Performance Antipatterns of One To Many Association in Hibernate

Patrycja Węgrzynewicz
Institute of Informatics
University of Warsaw
Banacha 2, 02-097 Warsaw, Poland
Email: patrycja@mimuw.edu.pl

Abstract—Hibernate is the most popular ORM framework for Java. It is a straightforward and easy-to-use implementation of Java Persistence API. However, its simplicity of usage often becomes mischievous to developers and leads to serious performance issues in Hibernate-based applications. This paper presents five performance antipatterns related to the usage of one-to-many associations in Hibernate. These antipatterns focus on the problems of the owning side of collections, the Java types and annotations used in mappings, as well as processing of collections. Each antipattern consists of the description of a problem along with a sample code, negative performance consequences, and the recommended solution. Performance is analyzed in terms of the number and complexity of issued database statement. The code samples illustrate how the antipatterns decrease performance and how to implement the mappings to speed up the execution times.

I. INTRODUCTION

Hibernate [1] is a remarkably popular implementation of Java Persistence API, i.e., the official standard of object-relational mapping in Java. However, many developers, especially of enterprise systems, complain on the correctness and the performance of Hibernate-based applications. It turns out that even relatively simple Hibernate applications impose a significant overhead on communication with a database, producing too many and/or inefficient SQL statements.

In this paper, we analyze the subject of the most common mapping, namely one-to-many associations of entities. Because of its wide usage as well as popularity of Hibernate, it seemed that the implementation of this association should have been well optimized. Our research revealed that even simple applications of one-to-many associations can result in (1) unexpected SQL statements being executed, (2) too many SQL statement being executed, and/or (3) too many objects being loaded into memory. Therefore, naïve mapping of one-to-many associations can introduce a significant performance overhead in a Hibernate-based application.

The main contribution of this paper consists of five performance antipatterns (i.e., bad practices with a significant impact on performance) related to the usage of one-to-many associations in Hibernate. Each antipattern consists of the description of a problem, performance consequences, the recommended solution, and a sample code.

II. RELATED WORK

Antipatterns are conceptually similar to design patterns as they describe recurring problems and provide solutions to them. A performance antipattern describes a bad practice that has a significant impact on performance. The articles [2] and [3] provide a good explanation of performance antipatterns along with the definition of 14 performance antipatterns. These papers have formed a good basis for further research on performance antipatterns, mainly in the field of automated detection and fixing of performance problems. The paper [4] introduces a framework for automated detection and assessment of performance antipatterns in component based systems. The authors of [5] focus on detection of performance antipatterns in architectural models. The article [6] explains the method to remove performance antipatterns from software by analyzing UML models. The paper [7] describes modelling and analysis of software performance antipatterns along with their constraints and solvability using different models.

As mentioned above there has been work done in the field of performance antipatterns, defining solid means in terms of their definitions, detection, and fixing. These efforts focus on generic and domain-independent antipatterns, mostly in the application layer. However, they do not touch the data layer. Numerous performance problems have their source in an inefficient way of retrieving or storing large amount of data in a database. Therefore, this area is important to correctly identify the source of performance issues in software. It is especially true nowadays, when the amount of data processed continually increases and the use of automated frameworks for object-relational mapping becomes more and more common.

There is a number of popular ORM libraries, like Hibernate [1], EclipseLink [8], Open JPA [9], and Data Nucleus [10] to name a few. Even though they implement the same specification — Java Persistence API, they differ in mapping policies, generate different schemata and SQL statements that have different performance characteristics. The article [11] analyzes the influence of optimizations on the performance of Hibernate. In another paper [12], the same authors compare the performance of an object-relational mapping tool (Hibernate) vs. an object-oriented database (db4o) using OO7 benchmark. Another comparative study of the performance of object-relational mapping tools for .NET platform can be found.
III. ANTI-PATTERN: INADEQUATE COLLECTION TYPE ON OWNING SIDE

A. Description

In JPA, @OneToMany annotation used on the owning side (i.e., the side used to manage persistency of elements) is one of the most common implementations of a one-to-many association between persistent entities. Such an approach is commonly used in enterprise development mainly due its simplicity and little coding overhead. However, it can introduce a serious performance overhead in Hibernate when combined with an inadequate Java collection type.

Table I presents three types of semantics available in Hibernate, dependent on the combination of a Java collection type and JPA annotations. Each of these semantics has a different performance characteristic. Table II shows the numbers of statements issued while persisting a collection of a given semantics after an addition or removal of a single element.

The bag semantics has the worst performance when it comes to the number of operations since it always re-creates the entire collection. Hibernate issues a delete statement to remove all associations of the old collection from the association table. Then, it issues \( N \) inserts to add all associations representing the new collection to the association table. Hibernate does not analyze how many elements have been changed in the collection.

