EIRES: Efficient Integration of Remote Data in Event Stream Processing

Bo Zhao¹, Han van der Aa², Thanh Tam Nguyen ³, Quoc Viet Hung Nguyen⁴, Matthias Weidlich¹

¹Humboldt-Universität zu Berlin ²Universität Mannheim ³Leibniz Universität Hannover ⁴Griffith University
bo.zhao@hu-berlin.de; han@informatik.uni-mannheim.de; tamnguyen@l3s.de; quocviethung.nguyen@griffith.edu.au; weidlima@hu-berlin.de

ABSTRACT
To support reactive and predictive applications, complex event processing (CEP) systems detect patterns in event streams based on predefined queries. To determine the events that constitute a query match, their payload data may need to be assessed together with data from remote sources. Such dependencies are problematic, since waiting for remote data to be fetched interrupts the processing of the stream. Yet, without event selection based on remote data, the query state to maintain may grow exponentially. In either case, the performance of the CEP system degrades drastically.

To tackle these issues, we present EIRES, a framework for efficient integration of static data from remote sources in CEP. It employs a cost-model to determine when to fetch certain remote data elements and how long to keep them in a cache for future use. EIRES combines strategies for (i) prefetching that queries remote data based on anticipated use and (ii) lazy evaluation that postpones the event selection based on remote data without interrupting the stream processing. Our experiments indicate that the combination of these strategies improves the latency of query evaluation by up to 3,752× for synthetic data and 47× for real-world data.

CCS CONCEPTS
- Information systems → Data streams.

KEYWORDS
complex event processing; data prefetching; query result caching

ACM Reference Format:

1 INTRODUCTION
Complex event processing (CEP) systems evaluate queries over continuous streams of events to detect patterns [1]. They provide the ability to identify a situation of interest with low latency, thereby supporting reactive and predictive applications [2]. A CEP query describes an event pattern as the joint occurrence of events of particular types, potentially in a specific ordering, correlated through predicates over their data payload and a time window. By evaluating such a query in an online manner, an event pattern is detected in near-real-time, as soon as it materializes.

The evaluation of CEP queries is computationally challenging. A CEP system needs to maintain a set of partial matches per query [3–5], which may grow exponentially in the number of processed events. Hence, common evaluation algorithms show an exponential (worst-case) time complexity [6]. Recognising this, optimization techniques based on state sharing [8, 7], query rewriting [8, 9], or load shedding [10, 11] have been proposed. These optimizations assume that the occurrence of a pattern is fully characterised by the events and their payload data. Yet, many applications require the combination of the events’ payload with data from remote sources to determine whether a partial match shall be processed further.

Consider the detection of fraud in credit card usage [12]. Here, events denote financial transactions or indicate status changes, such as an increase in spending limit. To block malicious transactions, CEP queries aim to identify patterns of suspicious use. Such patterns must be detected under tight latency bounds since, e.g., a credit card transaction needs to be cleared within 25ms [13]. Yet, the detection of suspicious patterns often depends on contextual information, such as a user’s spending history, transactional volume at a specific location, or the behaviour of other users. Since this information must be retrieved from remote data sources, incorporating it while still meeting the latency bounds becomes highly challenging.

A naive integration of remote data reduces the performance of a CEP system drastically. Fetching data only once it is needed for query evaluation, see Figure 1 (top), interrupts the processing of the stream. As we later demonstrate in our experiments, such an interruption temporarily increases the latency by orders of magnitude. Even a small latency of dozens of milliseconds to look up remote data is problematic, given that CEP systems achieve microsecond latencies and many applications enforce tight latency bounds. Although the assessment of remote data is mandatory to maintain result correctness, only incorporating it in a post-processing step, while ignoring it as part of event selection during query evaluation, is also not a viable option: The resulting non-determinism in query evaluation would add further to the exponential growth of partial matches, thereby increasing the processing latency.

In this paper, we present EIRES, a framework to address use cases such as the above one through an efficient integration of remote data in event stream processing. As shown in Figure 1, our idea is to decouple (i) the fetching of remote data from (ii) its use in query evaluation. While under a naive model, data is fetched once it is needed and then used immediately, EIRES employs a cache to decouple both operations.
2.1 A Model for CEP with Remote Data

Event streams. We adopt a common, relational model of event streams. An event is an occurrence of interest at a specific point in time, which is instantaneous, unique, and atomic. Its structure is defined by a schema, as was first proposed in traditional data stream processing [14]. An event schema is given by a sequence of attributes \( A = (A_1, \ldots, A_n) \), where each attribute is assigned a primitive data type. Then, an event \( e = (a_1, \ldots, a_n) \) is an instance of this schema, where \( a_i \) is the value of attribute \( A_i \). Moreover, events carry timestamps from a discrete, totally-ordered domain, defined by \( \mathbb{N} \). The timestamp of an event \( e \) is referred to by \( e.t \). Without loss of generality, we assume all events to have the same schema. Note that different types of events, as used later in our running example, are incorporated by predicates over a distinguished attribute.

An event stream is an infinite sequence of events, \( S = \langle e_1, e_2, \ldots \rangle \), that respects the order of event timestamps. For two events \( e_i \) and \( e_j \), it holds that \( i < j \) implies \( e_i.t \leq e_j.t \). We further define the notion of a finite stream prefix, up to index \( k \), as \( S(k) = \langle e_1, \ldots, e_k \rangle \).

Remote data. The evaluation of a CEP query may require data from remote sources. Here, we focus on what data elements are fetched from such sources, rather than how they are retrieved. That is, we abstract from the specific look-up queries executed at remote sources and use data elements as the basis for our model. A data element may be thought of as a key-value pair or a relational tuple. Formally, we use a set \( D = \{d_1, \ldots, d_n\} \) to capture the remote data elements. Moreover, we write \( |d| \in \mathbb{N} \) for the size of a data element.

Notably, data models are hierarchical in many applications, which means that there exists a containment relation between data elements. Therefore, we also consider a partial function \( \rho : D \rightarrow D \) that maps an element to another one if the former is contained in the latter. The size of the containing element is then given as the sum of the contained elements, i.e., \( |d| = \sum_{d' \in D, \rho(d') = d} |d'| \).

Since fetching data elements from different sources can result in different transmission latencies, we here assume that latency is monitored per data element, and denoted as \( \ell_{remote}(d) \).

CEP queries. Common query languages describe event patterns through operators (e.g., sequence), conditions based on data of the events’ payload or from remote sources, and a time window \([1, 15]\).

For the use case of financial fraud detection, an example query is shown in Listing 1, using a syntax similar to the SASE language [3]. It abstracts from the specific queries sent to remote databases, but highlights which predicates involve remote data and which payload data of events is used for the respective lookup. In this scenario, remote data captures the known locations of credit card usage per client, the card limits, and the set of pre-authorized clients for an organization. The latter is organized hierarchically, i.e., it can be fetched per credit card, per user, or for the whole organization.

The query is triggered by a transaction event \( t_1 \) (of type \( t \)) of high volume (>10k). A suspicious pattern may emerge in two ways based on other events recorded for the same credit card (SAME[cc]):

First, the event may be followed by a denied transaction \( s \) (type \( d \)) and another high-volume transaction \( t_2 \) that occurred at a location that both differs \((t_1.loc \neq t_2.loc)\) and is not in the set of known locations for the user, which is fetched from a remote database \((REMOTE[t_1.user])\).
Listing 1: Query to detect fraudulent transactions.

\[
\begin{align*}
\sigma_0: & \quad x.\text{time} < q_1.\text{time} + 5\text{min} \wedge x.\text{cc} = q_1.\text{cc} \\
\sigma_1: & \quad x.\text{type} = T \wedge x.\text{vol} > 10\text{k} \\
\sigma_2: & \quad \sigma_0 \wedge x.\text{type} = D \\
\sigma_3: & \quad \sigma_0 \wedge \sigma_1 \wedge r[x.\text{org}] \\
\sigma_4: & \quad \sigma_0 \\
\sigma_5: & \quad \sigma_0 \wedge \sigma_1 \\
\neg \sigma_2 & \quad \neg \sigma_4 \\
\neg \sigma_3 & \quad \neg \sigma_5 \\
\end{align*}
\]

Figure 2: Evaluation model for the query of Listing 1.

