

Data Oriented Design

for business applications



leveluppp

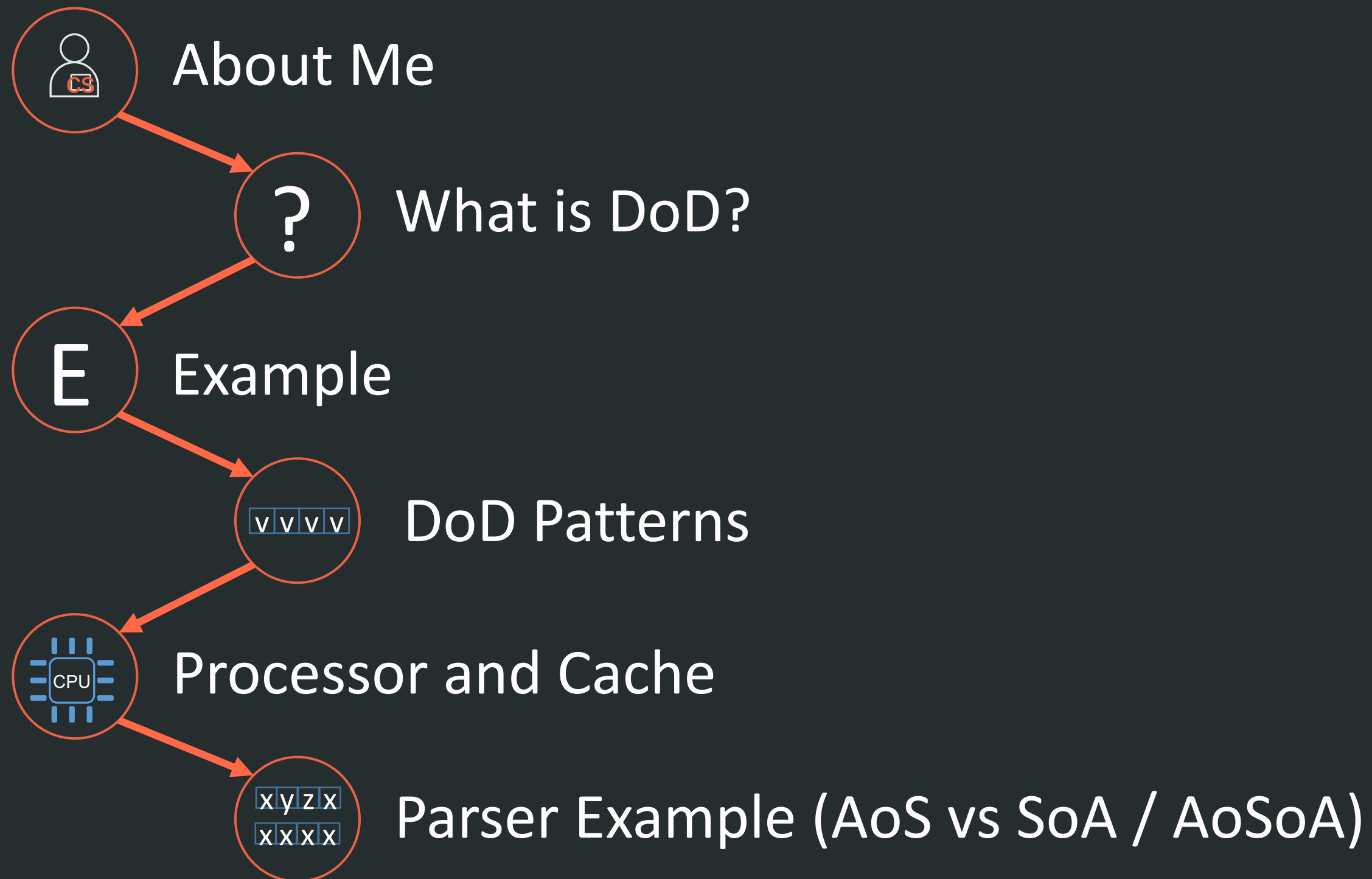


leveluppp.ghost.io



badamczewski01

Agenda



About Me



R&I Director in SDI Media

- Managing Research Teams
- Performance Freak



Building Mass Scale Index for Files

- Mass scale file parsing
- Mass scale statistics & data collection



Search the Index 🔍

Select option ▼

[Export](#) [Feedback](#) [Package All](#) [Cart \(66 010\)](#)

Showing 20 of 47 428 605 assets (1745 ms). Total Size: 96 215.45 GB

Add all to cart

- Hello World - Test1**
Technical Medical Demo

HCdn [cdn\vol1\assets\client\title\210-Media\File1.docx](#)

Language:	FPS:	Language Attributes:	Standard:	Folder Location:	Client:	Size:
Polish	N\A	N\A	Demo	N\A	Fake-Client-X	23.71 kb

[File Contents](#)
- Hello World - Test2**
Demo

HCdn [cdn\vol1\assets\client\title\210-Media\File2.pdf](#)

Language:	FPS:	Language Attributes:	Standard:	Folder Location:	Client:	Size:
English	N\A	N\A	Demo	N\A	Fake-Client-Y	113.71 kb

[File Contents](#)
- Hello World - Test3**
Demo

HCdn [cdn\vol1\assets\client\title\211-Media\File3.mov](#)

Language:	FPS:	Language Attributes:	Standard:	Folder Location:	Client:	Size:
Polish	N\A	N\A	QA-Test	N\A	Fake-Client-X	433.11 mb

[File Contents](#)



What is Data Oriented Design?

