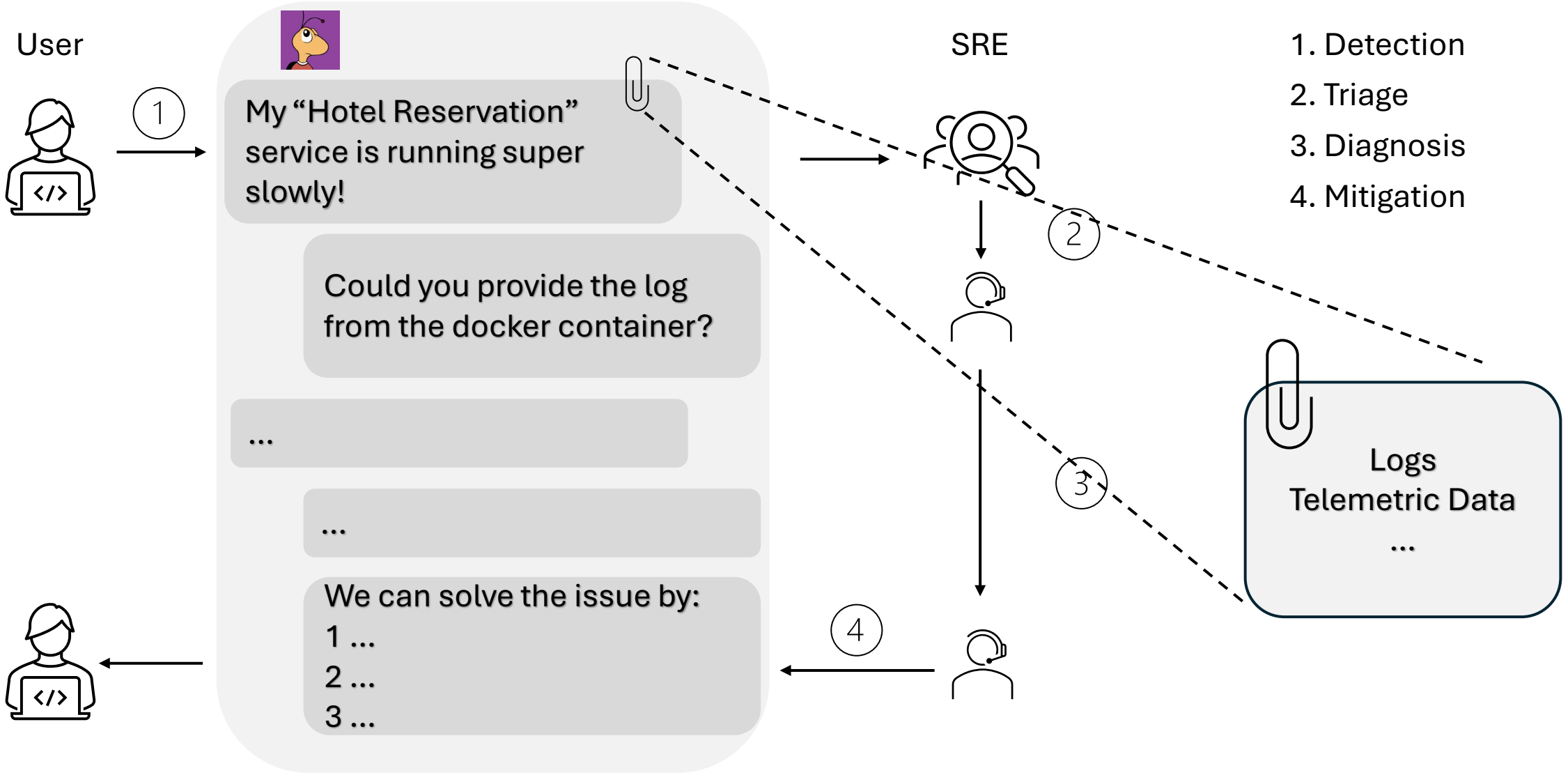
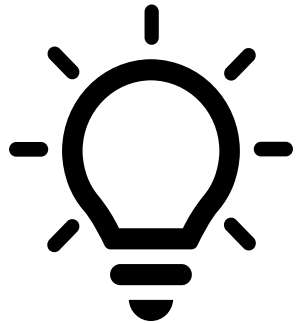
ARCA: AI for Root Cause Analysis

