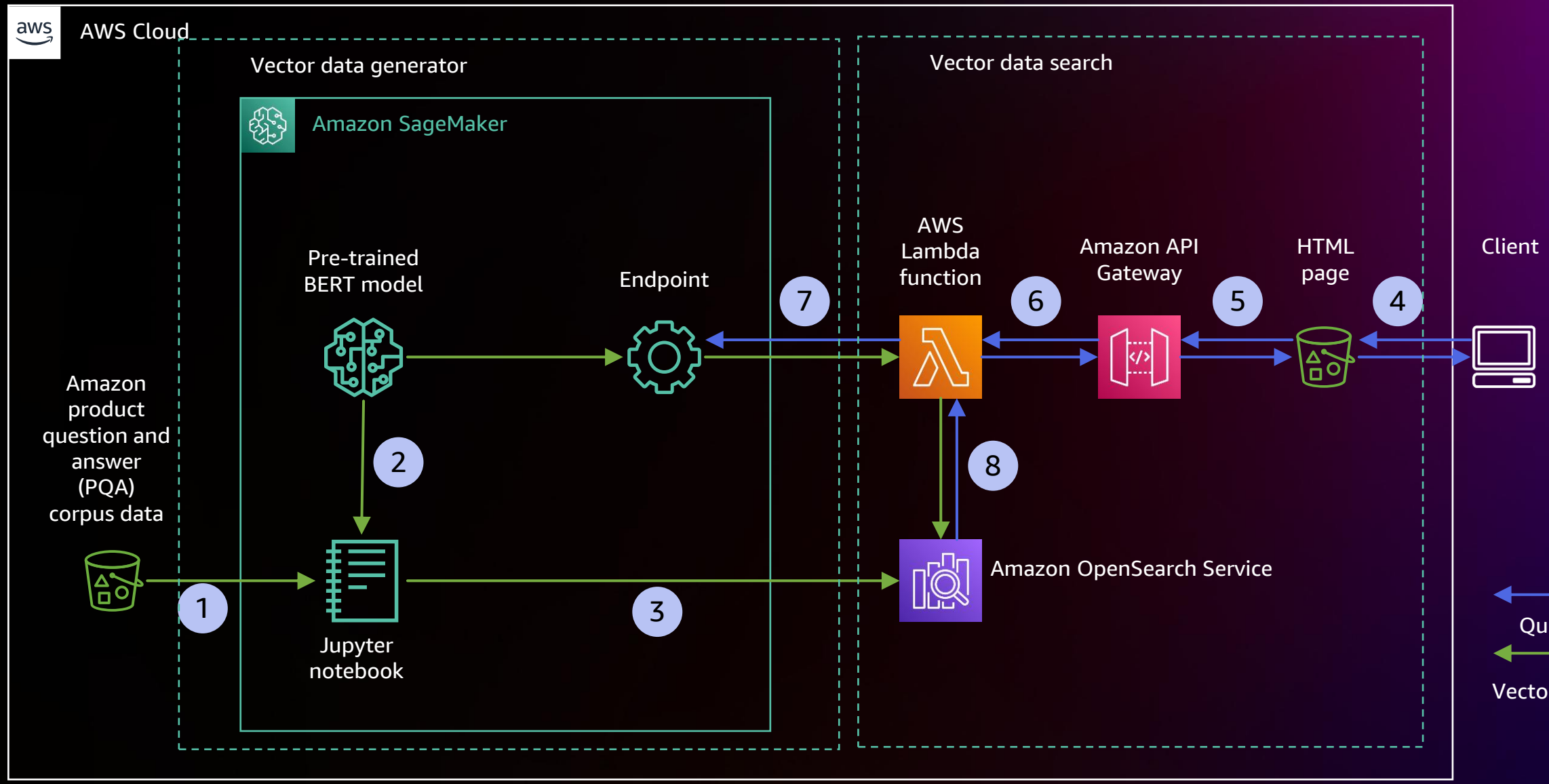
Text search



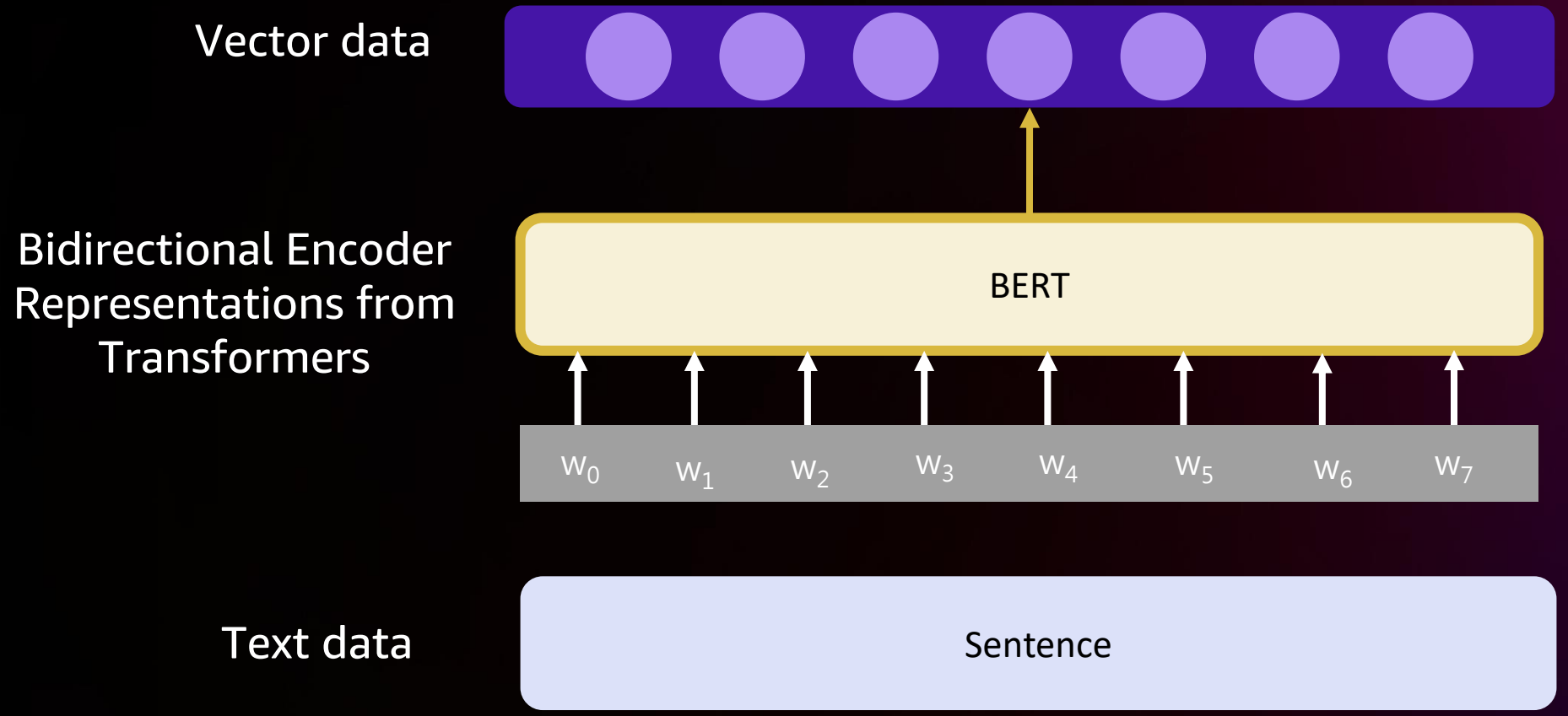
Semantic search



Semantic search architecture



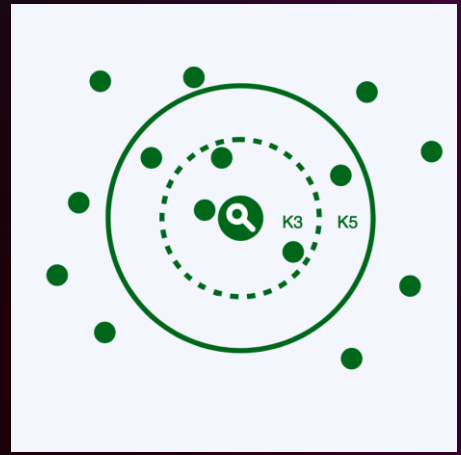
Vector data generator

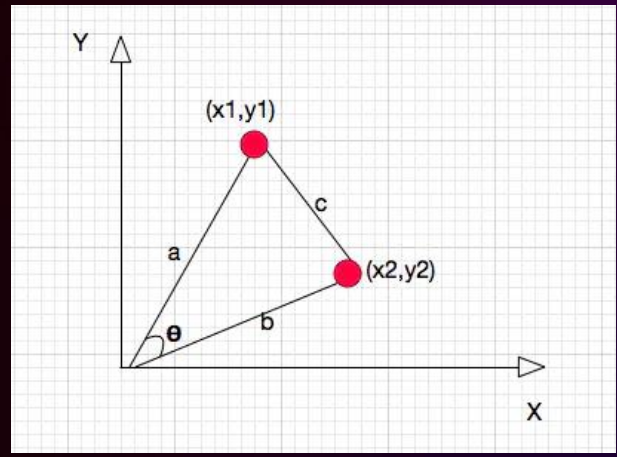


Vector data search: kNN

- Create vector data with feature data
- Calculate vector data similarity in the vector space
- Euclidean distance or cosine similarity