Object-Oriented Design focuses on modeling the world using Objects.

What is Data Oriented Design?

Object-Oriented Design focuses on modeling the world using Objects.

Data-Oriented Design focuses on the data and the algorithms that manipulate the data where:

data dependency, **alignment**, and memory access patterns are the most important things.

What is Data Oriented Design?

Object-Oriented Design focuses on modeling the world using Objects.

Data-Oriented Design focuses on the data and the algorithms that manipulate the data where:

data dependency, **alignment**, and memory access patterns are the most important things.

This in turn can increase performance when applied correctly.

What is Data Oriented Design?

Three principles:

1. Data is not the problem domain

What is Data Oriented Design?

Three principles:

1. Data is not the problem domain
2. Data is not about the structure and abstraction
3. Data can change, and algorithms will reflect that change

Data-Oriented Design leads to correct information

What is Data Oriented Design?

Three principles:

1. Data is not the problem domain
2. Data is not about the structure and abstraction
3. Data can change, and algorithms will reflect that change

Data-Oriented Design leads to correct information

Example



leveluppp



leveluppp.ghost.io



badamczewski01

Processor and Cache



leveluppp



leveluppp.ghost.io



badamczewski01

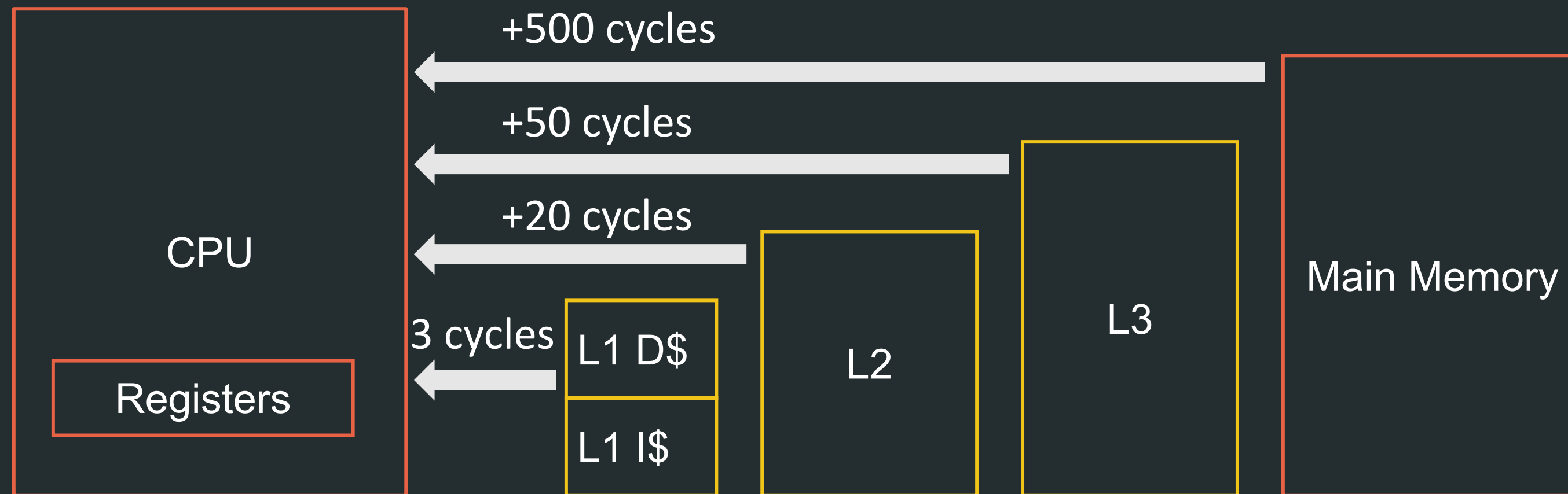
Processor and Cache

Artist Impression

Core 1	Core 2	Core 3	Core 4
Registers	Registers	Registers	Registers
Execution Units	Execution Units	Execution Units	Execution Units
L1 I/D	L1 I/D	L1 I/D	L1 I/D
L2	L2	L2	L2
L3 Shared Cache			