Yifan Wang, Kenneth P. Birman

Cornell University

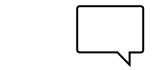
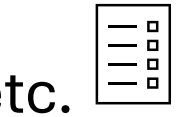
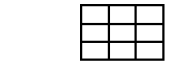
AI-Ops

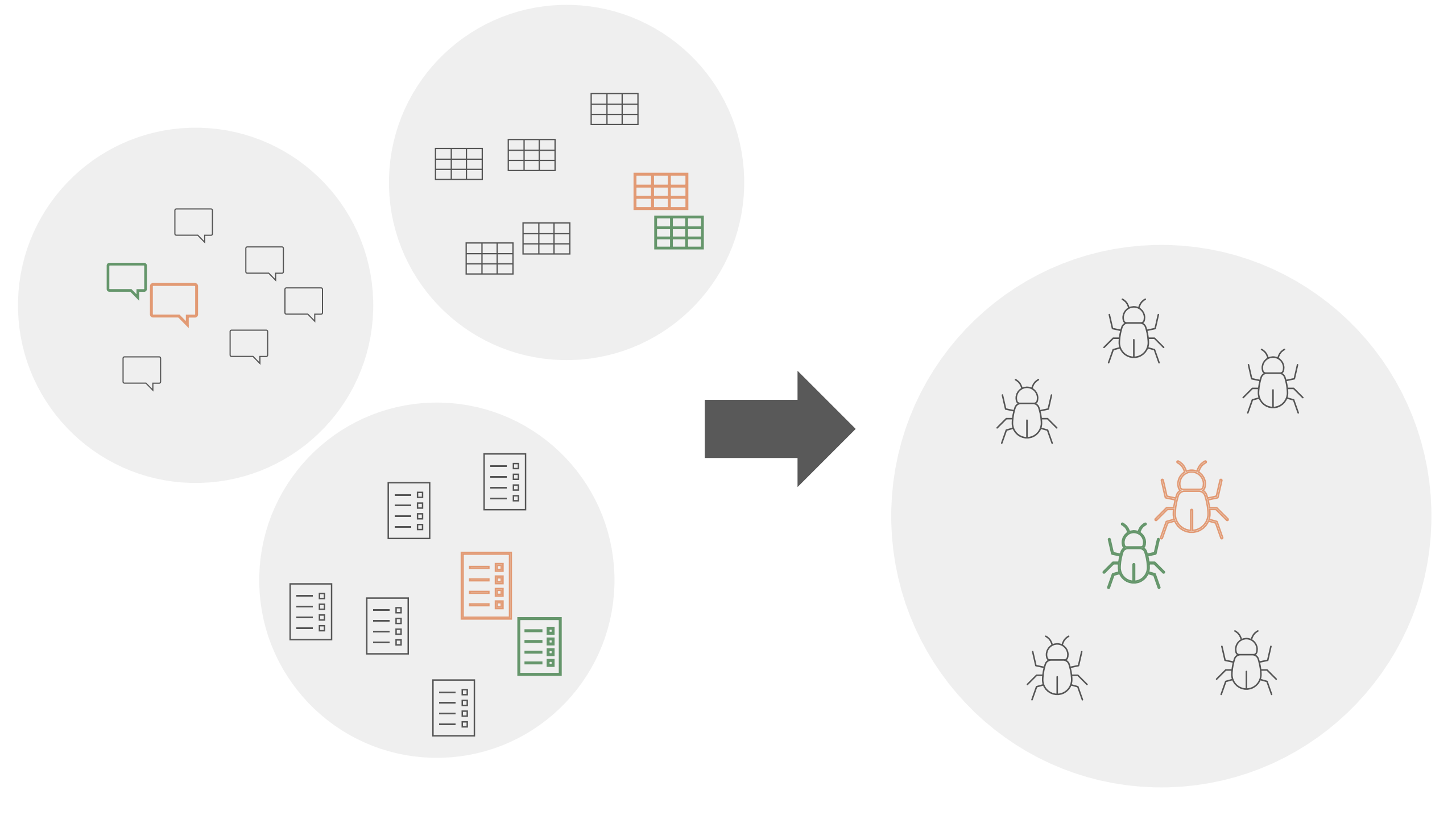

User
SRE
1. Detection
2. Triage
3. Diagnosis
4. Mitigation
