CS2 Lab Exercise: Merge Sort Performance Sampling

Objective

Implement the Merge Sort algorithm to arrange sequences of numbers in non-decreasing order. Instead of measuring execution time, count the **number of statements executed** while sorting. Use Excel to analyze the growth pattern and fit the best curve.

Part 1. Merge Sort with Statement Counting

Pseudocode with Statement Counting

```
global count = 0
```

```
mergeSort(A, left, right)
    count += 1                          // function call
    if left < right
        count += 1                      // if comparison
        mid = (left + right) / 2
        count += 1                      // assignment

        mergeSort(A, left, mid)
        mergeSort(A, mid + 1, right)

        merge(A, left, mid, right)
```

```
merge(A, left, mid, right)
    n1 = mid - left + 1
    n2 = right - mid
    count += 2                          // assignments

    create arrays L[0...n1-1] and R[0...n2-1]
    count += n1 + n2                    // assume each element copy counts as 1

    for i = 0 to n1 - 1
        L[i] = A[left + i]
        count += 1                      // assignment

    for j = 0 to n2 - 1
        R[j] = A[mid + 1 + j]
        count += 1                      // assignment

    i = j = 0
    k = left
    count += 3                          // assignments

    while i < n1 and j < n2
        count += 1                      // while condition
        if L[i] <= R[j]
            A[k] = L[i]
            i += 1
            count += 2                  // assignment and increment
        else
            A[k] = R[j]
            j += 1
            count += 2                  // assignment and increment
        k += 1
        count += 1                      // increment

    while i < n1
        A[k] = L[i]
        i += 1
        k += 1
        count += 3

    while j < n2
        A[k] = R[j]
        j += 1
        k += 1
        count += 3
```

Instructions

1. **Implement** the pseudocode in C++ and declare a global variable `long long count = 0` to track operations.
2. After sorting, print the total value of count.
3. Ensure the algorithm works correctly by testing on a small array and displaying its sorted version.

Data Sample

Start testing with this array:

{5, 2, 8, 6, 1, 3, 9, 7, 4}

Part 2. Benchmarking via Statement Count

Objective

Run Merge Sort on arrays of increasing sizes and **record the number of statements executed**.

Sample Sizes

Sample size	Number of instructions
100	
200	
400	
800	
1600	
3200	
6400	
12800	
25600	

Use **rand()** to generate random integers between 1 and 10000, as in the original lab.
e.g. `int aValue = rand() % 10000;`

Part 3. Data Visualization and Trendline Analysis (Excel)

Instructions

1. Open Excel.
2. Create a scatter plot with:
    - **X-axis**: Sample size
    - **Y-axis**: Statement count
3. Add trendlines:
    - Try **linear**, **logarithmic**, **polynomial (order 2 or 3)**, and **power** fits.
4. Enable and display:
    - **Trendline Equation**
    - **R^2 Value**
5. Compare R^2 values and decide which function best models the algorithm's growth behavior.

Expected Outcome

Merge Sort has **O(n log n)** complexity. The trendline with best R^2 should reflect this (likely a power or polynomial trend).

Definition:

R^2 represents the proportion of the variance in the dependent variable (e.g., statement count) that is predictable from the independent variable (e.g., sample size).

Key Points:

- R^2 ranges from **0 to 1**.
- **$R^2 = 1$** means the trendline **perfectly fits** the data.
- **$R^2 = 0$** means the trendline **does not explain** any of the data's variability.

Interpretation:

- An **R^2 of 0.95** means **95% of the variation** in the output is explained by the input via the model.
- Used in Excel (and statistics) to judge which trendline most accurately models the data.

Using Excel – Notes from Copilot

1. Create a Scatter Plot

1. Enter your data in two columns (X-values in one, Y-values in another).
2. Select the data.
3. Click **Insert** > **Scatter Chart** > **Scatter with Markers**.

2. Add a Trendline

1. Click on the scatter plot.
2. Right-click on any data point and select **Add Trendline**.
3. In the **Format Trendline** pane, choose the type of curve:

- **Linear** (straight-line fit)
- **Polynomial** (curved fit, adjust order for complexity)
- **Exponential** (rapid growth/decay)
- **Power** (power-law relationship)