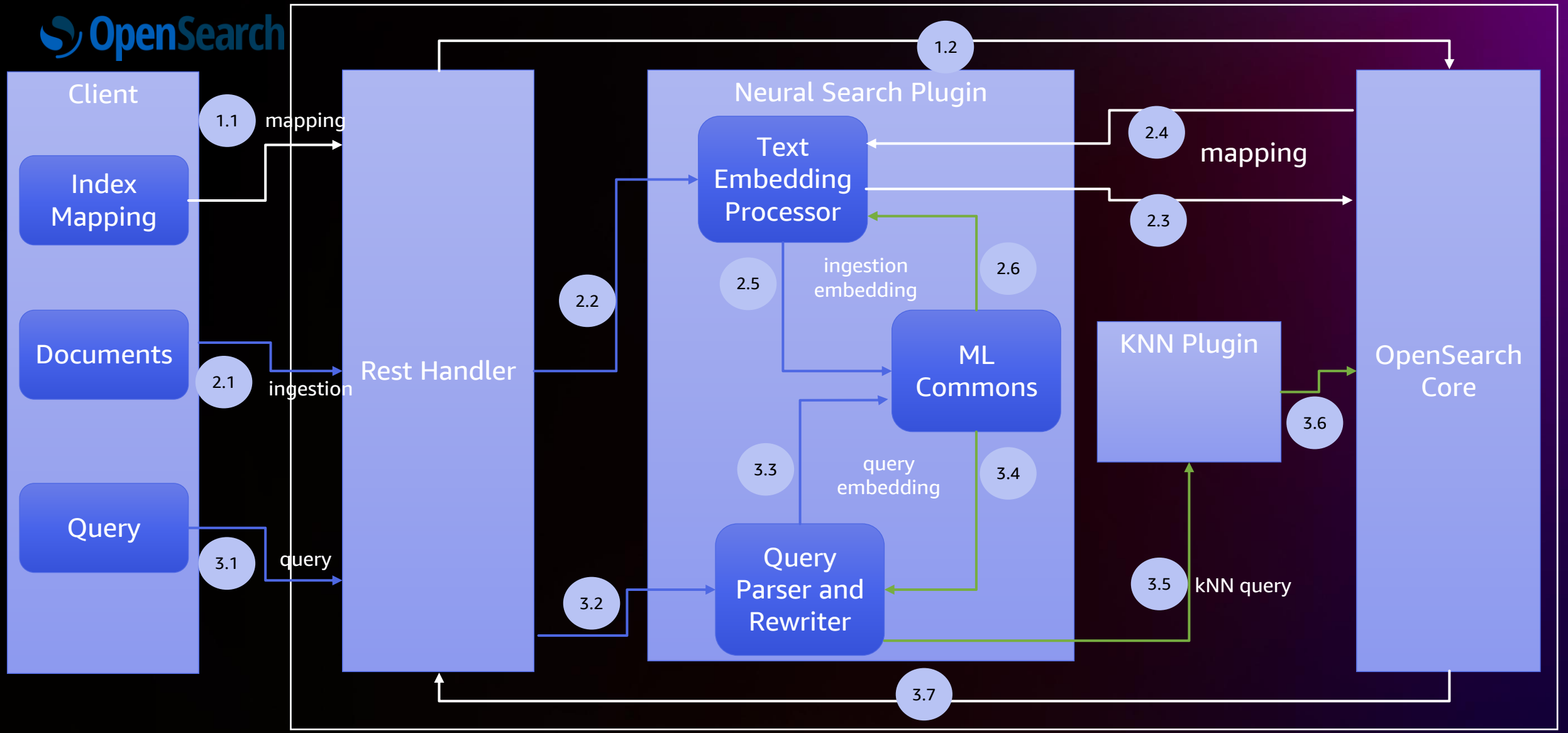
$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}}.$$