Second, the initial event may be followed by a change in the spending limit (event 1 of type 1), where the new limit is larger than the maximum limit of all credit cards within the same organization (queried from a remote database, \(\text{REMOTE}([t.\text{org}])\)). Afterwards, another transition is observed with a very high volume (>5k), for which the beneficiary ([t.\text{ben}]) is not in the set of pre-authorized clients, as stored at a remote data source (\(\text{REMOTE}([t.\text{user}])\)).

To evaluate such CEP queries, various formalisms have been proposed, many of them being based on automata [5, 16, 17]. Figure 2 exemplifies an evaluation model for the query of Listing 1. It defines states and state transitions to describe how a CEP system constructs matches of the query when processing a stream event by event. To this end, it maintains a set of partial matches (i.e., partial runs of the automaton). However, due to non-determinism in the automaton, the number of partial matches is potentially exponential in the number of processed events. Given the current set of partial matches, a CEP system checks how a new event of the stream changes the partial matches, i.e., whether it leads to a partial match being discarded, extended, or split up.

The evolution of partial matches is influenced by the time window, the predicates, and event processing policies of the query. Transitions in an automata are guarded and link the current input event (denoted by \(x\)), events of partial matches (denoted by state identifiers, such as \(q_1\)), and data from remote sources (denoted by a parametrised variable, such as \(r[q_1.\text{user}]\)). For example, predicate \(\sigma_1\) in Figure 2 checks the time window, \(x.\text{time} < q_1.\text{time} + 5\text{min}\) (i.e., \(\sigma_1\)), a condition over the event's payload data, \(x.\text{type} = T \wedge x.\text{vol} > 10\text{k}\) (i.e., \(\sigma_1\)), and a condition using remote data, \(x.\text{loc} \notin r[q_1.\text{user}]\).

Event processing policies fine-tune the consumption and selection of events [18]. Specifically, we consider a greedy policy, under which a partial match that, based on the transition guards in the automaton, can be extended with an input event from the stream is always split up into at least two partial matches, one extended with the event and one left unchanged. The latter models the case that an input event may be skipped, to derive matches constructed from any set of events that satisfy the predicates. This semantics is known as unconstraint [19] or skip-till-any-match [3].

We also consider a non-greedy policy, also known as continuous or skip-till-next-match. Here, the partial match would only be extended, i.e., only events that do not satisfy the predicates are skipped to always select the next event that satisfies them.

**Formal evaluation model.** We formalise the above model for the evaluation of CEP queries as follows. Let \(Q\) be a query and \(\tau_0\) its time window. Then, the result of evaluating \(Q\) over a stream \(S = (e_1, e_2, \ldots)\) are matches. A match is a finite sequence of events \((e'_1, e'_2, \ldots, e'_m)\) of the stream that is order-preserving, i.e., for \(e'_i \equiv e_k\) and \(e'_j = e_l\) it holds that \(i < j \leftrightarrow k < l\), and respects the time window of the query, i.e., \(e'_m \cdot t \leq \tau_0\).

The set of partial matches that is maintained by the CEP system for query \(Q\), after processing a stream prefix \(S(\cdot, k)\), is denoted as \(P(k) = (e_1, e_2, \ldots, e_m)\) when processing event \(S(k+1)\) by a set \(P(k, p \cdot k + 1) \subseteq D\). All such elements are given by \(D(k+1) = \bigcup_{p \in P(k)} D(p, k+1)\). Based thereon, the functionality of a CEP system can be described as a function that takes the event \(S(k+1)\), the current partial matches \(P(k)\), and the required remote data \(D(k+1)\) as input and returns sets of new partial matches \(P(k+1)\) and (complete) matches \(C(k+1): \)

\[f_Q(S(k+1), P(k), D(k+1)) \mapsto P(k+1), C(k+1).\] (1)

Applying this function repeatedly for a stream constructs a stream of sets of matches, \(R = (C(1), C(2), \ldots)\). To achieve compositional-ity of the model, one may order the matches per set and construct a single event per match. This way, \(R\) is transformed into an event stream again. Table 1 summarises our notations.

**Table 1: Notations.**

<table>
<thead>
<tr>
<th>Notation</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(e = (a_1, \ldots, a_n))</td>
<td>Event</td>
</tr>
<tr>
<td>(t)</td>
<td>Event timestamp</td>
</tr>
<tr>
<td>(S = (e_1, e_2, \ldots))</td>
<td>Event stream</td>
</tr>
<tr>
<td>(S(\cdot, k))</td>
<td>Event stream prefix at up to the (k)-th input event</td>
</tr>
<tr>
<td>(D = {d_1, \ldots, d_n})</td>
<td>Remote data elements</td>
</tr>
<tr>
<td>(\rho : D \mapsto D)</td>
<td>Part-of relation for data elements</td>
</tr>
<tr>
<td>(P(k))</td>
<td>Partial matches up to the (k)-th input event</td>
</tr>
<tr>
<td>(D(p, k), D(k))</td>
<td>Remote data needed by partial match (p), or all partial matches, to process the (k)-th input event</td>
</tr>
<tr>
<td>(C(k))</td>
<td>Complete matches up to the (k)-th input event</td>
</tr>
<tr>
<td>(R)</td>
<td>Output stream of complete matches</td>
</tr>
<tr>
<td>(\ell(k))</td>
<td>Query evaluation latency after the (k)-th input event</td>
</tr>
<tr>
<td>(\ell_{\text{remote}}(d))</td>
<td>Transmission latency for remote data element (d)</td>
</tr>
</tbody>
</table>

**2.2 Problem Statement.**

Query evaluation incurs latency—the time between the arrival of the last event of a match at the CEP system and the actual detection of the match. In our model, this corresponds to the time needed to evaluate function \(f_Q\). We denote the latency observed for a match \(e\) by \(\ell(e)\). For a particular point in time, i.e., when all matches \(C(k)\) have been generated, the latency is derived through some aggregation function \(a\) (e.g., median) from all these matches, \(\ell(k) = a(\ell(e) \mid e \in C(k))\). In practice, it is common to incorporate some smoothing in the latency assessment (e.g., using a sliding window).

We assume such smoothing to be part of the definition of \(\ell(k)\).

Based thereon, we formulate the problem of efficiently evaluating a CEP query as the minimization of the respective latency.

**Problem 1 (Efficient Query Evaluation).** Let \(Q\) be a CEP query and let \(S(\cdot, k)\) be a stream prefix. The problem of efficient query evaluation is to compute all matches \(R = (C(1), \ldots, C(k))\) of \(Q\), while minimizing the latency \(\ell(k)\), with \(k \rightarrow \infty\).
3 THE EIRES FRAMEWORK

To address Problem 1, we propose the EIRES framework. Below, we summarize its intuition, before outlining its components.

**Intuition.** We first reflect on the factors that constitute the latency $\ell(c)$ of match $c$, and hence, the aggregated latency $\ell(k)$. A first aspect is the latency that is inherent to the evaluation of partial matches when processing all events from the first to the last event of match $c$, denoted by $\ell_{\text{match}}(c)$. In addition, there is the latency of fetching remote data that was required to process these events, denoted by $\ell_{\text{fetch}}(c)$. It is the sum of the latencies of all fetch operations required for $c$, each defined by two timestamps $(t_n, t_a)$, i.e., the time $t_n \in \mathbb{N}$ at which the need for remote data is detected during query evaluation and the time $t_a \in \mathbb{N}$ at which the data is available and processing can continue. With $F(c) \subseteq \mathbb{N} \times \mathbb{N}$ as the set of the fetch operations during the construction of match $c$, we have:

$$\ell(c) = \ell_{\text{match}}(c) + \ell_{\text{fetch}}(c) = \ell_{\text{match}}(c) + \sum_{(t_n, t_a) \in F(c)} t_a - t_n \quad (2)$$

Existing optimizations for CEP, such as those based on state sharing [6, 7], rewriting [8, 9], or load shedding [10, 11], will ultimately improve the inherent latency of query evaluation, i.e., $\ell_{\text{match}}(c)$.

Figure 3: Intuition of the proposed strategies.

