



Jernej Kavka (JK)

SSW Solution Architect



@Jernej kavka



github.com/jernejk



<u>jkdev.me</u>



linkedin.com/in/jernejkavka/



Brisbane Full Stack User Group



Host @ Global Al The Podcast



Microsoft AI MVP and AI generalist

Join the Conversation @SSW_TV @jernej_kavka #EFCore







APR MAR JAN FEB MONDAY SUNDAY MONTHLY TIP PLAN SMART. FILL OUT YOUR WEEKLY AGENDA NOTING WHEN YOU WILL DO WHAT AND Intro NOTES The 7 Deadly Sins Mini Best Practices

Agenda

We have lots of DB engines data access landscape



SQL Server



Cosmos DB



PostgreSQL



MySQL

Various ORM



EF Core (multiple DBs)



Dapper (multiple DBs)



Marten (PostgreSQL)



Cosmos DB SDK

Why I F Core

- ☑ Great balance
 - Features
 - **⋠** Flexibility
 - Performance
- Awesome for relational DBs
- Built-in migrations, DB scaffolding and other tooling

Avoid the pain

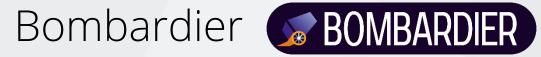
- The EF Core issues that are solved
- Simplifying your code
- Benchmark your code with tests
- The horror I have seen...



Benchmark tools

BenchmarkDotNet BenchmarkDotNet
Powerful .NET library for benchmarking

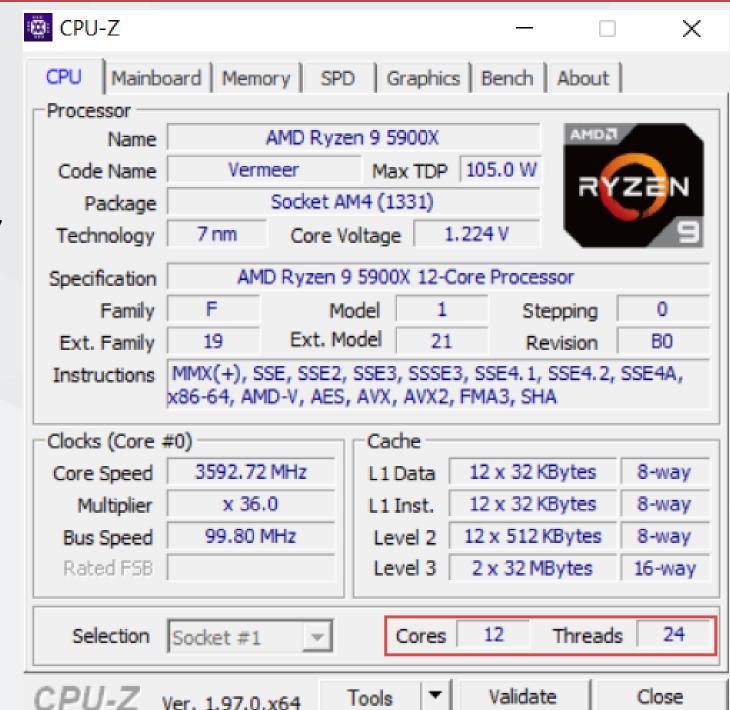
Memory consumption and time difference



Simple CLI load testing tool

My testing environment

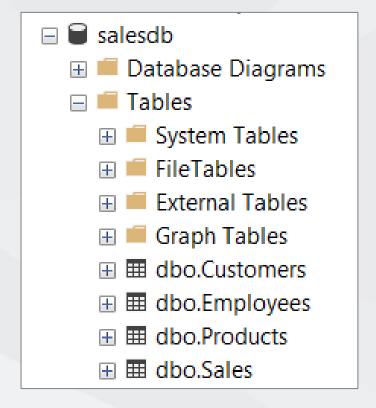
- .NET 7 with EF Core 7
- My PC
 - CPU: Ryzen 9 5900X
 - Cores: 12
 - Threads: 24
 - Ram: 32GB



Test data



- 7 million sales
- 504 products
- 23 employees
- 20,000 customers

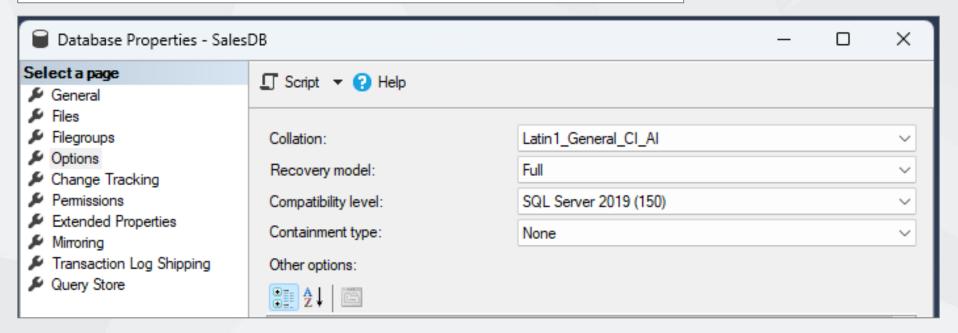


sqlskills.com/sql-server-resources/sql-server-demos/

Update to SQL Server 2016+

Compatibility level: 130+ (SQL Server 2016+)