OpenSearch neural-search plugin

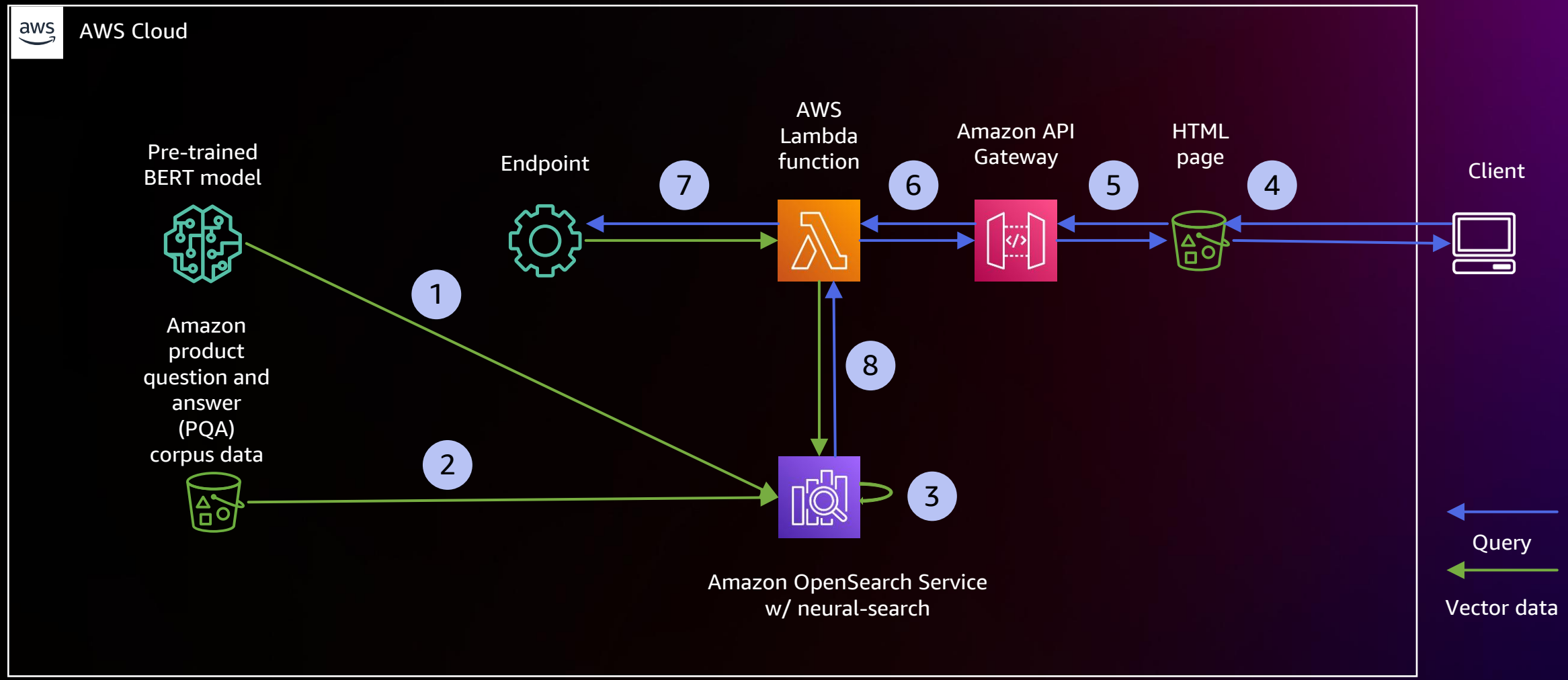
New in OpenSearch 2.4



GitHub RFC

Semantic search with Neural Search

New in OpenSearch 2.4



Workshop content

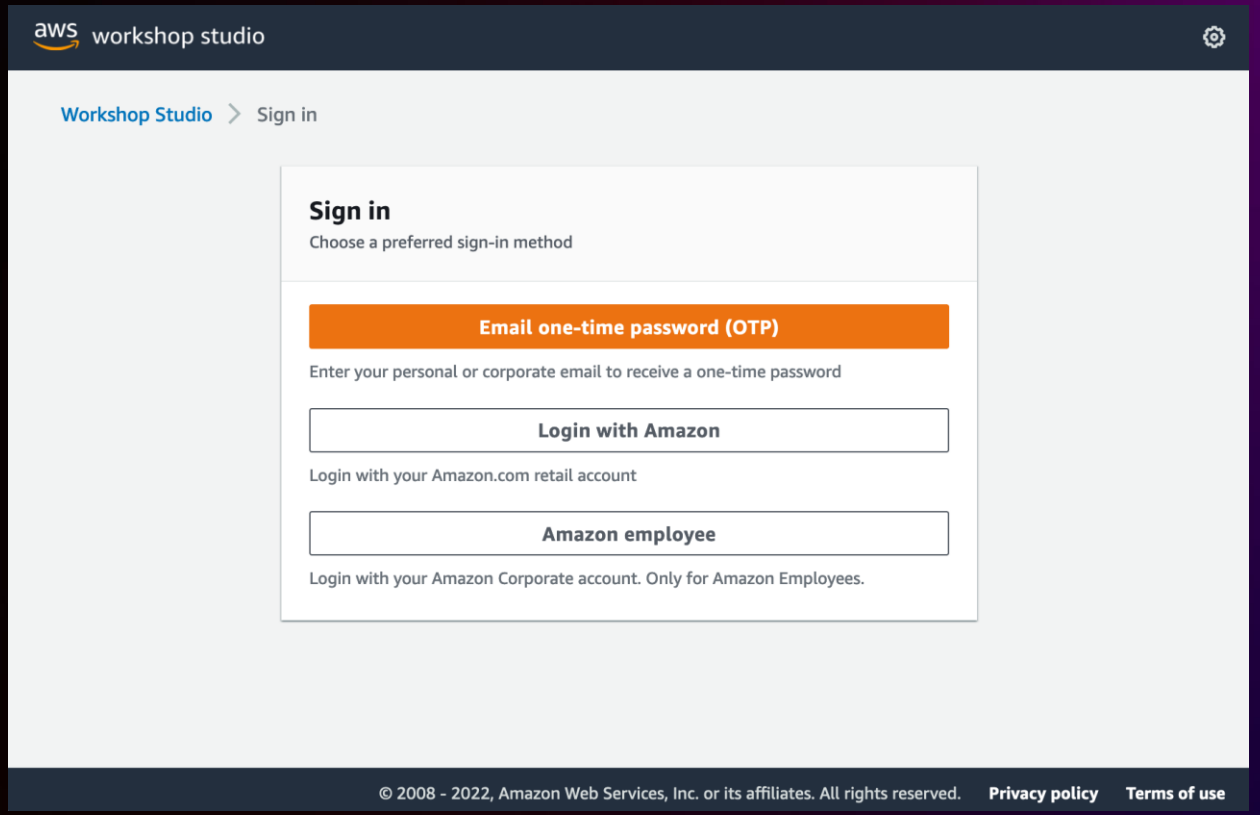
- Module 1: Fundamental search concepts
- Module 2: Text search
- Module 3: Semantic search
- Module 4: Full-stack semantic search
- Module 5: Semantic search with fine-tuned model (optional)