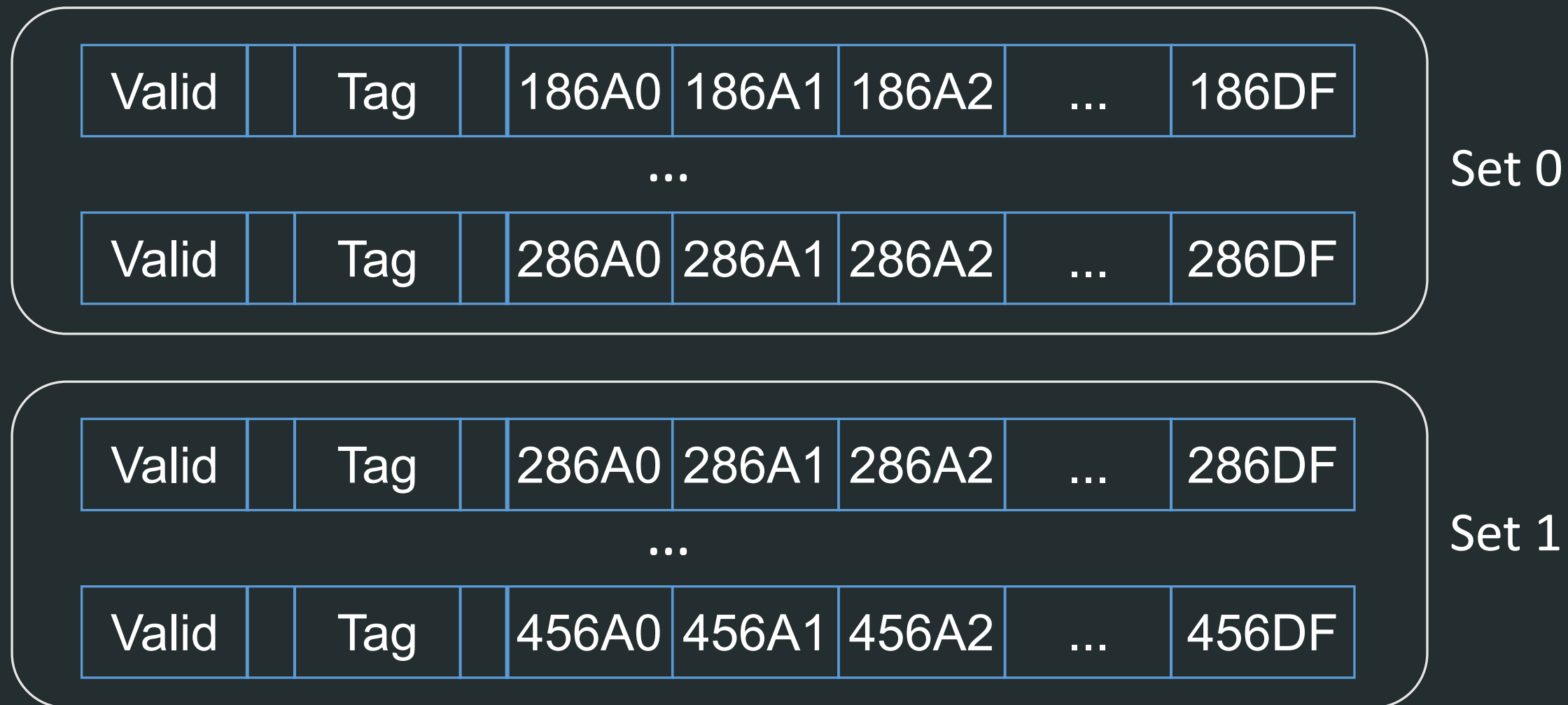
Processor and Cache

Artist Impression



Cache Lines

Artist Impression



$E = 2^e$ lines per Set

$S = 2^s$ Sets

What is Data Oriented Design?

Which program will execute faster?

```
void Rows()  
  for(int i = 0; i < cols; i++)  
    for(int j = 0; j < rows; j++)  
      array[ (i * cols) + j ]++;
```

```
void Cols()  
  for(int i = 0; i < rows; i++)  
    for(int j = 0; j < cols; j++)  
      array[ (j * cols) + i ]++;
```


Data Access Patterns

Rows()

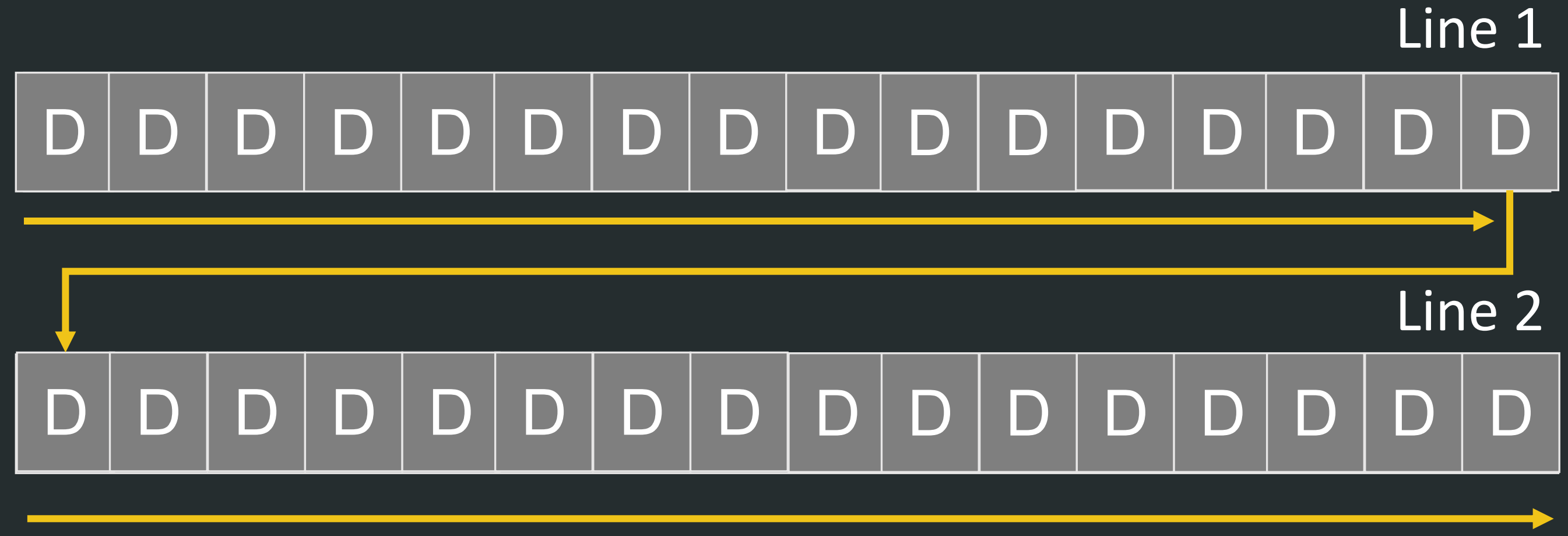


Cols()