ALTER DATABASE SalesDB SET COMPATIBILITY_LEVEL = 150;





The 7 Deadly EF Core Sins (easy to hard)

- 1. Casting **IQueryable** -> **IEnumerable**
- 2. Not using **AsNoTracking** (when appropriate)
- 3. Explicit joins
- 4. Getting all the columns
- 5. No pagination
- 6. Non-cancellable queries
- 7. Inefficient updates/deletes





(you never want the answer to this 🔒)

- Literally everyone before discovering the horrifying truth

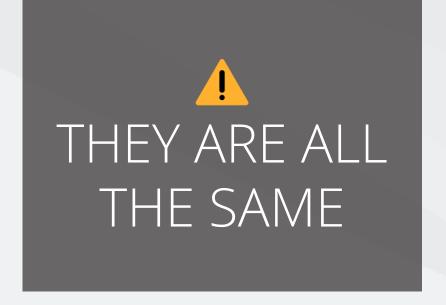
#1 IQueryable -> IEnumerable

```
[Benchmark]
0 references | 0 changes | 0 authors, 0 changes
public int NaiveCount()
{
    return SalesDbContext.Sales
    .ToList()
    .Count;
}
```

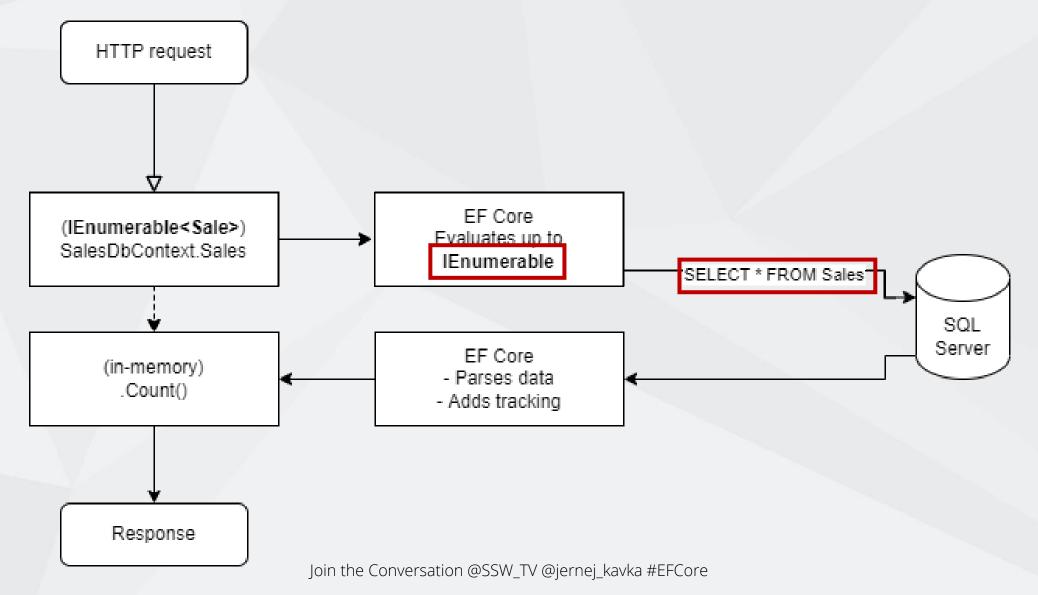
```
[Benchmark]
0 references | 0 changes | 0 authors, 0 changes
public int QuerableToEnumerableCount()
{
    IEnumerable<Sale> sales = SalesDbContext.Sales;
    return sales.Count();
}
```

```
[Benchmark]
0 references | 0 changes | 0 authors, 0 changes
public int ImplicitEnumerableCount()
{
    return GetSalesEnumerable().Count();
}

1 reference | 0 changes | 0 authors, 0 changes
private | IEnumerable<Sale> GetSalesEnumerable()
{
    return SalesDbContext.Sales;
}
```