By contrast, we focus on the latency induced by fetching of remote data, $\ell_{\text{fetch}}(c)$. As this latency corresponds to the time interval between the need $(t_n)$ and availability of remote data $(t_a)$, Figure 3 illustrates that it may be reduced from either end. Here, we assume that the latency of the actual transmission of a data element $d$, i.e., $\ell_{\text{remote}}(d)$, as induced by the requests and responses sent over the network, is monitored. The latency $\ell_{\text{remote}}(d)$ may be hidden by either moving $t_n$ closer to $t_a$, i.e., by fetching the data earlier than it is actually needed, or by moving $t_a$ closer to $t_n$, by postponing the evaluation of the predicates that are based on the remote data. We realize both ideas with a strategy for prefetching, coined PFetch, and a strategy for lazy evaluation, referred to as LzEval.

PFetch fetches remote data before it is actually needed, thereby hiding the data transmission latency. Data is then kept in a local cache at the CEP engine, from which it may be retrieved with negligible latency. PFetch needs to realize two main operations:

(L1) Decide on the partial matches for which lazy evaluation is applied, i.e., for which the evaluation of predicates is postponed.

(L2) Adapt the evaluation procedure for the partial matches with lazy evaluation. Verify the predicates based on remote data once the remote data is available in the cache.

Here, the challenge is that although postponing the evaluation of predicates can hide the transmission latency of remote data, $\ell_{\text{remote}}(d)$, it may simultaneously increase the inherent evaluation latency, $\ell_{\text{match}}(c)$, since postponement makes event selection less strict and thus results in the creation of additional partial matches. Therefore, LzEval shall only be applied as long as its benefits outweigh the associated increase in the evaluation latency.

Both strategies have in common that they require cache management, which corresponds to the following operation:

(C1) Maintain a set of data elements in the cache based on its current content and newly fetched elements.

Cache management calls for a mechanism to use the available storage optimally, which again requires to assess the impact that data elements have on the performance of query evaluation.

**EIRES components.** To realize the above ideas, the EIRES framework comprises three main components. As shown in Figure 4, these components extend the functionality of a CEP engine.

![Figure 4: Components of the EIRES framework.](image)

Utility modelling (§4) provides a cost model to assess the (expected) utility of remote data elements for query evaluation. As such, it provides the basis to select data elements for prefetching (P2); to decide on the partial matches for which lazy evaluation is applied (L1); and to govern the cache management (C1).

Remote data fetching (§5) realizes the PFetch and LzEval strategies and manages the trade-offs induced by operations (P1) and (P2), and by (L1) and (L2), respectively, using the cost model. The trade-offs are managed in an online manner, i.e., based on the information available at a specific point in time. The reason being that an optimal plan to combine PFetch and LzEval for some future states would require the utility of data elements to be stable. Yet, since the utility is based on the maintained partial matches, it may differ at the time a (pre)fetching decision is taken and at the time the data arrives at the CEP engine. Hence, the operations for prefetching and lazy evaluation are decoupled from the cache management.

Cache management (§6) realizes operation (C1), i.e., it retains the data elements in the cache that are most beneficial according to the cost model. Following the above argument on the infeasibility of optimal plans, cache management is also conducted online, using the information currently available.
We first present measures to assess the utility of remote data which is further detailed in Alg. 2. It then computes remote data (line 3), EIRES performs operation P2 in PFetch, prefetching remote required as input for processing. For a data element $d \in D$, at this point in time, the urgent utility is defined as the number of partial matches that require $d$ or one of its constituents (with $p^*$ as the reflexive transitive closure of the part-of-relation $\rho$), weighted by its transmission latency $\ell_{\text{remote}}(d)$:

$$UU(d, k) = \ell_{\text{remote}}(d) \cdot |\{p \in P(k) \mid \exists d' \in D(p, k + 1) : d \in \rho^*(d')\}|.$$  (3)

In the same vein, the number of partial matches that require $d$, or its constituents, can be considered for some future state of query evaluation. Given perfect information about the future stream up to a stream index $k' > k$, the future utility of $d$ is given by the urgent utilities of future partial matches in $k < i \leq k'$:

$$FU(d, k, k') = \sum_{k < i \leq k'} UU(d, i)$$  (4)

However, since information about future states is inherently uncertain, we actually need to compute an estimate for the future utility up to a time point $k'$, which we denote as $\hat{FU}(d, k, k')$. As such, in the remainder we rely on an overall utility that is defined as the weighted sum of the above measures, defined, with $\omega \in [0, 1]$, as:

$$U(d, k, k') = \omega \cdot UU(d, k) + (1 - \omega) \cdot \hat{FU}(d, k, k')$$  (5)

Here, the rational is that the weighting enables tuning of the respective importance of the known utility $UU$ and the estimated utility $\hat{FU}$. Moreover, different weighting schemes are applied when incorporating the utility assessment for the realization of PFetch, LzEval, and cache management. Specifically, the strategies for fetching remote data, PFetch and LzEval, assign higher weights to the urgent utility compared to the cache management. The latter is more effective when catering for the requirements of partial matches in terms of remote data over a longer time span.

Note that the greediness of event selection ($\S 2.1$) is implicitly incorporated in the above utility model, because its semantics is reflected in the number of partial matches, which is directly captured by the urgent utility. While the estimation of future utility does not directly capture greediness, it is indirectly taken into account: The estimation implicitly incorporates dependencies between counts of partial matches at different states, at least on an aggregated level (aggregation from stream index $k$ to $k'$ in Eq. 4).

Furthermore, our utility model is able to cope with multiple queries in a straightforward manner: The utility of a data element is assessed based on its related current and future partial matches, regardless of the query for which these partial matches have been created. Sharing of data elements among queries is thereby captured directly in our cost model. If queries are assigned priorities, these need to be used as weights in the utility definition in Eq. 3.

\textbf{4 UTILITY MODELLING}

We first present measures to assess the utility of remote data elements ($\S 4.1$). Since utility is in part determined by future system states, we also present a method for its efficient estimation ($\S 4.2$).

\textbf{4.1 Utility Definition}

During query evaluation, the role of a remote data element is to enable the evaluation of some query predicate to determine whether a partial match shall be discarded, extended, or split up. Therefore, the utility of a data element is primarily based on the number of partial matches for which the element is required in their evaluation. Since a fetched data element is cached, it may be used to evaluate predicates for upcoming partial matches as well. Hence, our utility assessment considers both, the current partial matches, which induce the urgent utility, and future partial matches potentially derived from them, which induce the future utility.

Consider a point in time when a stream prefix $S(.k)$ has been processed, so that $P(k)$ is the set of current partial matches. The CEP engine needs to handle event $S(k + 1)$. Recall from $\S 2.1$ that, for a partial match $p \in P(k)$, $D(p, k + 1)$ is the set of data elements required as input for processing. For a data element $d \in D$, at this point in time, the urgent utility is defined as the number of partial matches that require $d$ or one of its constituents (with $p^*$ as the reflexive transitive closure of the part-of-relation $\rho$), weighted by its transmission latency $\ell_{\text{remote}}(d)$:

$$UU(d, k) = \ell_{\text{remote}}(d) \cdot |\{p \in P(k) \mid \exists d' \in D(p, k + 1) : d \in \rho^*(d')\}|.$$  (3)

In the same vein, the number of partial matches that require $d$, or its constituents, can be considered for some future state of query evaluation. Given perfect information about the future stream up to a stream index $k' > k$, the future utility of $d$ is given by the urgent utilities of future partial matches in $k < i \leq k'$:

$$FU(d, k, k') = \sum_{k < i \leq k'} UU(d, i)$$  (4)

However, since information about future states is inherently uncertain, we actually need to compute an estimate for the future utility up to a time point $k'$, which we denote as $\hat{FU}(d, k, k')$. As such, in the remainder we rely on an overall utility that is defined as the weighted sum of the above measures, defined, with $\omega \in [0, 1]$, as:

$$U(d, k, k') = \omega \cdot UU(d, k) + (1 - \omega) \cdot \hat{FU}(d, k, k')$$  (5)

Here, the rational is that the weighting enables tuning of the respective importance of the known utility $UU$ and the estimated utility $\hat{FU}$. Moreover, different weighting schemes are applied when incorporating the utility assessment for the realization of PFetch, LzEval, and cache management. Specifically, the strategies for fetching remote data, PFetch and LzEval, assign higher weights to the urgent utility compared to the cache management. The latter is more effective when catering for the requirements of partial matches in terms of remote data over a longer time span.

