



COMP 6521

JOIN

Project Report

Submitted To

Dr. N. Shiri

Student Name	Student Id
Akhila Chilukuri	40071241
Alekya Karicherla	40059347
Ezhilmani Aakaash	40071067
Alla Sai Krishna	40069746

CALCULATION:

The total number of tuples in the output: 5549888

This is the collection of T1 and T2.

SORT-BASED

File Size (MB) : 140MB for Sorted Join, 1.12MB For Sort Based GPA Calculation

Coming over to the DISK I/O Part,

We have defined the block size to be 10000 manually for easier computation. Hence, we produce 550 sublists while computing the sort/merge of Phase-1 as per the first Mini-Project.

Number of Input Blocks for GPA using Sort: 27511

Number of Output Blocks for GPA using Sort: 1250

While computing the sort based sort-join, we change the block size as per the given question to be of 40 tuples/each. Therefore,

$5549888/40=138747$ Blocks needed.

Total Disk I/O for Phase 1 and Phase 2 – reading: 196466

Total Disk I/O for Phase 1 and Phase 2 – writing: 170093

Total Disk I/O for Phase 1 and Phase 2: 366559

Total Time to sort the T1 and T2 is 11389ms

Time Taken for Sort-based Join: 16272ms

Total = 27661ms

NESTED JOIN

File Size (MB) : 155MB for Sorted Join, 1.18MB For Nested-Join GPA Calculation

For the DISK I/O,

$5549888/40=138747$ Blocks needed.

I/O for nested reading is: 687682851

The reason for this large number is the fact that we are reading an enormous number of tuples from the T2 which runs as the inner loop. We have done this because this is the theoretical definition of nested loop join. For better performance, we have used a sorted file to reduce the number of reads.

In this way, we achieve a One and a Half Pass where-in, we assume that we end up searching half the records of T2 before computation.

Number of Disk I/O for Writing: 27511

Total Disk I/O = 687711612

Number of Input Blocks for GPA using Sort: 687684109

Number of Output Blocks for GPA using Sort: 1250

Time Taken in total = 75mins

The GPA of every student has been calculated and stored in a separate file for both the join techniques and the I/O's have also been included in the calculation.

SCREENSHOTS:

The screenshot shows the Eclipse IDE with a Java file named `gpatest.java`. The code implements a sorting algorithm with multiple phases. The console output provides detailed performance metrics for the Sort Based Sort-Join operation.

```
148         studentidold = Integer.parseInt(studentid);
149     }
150 }
151
152 else
153 {
154     //System.out.println("The student "+studentidold+"average is :"+studentaverage);
155     //System.out.println(studentcreditsum);
156     //System.out.println(studentgradesum);
157     studentaverage = studentgradesum/studentcreditsum;
158     out1.write(studentidold+" "+String.format("%.2f", studentaverage));
159     out1.newLine();
160     outcounter++;
161     if(outcounter==400)
162     {
```