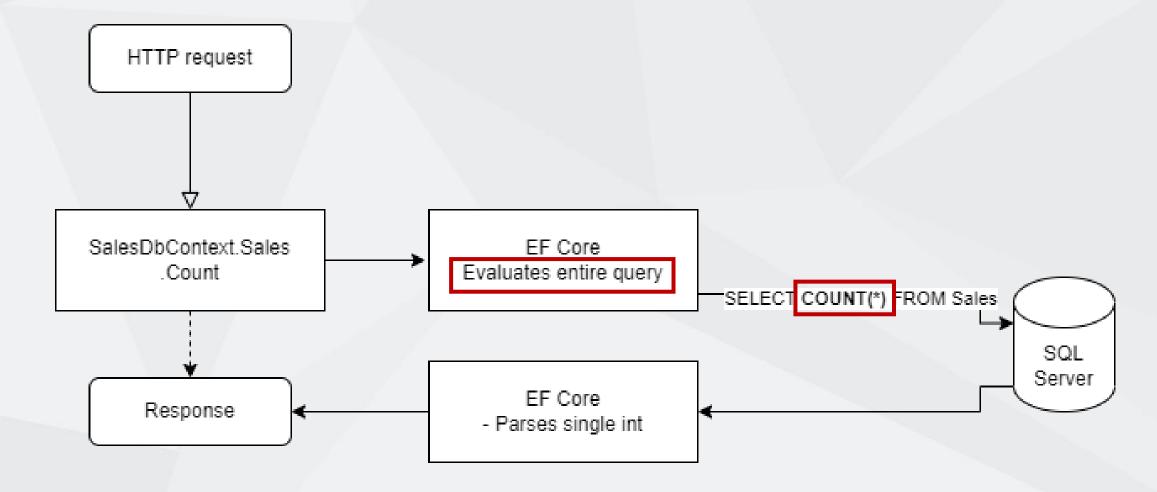
#1 What is going on?



#1 IQueryable

```
[Benchmark(Baseline = true)]
0 references | 0 changes | 0 authors, 0 changes
public int CountInDb()
    return GetSales().Count();
1 reference | 0 changes | 0 authors, 0 changes
private IQueryable<Sale> GetSales()
    return SalesDbContext.Sales;
```

#1 How is this better?



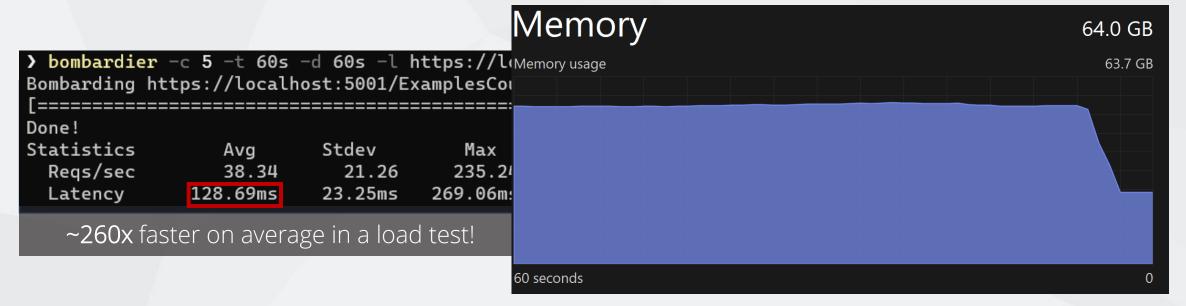
IQueryable + EF Core (EF Core LINQ)

#
IEnumerable (Standard LINQ)

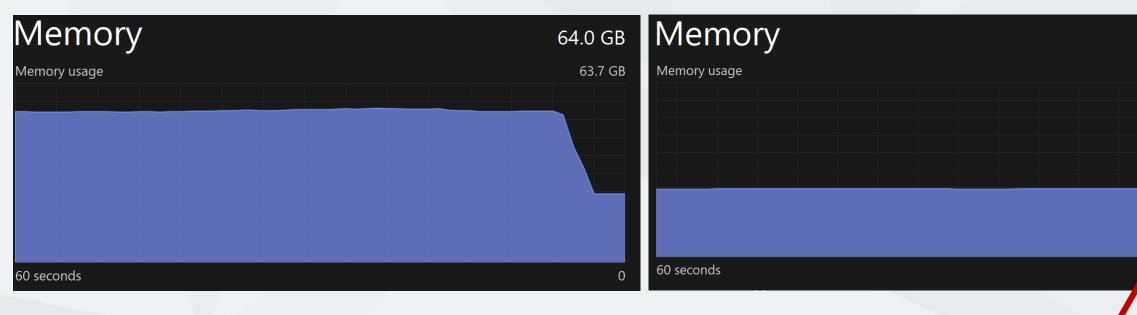
#1 What is the memory impact?



#1 What about load testing?



