Agents for Data Streaming: The Missing Pieces

Shreesha G. Bhat*, Landon Johnson*, Michael Noguera*, Aishwarya Ganesan, Ramnatthan Alagappan
University of Illinois Urbana-Champaign

Introduction

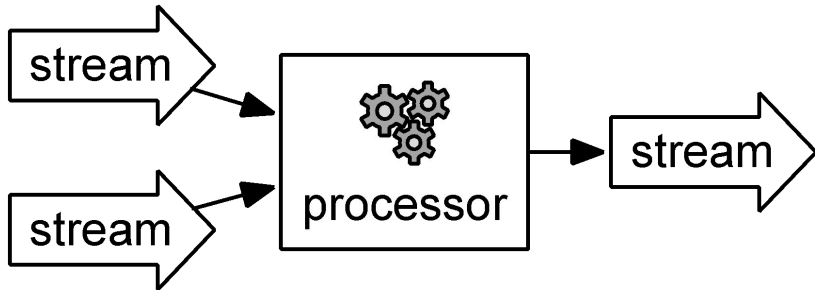
- Agents are increasingly using data systems
- An important class of data systems has not yet evolved

Real Time Inputs

Application metrics

IoT Sensor Data

User Actions (e.g. purchases)



Real Time Outputs

Dashboards

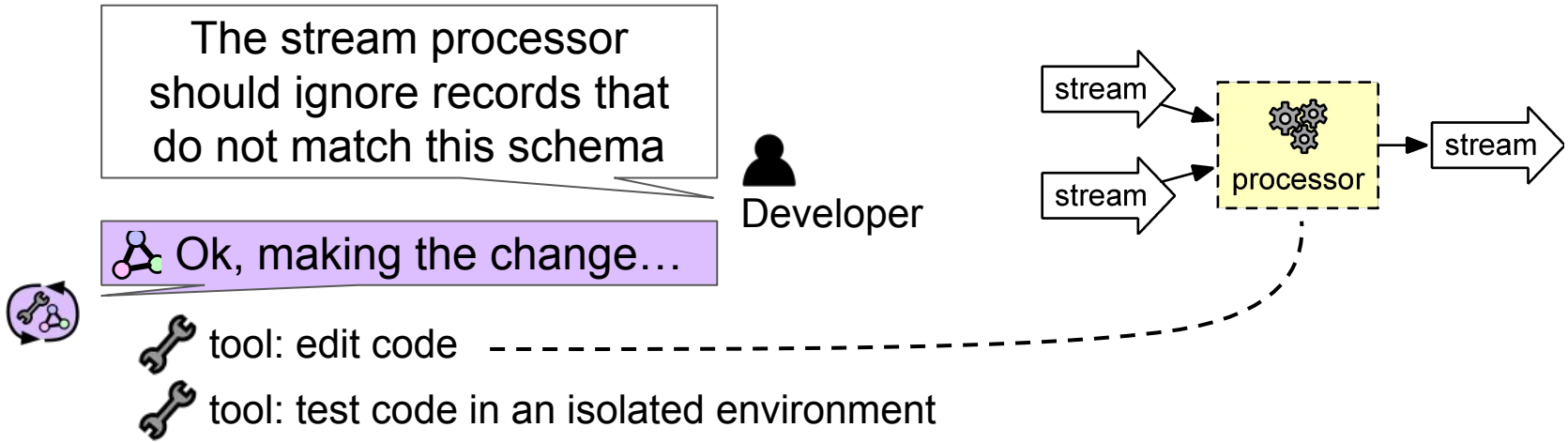
Live Inventory

Forecasts for Decision Making

- Widely used in Finance, IoT, Health…

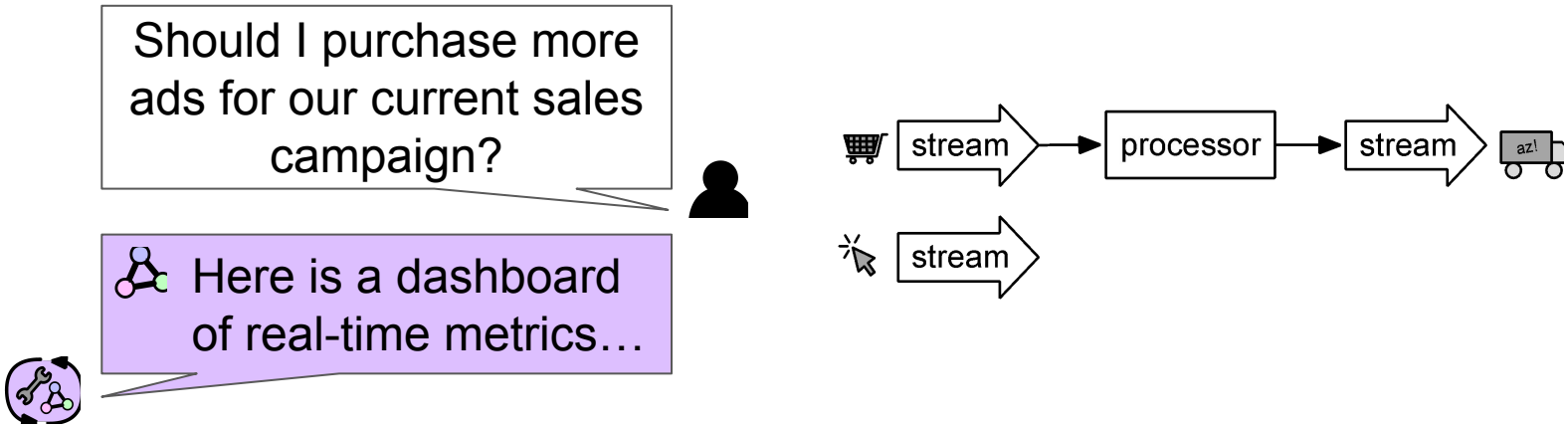
Modalities of Agent↔Stream Interaction

1 Software **developers can use agents** to edit existing stream pipelines and deploy new stream pipelines