Data Access Patterns

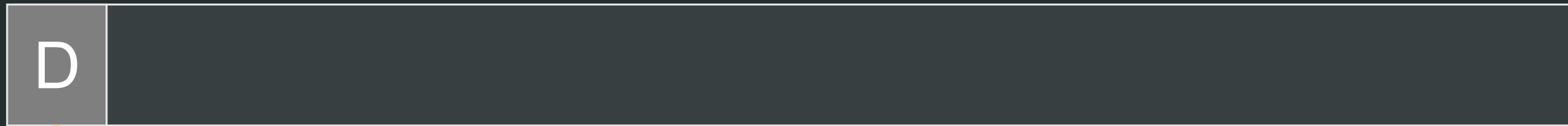
Rows()



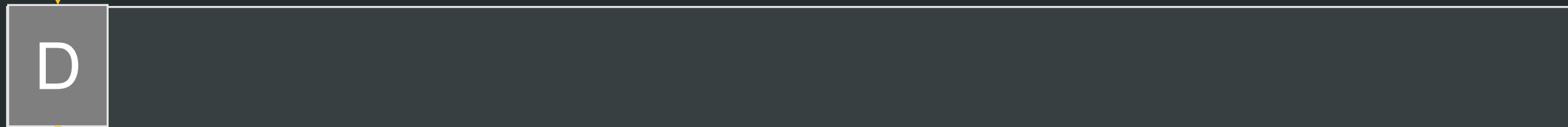
Data Access Patterns

Cols()

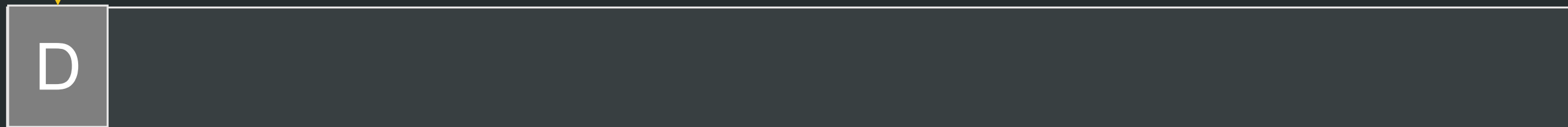
Line 1



Line 2



Line 3



Data Access Patterns

```

BenchmarkDotNet=v0.12.1, OS=Windows 10.0.18363.1198 (1909/November2018Update/19H2)
Intel Core i7-6700HQ CPU 2.60GHz (Skylake), 1 CPU, 8 logical and 4 physical cores
.NET Core SDK=5.0.100
  [Host] : .NET Core 3.1.9 (CoreCLR 4.700.20.47201, CoreFX 4.700.20.47203), X64 RyuJIT
  Job-NHPISD : .NET Core 3.1.9 (CoreCLR 4.700.20.47201, CoreFX 4.700.20.47203), X64 RyuJIT

InvocationCount=1  UnrollFactor=1

| Method | Mean | Error | StdDev |
|-----|-----:|-----:|-----:|
| Rows   | 305.3 ms | 6.05 ms | 8.49 ms |
| Cols   | 1,128.7 ms | 9.42 ms | 8.35 ms |

// * Hints *
Outliers
  RowsVsCols.Rows: InvocationCount=1, UnrollFactor=1 -> 1 outlier was removed (329.05 ms)
  RowsVsCols.Cols: InvocationCount=1, UnrollFactor=1 -> 1 outlier was removed (1.17 s)

// * Legends *
Mean : Arithmetic mean of all measurements
Error : Half of 99.9% confidence interval
StdDev : Standard deviation of all measurements
1 ms : 1 Millisecond (0.001 sec)