#1 Really bad vs good



64.0 GB

63.7 GB

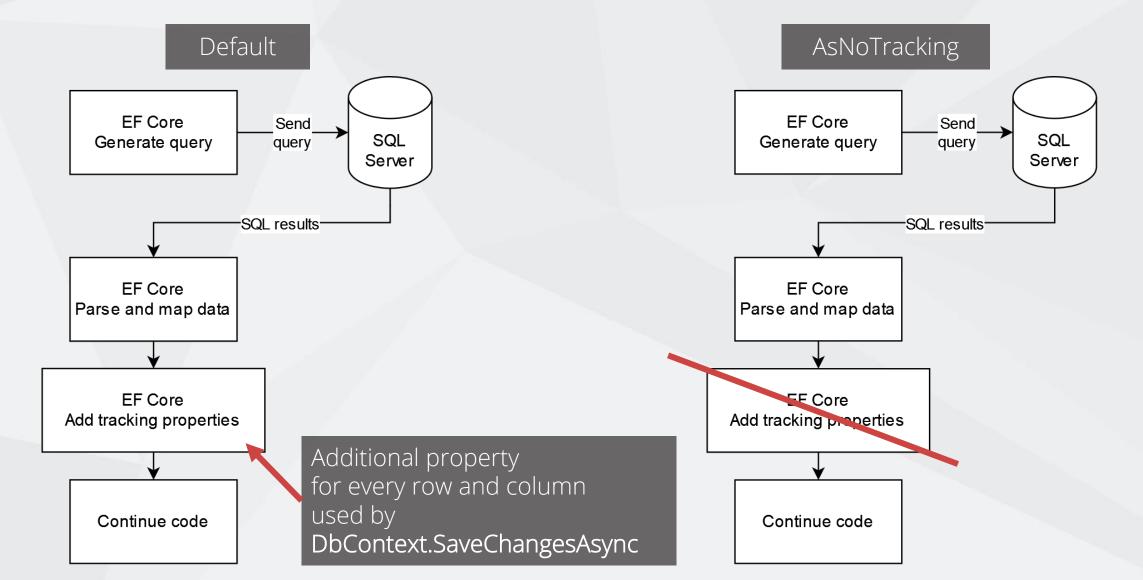
#2 AsNoTracking

Exclude change tracking on entities

Less memory and CPU usage

Changes to the entities are not tracked

#2 What is happening?



#2 What is the impact?



⚠ Results be wildly different between queries and DB providers! ⚠ Some found it has no effect on SQLite DB provider

#2 What about load testing?

```
> bombardier -c 5 -t 60s -d 60s -l https://localhost:5001/ExamplesWhere/whereNoTracking
Bombarding https://localhost:5001/ExamplesWhere/whereNoTracking for 1m0s using 5 connection(s)
Done!
Statistics
                                        Max
                  Avg
                           Stdev
  Reqs/sec
                             13.34
                  14.72
                                       249.94
 Latency
               358.92ms
                         105.40ms
                                        1.48s
     ~4x faster on average in a load test!
```

#3 Explicit includes

.Include(x => x.Customers)

They are always included even if not needed

Devs often forget to remove them

We get all columns even if we don't need them

#3 & #4 Implicit includes with Select

```
var query = _dbContext.Sales
                                                                   .AsNoTracking()
                                                                   .Where(x \Rightarrow x.SalesPersonId == 1)
                                                                   .Select(x \Rightarrow new SalesWithSalesPerson)
var query = _dbContext.Sales
    .AsNoTracking()
                                                                       CustomerId = x.CustomerId,
    .TagWithContext()
                                                                       SalesId = x.SalesPersonId,
    .Include(x \Rightarrow x.SalesPerson)
                                                                       ProductId = x.ProductId,
    .Where(x \Rightarrow x.SalesPersonId == 1);
                                                                       Quantity = x.Quantity,
                                                                       SalesPersonId = x.SalesPersonId.
                                                                       SalesPersonFirstName = x.SalesPerson.FirstName,
                                                                       SalesPersonLastName = x.SalesPerson.LastName
```

#3 & #4 Load Test

```
> bombardier -c 5 -t 60s -d 60s -l https://localhost:5001/ExamplesJoin/implicitJoin
Bombarding https://localhost:5001/ExamplesJoin/implicitJoin for 1m0s using 5 connection(s)
Done!
Statistics
                 Avg
                          Stdev
                                       Max
 Reqs/sec
                  7.49
                            25.53
                                      263.10
 Latency
              748.18ms
                         575.21ms
                                       5.30s
   ~2.5x faster on average in a load test!
```

#5 Common Pagination issues

- Inefficient code
 - Filter data
 - Get all filtered data
 - Count them
 - Apply Take and Skip in-memory for pagination

