Data Stream Query Optimization Using Deep Reinforcement Learning

Final presentation

Apoorva Tamaskar

Boston University
Metropolitan College

01 February 2021

Outline

Section 1

Background Information

Databases

How do databases store, retrieve and manipulate data?

1. Where is the data stored? Centralised, Distributed, Personal

Databases

How do databases store, retrieve and manipulate data?

1. Where is the data stored? Centralised, Distributed, Personal
2. Is the data format fixed? SQL vs NoSQL

Databases

How do databases store, retrieve and manipulate data?

1. Where is the data stored? Centralised, Distributed, Personal
2. Is the data format fixed? SQL vs NoSQL
3. What are the requirements? Scale, operations to be performed, QoS metrics to meet.

Databases

How do databases store, retrieve and manipulate data?

1. Where is the data stored? Centralised, Distributed, Personal
2. Is the data format fixed? SQL vs NoSQL
3. What are the requirements? Scale, operations to be performed, QoS metrics to meet.
4. What operations are going to be performed? E.g. Updates, Deletes, adding new fields, queries to be answered

Databases

How do databases store, retrieve and manipulate data?

1. Where is the data stored? Centralised, Distributed, Personal
2. Is the data format fixed? SQL vs NoSQL
3. What are the requirements? Scale, operations to be performed, QoS metrics to meet.
4. What operations are going to be performed? E.g. Updates, Deletes, adding new fields, queries to be answered
5. Online system or offline system? Handling static data Vs Data streams.

SQL Vs NoSQL

Some of the differences between these two are:-

1. SQL databases are table based databases whereas NoSQL databases can be document based, key-value pairs, graph databases.

SQL Vs NoSQL

Some of the differences between these two are:-

1. SQL databases are table based databases whereas NoSQL databases can be document based, key-value pairs, graph databases.
2. SQL databases are vertically scalable while NoSQL databases are horizontally scalable.

SQL Vs NoSQL

Some of the differences between these two are:-

1. SQL databases are table based databases whereas NoSQL databases can be document based, key-value pairs, graph databases.
2. SQL databases are vertically scalable while NoSQL databases are horizontally scalable.
3. SQL databases have a predefined schema whereas NoSQL databases use dynamic schema for unstructured data.

SQL Vs NoSQL

Some of the differences between these two are:-

1. SQL databases are table based databases whereas NoSQL databases can be document based, key-value pairs, graph databases.
2. SQL databases are vertically scalable while NoSQL databases are horizontally scalable.
3. SQL databases have a predefined schema whereas NoSQL databases use dynamic schema for unstructured data.
4. SQL requires specialized DB hardware for better performance while NoSQL uses commodity hardware.

SQL Vs NoSQL

Some of the differences between these two are:-

1. SQL databases are table based databases whereas NoSQL databases can be document based, key-value pairs, graph databases.
2. SQL databases are vertically scalable while NoSQL databases are horizontally scalable.
3. SQL databases have a predefined schema whereas NoSQL databases use dynamic schema for unstructured data.
4. SQL requires specialized DB hardware for better performance while NoSQL uses commodity hardware.
5. SQL is an ideal choice for the complex query intensive environment and NoSQL is a best used for solving data availability problems.

SQL

In the thesis we focus on Structured Query Language.
How does a SQL database look like?
An SQL database is a collection of tables of data.

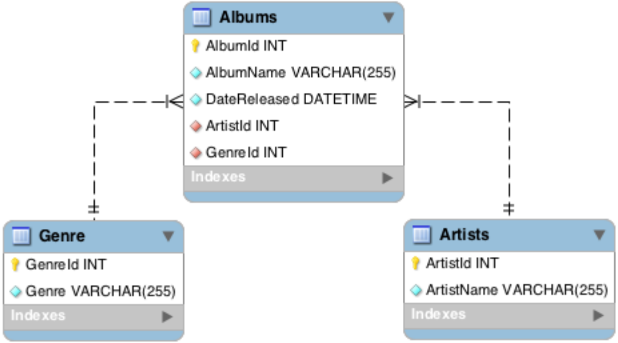


Figure: Album database

SQL

How does a table in SQL look like?
A table can be thought of a matrix, with each row representing a data point and column representing the attribute value for the data point.