Note that the greediness of event selection ($\S 2.1$) is implicitly incorporated in the above utility model, because its semantics is reflected in the number of partial matches, which is directly captured by the urgent utility. While the estimation of future utility does not directly capture greediness, it is indirectly taken into account: The estimation implicitly incorporates dependencies between counts of partial matches at different states, at least on an aggregated level (aggregation from stream index $k$ to $k'$ in Eq. 4).

Furthermore, our utility model is able to cope with multiple queries in a straightforward manner: The utility of a data element is assessed based on its related current and future partial matches, regardless of the query for which these partial matches have been created. Sharing of data elements among queries is thereby captured directly in our cost model. If queries are assigned priorities, these need to be used as weights in the utility definition in Eq. 3.

\textbf{4.2 Utility Estimation}

To estimate the future utility $\hat{FU}$ of a data element, we determine the expected number of partial matches for which the element is relevant to its evaluation. Since the utility of data elements needs to be materialized in an online manner, this estimation should not incur significant overhead. Therefore, we estimate the number of relevant partial matches by considering two aspects: (i) how many partial matches of a particular class are expected at a time point and (ii) to what fraction of these partial matches an element is relevant. Algorithm 2 demonstrates the estimation procedure. It should be noted that urgent utility $UU$ is directly monitored (line 2 and 6).
Number of partial matches. We partition partial matches into classes, where the partitioning is determined by the adopted computational model. That is, in an automata-based model, each state of the automaton denotes a class. Intuitively, the class of a partial match, through a set of query predicates, induces a class of data elements that may be needed for the evaluation of its predicates. E.g., in Figure 2, all partial matches of state $q_3$ of the automaton require the evaluation of predicate $p_3$ when processing an event. The evaluation of $p_3$ refers to the remote data element $r[q_1, user]$, i.e., the set of known locations associated with a particular user. Therefore, a data element $d$ that refers to such a set of locations is potentially relevant to any such partial match.

Let $\{1, \ldots, n\}$ be the identifiers of the classes of partial matches (i.e., automaton states), then we denote the expected number of matches of a class $j$ at time point $k$ as $\#P_j^k(k)$. To efficiently estimate $\#P_j^k(k)$, we compute it as the average number of partial matches of class $j$ over a time window of fixed size, as shown in line 9, Alg. 2.

Partial match relevance. However, a specific element is only truly relevant to a subset of partial matches of a particular class, which is determined based on the payload of a partial match’s events. E.g., continuing on the above example, a data element $d$, capturing the locations associated with a particular user, is only relevant to those partial matches of which the first event relates to that same user.

To estimate the fraction of partial matches of a class for which a particular data element is relevant, we employ a stochastic model. Given a class of partial matches $j$ and a data element $d \in D$, we use $Pr(j, d, k)$ to capture, at time point $k$, the probability that element $d$ is required to evaluate the predicates of partial matches of class $j$ when processing an input event. We assume that such a probability is relatively stable in the short term. The probability distribution may be derived from the value distribution of the events’ attributes that serve as a reference for the selection of data elements. Here, the value distribution and, hence, the probability distribution, again, is computed adopting a sliding window. In Alg. 2, line 3 and 8 show how to monitor and maintain it online.

Future utility. Based on the above, we estimate the future utility at the time the stream prefix $S_k$ has been processed, up to some future stream index $k' > k$ for a data element $d \in D$, as follows:

$$\hat{FU}(d, k, k') = (k' - k) \sum_{d' \in D, d \in P(d')} \#P_j^k(k) \cdot Pr(j, d', k)$$

Both, the average number of partial matches per class and the distribution of attribute values (as the basis for the probability distribution of the data access per class), are based on simple counts, which is why they can be maintained incrementally and relatively efficiently for a sliding window (line 3-4 and 8-10 in Alg. 2). Nevertheless, since their computation still introduces overhead, instead of recomputing all utility values every time a single event is processed, a lower frequency may be employed to trade off precision of the utility model versus the incurred computational overhead.

5 REMOTE DATA FETCHING

The EIRES framework defines PFetch and LzEval as two strategies to fetch remote data, which we describe in §5.1 and §5.2, respectively. As illustrated in line 5-10, Alg. 1, EIRES always performs prefetching.

Algorithm 2: Utility estimation.

```
// Input: Query Q with time window τQ; partial matches $P(k), \ldots, P(k - τQ)$, partitioned in classes $1 \leq j \leq n$, i.e., $P_j^k(k), P_j^k(k - 1), \ldots, P_j^k(k - τ_j)$. System configuration parameter: Weighting factor ω;
// State: Maintained transition counts per remote reference key tranKey and per class of partial matches tranClass;
// Output: Utility function $U$; estimated number of partial matches $\#P_j^k(k)$ partitioned in classes $\#P_j^k(k), 1 \leq j \leq n$.
1 foreach $d \in D$ required by a partial match $p \in P(k \setminus P(k - 1))$ of class $j$ do
   // Increase urgent utility because of new partial match
   2 $UU(d, k) ← UU(d, k) + 1;
   // Update auxiliary counts for future utility estimation
   3 tranKey(d, j, k) ← tranKey(d, j, k) + 1;
   4 tranClass(d, j, k) ← tranClass(d, j, k) + 1;
5 foreach $d \in D$ required by a partial match $p \in P(k - 1) \setminus P(k - 2)$ do
   // Decrease urgent utility due to timed out partial matches
   6 $UU(d, k) ← UU(d, k) - 1;
7 foreach $d \in D$ required by partial matches of class $j$ do
   // Estimate future utility
   8 $Pr(j, d, k) ← \frac{\sum_{i=k+1}^{k} tranKey(d, j, i)}{\sum_{i=k+1}^{k} tranClass(j, i)}$;
   9 $P_j^k(k) ← AVG(\{ P_j^k(k - τ_j), \ldots, P_j^k(k - 1 + τ_j), \ldots, P_j^k(k) \})$;
   10 $\hat{FU}(d, k, k + τ_j) = \sum_{d' \in D, d \in P(d')} \#P_j^k(k) \cdot Pr(j, d', k)$;
   // Compute overall utility
   11 $U(d, k, k + τ_j) = ω \cdot UU(d, k) + (1 - ω) \cdot \hat{FU}(d, k, k + τ_j)$;
   return $U, \#P_j^k(k)$;
```

If prefetching did not prepare required remote data on time, EIRES performs lazy evaluation (line 11).

5.1 Prefetching

As explained in §3, two operations need to be instantiated for PFetch: deciding when prefetching is beneficial for a data element (P1) and selecting the elements to prefetch (P2).

(P1): Prefetch timing. The best time to prefetch a data element $d$ is such that it arrives right before it is needed, i.e., ideally, prefetching is triggered at $t_p = t_n - t_{remote}(d)$. Prefetching too early should be avoided, since this will be based on a less accurate estimation of the future utility of $d$, given that it is generally more complex to make predictions further in advance. Hence, earlier prefetching increases the chance that $d$ is prefetched, consumes space in the cache, but is never used in query evaluation. On the contrary, late prefetching ($t_1 < t_p$) generally benefits from a more accurate utility estimation, but hides the transmission latency only partially. To estimate the appropriate time to prefetch in light of this trade-off, we propose complementary techniques: lookahead timing and estimated-arrival timing, with workflow shown in Alg. 3.

Lookahead timing. We strive to identify partial match classes that provide useful indicators of when to prefetch a data element related to another match class, i.e., a lookahead class. To do this, we exploit the partial order of the classes of partial matches, derived from the query predicates. This partial order corresponds to the order of an automaton’s states, which captures that partial matches of one class may become those of another class during query evaluation.

Let $\{1, \ldots, n\}$ be the identifiers of classes of partial matches with $m$ being a class of matches that require a data element $d$ (line 2 in Alg. 3). Then, we determine a lookahead class $j$ that is a predecessor of $m$ in the partial order of classes, denoted $j < m$, meaning that partial matches of class $j$ potentially develop into those of class $m$. 


Algorithm 3: Prefetch timing (P1 in PFetch)

State: Monitored event input rates $\lambda(m)$, latest prefetch cache hit history $H$.
Output: Prefetch timing function $T$, offset timing function $O$.