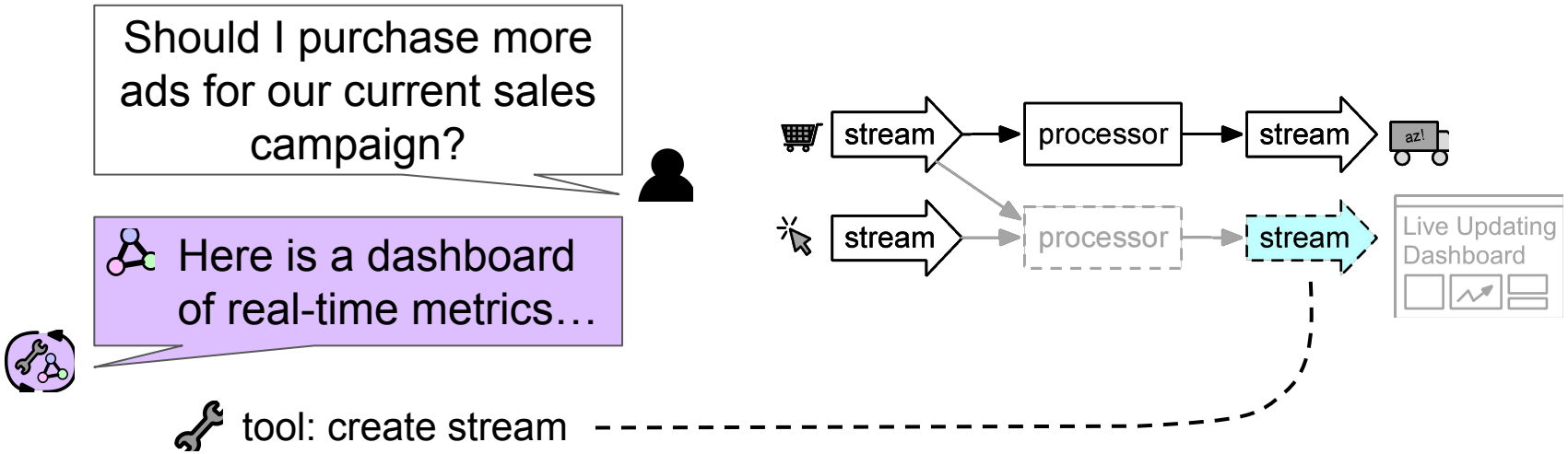
Modalities of Agent↔Stream Interaction

2 An agent can **use stream infrastructure as tools** to answer user questions



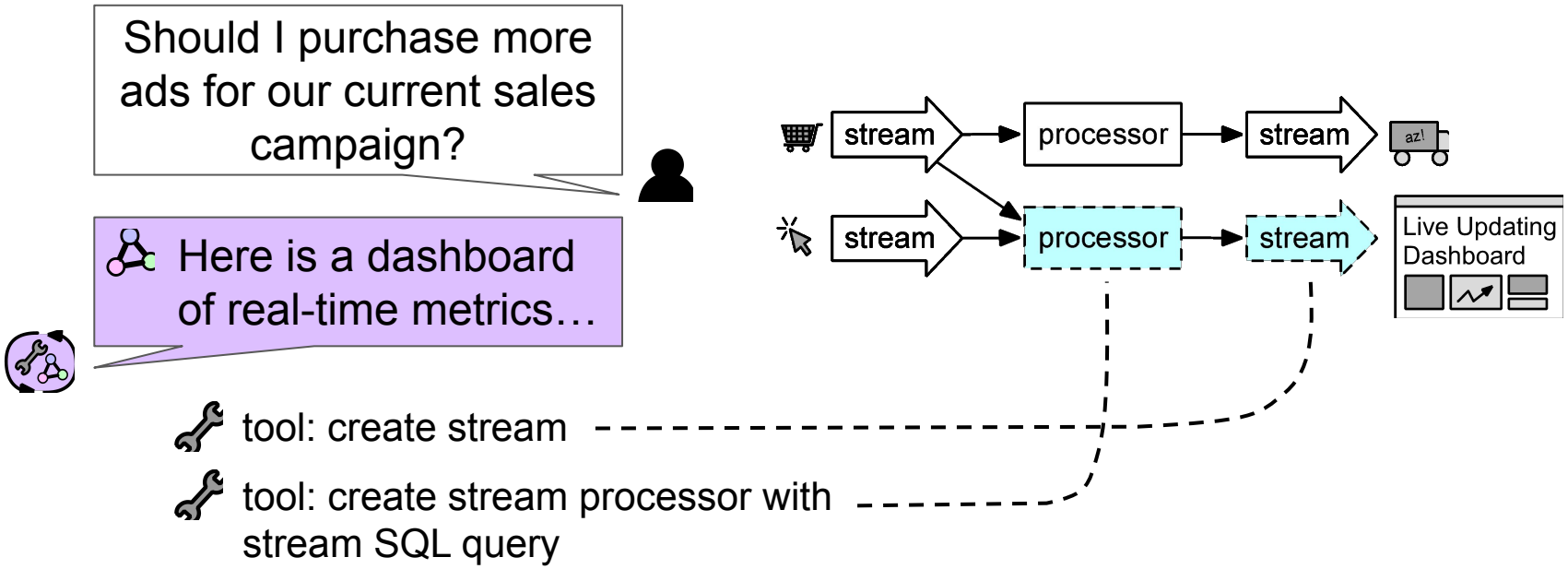
Modalities of Agent↔Stream Interaction

2 An agent can **use stream infrastructure as tools** to answer user questions