Table I: Three semantics for collections with @OneToMany annotation in Hibernate.

<table>
<thead>
<tr>
<th>Semantics</th>
<th>Java Type</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bag semantics</td>
<td>java.util.Collection</td>
<td>@OneToMany</td>
</tr>
<tr>
<td></td>
<td>java.util.List</td>
<td>@OneToMany</td>
</tr>
<tr>
<td>List semantics</td>
<td>java.util.List</td>
<td>@OneToMany / @IndexColumn / @OrderColumn</td>
</tr>
<tr>
<td>Set semantics</td>
<td>java.util.Set</td>
<td>@OneToMany</td>
</tr>
</tbody>
</table>

TABLE II

The numbers of DML statements issued while persisting a collection of \( N \) elements with a given semantics (default table mapping with an association table; no cascade option).

<table>
<thead>
<tr>
<th>Semantics</th>
<th>One Element Added</th>
<th>One Element Removed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bag semantic</td>
<td>1 delete, ( N ) inserts</td>
<td>1 delete, ( N ) inserts</td>
</tr>
<tr>
<td>List semantic</td>
<td>1 insert, ( M ) updates</td>
<td>1 insert, ( M ) updates</td>
</tr>
<tr>
<td>Set semantic</td>
<td>1 insert</td>
<td>1 delete</td>
</tr>
</tbody>
</table>

B. Consequences

The performance consequences of the usage of an inadequate collection type on the owning side of a one-to-many association include an increased workload on the database engine because:

- For usage patterns of collections where in a single transaction the collection is usually left unmodified or only a few elements are added or removed, the usage of the bag semantics (i.e., java.util.Collection or java.util.List without the index or order annotations) is notably inefficient.
- For usage patterns of collections where in a single transaction most of the elements of the collection are removed, the usage of the set or list semantics (i.e., java.util.Set or java.util.List with the index or order annotation) is inefficient.

- For usage patterns of collections where in a single transaction the collection is usually left unmodified or only a few elements are added or removed, the usage of the bag semantics (i.e., java.util.Collection or java.util.List without the index or order annotations) is notably inefficient.
- For usage patterns of collections where in a single transaction most of the elements of the collection are removed, the usage of the set or list semantics (i.e., java.util.Set or java.util.List with the index or order annotation) is inefficient.

- For usage patterns of collections where in a single transaction the collection is usually left unmodified or only a few elements are added or removed, the usage of the bag semantics (i.e., java.util.Collection or java.util.List without the index or order annotations) is notably inefficient.
- For usage patterns of collections where in a single transaction most of the elements of the collection are removed, the usage of the set or list semantics (i.e., java.util.Set or java.util.List with the index or order annotation) is inefficient.
C. Solution

The solution to this antipattern is to analyze the usage profile of collections in an application and adjust the type of Java collections accordingly:

- If a collection is constant or is the subject to minimal changes or is relatively small, the recommended collection type is `java.util.Set` with the set semantics.
- If a collection is heavily modified, the recommended collection type is `java.util.Collection` or `java.util.List` with the bag semantics.

D. Sample Code

Listing 1 presents an example of the antipattern related to an inadequate collection type on the owning side. The sample code has two persistent entities (Forest and Tree), which are connected by an unidirectional association (Forest has a collection of Tree objects). The classes meet all requirements imposed by JPA and Hibernate on persistent entities (e.g., no-arg constructor). We use a minimal set of additional annotation and configuration parameters, instead relying on default values.

To test the persistency mechanism implemented in Hibernate, we execute two transactions. To ensure the same runtime environment (e.g., clear caches), each transaction is executed with a new EntityManager instance. The first transaction creates a Forest instance and 10000 Tree instances planted in the newly created Forest, whereas the second transaction finds the previously created Forest instance, creates a new Tree instance and plants it in the found Forest. It turns out that Hibernate for such a piece of code as in the second transaction re-creates the entire collection (i.e., executes one delete and 10,001 inserts). This behavior of Hibernate is not performance-wise since in our example one insert would be enough to synchronize the state of the object in the main memory with the state of the records in the database. For large collections with only a few changes it imposes a significant performance overhead.

IV. ANTIPATTERN: ONE-TO-MANY AS OWNING SIDE

A. Description

The antipattern relates to the usage of the collection side (i.e., `@OneToMany`) as the owning side of an association for large collections, especially the ones which expect only a few changes in a single transaction.

According to Section III, for such a use case we should use `java.util.Set` to minimize the performance overhead. However, even usage of `java.util.Set` does not guarantee the optimal performance.

First, due to the Java set semantics the entire collection needs to be loaded into the main memory in order to enforce a uniqueness check in case of addition of elements to the collection. It results in an additional database query (or queries due to batch loading) issued and additional processing dedicated to the transformation of each returned row into an object. These additional queries and processing happen even if the application logic does not access the elements of the collection in question.