Console Output:

```
<terminated> PhaseOne (2) [Java Application] C:\Program Files\Java\jre1.8.0_45\bin\javaw.exe (Apr 10, 2018, 11:26:27 AM)
Time take to sort relation T1 is 413ms (~approx 0.413sec)
Time take to sort relation T2 is 12876ms (~approx 12.076sec)
Total Time to sort the T1 and T2 is 12489
Total number of records 5549888
Number of sublist 550
Total number of blocks used to store the Tuples is 138747
iteration 0 in phase 2 merging time taken for this iteration is 130ms (~approx 0.13sec)
iteration 1 in phase 2 merging time taken for this iteration is 61ms (~approx 0.061sec)
iteration 2 in phase 2 merging time taken for this iteration is 53ms (~approx 0.053sec)
iteration 0 in phase 2 merging time taken for this iteration is 2962ms (~approx 2.962sec)
iteration 1 in phase 2 merging time taken for this iteration is 2745ms (~approx 2.745sec)
iteration 2 in phase 2 merging time taken for this iteration is 3069ms (~approx 3.069sec)
iteration 3 in phase 2 merging time taken for this iteration is 2452ms (~approx 2.452sec)
iteration 4 in phase 2 merging time taken for this iteration is 2475ms (~approx 2.475sec)
iteration 5 in phase 2 merging time taken for this iteration is 2099ms (~approx 2.099sec)
iteration 6 in phase 2 merging time taken for this iteration is 2345ms (~approx 2.345sec)
iteration 7 in phase 2 merging time taken for this iteration is 1878ms (~approx 1.878sec)
iteration 8 in phase 2 merging time taken for this iteration is 1783ms (~approx 1.783sec)
iteration 9 in phase 2 merging time taken for this iteration is 2427ms (~approx 2.427sec)
done with the sorted join. Time Taken: 15716
Number of Input blocks for Gpa Calculation using Sort Join: 27511
Number of Output blocks for Gpa Calculation using Sort Join: 1250
Time taken is :1467millisecc
join started
```

Sort Based Sort-Join

The screenshot shows a Notepad++ window displaying the output of a Sort Based Join operation. The output is a table with two columns: StudentId and AverageGrade.

StudentId	AverageGrade
40000000	3.26
40000001	3.33
40000002	3.43
40000003	3.39
40000004	3.37
40000005	3.52
40000006	3.27
40000007	3.41
40000008	3.40
40000009	3.40
40000010	3.27
40000011	3.37
40000012	3.26
40000013	3.34
40000014	3.61
40000015	3.38
40000016	3.48
40000017	3.35
40000018	3.24
40000019	3.37
40000020	3.43
40000021	3.42
40000022	3.38
40000023	3.46
40000024	3.49
40000025	3.44
40000026	3.21
40000027	3.24
40000028	3.31
40000029	3.39
40000030	3.34
40000031	3.52
40000032	3.35
40000033	3.24
40000034	3.38
40000035	3.48
40000036	3.37
40000037	3.18
40000038	3.46
40000039	3.33
40000040	3.30

Sort Based Join – GPA Calculation

```

iteration 0 in phase 2 merging time taken for this iteration is 16629ms(~approx 16.629sec)
iteration 1 in phase 2 merging time taken for this iteration is 11957ms(~approx 11.957sec)
iteration 2 in phase 2 merging time taken for this iteration is 6919ms(~approx 6.919sec)
iteration 3 in phase 2 merging time taken for this iteration is 12177ms(~approx 12.177sec)
iteration 4 in phase 2 merging time taken for this iteration is 5695ms(~approx 5.695sec)
iteration 5 in phase 2 merging time taken for this iteration is 3233ms(~approx 3.233sec)
iteration 6 in phase 2 merging time taken for this iteration is 4401ms(~approx 4.401sec)
iteration 7 in phase 2 merging time taken for this iteration is 4081ms(~approx 4.081sec)
iteration 8 in phase 2 merging time taken for this iteration is 4253ms(~approx 4.253sec)
iteration 9 in phase 2 merging time taken for this iteration is 4840ms(~approx 4.84sec)
Time taken to read and write output to disk in Merging is :8.02 sec
Total number of the disk IO in Phase 1 & Phase 2 for reading is 196466
Total number of the disk IO in both Phase 1 & Phase 2 for writing is 170093
Total number of the disk IO in Phase 1 & Phase 2 for reading and writing is 366559

Number of Input blocks for Gpa Calculation using Sort Join: 27511
Number of Output blocks for Gpa Calculation using Sort Join: 1250
Time taken is :1166millisec