Get started with this workshop

- As a participant, you will have access to an AWS account with any optional pre-provisioned infrastructure and IAM policies needed to complete this workshop
- The AWS account will only be available for the duration of this workshop; you will lose access to the account thereafter
- The pre-provisioned infrastructure are deployed to a **us-east-1** Region
- Be sure to review the terms and conditions of the event; do not upload any personal or confidential information in the account

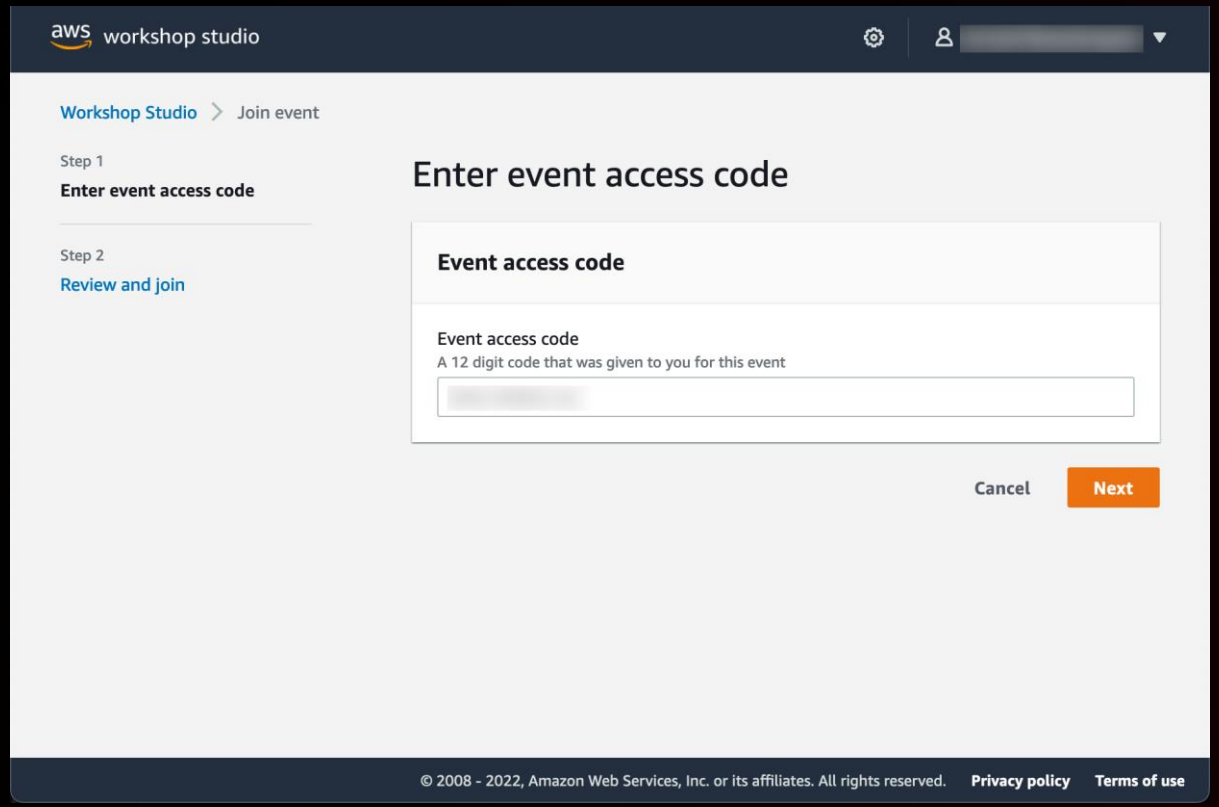
Step 1: Sign-in via OTP

https://catalog.workshops.aws/join



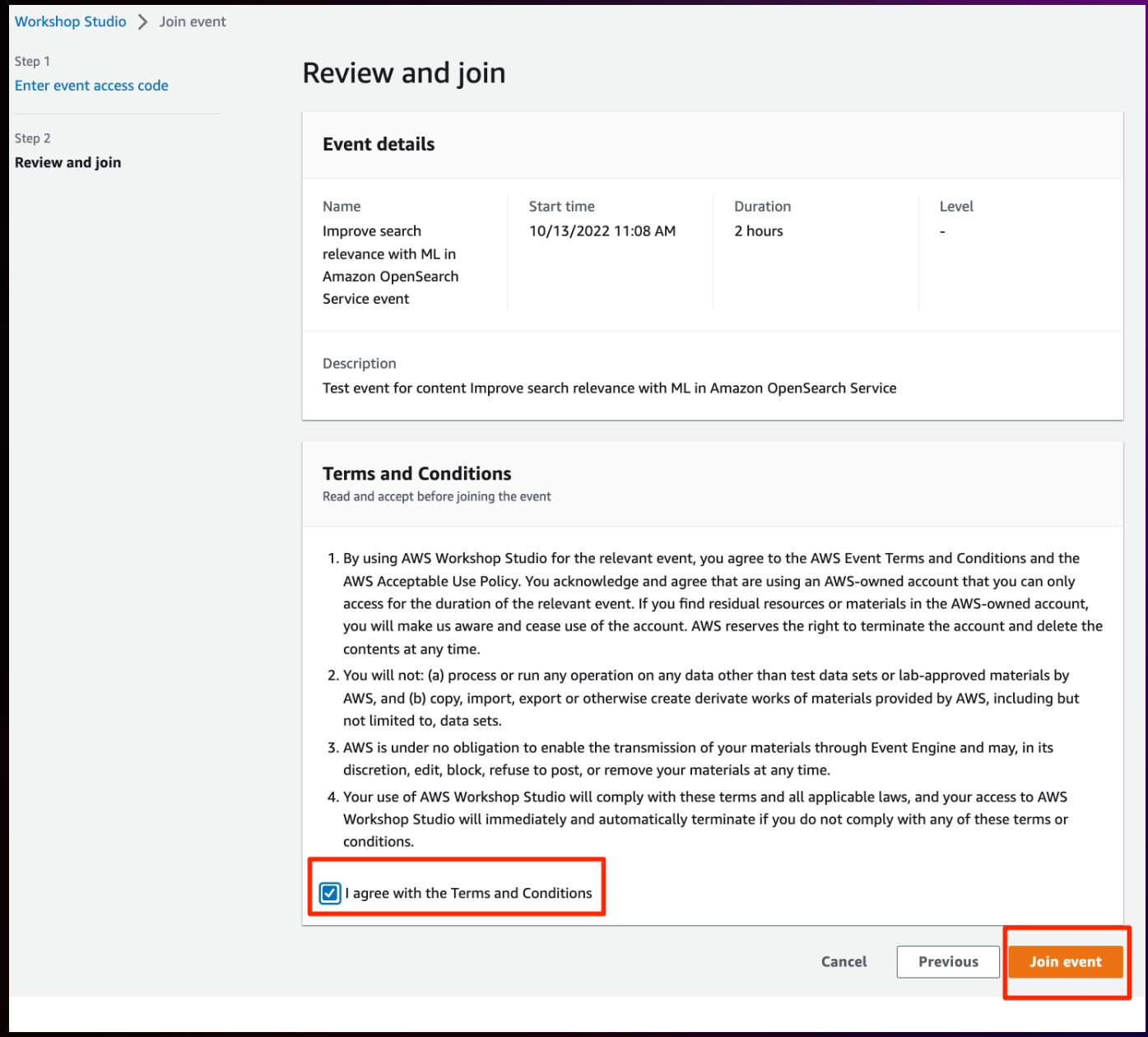
Step 2: Enter event access code

Enter 12-digit event access code



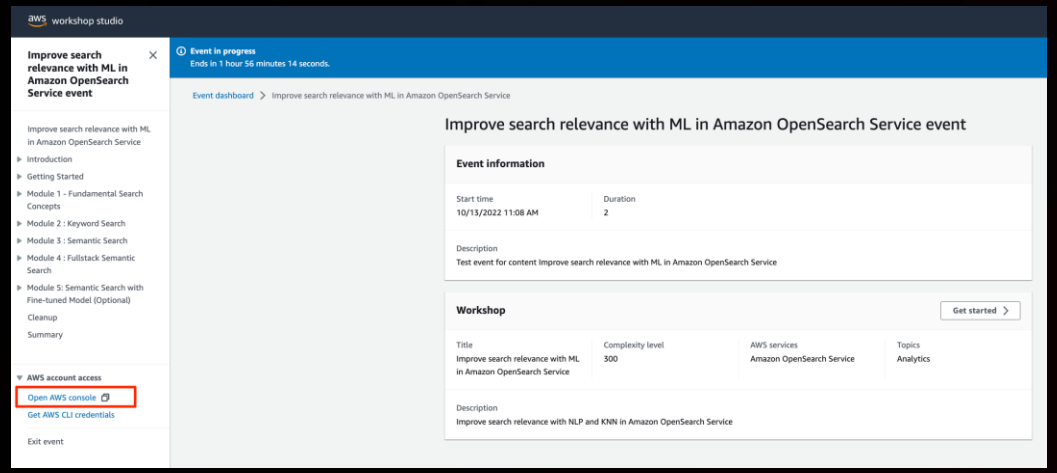
7fcc-064e80-15

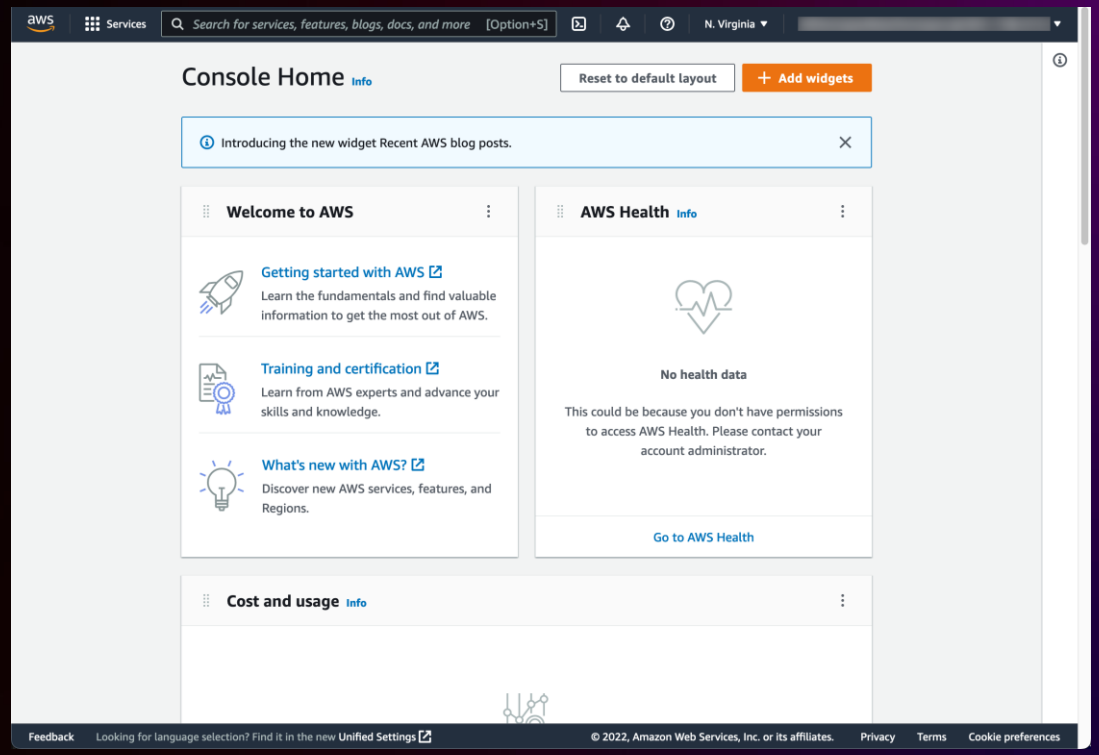
Step 3: Review terms and join event



