Log Visualization

Woojong Koh and Armin Samii

Log Visualization

Woojong Koh and Armin Samii

Overview

- The system administrator uses log files to understand what goes on underneath.

- These log files can be tens or hundreds of megabytes, and impossible to parse without tools.

➡ How to (1) display massive log data visually and (2) spot abnormalities with little effort

Goal

- Aggregate this data into a visualization that can be easily checked by the system administrator.

- With a large enough cluster, this will require focusing the user's attention on **"interesting" parts** of the system state.

- This will require pseudo-detecting anomalies, which in turn requires a constantly-evolving model of what the "normal" state of the system is.

Data Analysis

I1008 23:23:04.868943 10210 training_slave.cc:2295] Starting job autoclust_detect on node 5: [425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545]
I1008 23:23:04.905982 10210 training_slave.cc:2295] Starting job autoclust_detect on node 4: [426 431 436 441 446 451 456 461 466 471 476 481 486 491 496 501 506 511 516 521 526 531 536 541 546]
I1008 23:23:04.941817 10210 training_slave.cc:2295] Starting job autoclust_detect on node 3: [427 432 437 442 447 452 457 462 467 472 477 482 487 492 497 502 507 512 517 522 527 532 537 542 547]
I1008 23:23:04.988983 10210 training_slave.cc:2295] Starting job autoclust_detect on node 2: [428 433 438 443 448 453 458 463 468 473 478 483 488 493 498 503 508 513 518 523 528 533 538 543 548]
I1008 23:23:05.026718 10210 training_slave.cc:2295] Starting job autoclust_detect on node 1: [429 434 439 444 449 454 459 464 469 474 479 484 489 494 499 504 509 514 519 524 529 534 539 544 549]
I1008 23:23:05.071660 10210 training_slave.cc:1988]
Running Command: autoclust_detect
	Number Pending: 425
	Number Completed: 125 of 10000
	Available Processors: 0 of 17
	Running Processors: 17 of 17
I1008 23:23:10.071907 10210 training_slave.cc:1988]
Running Command: autoclust_detect
	Number Pending: 425
	Number Completed: 125 of 10000
	Available Processors: 0 of 17
	Running Processors: 17 of 17
I1008 23:23:10.455950 10210 training_slave.cc:1797] Job completed on node: 6
I1008 23:23:10.456029 10210 training_slave.cc:1812] Node 6 output indices: [5 22 39 56 73 90 107 124 141 158 175 192 209 226 243 260 277 294 311 328 345 362 379 396 413]

…

Extracting Data

I1008 23:23:05.071660 10210 training_slave.cc:1988]
Running Command: autoclust_detect
    Number Pending: 425
    Number Completed: 125 of 10000
    Available Processors: 0 of 17
    Running Processors: 17 of 17

I1009 00:12:41.683965 10210 training_slave.cc:1988]
Running Command: autoclust_detect
    Number Pending: 125
    Number Completed: 9875 of 10000
    Available Processors: 10 of 17
    Running Processors: 7 of 17

Extracting Data

I1008 23:23:05.071660 10210 training_slave.cc:1988]
Running Command: autoclust_detect
    Number Pending: **425**
    Number Completed: **125** of 10000
    Available Processors: **0** of 17
    Running Processors: **17** of 17

I1009 00:12:41.683965 10210 training_slave.cc:1988]
Running Command: autoclust_detect
    Number Pending: **125**
    Number Completed: **9875** of 10000
    Available Processors: **10** of 17
    Running Processors: **7** of 17