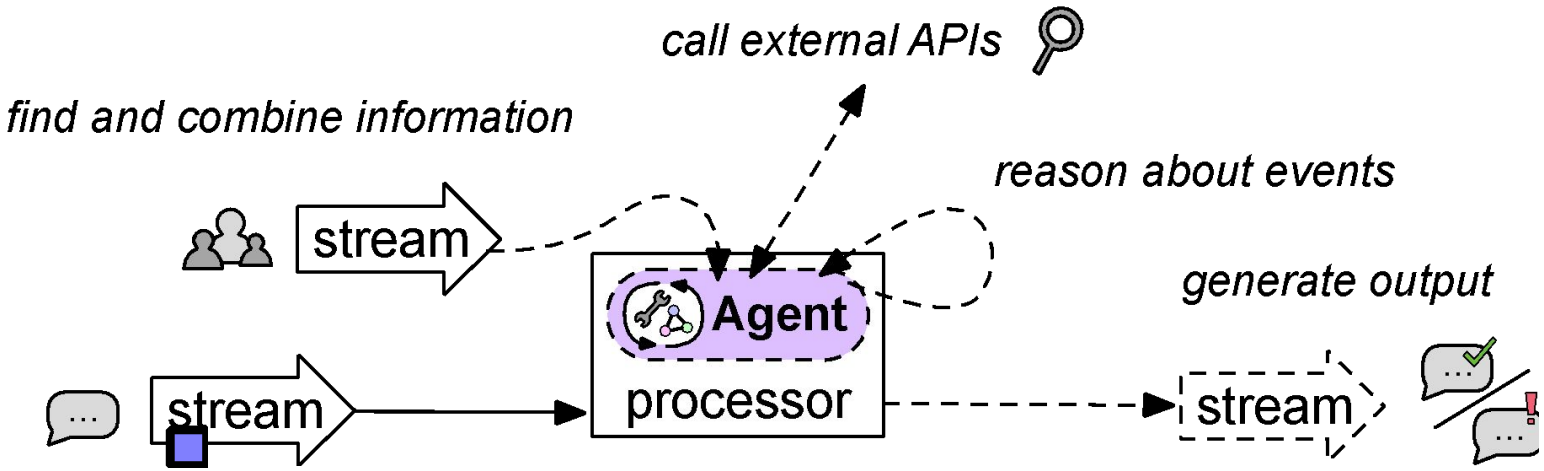
Modalities of Agent↔Stream Interaction

2 An agent can **use stream infrastructure as tools** to answer user questions



Modalities of Agent↔Stream Interaction

3 An **embedded agent** can autonomously process stream events



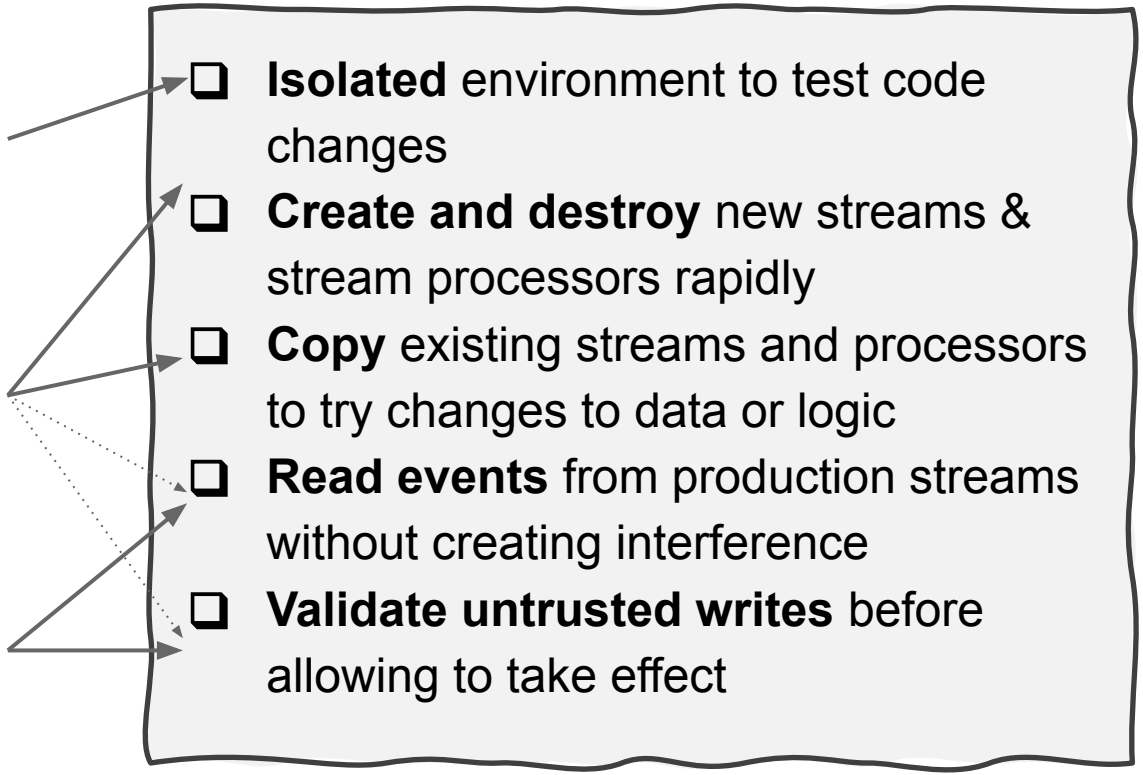
Ex. Identify user comments that violate the content policy