CustomerID	CustomerName	ContactName	Address	City	PostalCode	Country
1	Alfreds Futterkiste	Maria Anders	Obere Str. 57	Berlin	12209	Germany
2	Ana Trujillo Emparedados y helados	Ana Trujillo	Avda. de la Constitución 2222	México D.F.	05021	Mexico
3	Antonio Moreno Taquería	Antonio Moreno	Mataderos 2312	México D.F.	05023	Mexico
4	Around the Horn	Thomas Hardy	120 Hanover Sq.	London	WA1 1DP	UK
5	Berglunds snabbköp	Christina Berglund	Berguvsvägen 8	Luleå	S-958 22	Sweden

Figure: Customer Database

SQL

What is a SQL query?
An SQL query is a question or a request for answer on a database.

```sql
        SELECT * FROM Customers
        WHERE Country='Mexico';

```

Listing 1: SQL statement selects all the customers from the country "Mexico" in the "Customers" table

SQL

Hows is a SQL query executed?

```sql
        SELECT MovieTitle
        FROM StarsIn
        WHERE StarName IN(
            SELECT name
            FROM MovieStar
            WHERE birthdate LIKE '%1960'
        );

```

Listing 2: SQL query to convert

SQL: Pipeline

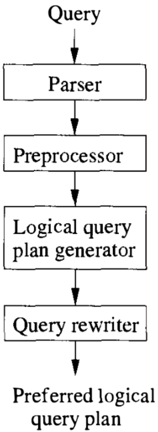


Figure: The pipeline for query processing

SQL: Parser

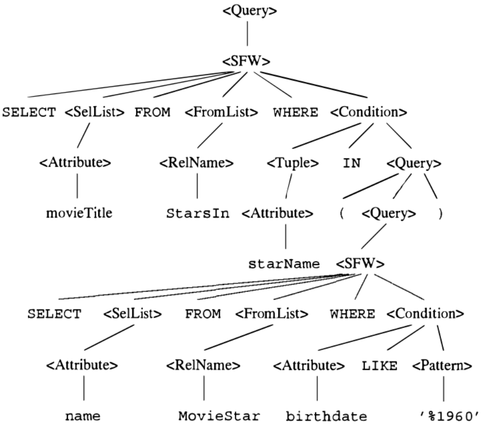


Figure: An example of parse tree

SQL: Relational Algebra: Selection

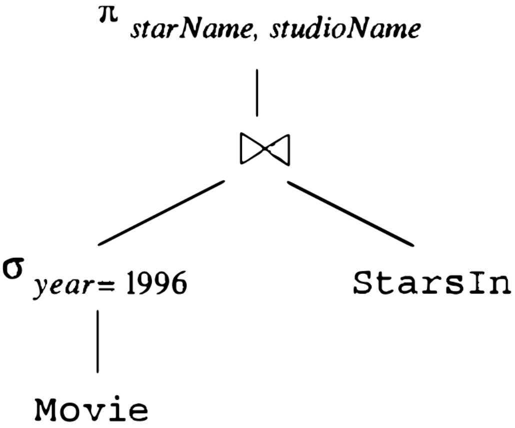


Figure: An example of selection being pushed down for optimization

SQL: Relational Algebra: Selection

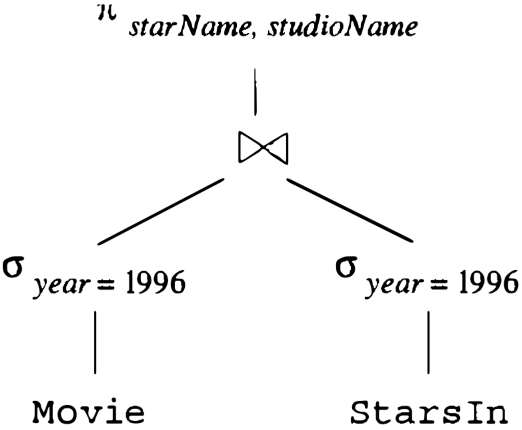


Figure: An example of selection being pushed down for optimization

SQL: Relational Algebra: Projection

$$\pi_{starName}$$

$$\sigma_{movieYear = 1996}$$

StarsIn

Figure: An example of projection being pushed down for optimization