Prev	Next
425	125
125	9875
0	10
17	7

Visualization

Visualization

1. Level of detail (zoom in and out)

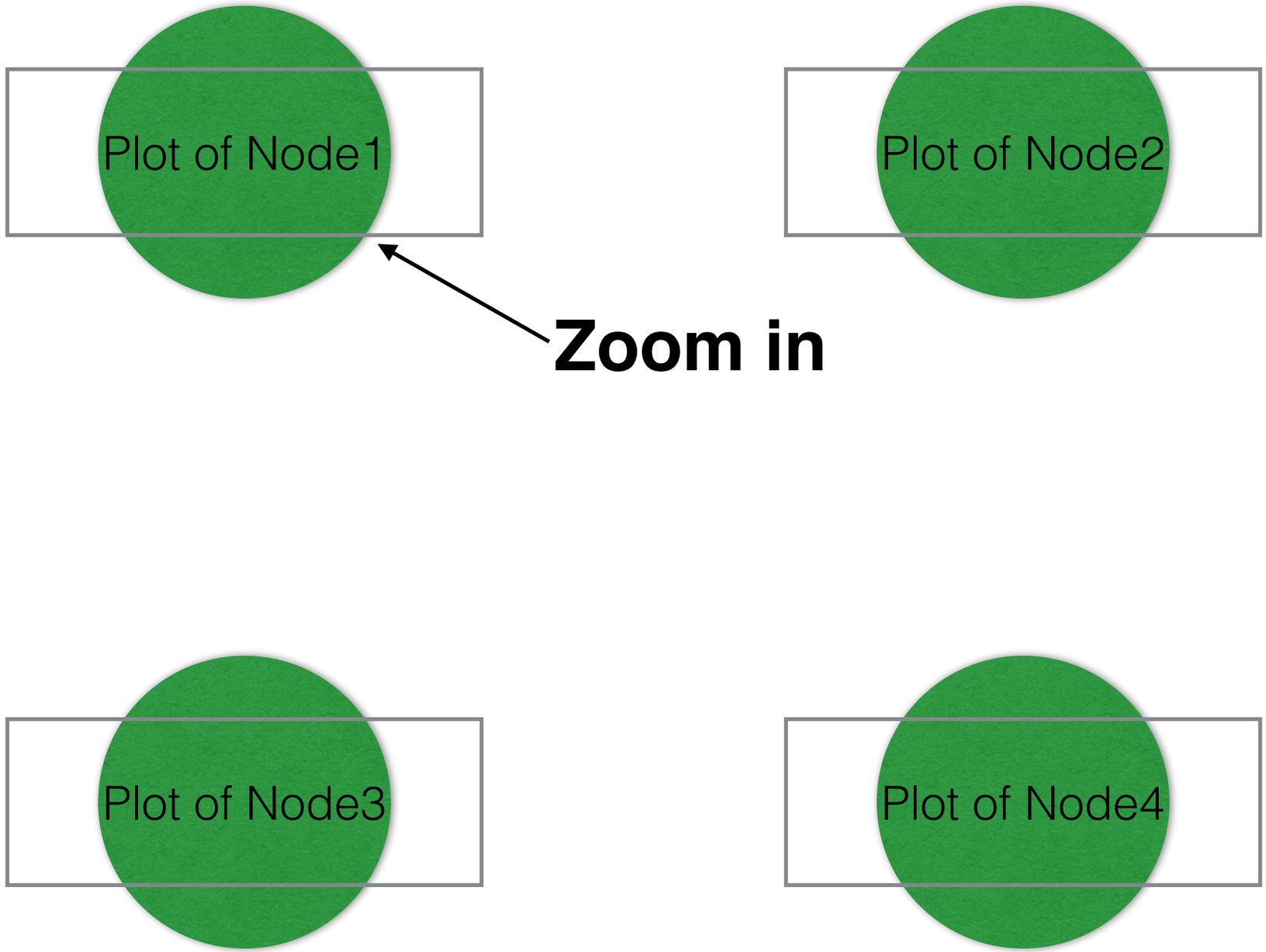
2. Reordered plots to show similar patterns