```

```

join started
nested time in :75 min
Total number of the disk IO for the nested loop join for reading is 687682851
Total number of the disk IO for the nested loop join for writing is 27511
Total number of the disk IO for writing the GPA for the nested loop join is 1250
Total number of the disk IO for the nested loop join for reading and writing is 687711612
Number of Input blocks for Gpa Calculation using Nested Join: 687684109
Number of Output blocks for Gpa Calculation using Nested Join: 1250

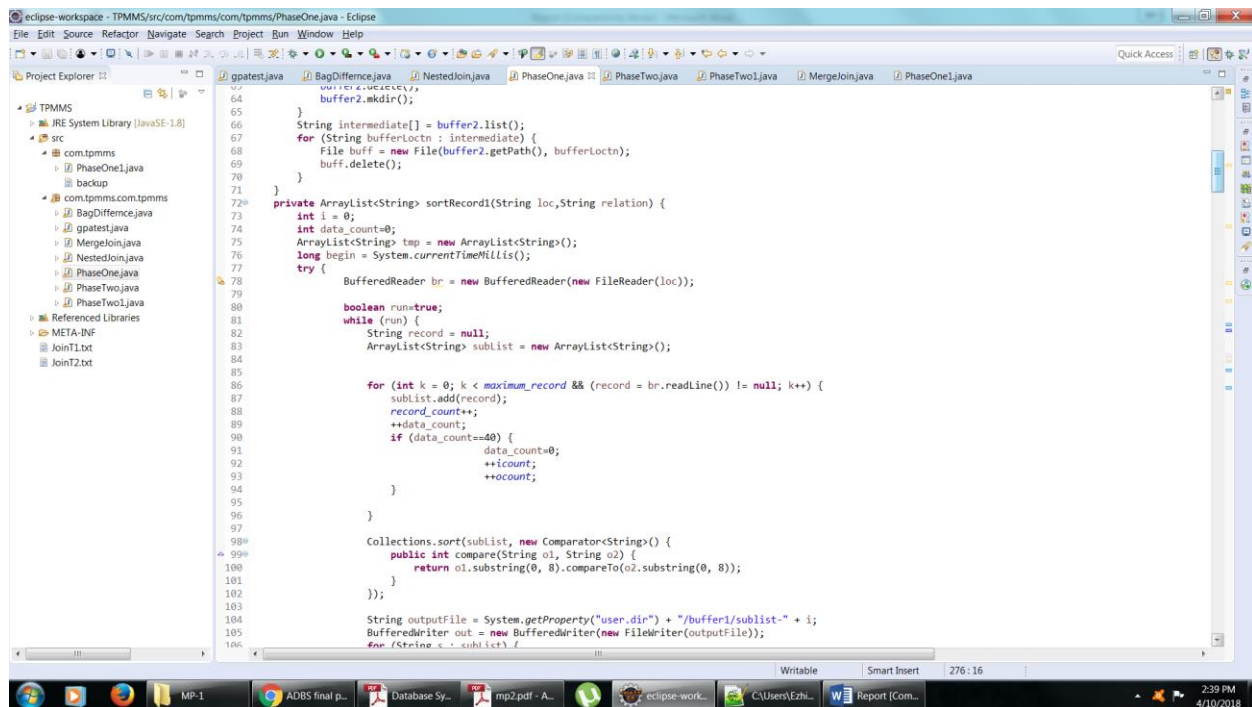
```

Nesteed-Loop Join

Studentid	AverageGrade
40046895	3.41
40023392	3.35
40003350	3.16
40017692	3.23
40041334	3.24
40025952	3.30
40016493	3.50
40022560	3.55
40001707	3.15
40017058	3.15
40026871	3.46
40036374	3.21
40032938	3.22
40038065	3.39
40003933	3.32
40009029	3.46
40004715	3.41
40049162	3.17
40027404	3.16
40014285	3.30
40022920	3.36
40008845	3.22
40013683	3.38
40005290	3.38
40017356	3.15
40014606	3.32
40005906	3.37
40042152	3.33
40045099	3.29
40044831	3.37
40036180	3.22
40000610	3.17
40024595	3.28
40034015	3.36
40026050	3.33
40039896	3.28
40011805	3.10