#5 Pagination code – naïve

```
IQueryable<Sale> query = _dbContext.Sales
    .AsNoTracking()
    // Very commonly used in combination of in-memory pagination, making things even worse
    .Include(x \Rightarrow x.SalesPerson)
    .Where(x \Rightarrow x.SalesPersonId == salesPersonId);
                                                          Gets all filtered data
List<Sale> dbResult = await guery.ToListAsync(ct);
int count = dbResult.Count;
                                                          Count in-memory
List<SalesWithSalesPerson> result = dbResult
                                                          Pagination is done
    .Skip(page * pageSize)
    .Take(pageSize)
                                                          In-memory
    .Select(x \Rightarrow new SalesWithSalesPerson
        CustomerId = x.CustomerId,
        SalesId = x.SalesPersonId,
        ProductId = x.ProductId,
        Quantity = x.Quantity,
        SalesPersonId = x.SalesPersonId,
        SalesPersonFirstName = x.SalesPerson.FirstName,
        SalesPersonLastName = x.SalesPerson.LastName
    .ToList();
```

#5 Pagination code – improved

```
We are preparing the query
IQueryable<SalesWithSalesPerson> query = _dbContext.Sales
                                                            For counting and pagination
    .AsNoTracking()
    .Where(x \Rightarrow x.SalesPersonId == salesPersonId)
    .Select(x ⇒ new SalesWithSalesPerson
        CustomerId = x.CustomerId,
        SalesId = x.SalesPersonId,
        ProductId = x.ProductId,
        Quantity = x.Quantity,
        SalesPersonId = x.SalesPersonId,
        SalesPersonFirstName = x.SalesPerson.FirstName,
        SalesPersonLastName = x.SalesPerson.LastName
    3);
// After all conditions are applied, count them in DB.
int count = await query.CountAsync(ct)
                                                           Count in SQL Server
// Apply paginations, sorts and more complex select statements.
query = query
                                                           Pagination is done
    .Skip(page * pageSize)
                                                           on SQL Server
    .Take(pageSize);
List<SalesWithSalesPerson> result = await query.ToListAsync(ct)
```

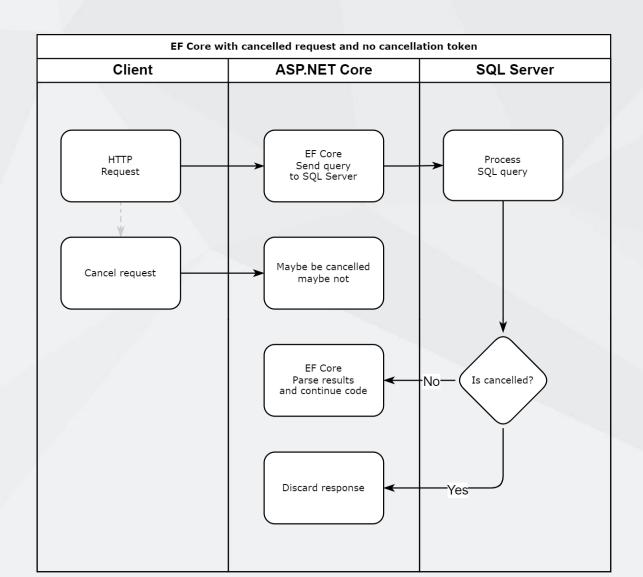
#5 Load test

```
> bombardier -c 5 -t 60s -d 60s -l https://localhost:5001/ExamplesPaginations/executedOnDB
Bombarding https://localhost:5001/ExamplesPaginations/executedOnDB for 1m0s using 5 connection(s)
Done!
Statistics
                 Avg
                          Stdev
                                       Max
 Reqs/sec
                 65.41
                            21.98
                                       263.10
              76.94ms
 Latency
                          13.79ms
                                    237.05ms
    ~18x faster on average in a load test!
```

#6 Non-cancellable queries

- When query is running on SQL Server
 - It will run until completion
- Complex queries can stuff up SQL Server

#6 What's the problem?