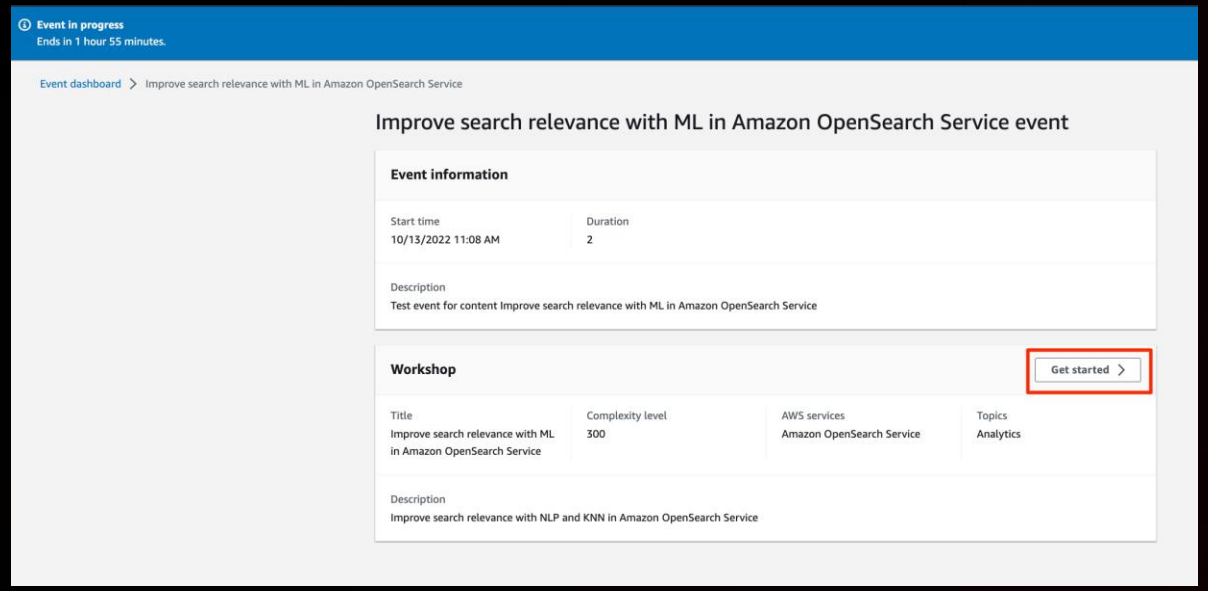
Step 4: Access AWS account

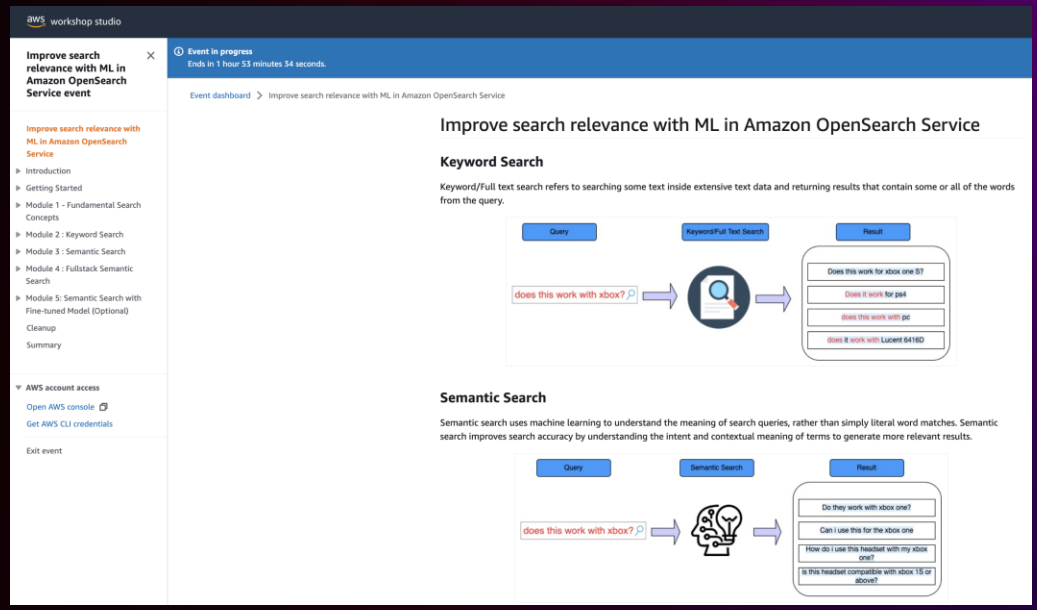
Access the AWS Management Console or generate AWS CLI credentials as needed





Step 5: Get started with the workshop





Workshop facilitators

- Jey Vell, Senior SA Manager
- Viral Shah, Principal GTM Specialist
- Josh Tow, Solution Architect
- Arun Lakshmanan, Analytics Specialist SA
- Mamta Vaidya, SA Manager
- Prashant Agrawal, Senior Analytics Specialist SA
- Dilip Rajan, Mgr, Partner Soln Architect
- Rafael Rodrigues, Enterprise Solutions Architect

Get started

Workshop for this event

https://tinyurl.com/2navfkaf

Event access code

7fcc-064e80-15

Workshop permanent link

https://tinyurl.com/592ej6kw

Thank you!

Jianwei Li
lijianwe@amazon.com

Muthu Pitchaimani
muthupmk@amazon.com

Please complete the session survey in the **mobile app**