```

The more the data
The worse the performance

DoD Patterns



leveluppp



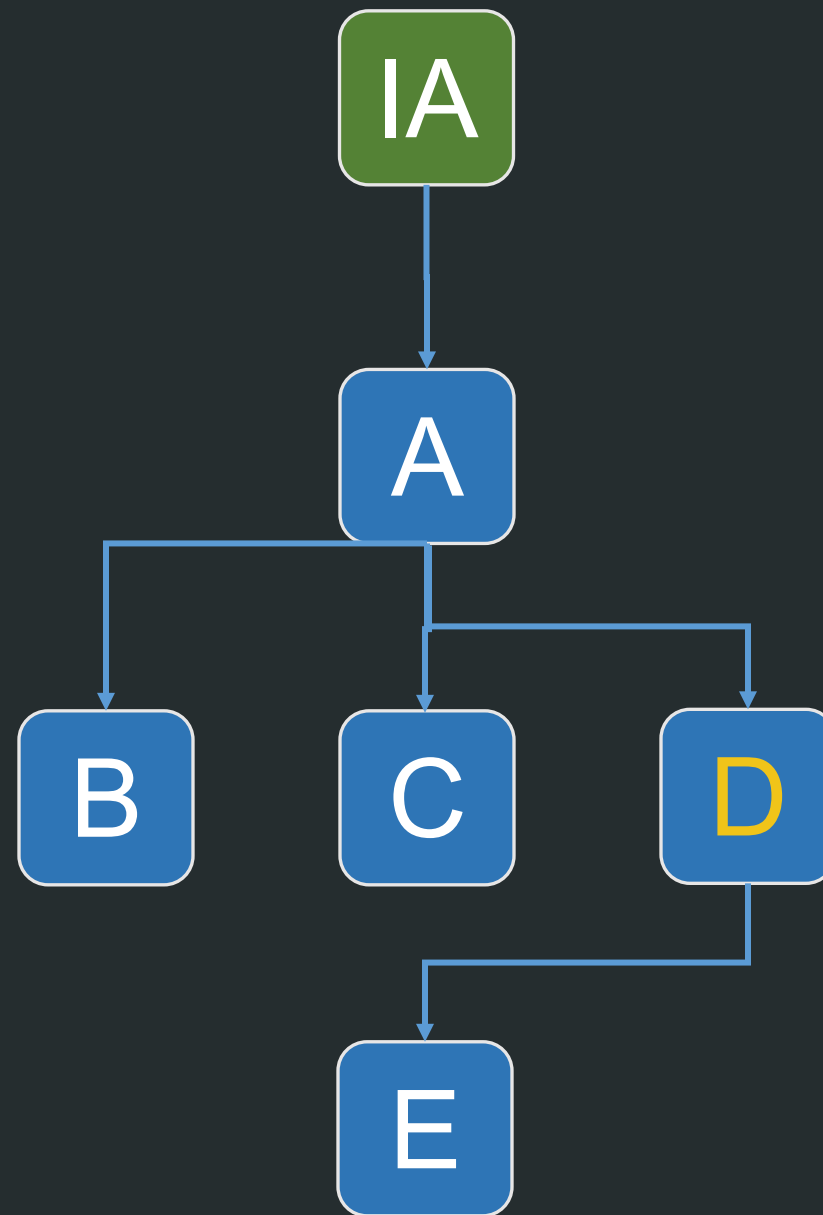
leveluppp.ghost.io



badamczewski01

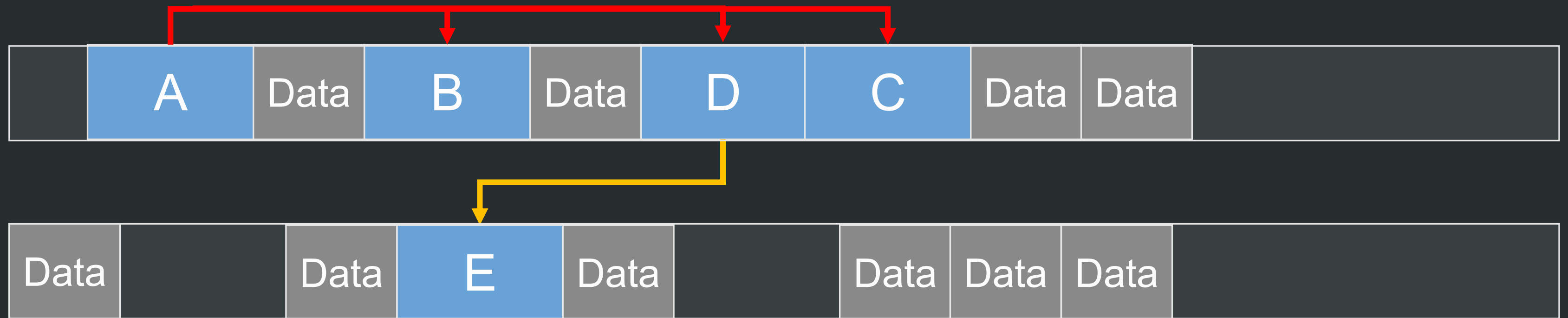
Object-Oriented vs Data-Oriented

Object-Oriented graph:



Memory and Data Layout

Object Graph in Memory:



Object-Oriented vs Data-Oriented

```
public class Transaction
  int Id { get; set; }
  string Title { get; set; }
  TransactionInfo Info { get; set; }
  DateTime Date { get; set; }
  bool IsDPD { get; set; }

  bool Send(string to);
  bool IsValid();
```