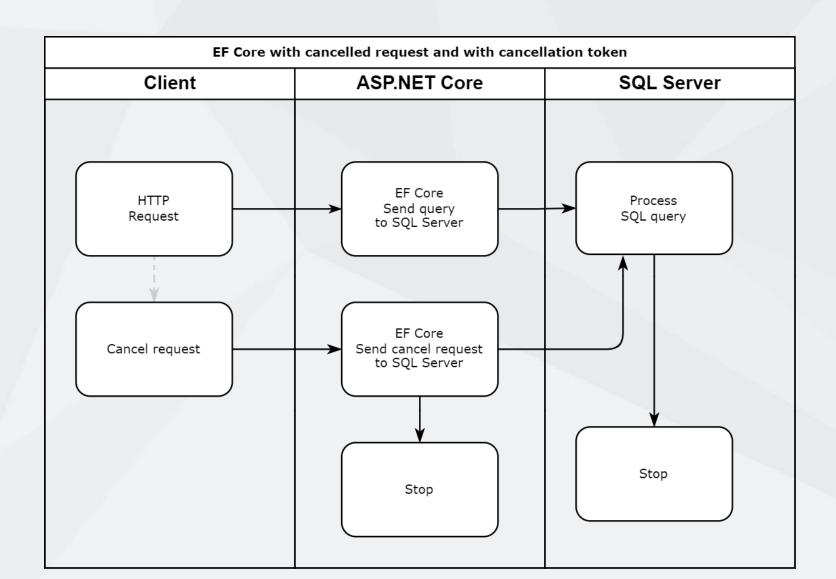


#6 Async and CancellationToken

Stop execution on the server when cancelled

```
[HttpGet("splitQuery")]
0 references | Jernej Kavka (JK), 65 days ago | 1 author, 1 change
public async Task<TestResult<int≫ SplitQuery(CancellationToken ct)
    var query = _dbContext.Sales
        .AsNoTracking()
        .AsSplitQuery()
         .Where(x \Rightarrow x.SalesPersonId == 1)
        .Select(x ⇒ new SalesWithSalesPerson
            CustomerId = x.CustomerId,
            SalesId = x.SalesPersonId,
            ProductId = x.ProductId,
            Quantity = x.Quantity,
            SalesPersonId = x.SalesPersonId,
            SalesPersonFirstName = x.SalesPerson.FirstName,
            SalesPersonLastName = x.SalesPerson.LastName
    List<SalesWithSalesPerson> result = await query.TagWithContext().ToListAsync(ct);
```

#6 How it works cancellation token?



#6 Load test

~2x faster on average in a load test! Also, responsiveness returns 40 seconds faster

- Before EF7 we had to fetch data before could update
- Entities need to be tracking
- Can be very inefficient

IQueryable<Employee> query = GetBaseQuery(isLoadFriendly);

```
var employees = query.ToList();
foreach (var employee in employees)
    string firstName = employee.FirstName;
    employee.FirstName = employee.LastName;
    employee.LastName = firstName;
                                  -- Get all employees that we want to swap first and last name.
_dbContext.SaveChangesAsync();
                                  -- We need to fetch all columns as it needs to be tracking!
                                  SELECT TOP(@__p_0) [e].[EmployeeID], [e].[FirstName], [e].[LastName], [e].[MiddleInitial]
                                  FROM [Employees] AS [e]
                                  SET NOCOUNT ON;
                                  UPDATE [Employees] SET [FirstName] = @p0, [LastName]
                                  OUTPUT
                                  WHERE [EmployeeID] = @p2;
                                  -- Lots of updates ...
                                  UPDATE [Employees] SET [FirstName] = @p63, [LastName]
                                                                                          @p64
                                  OUTPUT
                                  WHERE [EmployeeID] = @p65;
                                  UPDATE [Employees] SET [FirstName] = @p66, [LastName]
                                                                                          @p67
                                  OUTPUT
                                  WHERE [EmployeeID] = @p68;
```

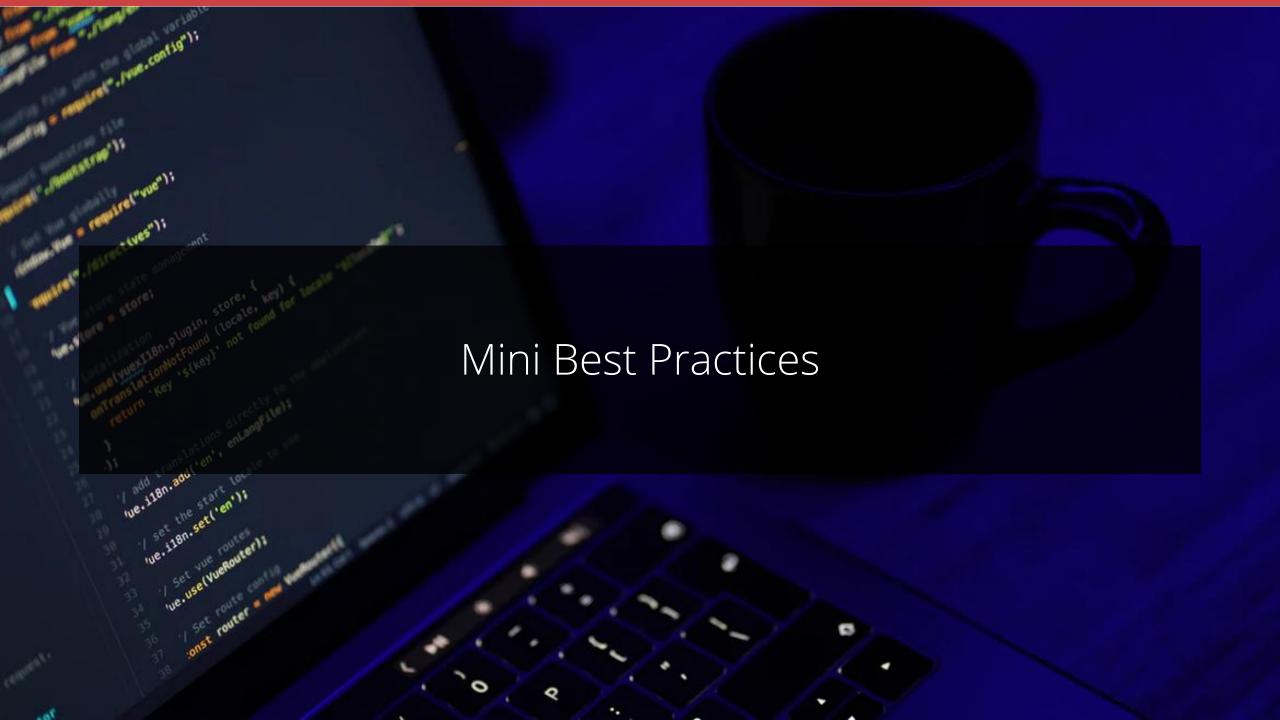
```
-- Swap first and last name

UPDATE [e]

SET [e].[LastName] = [e].[FirstName],
        [e].[FirstName] = [e].[LastName]

FROM [Employees] AS [e]
```