Group 2

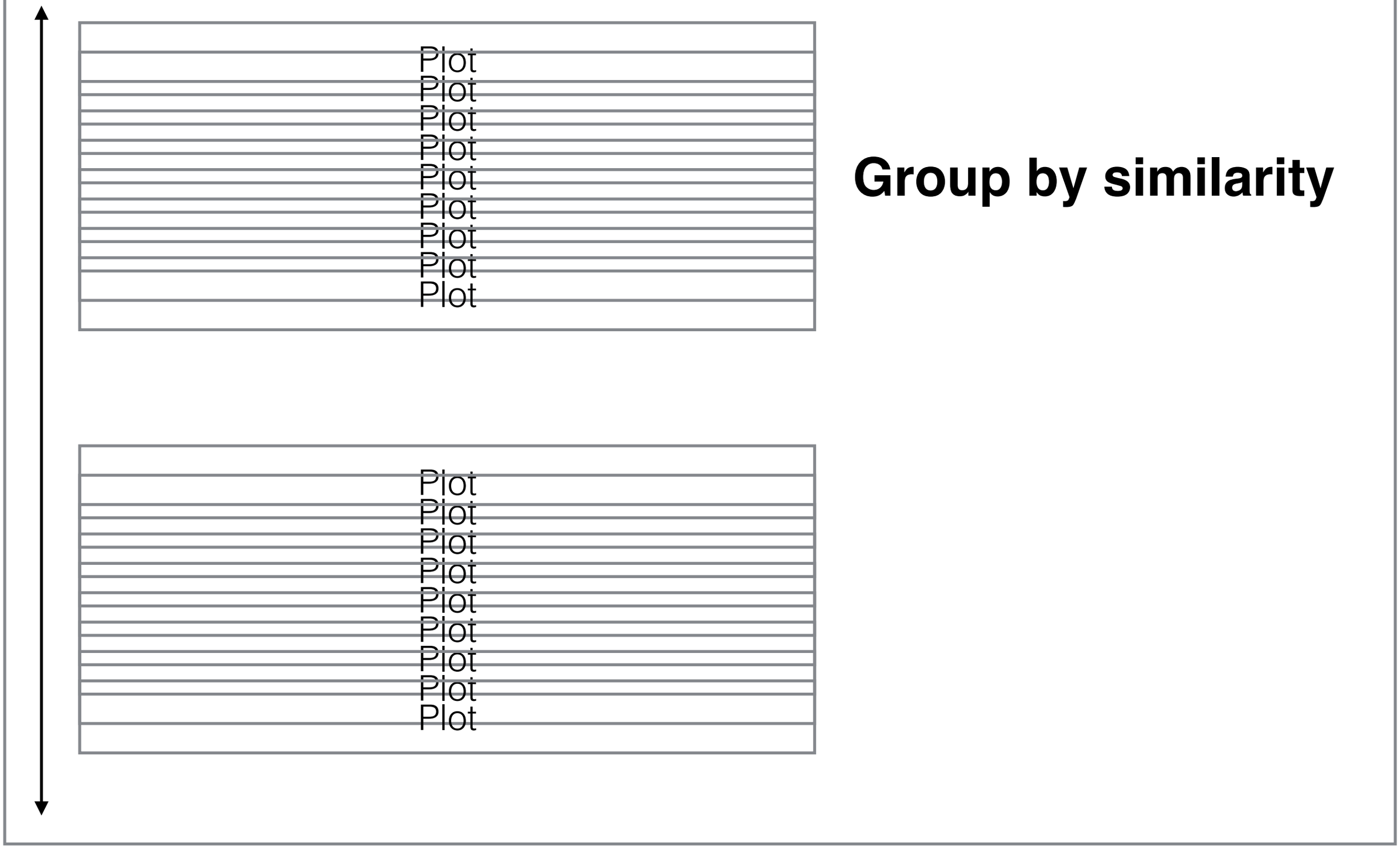
Group 1

Zoom in

Group 3


Plot of Node1
Plot of Node2
Zoom in
Plot of Node3
Plot of Node4

Screen


Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Plot
Group by similarity

Log Visualization for Distributed System

Woojong Koh and Armin Samii

Thank you