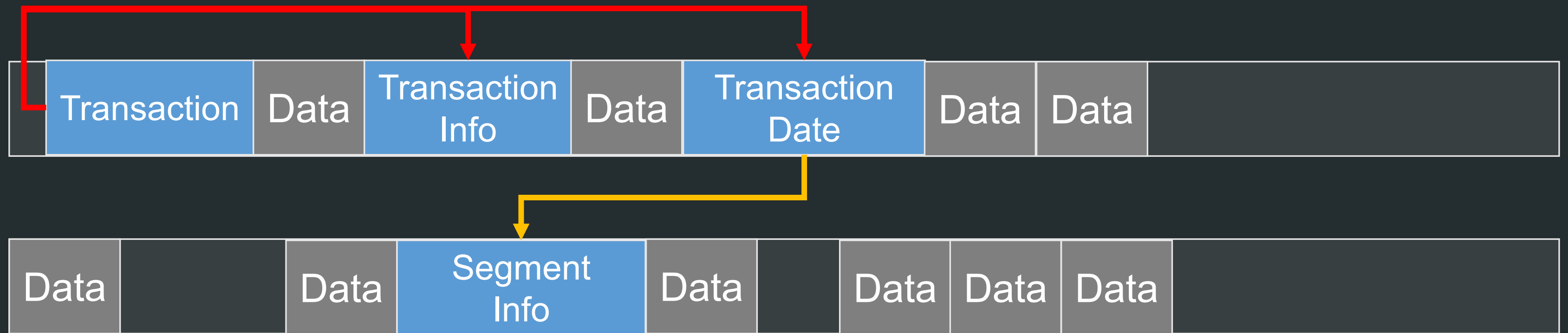
```
public class TransactionInfo
  double Amount { get; set; }
  string From { get; set; }
  string To { get; set; }
  SegmentInfo Segment { get; set; }

public class DateTime
  string Date { get; set; }
  string Time { get; set; }
```

Object-Oriented graph Example:

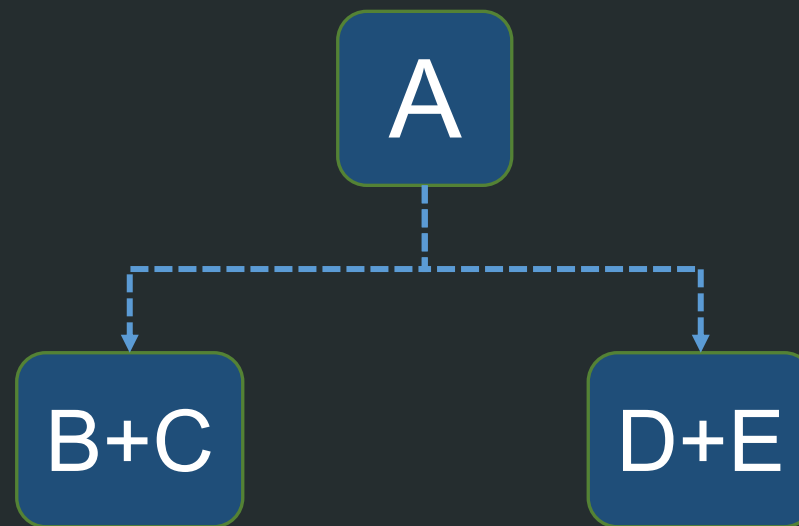
Memory and Data Layout

Object Graph in Memory:



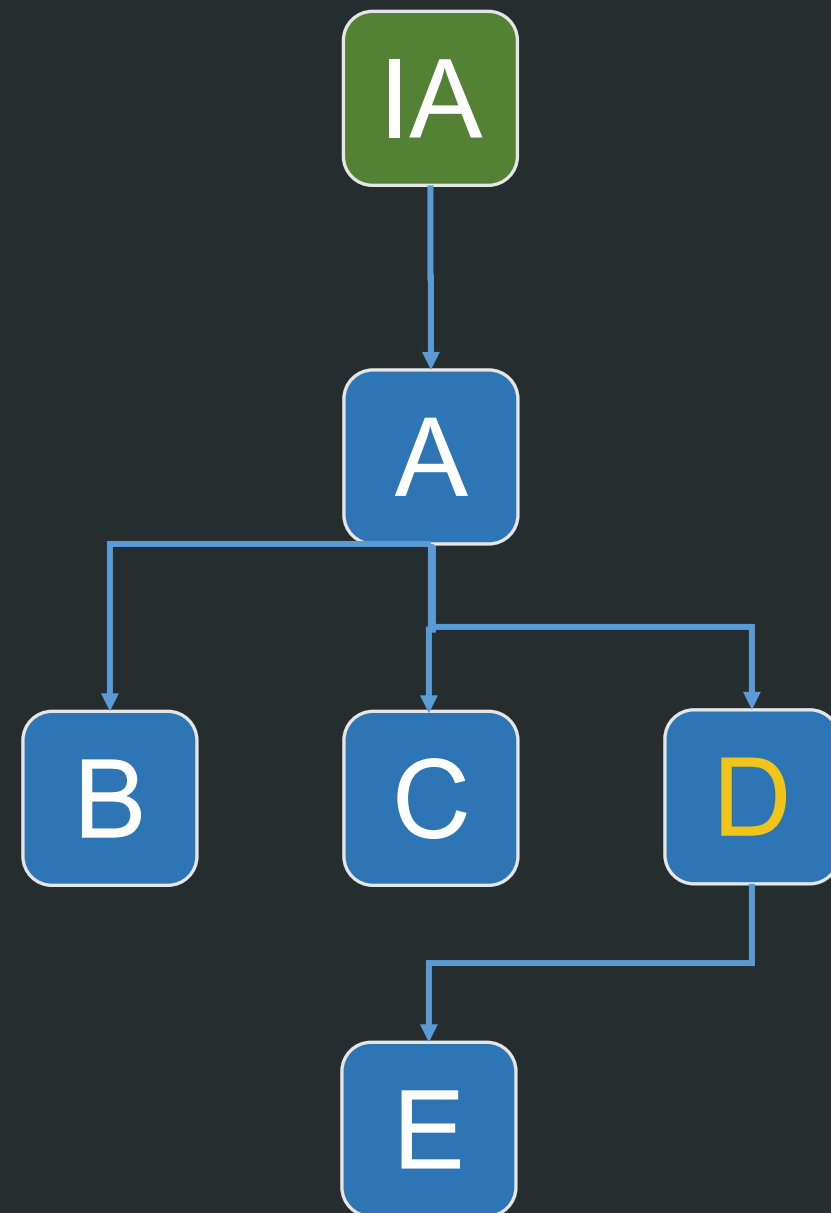
Object-Oriented vs Data-Oriented

Data-Oriented graph:

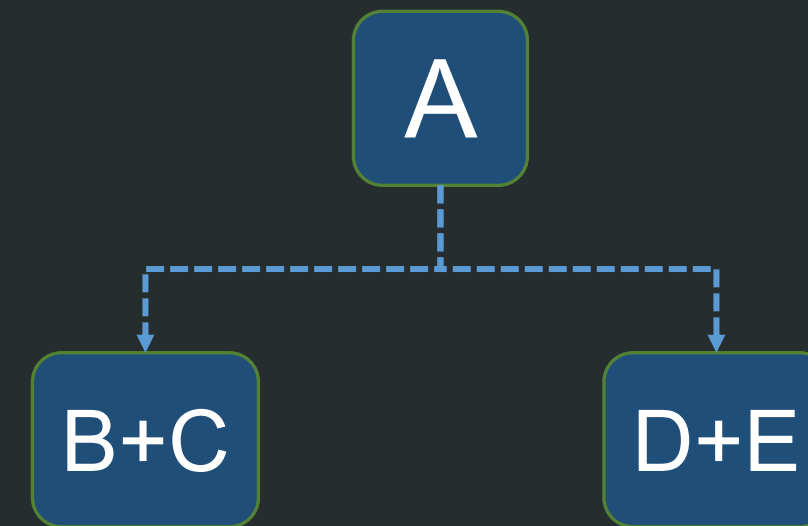


Object-Oriented vs Data-Oriented

Object-Oriented graph:

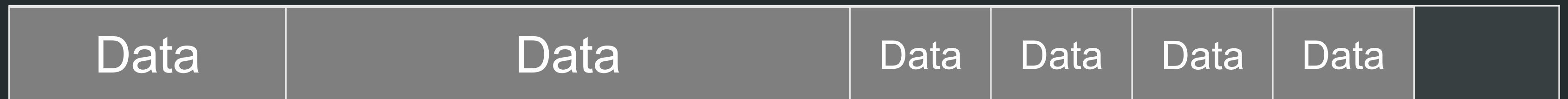
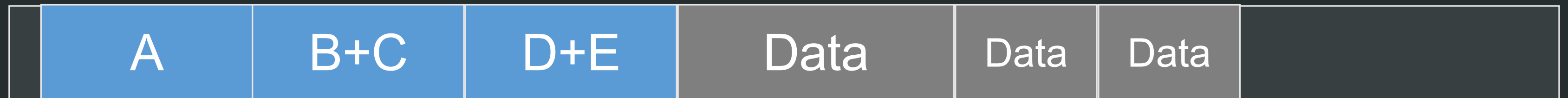


Data-Oriented graph:



Memory and Data Layout

Data Oriented Graph in Memory (Flat / Hot / Cold):



Object-Oriented vs Data-Oriented

Data-Oriented graph Example (Flat Data):

```
public struct Transaction
  int Id;
  double Amount;
  long From;
  long To;
  string Title;
  bool isDPD;
  ....
```

Object-Oriented vs Data-Oriented

Data graph Example (Hot/Cold Split):

```
public struct Transaction
```

```
    int Id;
```

```
    double Amount;
```

```
    long To;
```

```
    bool IsDPD;
```

```
    TransactionInfo Info;
```

```
public struct/class TransactionInfo
```

```
    string Title;
```

```
    double Amount;
```

```
    string From;
```

```
    string Date;
```

```
    string Time;
```

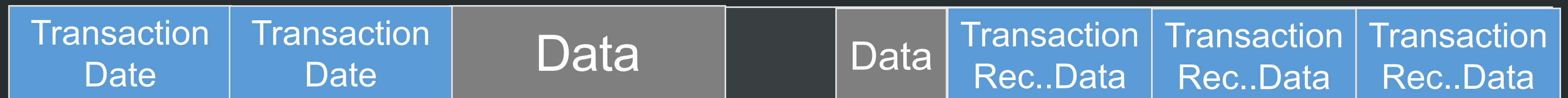
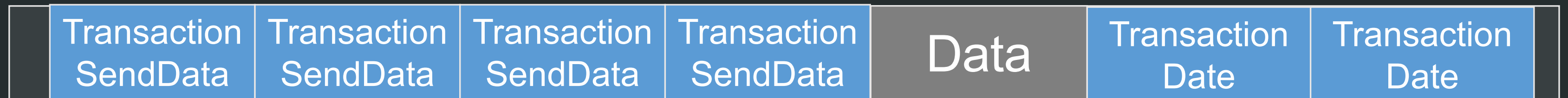
Object-Oriented vs Data-Oriented

Data-Oriented graph Example (Struct of Arrays):

```
public struct Transactions
  public TransactionSendData[] SendData;
  public TransactionReceiveData[] ReceiveData;
  public TransactionDate[] Dates;
  ....
```