1
My “Hotel Reservation” service is running super slowly!
2
Could you provide the log from the docker container?
...
3
Logs
Telemetric Data
...
...
We can solve the issue by:
1 ...
2 ...
3 ...
4



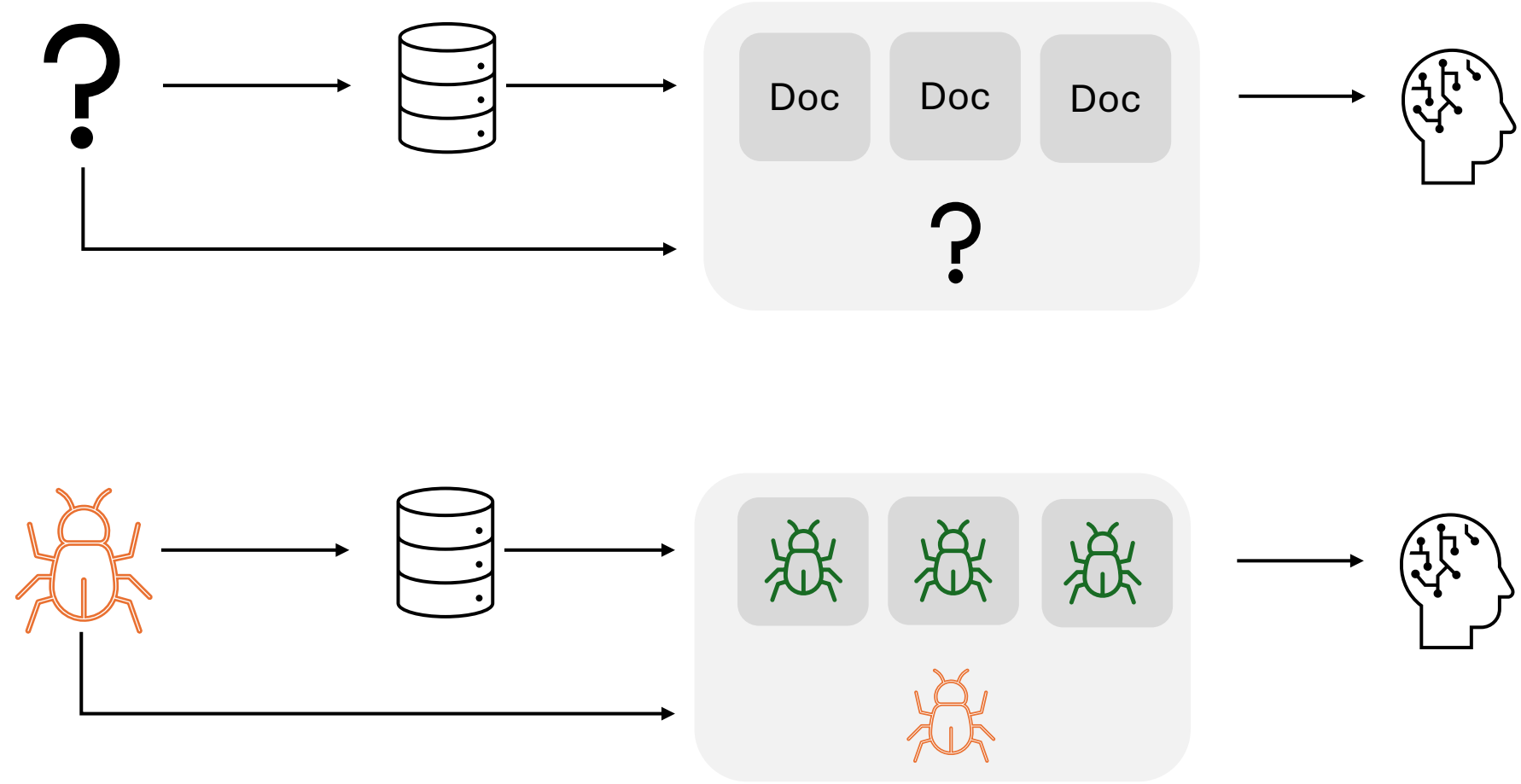
Similar bugs might share similar solution.