1. $T \leftarrow 0$, $O \leftarrow 0$.
2. foreach $d \in D$ required by partial matches of class $m$ do
3.   $j \leftarrow m$’s directly preceding class,
4.   while $j < m$ do
5.     // Check if lookahead timing can be applied
6.     if partial matches of class $j$ require $d$ & $j > 1$ then
7.       if $H(m, j, d) = true$ then // Prior cache hit successful
8.         $O(d) \leftarrow 0$, $T(d) \leftarrow j$; // Set time offset to be 0
9.         break;
10.    else $j \leftarrow j$’s directly preceding class; // Try next class
11.   else // Use estimated-arrival timing
12.     $O(d) \leftarrow 1/\lambda(j) - t_{remote}(d)$; $T(d) \leftarrow j$; // Set time offset
13.     break;

14. return $T$, $O$.

From the set of all preceding classes $\{1 \leq j \leq n \mid j < m\}$, we choose $j$ such that 1) its partial matches contain events of which the payload serves as a reference to identify $d$, and 2) it is closest to $m$ in the partial order while still allowing for timely and accurate prefetching. This process is detailed in line 3-9 in Alg. 3. Here, without loss of generality, we illustrate the situation that class $m$ has one directly preceding class, though in practice it may have multiple such classes.

If multiple classes contain references to the same data elements, multiple prefetches could be merged into a single fetch through semantic query rewriting. If they contain references to different data elements, a traverse order could be enforced over all these directly preceding classes, i.e., ordered by monitored transmission latencies. ERES incorporates both of these approaches.

We determine the lookahead class $j$ in a dynamic manner, based on the recent cache hit history from cache management (input $H$ in Alg. 3). Particularly, for a given data element $d$, $H$ maintains cache hit/miss information, where $H(i, i’, d)$ returns false if $d$ was prefetched upon the construction of a partial match of class $i$, but was not available when required during the evaluation of a partial match of class $i’$. Such a cache miss can have two causes: First, it may hold that $t_{remote} < t_{remote}(d)$, i.e., prefetching occurs too late because the time for a partial match to develop from class $i$ to $i’$ is shorter than the transmission latency. Second, when constructing a partial match of class $i$, the utility estimation may have been inaccurate, so that the wrong data elements have been fetched. Either way, such a recent cache miss indicates that the respective class is not suited to trigger prefetching.

In sum, for a partial match of class $m$ that requires a data element $d$, we select the preceding class $j$ closest to $m$ for which $H(d, j, m) = true$ (line 6). Then, the point in time to prefetch data element $d$ for partial matches of class $m$ is defined as the moment that a partial match for class $j$ is constructed. For example, if $H(d, j_1, m)$ and $H(d, j_2, m)$ are both true, and $j_1 < j_2$, we select $j_2$ as the lookahead class for $m$, avoiding the unnecessarily early fetching that using $j_1$ would yield. If partial matches of multiple classes $m$ reference $d$, the smallest index $j$, in terms of the partial order $<$, is chosen.

We deal with event stream fluctuations by internally maintaining counts of cache misses in $H$, so that a threshold can determine what to interpret as sufficient negative evidence. Also, values are reset to zero after a fixed time period after their last increment.

Note that there may be partial match classes for which lookahead timing is unsuitable, i.e., cache misses occur since there is no lookahead class enabling timely and accurate prefetching. This occurs, e.g., when the reference required to identify $d$ is part of the payload of the input event that lead to the creation of the partial match of class $m$. An example for that is the class $q_1$ in Figure 2, which requires the evaluation of a predicate using remote data on pre-authorized clients ($r(q_1.org)$) that can only be fetched based on the event that lead to the respective match. For those cases, we determine the time to prefetch based on estimated-arrival timing.

Estimated-arrival timing. When lookahead timing is not applicable for a specific partial match class $m$, we instead determine the time to prefetch by incorporating the inter-arrival time of events that satisfy the predicates to check for partial matches of class $m$ but are not based on remote data (line 10-12 in Alg. 3). Consider again class $q_1$ in Figure 2. The respective partial matches are extended when an event with $x.type=T$, $x.vol>50k$, and $x.ben \neq r(q_1.org)$ is received. Since lookahead timing is not applicable for the remote data, we estimate the expected time until an event satisfying $x.type=T$ and $x.vol>50k$ is received. This way, we obtain an estimate for the time between the creation of a partial match of class $q_1$ and the time the data element $r(q_1.org)$ is actually needed. To derive this inter-arrival time, various stochastic processes for event arrival may serve as a foundation. Selecting one of them, their parameters shall be learned from historic data or through monitoring.

Here, we illustrate the general procedure when event arrival follows a Poisson process [20], as observed in many domains where events correspond to requests triggered by people. Then, events are independent and occur with a constant mean arrival rate $\lambda$, which is the only parameter that needs to be monitored. This is similar to Akdere et al.’s work [21], but differs in granularity: Akdere et al. [21] works on the level of processing states, whereas our model estimates the arrival of events that have a certain payload.

Inter-arrival times are exponentially distributed with their expectation being $E = 1/\lambda$. Based thereon, at time $t$, we estimate the time the remote data is needed as $t + 1/j$, so that the time for prefetching becomes $t_p = t + 1/j - t_{remote}(d)$ (line 11 in Alg. 3). Taking up the previous example, events with $x.type=T$ and $x.vol>50k$ may be rare, such that the expected inter-arrival time is $E = 300ms$. Then, with a data transmission latency of $t_{remote} = 50ms$, we trigger prefetching only 250ms after the creation of the respective partial match (i.e., when processing the next input event after this time period).

(P2): Prefetch selection. Using the above techniques, we determine the point in time at which prefetching is beneficial for a particular data element. However, at a specific moment while processing the event stream, we still need to select those data elements for which prefetching shall actually be invoked. This decision is taken using our utility model: We only prefetch data elements, for which prefetching is beneficial at that point in time and for which the utility is higher than the minimum utility value of an element currently in the cache (line 9-10 in Alg. 1). At time (stream index) $k$, let $C(k) \subseteq D$ be the content of the cache and $D(k)$ the set of data elements for which prefetching at time $k$ would be beneficial. Then, we select $D’(k)$ for prefetching, defined as

$$D’(k) = \{d \in D(k) \mid U(d) > \min_{d’ \in C(k)} U(d’).\}$$

(7)
Algorithm 4: Lazy evaluation (LzEval)

<table>
<thead>
<tr>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Input: input event (S(k+1)); partial matches (P(k)); remote data (D(k+1)); cache (C); CEP engine (f_g); estimated number of partial matches (#P(k)) partitioned in classes (#P(k), 1 \leq i \leq n).</td>
</tr>
<tr>
<td>2</td>
<td>Output: Matches (C(k+1)); partial matches (P(k+1)).</td>
</tr>
<tr>
<td>3</td>
<td>foreach (d \in D(k+1)), (C) required by partial matches of class (j) do</td>
</tr>
<tr>
<td>4</td>
<td>foreach class (m, j \leq m) do</td>
</tr>
<tr>
<td>5</td>
<td>if ((m, f)) is not checked against (\hat{succ}(j, f)) then</td>
</tr>
<tr>
<td>6</td>
<td>(\lambda \gets \lambda +</td>
</tr>
<tr>
<td>7</td>
<td>(\Delta.\ell_{\text{remote}} \gets \min(E(j, m), \ell));</td>
</tr>
<tr>
<td>8</td>
<td>(\Delta.\ell_{\text{match}} \gets \ell_p, \prod_{1 \leq i \leq n} #P(k) \cdot \lambda[i + 1] \cdot E(j, m));</td>
</tr>
<tr>
<td>9</td>
<td>if (\Delta.\ell_{\text{remote}} &gt; \Delta.\ell_{\text{match}}) then</td>
</tr>
<tr>
<td>10</td>
<td>(\hat{succ}(j, f) \gets \hat{succ}(j, f) \cup {m});</td>
</tr>
<tr>
<td>11</td>
<td>else (\hat{succ}(j, f) \gets \hat{succ}(j, f) \setminus {m});</td>
</tr>
<tr>
<td>12</td>
<td>// benefit expected, postpone evaluation of (d)'s predicates</td>
</tr>
<tr>
<td>13</td>
<td>if ((m, f) \in \hat{succ}(j, f)) then</td>
</tr>
<tr>
<td>14</td>
<td>Fetch (d);</td>
</tr>
<tr>
<td>15</td>
<td>(P'' \leftarrow f_0(S(k+1), P(k))); // ignore (d)'s predicates</td>
</tr>
<tr>
<td>16</td>
<td>else Fetch (d), block stream processing until (d) arrives at (C);</td>
</tr>
<tr>
<td>17</td>
<td>(P''; C'' \leftarrow f_0(S(k+1), P(k) \cup P', D(k+1) \cap C');</td>
</tr>
<tr>
<td>18</td>
<td>(P(k+1) \leftarrow P(k+1) \cup P''; )</td>
</tr>
<tr>
<td>19</td>
<td>(C(k+1) \leftarrow C(k+1) \cup C'').</td>
</tr>
<tr>
<td>20</td>
<td>return (C(k+1), P(k+1)).</td>
</tr>
</tbody>
</table>