Unique Requirements of Stream Agents

1. **Developer uses agent** to edit existing/deploy new streaming pipelines

2. Agent uses **stream infrastructure as tools** to answer user's question

3. **Embedded agent** autonomously processes stream events



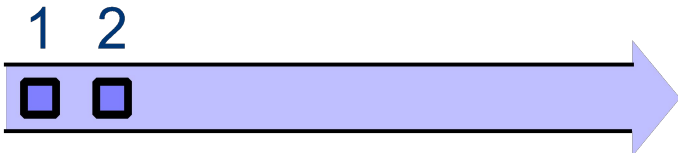
Stream Infrastructure for Stream Agents

- ❑ Isolated environments
- ❑ Create and destroy streams & stream processors rapidly
- ❑ Copy streams & SPs
- ❑ Read without interference
- ❑ Validate writes before allowing

Our answer to these requirements is a ***forkable stream***.

AgileLog [1] is the first step, adding forks to shared log stream storage.

- Each agent can operate on its own isolated fork



Next steps include forking stream processors with state, then whole pipelines.

[1] Shreesha G Bhat, Tony Hong, Michael Noguera, Ramnatthan Alagappan, and Aishwarya Ganesan. Agilelog: A Forkable Shared Log for Agents on Data Streams. *arXiv preprint arXiv:2604.14590*, 2026.

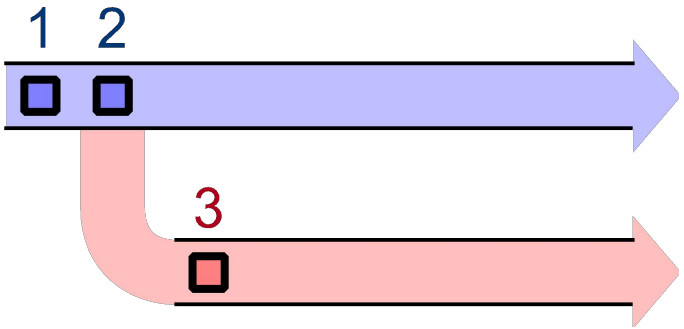
Stream Infrastructure for Stream Agents

- ❑ Isolated environments
- ❑ Create and destroy streams & stream processors rapidly
- ❑ Copy streams & SPs
- ❑ Read without interference
- ❑ Validate writes before allowing

Our answer to these requirements is a ***forkable stream***.

AgileLog [1] is the first step, adding forks to shared log stream storage.

- Each agent can operate on its own isolated fork



Next steps include forking stream processors with state, then whole pipelines.

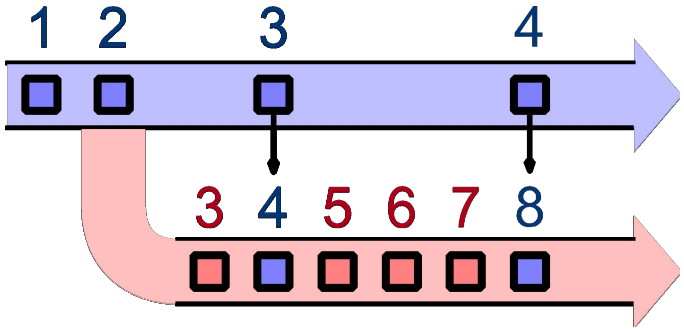
[1] Shreesha G Bhat, Tony Hong, Michael Noguera, Ramnatthan Alagappan, and Aishwarya Ganesan. Agilelog: A Forkable Shared Log for Agents on Data Streams. *arXiv preprint arXiv:2604.14590*, 2026.