Bug: multi-modal data object

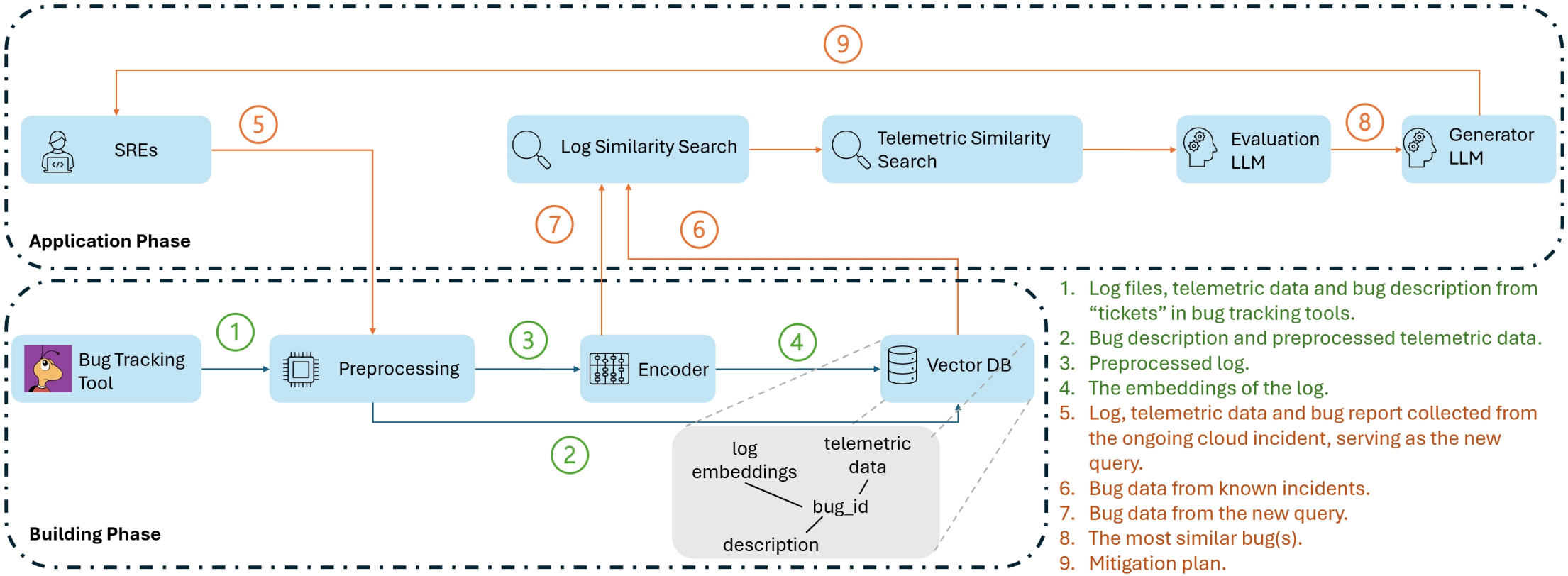
Components	Data Mode
Issue Description	Natural Language 
Log	Semi-structured language, numeric data, natural language, etc. 
Telemetric/Performance Data	Multi-variant time series numeric data 



Retrieval Augmented Generation (RAG)



ARCA Architecture



1. Log files, telemetric data and bug description from “tickets” in bug tracking tools.
2. Bug description and preprocessed telemetric data.
3. Preprocessed log.
4. The embeddings of the log.
5. Log, telemetric data and bug report collected from the ongoing cloud incident, serving as the new query.
6. Bug data from known incidents.
7. Bug data from the new query.
8. The most similar bug(s).
9. Mitigation plan.