For only 23 records, it's still ~3x faster!
Performance improvements can be massive depending on the query!



Bonus #1 - Other things I have seen

```
0 references | 0 changes | 0 authors, 0 changes
public bool HasAny()
{
    var employees = _dbContext.Employees
        .Where(x \Rightarrow x.EmployeeId == 1)
        .ToList();
    return employees.Count > 0;
}

0 references | 0 changes | 0 authors, 0 changes
public bool HasAny()
{
    return _dbContext.Employees
        .Any(x \Rightarrow x.EmployeeId == 1);
}
```

```
Oreferences | O changes | O authors, O changes
public List<SaleModel> GetSales(int? saleId)
{
    var sales = _dbContext.Sales
        .Include(x ⇒ x.Product)
        ToList();

    if (saleId ≠ null)
    {
        sales = sales
        .Where(x ⇒ x.SalesId == saleId.Value)
        .ToList();
}

    return sales.Select(x ⇒ new SaleModel
    {
        SaleId = x.SalesId,
        Quantity = x.Quantity,
        ProductName = x.Product.Name
    }).ToList();
}
```

```
public List<SaleModel> GetSales(int? saleId)
{
    IQueryable<Sale> salesQuery = _dbContext.Sales;

    if (saleId ≠ null)
    {
        salesQuery = salesQuery
            .Where(x ⇒ x.SalesId == saleId.Value);
    }

    return salesQuery.Select(x ⇒ new SaleModel
    {
        SaleId = x.SalesId,
        Quantity = x.Quantity,
        ProductName = x.Product.Name
    }).ToList();
}
```

Bonus #2

DbContextPool

- Reuses existing **DbContext**
- Fewer connections to SQL Server
- Improves resiliency

Bonus #3

TagWith

Adds comments into SQL query for easier debugging

```
var list = await _context.Tweets
    .TagWith("GetTweets")
    .ToListAsync(cancellationToken)
    .ConfigureAwait(false);

-- GetTweets

SELECT [t].[Id], [t].[CreatedUtc], [t].[Message], [t].[Username]
FROM [Tweets] AS [t]
```

Bonus #4 – ChatGPT



You're an EF Core expert and clean code developer.

<user's instructions>

DbContext:

<Optional project's DbContext, can be trimmed to include only the relevant parts>

Entities:

<Only necessary entities>

Return types:

<Optional return type we want to map>



ChatGPT may produce inaccurate information about people, places, or facts. ChatGPT May 12 Version

Bonus #4 – ChatGPT



You're an EF Core expert and clean code developer.

Get all employees details, for each get statistics on how many produthey earned and a list of sales they did, all paginated.

```
DbContext:

public partial class SalesDbContext : DbContext

{
    public SalesDbContext(DbContextOptions<SalesDbContext> opt
        : base(options)
        {
        }
        public virtual DbSet<Customer> Customers { get; set; }
```

```
public async Task<List<EmployerStats>> GetEmployeeStats(int pageNumber,
   return await _context.Employees
        .OrderBy(e => e.EmployeeId)
        .Skip((pageNumber - 1) * pageSize)
        .Take(pageSize)
        .Select(e => new EmployerStats
           FirstName = e.FirstName,
           LastName = e.LastName,
           TotalProductSold = e.Sales.Count,
            TotalSales = e.Sales.Sum(s => s.Quantity * s.Product.Price),
           TotalQuantity = e.Sales.Sum(s => s.Quantity),
           Sales = e.Sales.Select(s => new SaleModel
               SaleId = s.SalesId,
                ProductName = s.Product.Name,
               Quantity = s.Quantity
           3)
        3)
        .ToListAsync();
```

Before we end #1

When is something **not** premature optimization?