3. Display the Equation and R-Squared Value

1. In the **Format Trendline** pane, check:

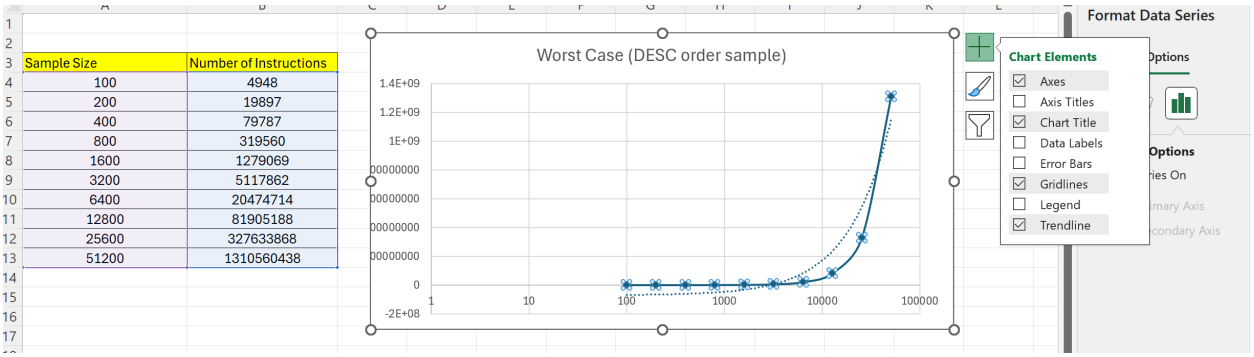
- **Display Equation on Chart** (shows the mathematical formula).
- **Display R-squared Value on Chart** (indicates how well the curve fits).

4. Adjust the Trendline for Better Fit

1. If using a **Polynomial** trendline, increase the order (e.g., 2, 3, or 4) for a better fit.
2. If using an **Exponential** or **Power** trendline, ensure your data follows the expected pattern.

5. Verify the Fit

1. Copy the equation from the chart.
2. Use it in Excel to calculate predicted values and compare them to actual data.


Sample Size
Number of Instructions
100
4948
200
19897
400
79787
800
319560
1600
1279069
3200
5117862
6400
20474714
12800
81905188
25600
327633868
51200
1310560438
Worst Case (DESC order sample)
1.4E+09
1.2E+09
1E+09
00000000
00000000
00000000
00000000
0
-2E+08
1
10
100
1000
10000
100000
Chart Elements
Axes
Axis Titles
Chart Title
Data Labels
Error Bars
Gridlines
Legend
Trendline
Format Data Series
Options
Options
ies On
mary Axis
condary Axis

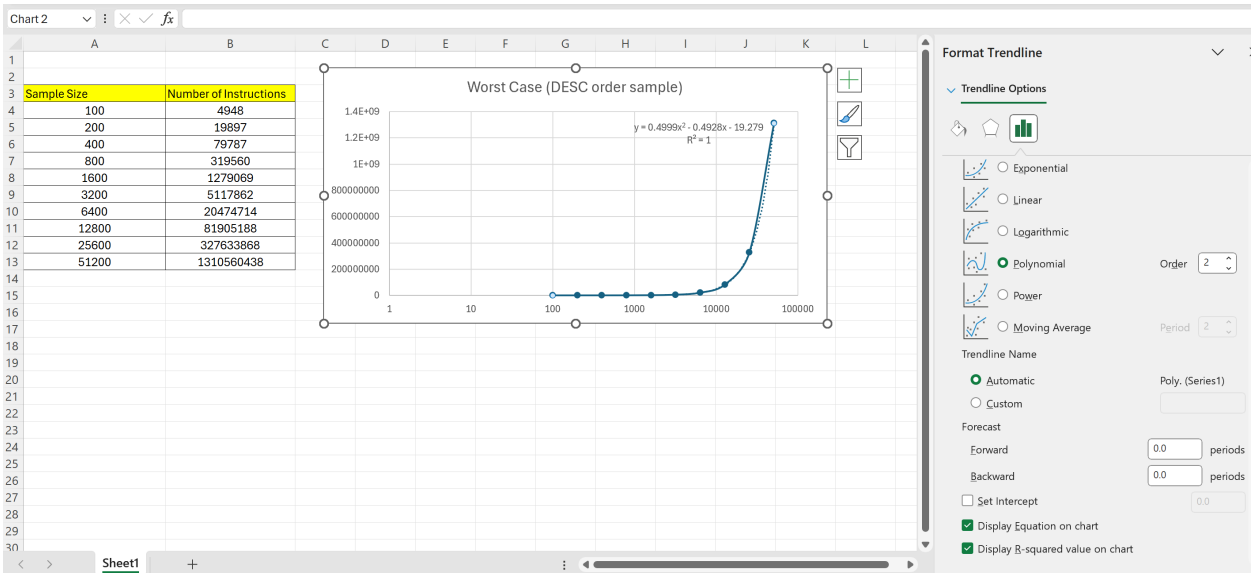


Chart 2
Sample Size
Number of Instructions
100
4948
200
19897
400
79787
800
319560
1600
1279069
3200
5117862
6400
20474714
12800
81905188
25600
327633868
51200
1310560438
Worst Case (DESC order sample)
y = 0.4999x² - 0.4928x - 19.279
R² = 1
1.4E+09
1.2E+09
1E+09
800000000
600000000
400000000
200000000
0
1
10
100
1000
10000
100000
Sheet1
Format Trendline
Trendline Options
Exponential
Linear
Logarithmic
Polynomial
Order
2
Power
Moving Average
Period
2
Trendline Name
Automatic
Poly. (Series1)
Custom
Forecast
Forward
0.0
periods
Backward
0.0
periods
Set Intercept
0.0
Display Equation on chart
Display R-squared value on chart