Log

Purging:

keeping error-related messages.

Embedding:
vectorizing the purged log file.

Searching:
ANN using cosine similarity search.

Telemetric Data

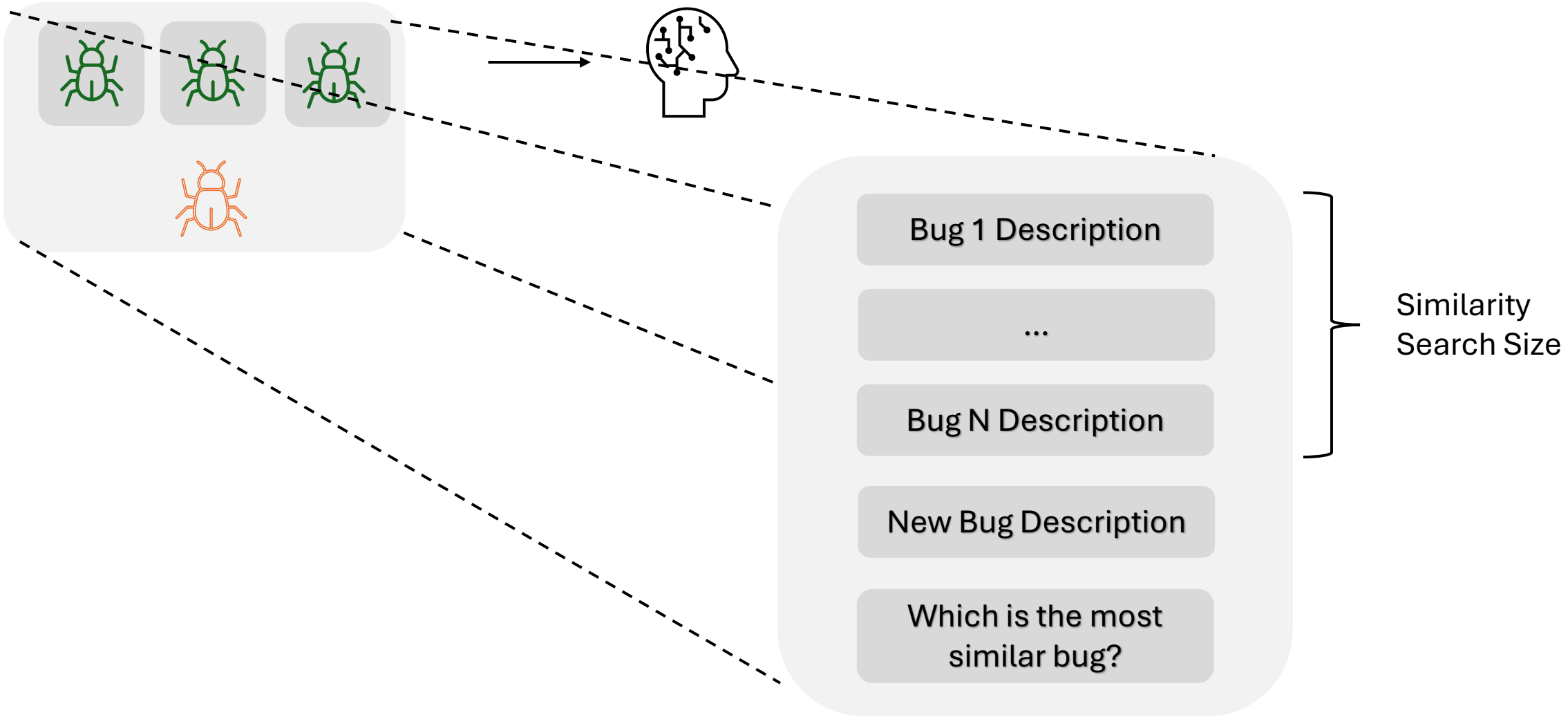
Feature Selection: docker performance counters

Embedding: vectorizing using statistics of each feature.