- Following best practices for given scenario
- When high-load or high-data is expected
 - i Talk to Product Owner and make it part of acceptance criteria
- When it already causes performance issues
 - 1 Talk to Product Owner and make it part of acceptance criteria or create tech debt PBI/issue and mention it in the code
 - ssw.com.au/rules/technical-debt/

Before we end #2

What should you not do?

- "We'll fix it later" with no follow-up
 - i Create a tech debt PBI/issue and a link to it in code as a TODO comment
 - ssw.com.au/rules/technical-debt/



- "Premature optimization" as an excuse for bad code
 - i There are valid reason to not optimize your code, bad code is not one of them!

Resources 📮

Source code for examples and benchmarking

github.com/jernejk/EFCoreSamples.StabilityAndPerformance

Microsoft resources for better EF Core performance

docs.microsoft.com/en-us/ef/core/performance/

Rules to Better Entity Framework

ssw.com.au/rules/rules-to-better-entity-framework

Rules to Better LINQ

ssw.com.au/rules/rules-to-better-ling

Nick Chapsas YouTube – Great breakdowns of features and best practices youtube.com/watch?v=Q4LtKa_HTHU

Key takeaways ⊶

- IEnumerable is dangerous when used with EF Core
- AsNoTracking has massive performance impact
 - Make sure you don't need to update/delete the entities
- Simpler code can result in better performance

The sequel – NDC Sydney 2023

NDC { Sydney }

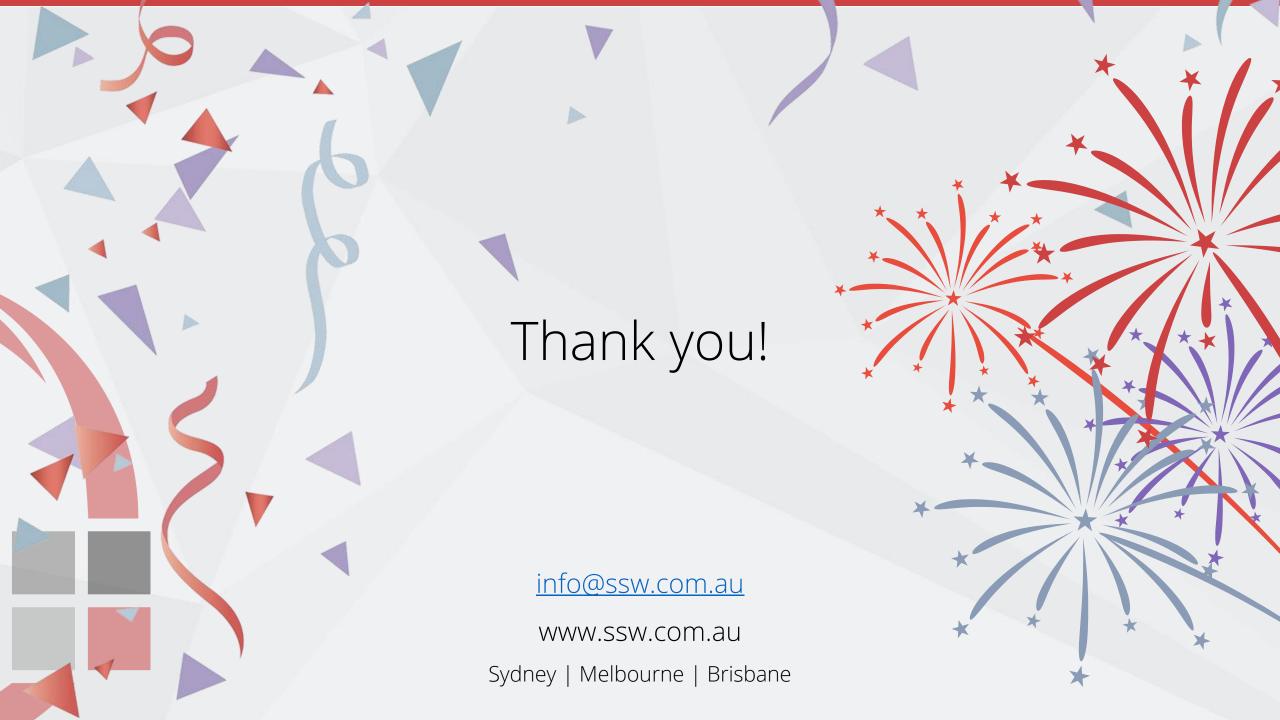
Thursday Room 4 17:40 - 18:40 (UTC+11) Talk (60 min)

Unlocking the Full
Potential: The
Harmonious Dance of
EF Core and SQL Server

Find this presentation on GitHub!



github.com/sswconsulting/presentations



Questions?

info@ssw.com.au

www.ssw.com.au

Sydney | Melbourne | Brisbane