Second, there might be an overhead connected with transactions and locking, when the entity containing the collection in question uses optimistic locking with versioning. In such cases, Hibernate locks not only the entity but also the collection, which lowers the capability of an application to serve concurrent requests.

B. Consequences

The performance consequences of the `@OneToMany` as the owning side antipattern are as follows:

- There is an increased workload on the database engine. Hibernate issues an additional database query (or a number of queries in case of batch loading) to retrieve the elements of a collection.
- There is an increased workload on the application server (CPU). Hibernate needs to convert each returned row into an object, even if the application logic does not access those objects.
- The memory footprint is significant, especially for large collections. It can result in slower performance due to insufficient memory available, leading to more frequent garbage collections or even page swaps.

Listing 1. The example of an inadequate collection type on the owning side.
```java
public class Tree {
    private @GeneratedValue
    private Long id;
    private String name;
    ...
}
```
```
// Transaction 1
// creates and persists a forest...
...
// Transaction 2
Tree tree = new Tree("oak");
em.persist(tree);
Forest forest = em.find(Forest.class, id);
forest.plantTree(tree);
```
C. Solution

The solution to this antipattern is to manage a collection from the @ManyToMany side instead of @OneToMany side, i.e., to make @ManyToMany the owning side of the association. In case we are not able to change the owning side of an association, we should at least exclude such a collection from locking by using a Hibernate-specific annotation: @OptimisticLock(excluded=true).

D. Sample Code

Listing 2 presents an example of @OneToMany as the owning side.

```java
@Entity
public class Forest {
    @Id @GeneratedValue
    private Long id;
    @Version
    private.Integer version;
    @OneToMany
    Set<Tree> trees = new HashSet<Tree>(){
        public void plantTree(Tree tree){
            trees.add(tree);
        }
    }
}
```

• There may be a decreased throughput. Due to transaction and locking issues, the capability of an application to serve concurrent requests may be significantly lowered.

V. ANTIPATTERN: INADEQUATE COLLECTION TYPE ON INVERSE SIDE

A. Description

This antipattern relates to the usage of java.util.Set on the inverse side (also known as the mapped collection) of a bidirectional one-to-many association in Hibernate.

In case of bidirectional associations, it is recommended to synchronize the state of the objects (i.e., the mapped collection and the element entities with @ManyToMany pointers) in memory. The common implementation to ensure the consistency of the objects relies on an automated update of the mapped collection by the setter method in an element entity responsible for setting the @ManyToMany pointer (see Section V-D).

Therefore, when the type of the mapped collection is java.util.Set, it means that the entire collections needs to be loaded into the main memory in response to each addition of new elements, even though the application does not access neither the collection nor its elements. It happens due to the Java set semantics mentioned earlier, which requires a uniqueness check on the elements of a set.

B. Consequences

The performance consequences of the usage of java.util.Set on the inverse side of a bidirectional one-to-many association with in-memory state synchronization are as follows:

• There is an increased workload on the database engine. Hibernate issues an additional database query (or a number of queries in case of batch loading) to retrieve the elements of a collection.

• There is an increased workload on the application server (CPU). Hibernate needs to convert each returned row into an object, even if the application logic does not access those objects.

• For large collections, there is a significant memory footprint. It can slow down performance of an application due to insufficient memory available, more frequent garbage collections, or even page swaps.

C. Solution

The recommended Java types to be used on the inverse side of a one o many association are java.util.Collection or java.util.List. These types do not force loading the elements into memory until the elements are accessed by client code.

D. Sample Code

Listing 2 presents an example of java.util.Set used on the @OneToMany inverse side with in-memory state synchronization. The example continues the previously described Forest and Tree classes presented in Sections III and IV. Here, Tree is the owning side of an association, while Forest.
is the inverse side. Therefore, to save changes in a database, we need to set a right Forest reference in a Tree instance. While setting the Forest instance in a Tree, the collection of Trees is automatically updated to include the new tree (the invocation of plantTree). Such a pattern is widely used and recommended to ensure up-to-date states of objects. However, it results in a significant performance overhead, as the second transaction, which only adds a new tree and does not access other trees in the forest, Hibernate has to load the entire collection into the main memory.

In such a case, Hibernate is not able to track what has been changed in a collection and its policy is to re-create the entire collection regardless of the actual modifications. Therefore, even if the elements of the collection have not been changed, Hibernate issues a delete to remove the associations from the association table and then performs as many inserts as there are elements in the collection.

B. Consequences

The performance consequences of a lost proxy on the owning side are as follows:

- There is an increased workload on the database engine.
- There are unnecessary database operations performed.
- Hibernate re-creates an entire collection, performing one delete to clear the collection and as many inserts as there are elements in the collection. Such a re-creation results in suboptimal performance due to the operations on data which actually have not been changed.