Stream Infrastructure for Stream Agents

- ❑ Isolated environments
- ❑ Create and destroy streams & stream processors rapidly
- ❑ Copy streams & SPs
- ❑ Read without interference
- ❑ Validate writes before allowing

Our answer to these requirements is a ***forkable stream***.

AgileLog [1] is the first step, adding forks to shared log stream storage.



- Each agent can operate on its own isolated fork
- Child streams can see parent changes using *continuous forks*

Next steps include forking stream processors with state, then whole pipelines.

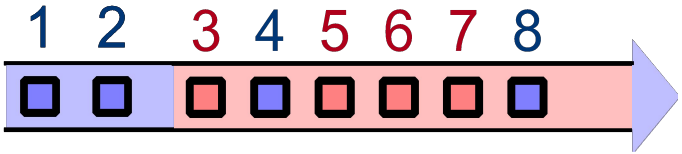
[1] Shreesha G Bhat, Tony Hong, Michael Noguera, Ramnatthan Alagappan, and Aishwarya Ganesan. Agilelog: A Forkable Shared Log for Agents on Data Streams. *arXiv preprint arXiv:2604.14590*, 2026.

Stream Infrastructure for Stream Agents

- ❑ Isolated environments
- ❑ Create and destroy streams & stream processors rapidly
- ❑ Copy streams & SPs
- ❑ Read without interference
- ❑ Validate writes before allowing

Our answer to these requirements is a ***forkable stream***.

AgileLog [1] is the first step, adding forks to shared log stream storage.



- Each agent can operate on its own isolated fork
- Child streams can see parent changes using *continuous forks*
- Optionally *promote* children (for details see AgileLog)

Next steps include forking stream processors with state, then whole pipelines.

[1] Shreesha G Bhat, Tony Hong, Michael Noguera, Ramnatthan Alagappan, and Aishwarya Ganesan. Agilelog: A Forkable Shared Log for Agents on Data Streams. *arXiv preprint arXiv:2604.14590*, 2026.

Benchmarks

- Infrastructure changes

Benchmarks

Infrastructure changes

Benchmarks

Infrastructure changes

Measurements

Benchmarks

Infrastructure changes

Measurements

	Spider	KaggleDBQA	BIRD
Exploratory	○	○	○

Benchmarks

Infrastructure changes

Measurements

		Spider	KaggleDBQA	BIRD	Spider 2.0	BEAVER	CoSQL	DAB	AgentFuel
Feature	Exploratory	○	○	○	●	○	○	●	○
	Conversational	○	○	○	○	○	●	○	○
	Enterprise Data	○	○	○	○	●	○	●	●
	Real-Time Queries	○	○	○	○	○	○	○	○

Benchmarks

Infrastructure changes

Measurements

Feature	Spider	KaggleDBQA	BIRD	Spider 2.0	BEAVER	CoSQL	DAB	AgentFuel
Exploratory	○	○	○	●	○	○	●	○
Conversational	○	○	○	○	○	●	○	○
Enterprise Data	○	○	○	○	●	○	●	●
Real-Time Queries	○	○	○	○	○	○	○	○

Benchmarks

Infrastructure changes

Measurements

Feature	Spider	KaggleDBQA	BIRD	Spider 2.0	BEAVER	CoSQL	DAB	AgentFuel
Exploratory	○	○	○	●	○	○	●	○
Conversational	○	○	○	○	○	●	○	○
Enterprise Data	○	○	○	○	●	○	●	●
Real-Time Queries	○	○	○	○	○	○	○	○

Windows

Benchmarks

Infrastructure changes

Measurements

		Spider	KaggleDBQA	BIRD	Spider 2.0	BEAVER	CoSQL	DAB	AgentFuel
Feature	Exploratory	○	○	○	●	○	○	●	○
	Conversational	○	○	○	○	○	●	○	○
	Enterprise Data	○	○	○	○	●	○	●	●
	Real-Time Queries	○	○	○	○	○	○	○	○