SQL: Relational Algebra: Projection

$$\pi_{starName}$$

$$\sigma_{movieYear = 1996}$$

$$\pi_{starName,\ movieYear}$$

`StarsIn`

Figure: An example of projection being pushed down for optimization

SQL: Relational Algebra: Join

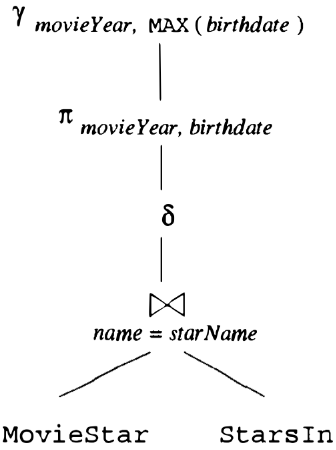


Figure: An example of join being optimized

SQL: Relational Algebra: Join

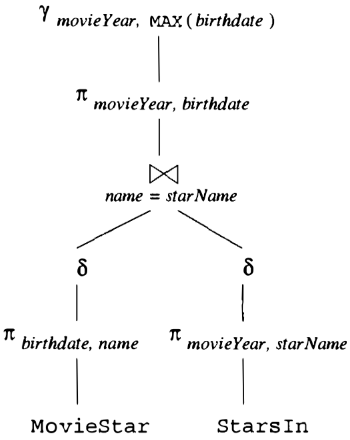


Figure: An example of join being optimized

SQL: Grouping operators

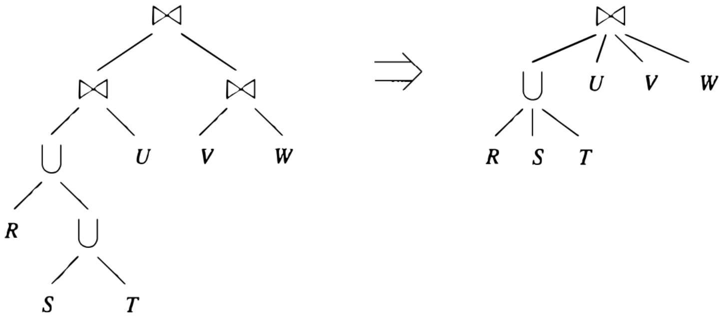


Figure: An example associative operators being grouped

SQL

a

SQL

a

SQL

a

SQL

a

Data Streams

Stream query optimization is the process of modifying a stream processing query, often by changing its graph topology and or operators, with the aim of achieving better performance (such as higher throughput, lower latency, or reduced resource usage), while preserving the semantics of the original query.
Stream query optimizations are best understood with respect to stream graphs. A stream graph is a directed graph whose edges are streams and whose nodes are operators. Root and leaf nodes are called sources and sinks, respectively.

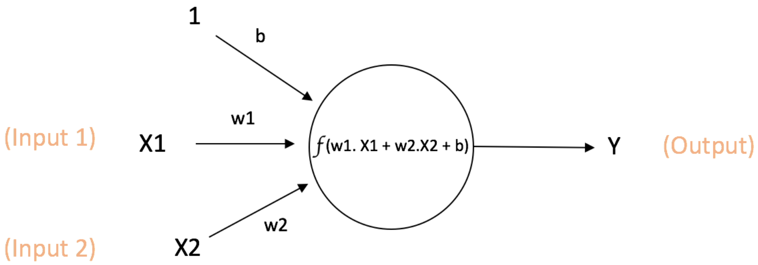
Possible Optimizations

- Batching
- Placement
- State sharing
- Load Balancing
- Algorithm selection
- Load Shedding
- Fusion
- Operator Separation
- Operator Reordering
- Redundancy elimination
- Fission

We focus on Operator Reordering.

Deep Neural Networks

To understand a neural network we should first look at a neuron.



Output of neuron $= Y = f(w1.X1 + w2.X2 + b)$

Figure: Example of a neuron

f is generally taken to be a non linear function. This non linearity grants neural networks additional flexibility.