Nesteed-Loop Join GPA

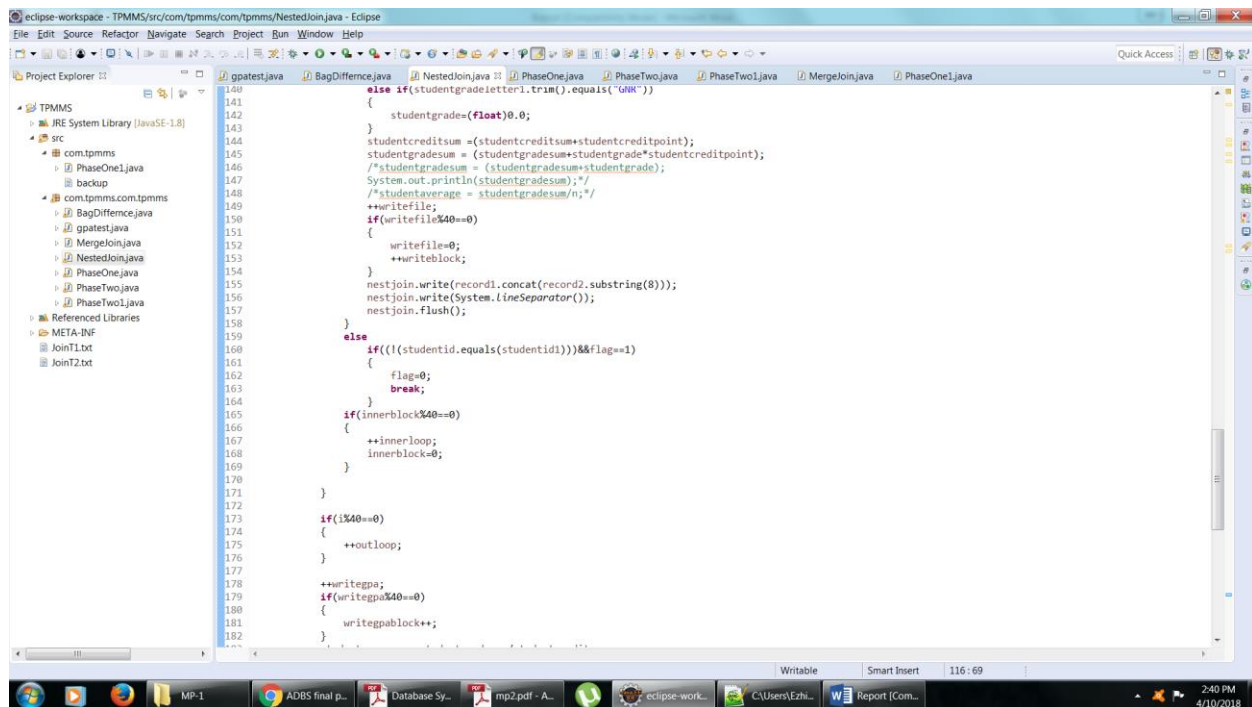
CODING



The screenshot shows the Eclipse IDE with the 'TPMMS' project open. The 'PhaseOne.java' file is selected in the Project Explorer and is open in the editor. The code implements a sort-based join algorithm. It starts by reading records from a file into an ArrayList. Then, it sorts the records using a custom comparator. Finally, it writes the sorted records to an output file. The code is as follows:

```
113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000
```

Sort-Based Join



The screenshot shows the Eclipse IDE with the 'TPMMS' project open. The 'NestedJoin.java' file is selected in the Project Explorer and is open in the editor. The code implements a nested loop join algorithm. It reads records from a file into an ArrayList. Then, it iterates over each record in the first table and compares it with every record in the second table. If a match is found, the joined record is written to the output file. The code is as follows:

```
113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000
```

Nested Loop Join