Windows

Time-Sensitive Processing

Benchmarks

Infrastructure changes

Measurements

Feature	Spider	KaggleDBQA	BIRD	Spider 2.0	BEAVER	CoSQL	DAB	AgentFuel
Exploratory	○	○	○	●	○	○	●	○
Conversational	○	○	○	○	○	●	○	○
Enterprise Data	○	○	○	○	●	○	●	●
Real-Time Queries	○	○	○	○	○	○	○	○

Windows

Time-Sensitive Processing

Reordering

Benchmarks

Infrastructure changes

Measurements

Feature	Spider	KaggleDBQA	BIRD	Spider 2.0	BEAVER	CoSQL	DAB	AgentFuel
Exploratory	○	○	○	●	○	○	●	○
Conversational	○	○	○	○	○	●	○	○
Enterprise Data	○	○	○	○	●	○	●	●
Real-Time Queries	○	○	○	○	○	○	○	○

Windows

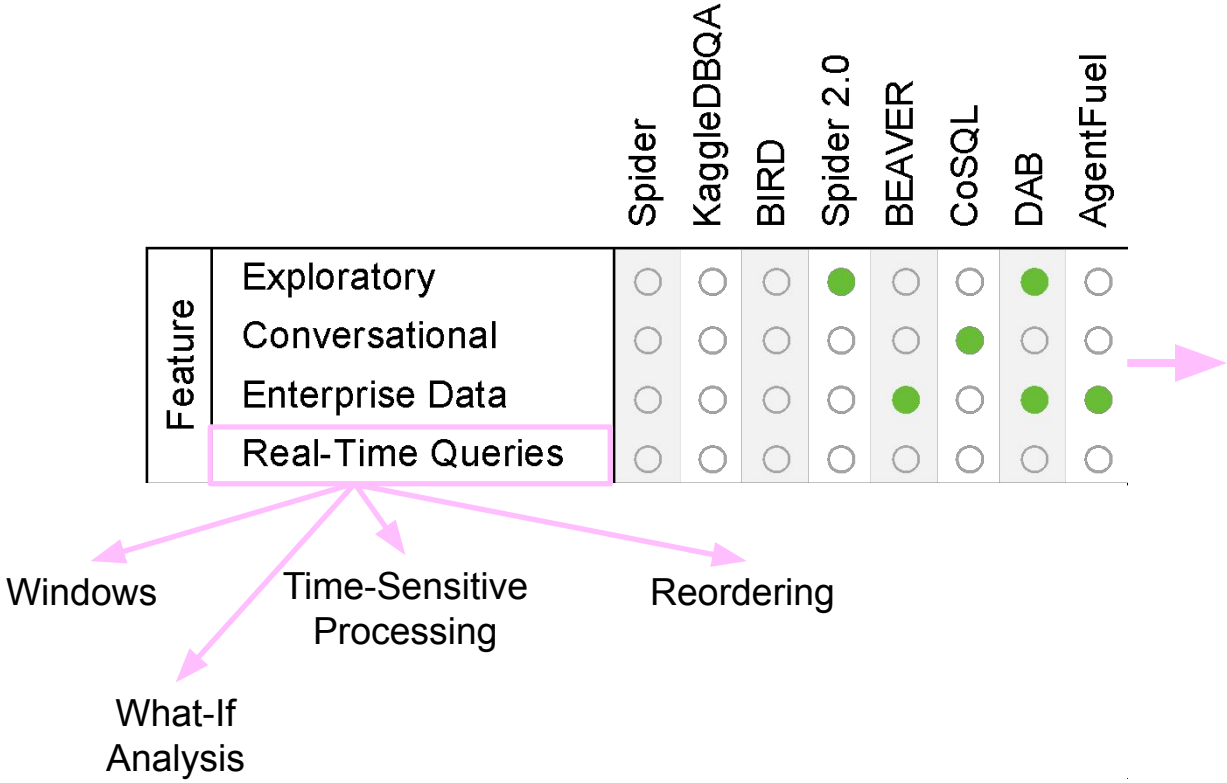
Time-Sensitive Processing