5.2 Lazy Evaluation

For LzEval, in turn, two operations need to be instantiated: selecting the partial matches for which the strategy is applied (L1) and adapting the evaluation procedure for those matches (L2). The workflow of LzEval is sketched in Alg. 4.

(L1): Selection of partial matches. LzEval triggers a fetch operation for remote data \(d\) when it is required, but postpones the actual evaluation of the respective predicates. While this hinders the transmission latency \(\ell_{\text{remote}}(d)\), it also makes event selection less strict, possibly resulting in an exponential growth of the number of partial matches. As such, the inherent evaluation latency \(\ell_{\text{match}}(C)\) may increase, which may thwart the benefit of the reduction of \(\ell_{\text{remote}}(d)\). Recognizing this trade-off, we therefore only apply LzEval to those partial match classes where an actual benefit is expected.

To determine these classes, we estimate the time gained by hiding the transmission latency, denoted as \(\Delta.\ell_{\text{remote}}\), and the overhead incurred by the additional partial matches, \(\Delta.\ell_{\text{match}}\). The latency gain, \(\Delta.\ell_{\text{remote}}\), depends on the difference between the time \(t_n\) at which a data element is needed (and fetching is triggered), and when the predicate is actually evaluated \(t_e\). Ideally, evaluation happens after the data is available (i.e., \(t_e < t_n\)), so that the entire transmission latency is hidden, i.e., \(\Delta.\ell_{\text{remote}} = \ell_{\text{remote}}(d)\). The overhead \(\Delta.\ell_{\text{match}}\) depends on the number of additional partial matches caused by lazy evaluation during \(t_e - t_n\) and the additional latency incurred per partial match. While the latter is assumed to be a known constant, \(\ell_{\text{pm}}\), the other parameters need to be estimated.

To estimate the number of additional partial matches, we follow an approach similar to the estimated-arrival timing (§5.1). Assume that the evaluation of a predicate of class \(j\) requires remote data, yet that this evaluation may be postponed until the creation of a partial match of a later class \(m, j < m\). Then, we estimate the average time \(t_{j\rightarrow m}\) for a partial match of class \(j\) to develop into one of class \(m\), when ignoring all predicates that are based on remote data. As before, we assume that, for each possible extension of a partial match, the respective event arrivals follow a separate Poisson process of a monitored rate. Then, we derive an estimate for \(t_{j\rightarrow m}\) based on a compound Poisson process induced by the sequence of intermediate classes \([r_1, \ldots, r_m]\) of partial matches, \(j < r_1 < \ldots < r_m < m\). With \(\lambda_i\) as the rate of the process describing the arrivals that construct partial matches of class \(i\), the expectation of the compound Poisson process is \(E(j, m) = 1/\sum_{r_i \leq j \leq m} \lambda_i\) (line 6 in Alg. 4). Although the estimate assumes all Poisson processes to be independent, i.e., it ignores sequential dependencies between the classes, it suffices as an upper bound for our purposes.

Using \(E(j, m)\) as an estimate for \(t_{j\rightarrow m}\), we derive the remaining parameters. The hidden part of the transmission latency is given as \(\Delta.\ell_{\text{remote}}(j, m) = \min(E(j, m), \ell_{\text{remote}}(d))\) (line 7 in Alg. 4). The estimation of \(\Delta.\ell_{\text{match}}(j, m)\) is time-varying and incorporates \#\(P(k)\), i.e., the expected number of partial matches of class \(i\), as estimated at time point (or stream index) \(k\), see §4.2. For each class \(i\), this number is multiplied with the arrivals of events that may extend the match by satisfying all the predicates not based on remote data (\(\hat{\lambda}_{ri+1}\)) and the estimate for \(t_{j\rightarrow m}\) (line 8 in Alg. 4). While this estimates the number of additional partial matches during \(t_e - t_n\), multiplying it with the constant additional evaluation latency per partial match (\(\ell_{\text{pm}}\)) yields the estimate for accumulated increase in evaluation latency (to simplify notation, we define \(r_{s+1} = m\)).

\[
\Delta.\ell_{\text{match}}(j, k, \ell) = \ell_{\text{pm}} \sum_{1 \leq i \leq s} (\#P'\cdot \hat{\lambda}_{ri+1} \cdot E(j, m))
\]

For each class \(j\) that requires remote data during query evaluation, we determine the set of succeeding classes for which lazy evaluation is beneficial. This set is defined as \(\hat{\text{succ}}(j) = \{m \in \{1, \ldots, n\} \mid j < m \land \Delta.\ell_{\text{remote}} > \Delta.\ell_{\text{match}}\}\), i.e., the hidden part of the transmission latency is larger than the overhead by the increased evaluation latency. Then, all partial matches of a class \(j\), for which \(\hat{\text{succ}}(j)\) is non-empty, are selected for lazy evaluation (line 9-11 in Alg. 4).

Conceptually, \(\hat{\text{succ}}(j)\) must be computed for every different transmission latency, \(\ell_{\text{remote}}(d)\) (line 5 in Alg. 4). In practice, to improve efficiency and to reuse results, transmission latency may be lifted to coarser granularities, e.g. to a millisecond level.

(L2): Adapted evaluation procedure. Consider partial matches of a class \(j\) and \(\hat{\text{succ}}(j)\) as the succeeding classes for which lazy evaluation is beneficial. Also, let \(S(\cdot, k)\) be the time the stream prefix was processed and \(C(k) \subseteq D\) as the cache content. Then, query evaluation to process event \(S(k+1)\) is adapted for the partial matches of class \(j\), as follows. For each predicate that requires a remote data element \(d\), the availability in the cache is checked. If \(d \in C(k)\), the predicate is directly evaluated (line 16 in Alg. 4). If not, a fetch request for \(d\) is triggered and the predicate is marked as postponed (line 12-14 in Alg. 4). The same procedure is followed as long as the partial match of class \(j\) develops into one of a class \(m \in \hat{\text{succ}}(j)\). Upon construction of a respective partial match, the cache is checked and, upon data availability, the predicate is evaluated.

Once a partial match of class \(j\) develops into a class \(m' \notin \hat{\text{succ}}(j)\) for which lazy evaluation is not beneficial, a different strategy is implemented. Postponing the predicate evaluation further would increase the overall latency. Hence, query evaluation continues only once the data element \(d\) becomes available (line 15 in Alg. 4).
6 CACHE MANAGEMENT

Cache management in EIRES shall retain the data elements that are most beneficial in the cache, thereby realizing operation (C1) as introduced in §3. Although the benefit of a data element strongly relates to its expected utility (see §4), certain query semantics suggest to base cache management on a relatively simple policy. Recall the greediness of CEP queries in event selection (§2.1). Either semantics motivates a different policy for the cache management.

LRU policy. Under a greedy query semantics, a large number of partial matches that require the same data elements can be expected to materialize. This, in turn, will induce a large number of access requests to the cache for the respective data. In terms of our utility model, the urgent utility then becomes a good estimator for the future utility. We can exploit this effect, even without using any computed utility value, by adopting the widely-established least-recently-used (LRU) policy for cache management. Data elements in the cache are ranked by the time of last access to them, evicting those that have not been accessed for the longest time.