Memory and Data Layout

Data Oriented Graph in Memory (Struct of Arrays):



About Booleans

Data Oriented Graph in Memory (Struct of Arrays):

```
public struct Transaction
```

```
  int Id;
```

```
  double Amount;
```

```
  long From;
```

```
  long To;
```

```
  string Title;
```

```
  bool isDPD;
```

```
  ....
```

Packed 1b

NoN-Packed 4b

NoN-Packed 8b

Real Example



leveluppp



leveluppp.ghost.io



badamczewski01



Search the Index



Select option



Export Feedback Package All Cart (66 010)

Showing 20 of 47 428 605 assets (1745 ms). Total Size: 96 215.45 GB

Add all to cart

Hello World - Test1

Technical Medical Demo



HCdn [cdn\vol1\assets\client\title\210-Media\File1.docx](#)

Language:	FPS:	Language Attributes:	Standard:	Folder Location:	Client:	Size:
Polish	N\A	N\A	Demo	N\A	Fake-Client-X	23.71 kb

[File Contents](#)

Hello World - Test2

Demo



HCdn [cdn\vol1\assets\client\title\210-Media\File2.pdf](#)

Language:	FPS:	Language Attributes:	Standard:	Folder Location:	Client:	Size:
English	N\A	N\A	Demo	N\A	Fake-Client-Y	113.71 kb

[File Contents](#)

Hello World - Test3

Demo



HCdn [cdn\vol1\assets\client\title\211-Media\File3.mov](#)

Language:	FPS:	Language Attributes:	Standard:	Folder Location:	Client:	Size:
Polish	N\A	N\A	QA-Test	N\A	Fake-Client-X	433.11 mb

[File Contents](#)

Real Example

```
DoDSamples - Sample.po
Sample.po x
1 1
2 True
3 B
4 2020-11-22T23:31:01
5 431
6 purus nec ornare Maecenas et tempus dolor Aliquam non convallis quam Orci varius natoque
7 tellus Pellentesque lobortis fringilla orci vitae fermentum Nulla a mauris eu
8
9 2
10 False
11 C
12 2020-11-23T00:30:26
13 441
14 Quisque gravida dolor
15
16 3
17 True
18 C
19 2020-11-23T01:52:03
20 119
21 curae; Pellentesque bibendum
22 arcu vel massa bibendum porta Ut vehicula sed purus nec ornare Maecenas et tempus
23
```

Real Example: Sum Amount

Method	Categories	Samples	Mean	Error	StdDev	Ratio	RatioSD
OOB_Parser_StateMachine	OOB	10000	34.258 ms	0.9289 ms	2.7242 ms	1.08	0.07
OOB_Parser_Regex	OOB	10000	33.533 ms	0.6644 ms	1.3571 ms	1.00	0.00
DoD_Parser_ArrayOfStructs	DoD	10000	8.965 ms	0.1013 ms	0.0898 ms	1.00	0.00
DoD_Parser_ArrayOfStructs_Split	DoD	10000	8.889 ms	0.1537 ms	0.1362 ms	0.99	0.02
DoD_Parser_ArrayOfStructs_HotColdSplit	DoD	10000	8.863 ms	0.0841 ms	0.0745 ms	0.99	0.01
DoD_Parser_StructOfArrays	DoD	10000	8.772 ms	0.1152 ms	0.1078 ms	0.98	0.02
DoD_Parser_StructOfArrays_SIMD	DoD	10000	4.260 ms	0.0568 ms	0.0531 ms	0.48	0.01

Real Example: Parse + Sum Amount

```
// * Summary *
```

```
BenchmarkDotNet=v0.12.1, OS=Windows 10.0.18363.1198 (1909/November2018Update/19H2)
```

```
Intel Core i7-6700HQ CPU 2.60GHz (Skylake), 1 CPU, 8 logical and 4 physical cores
```

```
.NET Core SDK=5.0.100
```

```
[Host] : .NET Core 3.1.9 (CoreCLR 4.700.20.47201, CoreFX 4.700.20.47203), X64 RyuJIT
```

```
DefaultJob : .NET Core 3.1.9 (CoreCLR 4.700.20.47201, CoreFX 4.700.20.47203), X64 RyuJIT
```

Method	Categories	Samples	Mean	Error	StdDev	Ratio	RatioSD
OOP_Parser_Regex	OOP	10000	41.259 ms	0.8142 ms	1.9350 ms	1.00	0.00
OOP_Parser_StateMachine	OOP	10000	33.625 ms	0.6709 ms	1.8702 ms	0.81	0.06
DoD_Parser_ArrayOfStructs_Split	DoD	10000	29.913 ms	0.5941 ms	0.9423 ms	1.06	0.04
DoD_Parser_ArrayOfStructs	DoD	10000	28.177 ms	0.4419 ms	0.5900 ms	1.00	0.00
DoD_Parser_ArrayOfStructs_HotColdSplit	DoD	10000	15.129 ms	0.2872 ms	0.2686 ms	0.54	0.01
DoD_Parser_StructOfArrays	DoD	10000	9.302 ms	0.1791 ms	0.1759 ms	0.33	0.01
DoD_Parser_StructOfArrays_SIMD	DoD	10000	4.538 ms	0.0442 ms	0.0369 ms	0.16	0.00

Thank You



leveluppp



leveluppp.ghost.io



badamczewski01