C. Solution

The recommended solution to the antipattern is to operate on collection objects returned by Hibernate as in most cases it is a much more efficient approach. However, it might be performance-wise to re-create the entire collection in cases where most elements of the collection have been removed. On the other hand, it is one of the places where Hibernate could apply a smarter policy, especially as it has all required data available.

D. Sample Code

Listing 4 presents an example of a lost (according to Hibernate) collection proxy on the owning side. The example consists of two persistent entities: Hydra and Head. Hydra is a mythical creature that re-grows three heads in place of one head cut off. In order to model this feature, we need to provide strict encapsulation of heads. Therefore, getHeads returns an unmodifiable wrapper over the mutable collection of heads. In the first transaction, we create and persist a Hydra instance with three Heads. In the second transaction, we simply read the previously stored instance. It turns out that for this piece of code Hibernate executes two selects, one delete and three inserts, even though the code is purely procedural. The method returns an unmodifiable wrapper over the original proxy returned by Hibernate. Obviously, it returns a different Java object that the original one loaded by Hibernate. Thus, Hibernate decides that the collection has been changed and re-creates the entire collection.

Listing 5 presents a more straightforward example of a lost collection proxy on the owning side. Again we implement two
Listing 4. The example 1 of a lost collection proxy on the owning side.

```java
@Entity
public class Hydra {
    private Long id;
    private List<Head> heads = new ArrayList<Head>();
    ...
    @Id @GeneratedValue
    public Long getId() {...}
    protected void setId() {...}
    @OneToMany(cascade=CascadeType.ALL)
    public List<Head> getHeads() {
        return Collections.unmodifiableList(heads);
    }
    protected void setHeads(List<Head> heads) {
        this.heads = heads;
    }
}
...
// Transaction 1
// creates and persists the hydra...
// ...with 3 heads
...
// Transaction 2
Hydra found = em.find(Hydra.class, id);
```

persistent classes: Hydra and Head. However, we do not introduce strict encapsulation. Instead we implement simple getters and setters for all fields. In the second transaction, we create a new collection containing the current heads of our Hydra instance. In terms of our business model, nothing has changed — the heads are the same heads as originally loaded. However, Hibernate observes a different collection reference and applies the policy of re-creation of the entire collection.

VII. ANTIPATTERN: ONE-BY-ONE PROCESSING OF COLLECTION

A. Description

The antipattern refers to a sequential processing of a persistent collection, i.e., a piece of code iterates over the collection and for each element in a collection, it may perform a database operation.

Best practices of database programming highlight the need to operate on the sets of records instead of single records. While SQL provides the means to such a paradigm switch in programming, Java is an object-oriented language without support to relational algebra. Therefore, it is a common approach in the Java world to iterate over persistent collections and process their elements one-by-one. Such an approach often leads to a significant performance overhead, considering the number of database round-trips and the volume of data passed between a database and an application.

Listing 5. The example 2 of a lost collection proxy on the owning side.

```java
@Entity
public class Hydra {
    @Id @GeneratedValue
    private Long id;
    @OneToMany(cascade=CascadeType.ALL)
    private List<Head> heads = new ArrayList<Head>();
    ...
    public Long getId() {...}
    protected void setId() {...}
    public List<Head> getHeads() {
        return heads;
    }
    public void setHeads(List<Head> heads) {
        this.heads = heads;
    }
}
...
// Transaction 1
// creates and persists the hydra...
// ...with 3 heads
...
// Transaction 2
Hydra found = em.find(Hydra.class, id);
List<Head> currentHeads = new ArrayList<Head>(found.getHeads());
found.setHeads(currentHeads);
```

B. Consequences

The performance consequences of one-by-one processing of a persistent collection are as follows:

- A high number of database operations is executed. It is proportional to the size of the collection.
- RDBMS engine is used ineffectively.
- Network latency can sum up to a significant performance overhead.

C. Solution

The solution to this antipattern is to utilize the capabilities of a relational database by the usage of bulk statements and aggregate functions. Frequently it requires a different object model.

D. Sample Code

Listing 6 presents an example of a one-by-one processing of a collection. The example continues the previous examples consisting of Forest and Tree. Here, in the second transaction we want to delete the entire Forest. A simple remove causes a constraint violation exception since there are trees associated with the forest to be removed. Therefore, we need to unbind the trees first. Unfortunately, in Hibernate there is no other way to do this than setting the Forest reference in each Tree instance to null. In our example it results in
impact on the performance as it usually leads to more frequent garbage collection or even page swaps.

The presented antipatterns show that the usage of Hibernate is not as simple as it looks at first glance. Even plain use cases can significantly decrease the performance of an application. The antipatterns explain how to use Hibernate efficiently and what policies should be improved in Hibernate in order to shorten the execution time.

REFERENCES