Cost-based policy. With non-greedy query semantics, individual data elements are expected to be accessed less often and in a diminishing manner. Hence, the current access frequency, i.e., the urgent utility, no longer provides a good estimator for the future utility. In that case, cache management shall exploit the computed utility.

In our cost-based policy, we separate the handling of data elements based on the fetching strategy that led to their retrieval. The reason being that, while any element requested through LzEval will certainly be required by some partial match, this is not necessarily the case for data elements requested through, possibly inaccurate, predictions of PFetch. Against this background, we adopt two, purely conceptual, cache tiers, $T_1$ and $T_2$, to separate elements that will certainly be used ($T_1$) from those for which usage is uncertain ($T_2$). Then, elements in $T_1$ will be retained over all elements of $T_2$, but are moved to $T_2$ after a first access.

Once the cache capacity is reached, data elements (from $T_2$ before $T_1$) are evicted based on their utility. Let $C \subseteq D$ denote the current content of the cache and $b \in N$ its capacity. Then, incorporating the size of data elements, the selection of the data elements $R \subseteq C$ to retain can be formulated as a knapsack problem:

$$\text{select } R \subseteq C \text{ that maximizes } \sum_{d \in k} U(d) \text{ subject to } \sum_{d \in k} |d| \leq b.$$

The above is a standard, NP-hard, knapsack problem [22]. Yet, it may be approximated [23], e.g., by selecting data elements in the order of their utility and size ratios, until the capacity is reached.

7 EXPERIMENTAL EVALUATION

We evaluated the EIRES framework on various scenarios. §7.1 first outlines the setup. For controlled experiments, §7.2 reports on the overall efficiency and effectiveness, while §7.3 explores the sensitivity. Finally, §7.4 discusses two real-world scenarios. Note that an extended version of this paper including query specifications and additional experiments as well as the employed implementation and datasets are publicly available.¹

7.1 Experimental Setup

Dataset and queries. For controlled experiments, we generated synthetic data with a Type ($\mathbb{U}(\{A, B, C, D\})$), an ID ($\mathbb{U}(\{1, 100\}$), and two numeric attributes $V_1$ ($\mathbb{U}(1, 10000)$), and $V_2$ ($\mathbb{U}(1, 100000)$). This dataset enables us to evaluate common queries that test for sequences of events of different types, which are correlated by an ID. Further correlation predicates and references to remote data may be defined for attributes $V_1$ and $V_2$.

For the synthetic dataset, we evaluate the queries Q1 and Q2 of Listing 2. They differ in their structure (pure sequence vs. disjunction of sequences) and the integration of remote data in predicates.

<table>
<thead>
<tr>
<th>Query</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1: SEQ(A a, B b, C c, D d, E e, F f) WHERE a.v1=REMOTE[d.v1] AND a.v2=e.v2 AND b.v1=REMOTE[h.v1]</td>
<td>5th within 8 min</td>
</tr>
<tr>
<td>Q2: SEQ(A a, SEQ(B b, C c, D d, E e)) OR SEQ(C c, B e) WHERE a.v1=b.v1 AND a.v2=e.v1 AND d.v1=REMOTE[a.v1] AND c.v2=REMOTE[a.v2]</td>
<td>5th within 50K</td>
</tr>
</tbody>
</table>

**Listing 2:** Queries for the synthetic dataset.

Baselines. We compare our approaches for data fetching, PFetch, LzEval, and their combination (Hybrid) against three baselines. BL1 denotes the naive integration of data fetching. It interrupts query evaluation when remote data is needed and continues once it has been fetched. BL2 employs a cache of remote data to improve the efficiency of query evaluation. For this cache, we consider both policies discussed in §6, i.e., LRU and the cost-based policy. Finally, BL3 first ignores all predicates based on remote data. Upon reaching a final state of the evaluation model, it fetches the remote data and conducts the respective event selection.

Measures. Our focus is the latency of query evaluation, i.e., the time between the ingestion of the last event needed to construct a match and its actual detection. Depending on the experimental setting, our latency measurements are based on 100k to 5 million matches. Specifically, we report the 5th, 25th, 50th, 75th, and 95th percentiles of the latency values. Moreover, we report the throughput, i.e., the number of events processed per second.

Implementation and environment. For the evaluation, we used an automata-based CEP engine, written in C++. All reported results are averaged over 20 runs on a NUMA node with 4 Intel Xeon E5-4880 CPUs (60 cores) and 1TB RAM, running openSUSE 15.0.

7.2 Overall Effectiveness and Efficiency

Figures 5 and 6, respectively, depict the overall performance obtained for Q1 and Q2 on the synthetic dataset. We compare the PFetch, LzEval, and Hybrid strategy against the three baselines under both greedy and non-greedy selection for two cache eviction policies. Here, the cache capacity is set to 10% of a remote key’s size range, i.e., 10,000 items, while the transmission latency of remote data is uniformly distributed between 10µs and 100µs.

As shown in the figures, the Hybrid strategy consistently outperforms all other approaches. Furthermore, both PFetch and LzEval also always outperform all three baselines considerably, though it depends on the context which of these achieves better results.

¹https://github.com/zbjob/EIRES
Selection strategies. For non-greedy selection in Q1, the median latencies of Hybrid are 10µs (cost-based cache) and 16µs (LRU). Since the median latencies for the baselines range between 264µs (BL3) and 314µs (BL1), Hybrid reduces the median latency by at least 26×. For the 95th percentiles, the reduction is at least 4× (cost-based cache) and 2.5× (LRU). For Q2, we observe smaller gains in median latencies, yet bigger gains for the 95th percentile. Among the baselines, BL3 outperforms both BL1 and BL2 for Q1, but this is the opposite for Q2. This is because Q1 has two states requiring different remote data in the CEP engine, whereas Q2 has one at each branch of disjunction sequences. Since BL3 fetches different data items at once, its aggregated transmission latency is the maximal latency of all required remote data, instead of their sum (as for BL1 and BL2). For Q1, this benefit outweighs the overhead caused by extra partial matches under non-greedy selection.

The results show greater variety for greedy selection policies. For Q1, Hybrid reduces the median latency by 111× (cost-based cache) and 63× (LRU) when compared to BL1. Furthermore, the reductions for BL3 are 283× (cost-based cache) and 160× (LRU), yet the reductions are only 6× (cost-based cache) and 2.8× (LRU) when compared to BL2. The reductions for the 95th percentiles are 62× (BL1, cost-based), 44× (BL1, LRU), 6× (BL2, cost-based), 2.8× (BL2, LRU), 558× (BL3, cost-based) and 392× (BL3, LRU). For Q2, gains are even more extreme, i.e., Hybrid reduces median latencies by up to 2,726× and the 95th percentiles latency by up to 3,752×.

Caching policies. The impact of a local cache differs considerably per scenario. Since the reusability of data elements is low for non-greedy selection, a cache contributes little here, as illustrated by the comparable performance of BL1, BL2, and BL3 in Figure 5a-5b and Figure 6a-6b. For greedy selection, in turn, adding a cache reduces latencies by at least two orders of magnitude, see BL1, BL2, and BL3 in Figure 5c-5d and Figure 6c-6d. In such settings, LRU outperforms a cost cache for BL2 due to its small computation overhead.

By contrast, when combining a cache with PFetch or LzEval, the cost-based policy achieves better performance. For instance, LzEval with a cost-based cache (Figure 5a) outperforms its counterpart with LRU (Figure 5b), especially for the median latency. While both employ the same LzEval procedure, the LRU policy ignores utilities, so that promising data elements may be fetched, but not kept in cache for sufficient time. This same trend is observed for PFetch.

Benefits of Hybrid. The Hybrid strategy always outperforms the individual PFetch and LzEval strategies, as their combination helps to overcome their individual weaknesses. For instance, the performance of PFetch is closely related to the quality of the made prefetching predictions. When PFetch fails to prefetch a data element, query evaluation is interrupted, resulting in high tail latencies (95th percentile), see Figure 5d. LzEval generally has lower median
and 75\textsuperscript{th} percentile latencies (Figure 5a-5b), because its fetching decision is always accurate. Yet, additional partial matches affect the latency under greedy selection, see Figure 5c-5d and Figure 6c-6d.