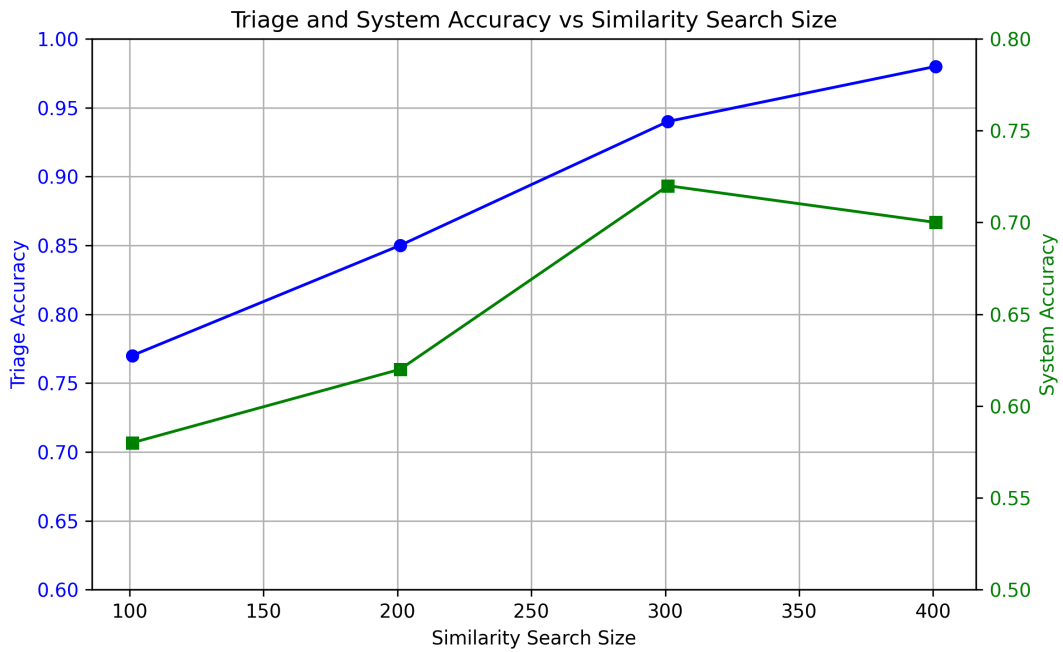
Searching:

KNN using cosine similarity.

Issue Descriptions



Findings


Triage and System Accuracy vs Similarity Search Size
Triage Accuracy
System Accuracy
Similarity Search Size
1.00
0.95
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.80
0.75
0.70
0.65
0.60
0.55
0.50
100
150
200
250
300
350
400

Avg. Single Query Costs and Time vs Similarity Search Size
Avg. Single Query Costs (USD)
Avg. Single Query Time (Seconds)
Similarity Search Size
0.040
0.035
0.030
0.025
0.020
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
100
150
200
250
300
350
400

ARCA as a Log Clustering Tool

- ARCA's log processing module alone.
- Tested on super computing site logs.

Data Set	F1-Score	Recall	Precision
BGL	**0.995**/0.976	**0.99**/0.982	**1**/0.970
Thunderbird	**0.984**/0.97	0.975/**0.99**	**1**/0.97
Spirit	**0.993**/0.992	0.986/**0.999**	**1**/0.984
Liberty	0.986/*	0.986/*	0.986/*

More challenges

Different organizations of the knowledge base.

Coverage for more data modes.

Agentic design.