Reordering

Benchmarks

Infrastructure changes

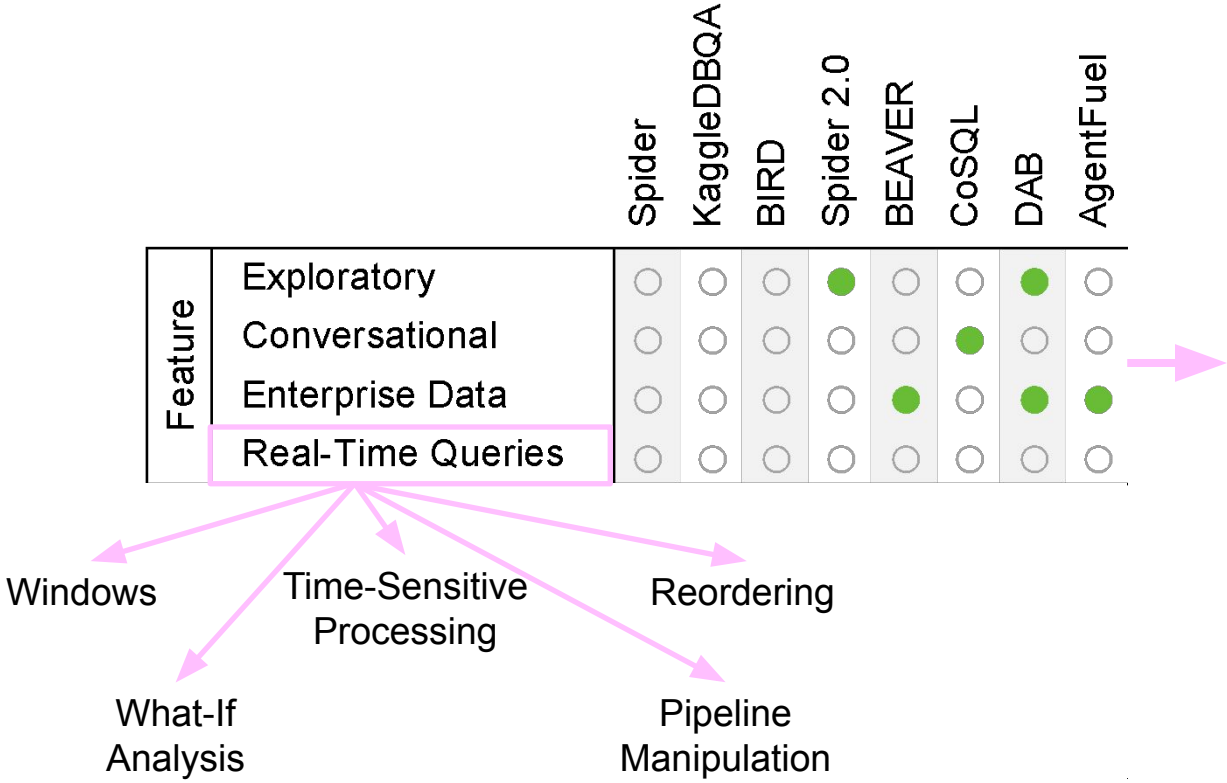
Measurements



Benchmarks

Infrastructure changes

Measurements



Benchmarks

Infrastructure changes

Measurements

		Spider	KaggleDBQA	BIRD	Spider 2.0	BEAVER	CoSQL	DAB	AgentFuel
Feature	Exploratory	○	○	○	●	○	○	●	○
	Conversational	○	○	○	○	○	●	○	○
	Enterprise Data	○	○	○	○	●	○	●	●
	Real-Time Queries	○	○	○	○	○	○	○	○
Evaluation	Accuracy	●	●	●	●	●	●	●	●
	Time-to-Plan	○	○	○	○	○	○	●	○
	Plan Performance	○	○	●	○	○	○	●	○
	Tokens	○	○	○	○	○	○	●	○
	Resource Utilization	○	○	○	○	○	○	○	○

ILLINOIS

DASSL: Distributed And Storage Systems Lab

Read our workshop paper!