Hybrid combines alleviates the aforementioned issues. If inaccurate prediction lead to non-effective prefetching, Hybrid still (generally) performs lazy evaluation, thereby avoiding the interruption of stream processing. However, the overhead is smaller than using LzEval alone, since, due to prefetching, many matches are already handled by prefetching.

**Throughput.** We also investigate the throughput performance. Due to limited space, Figure 7 focuses on the throughput for Q1 under non-greedy selection with either policy for cache management. Throughput performance is largely in line with the observed latencies, with a few deviations. For instance, while LzEval with a cost-based cache has a lower median latency compared to LzEval with LRU, the observed throughput is virtually equivalent.

### 7.3 Sensitivity Analysis

We assess the sensitivity of the proposed strategies with respect to: the utility estimation quality, the cache size, the remote data transmission latency, and the weighting factor \( \omega \) which tunes the share of urgent utility and future utility in Eq. 5. All results were obtained for query Q1, with a cost-based cache and greedy selection.

**Utility estimation quality.** The quality of the utility model affects the efficacy of all strategies. We assess the impact of the estimation quality by injecting noise into the employed estimations, where a noisy estimation means that an expected partial match will not actually materialize. We compare the corresponding latency variations by injecting noise into 10\% to 90\% of the estimations.

PFetch is sensitive to such noise, see Figure 8a, since its median latency grows for higher noise levels. The reason is twofold: PFetch prefetches the wrong data elements, while poor utility estimation also negatively impacts the elements that are evicted from the cache. Above a 50\% noise ratio, its median latency already exceeds LzEval’s 75\textsuperscript{th} percentile and Hybrid's 95\textsuperscript{th} percentile.

LzEval is less sensitive to noise, even obtaining stable latencies across noise levels for the 5\textsuperscript{th} and 25\textsuperscript{th} percentile, as well as the median. This is because LzEval's initial decision about what to fetch is not affected by utility, unlike for PFetch. Still, low quality utility estimations can lead to poor decision on which partial match evaluations to postpone and, thus, to additional partial match growth and associated overhead. Again, the cache management also deteriorates for higher noise levels. As a result, especially LzEval’s 95\textsuperscript{th} percentile latency grows along with an increasing noise ratio.

Since Hybrid combines the advantages of both PFetch and LzEval, it generally outperforms the individual strategies. However, an exception is observed at the 90\% noise ratio, where Hybrid’s 75\textsuperscript{th} percentile latency is higher than LzEval’s. Given such inaccurate utility estimations, the downsides of PFetch outweigh its benefits, resulting in a better latency for LzEval than for Hybrid.

**Cache size.** The size of the employed cache naturally affects all strategies, where a larger cache size is beneficial, see Figure 8b. In these results, we also observe that PFetch is more sensitive than the other strategies. This is because a larger cache allows for more tolerance in terms of incorrectly prefetched data elements, whereas a smaller cache size will be clogged by them.

**Remote data transmission latency.** We consider that for higher transmission latencies, failing to fetch a data element leads to longer delays. We tested this aspect by evaluating the performance for different transmission delays. As shown in Figure 8c, the latency of all strategies increases along with the transmission latency. However, PFetch is again most sensitive here. This is because prefetching needs to occur earlier for increased transmission latencies, which results in less accurate prefetch decisions.

**Weighting factor in utility definition.** Lastly, we consider the impact of \( \omega \), which balances the urgent and future utility in our model (§4.1, Eq. 5). We consider the impact of \( \omega \) in the utility to guide the fetching of remote data (§5.1) and to manage the cache (§6). We refer to these, respectively, as \( \omega_{\text{fetch}} \) and \( \omega_{\text{cache}} \).

Figure 9a shows the results obtained when varying \( \omega_{\text{fetch}} \) while fixing \( \omega_{\text{cache}} \) at 0.3. Here, increasing the weight reduces the 75\textsuperscript{th} and 95\textsuperscript{th} percentile latencies. This is expected, as a low weight results in the current demand for remote data (urgent utility) being largely ignored. Yet, beyond \( \omega_{\text{fetch}} = 0.7 \), performance starts to deteriorate. For such high weights, decisions are strongly based on the current demand for remote data, but ignore the future demand. As such, we observe optimal results for \( \omega_{\text{fetch}} = 0.7 \). However, the utility model turns out to be robust: For any factor that emphasizes the urgent demand, but does not ignore the future usage of remote data, i.e., \( \omega_{\text{fetch}} \in [0.5, 0.9] \), the performance is close to the optimal, with especially stable median latencies.

With \( \omega_{\text{cache}} \) set to 0.7, varying \( \omega_{\text{cache}} \) shows a similar trend, as shown in Figure 9b. Here, the optimal value turns out to be 0.5, which indicates that cache management considers the current demand and the future usage of remote data equally. Values in the range [0.3, 0.7] achieve comparable performance, though, which again points to a certain robustness of our utility model.

### 7.4 Case Studies

Finally, we applied our approach in two real-world scenarios, using a cost-based cache under greedy selection.

**Bushfire detection.** This case uses a real-world dataset of a bushfire detection system. The system’s geostationary operational environmental satellite, GOES-16 [24], is capable of detecting heat signatures produced by fires [25]. While it tracks these in real-time, the obtained information is combined with data from a ground-based sensor network for fine-grained validation. For this dataset, we employ a query that detects bushfires during daytime [25]. In essence, the query identifies the repeated occurrence of a specific radiation pattern for a geographical area in the satellite data. However, to evaluate this query, the receiver handling the satellite data needs to fetch remote data from temperature and humidity sensors.

As shown in Figure 10a, the results obtained for this case follow the trend observed for the synthetic dataset, with all our approaches outperforming the baselines and Hybrid performing best. Hybrid reduces median latencies for BL1, BL2 and BL3 by 206\%×, 21\%× and 200\%×, respectively. For the 95\textsuperscript{th} percentile latencies, the improvements are 18\%×, 13\%× and 14\%×. We also observe that PFetch has similar performance as Hybrid, except for the 95\textsuperscript{th} percentile latency, which shows that PFetch accurately anticipates the need for remote data.

**Cluster monitoring.** This case uses data from Google Cluster Traces [26]. It contains events that indicate the lifecycle of tasks...
running in a large-scale cluster. For this dataset, we employ a query that detects the following pattern: A task is submitted, scheduled, and evicted on one machine; later, in a different region, it is rescheduled and evicted again; finally it is rescheduled on a third machine, in another region, but fails execution. Here, information on the regions needs to be fetched from a remote database. We adopted a data transmission latency drawn uniformly from 1ms to 10ms.

The results in Figure 10 highlight that Hybrid outperforms all other approaches. It reduces the median latencies by $73\times$, $47\times$ and $11,879\times$ for BL1, BL2 and BL3. For the 95th percentile latencies, Hybrid improves $13\times$, $7\times$ and $1,335\times$.

Discussion. We observe more than an order of magnitude difference in performance benefits achieved for synthetic and real-world datasets. This is because the queries for the real-world datasets contain many compute-intensive predicates (e.g., the bushfire detection query computes the spatial overlap of geographic areas). Also, these queries have larger time windows and, hence, tend to generate more partial matches.

8 RELATED WORK

Prefetching and caching strategies have been incorporated in data processing systems for decades. Below, we review some important related techniques in diverse contexts and application scenarios.

9 CONCLUSIONS

We proposed Eieres, a framework for efficient integration of remote data in the evaluation of queries over event streams. Our core idea is to decouple fetching of remote data and its use in query evaluation. Data elements may be fetched before the need for them materializes and the evaluation of partial matches may be postponed until after the data is available. The Eieres framework facilitates these ideas through a cost model to evaluate data utility; strategies for fetching remote data, either through prefetching or lazy evaluation; and policies for cache management. Our experimental results show that Eieres improves the latency of query evaluation by up to $3,752\times$ for synthetic data and $47\times$ for real-world data.

In future work, we intend to explore the instantiation of Eieres for tree-based execution models [5] that define an order of operator evaluation and a hierarchy of buffers. Since our utility modelling is based on present and future partial matches, we expect to confirm our experimental results obtained for automata-based models.

ACKNOWLEDGMENTS

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with project-ID 246594964 and by the Federal Ministry of Education and Research (BMBF), Germany under the project LeibnizKILabor with grant No. 01DD20003.
REFERENCES