Deep Neural Networks

Deep neural networks is a layer wise combinations of neurons
Building up on the neuron seen in the last slide. We have

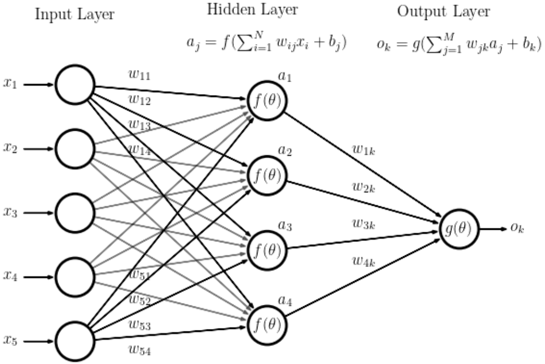


Figure: Example of a deep neural network

Backpropogation

How does a neural network train on these parameters? First look at how a single node back propogates.

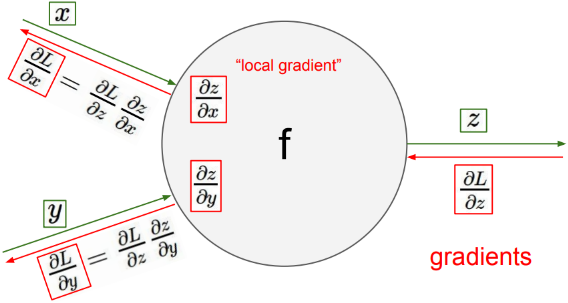


Figure: Example of backpropogation on a node

Backpropogation

By doing backpropogation on each node, we finish the process.

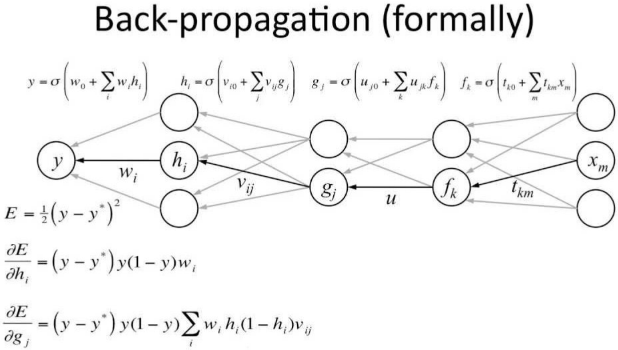


Figure: Example of backpropogation

Reinforcement Learning

What is reinforcement learning? How is it different from supervised and unsupervised learning?

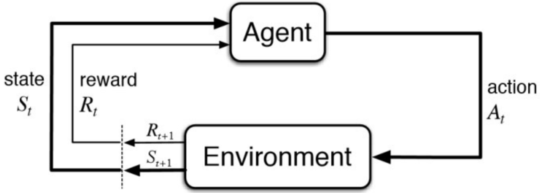


Figure: Example of a framework of a reinforcement learning agent

Value Iteration

```
for s in S:
    V(s)=0
while(not converged):
    for s in S:
        V(s)=R(s)+max over all action[gamma*(sum(P(s,a,s')V(s')))]

```

Listing 3: value iteration algorithm

Policy Iteration

```
initialize random pi
while(not converged):
    V=V(pi)
    for s in S:
        pi(s)=max over all actions[sum(P(s,a,s')V(S'))]
```

Listing 4: Policy iteration algorithm

Deep Reinforcement Learning

Deep reinforcement learning combines Deep learning and reinforcement learning

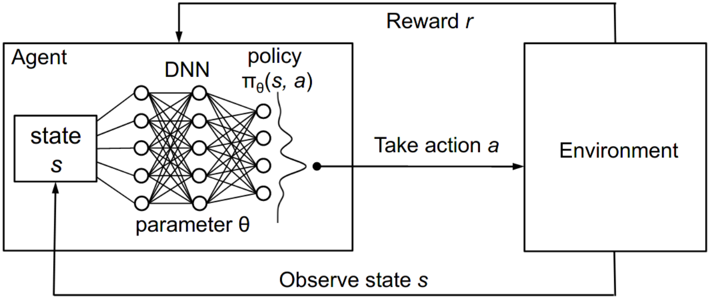


Figure: Deep Reinforcement learning(DQN) framework