Homework 2

for the course Parallel and distributed processing

Aleksandar Ivanovski (Student ID: 186063)

31 October 2020

1 What is a distributed system?

Distributed system is a group or collection of components which are located in different machines, but work together to achieve a common goal, i.e. to execute the user's task, and provide the feeling to the user that the whole system is one homogeneous unit, rather than multiple components.

These components can be geographically separated, but connected together via network and still provide the same feeling to the user.

2 Specify which architectures are for distributed systems and what are their advantages and disadvantages.

The architecture of the distributed systems can be interpreted from different aspect. One aspect is from higher perception and contains the following architectures: client–server, three-tier, n-tier, or peer-to-peer, this differentiation is more logical.

The advantages and disadvantages of the mentioned architectures are the following. [6]

Client-server

Advantages

- Scalability: The capacity of the clients and the servers can be increased separately. New nodes can be added to the network at any time those nodes can be clients or servers.
- Ease of maintenance because each node has standalone role and responsibility and can be repaired, upgraded or replaced without affecting the overall system if done correctly.
- Centralization of control: A dedicated server controls access, resources and the integrity of the data, and unauthorized client cannot damage or misuse the system.

Disadvantages

- Specialized networking operating system is required in order for the server side of the system to work.
- The server side of the system is single point of failure.

- The server side of the system requires special managing staff.
- Congestion can occur, i.e. the server side of the system has reached the maximum capacity, so new clients cannot be served.

Three-tier

Advantages

- Total decoupling: All of the three logical layers, i.e. Presentation Layer, Application-logic layer and Data Management layer are fully decoupled and services are provided to the upper layers via predefined and well documented API-s.
- Horizontal-Scalability can be achieved by utilizing each layer independently.
- Reusability can be achieved because each layer and component in that layer is independent unit which is already tested.
- The fault tolerance can be improved, because each layer will have it's own responsibilities and mechanisms for fault-tolerance which lowers the overall fall rate of the system.

Disadvantages

- Three-tier system is more complex do develop because each of the tree layers must be independently developed, tested and integrated to the whole system.
- A change in the requirements affects all the layers and can increase the overall change adaptation time - this process is also known as cascade change.
- Because of the more layers the amount of resources needed is bigger, and that increases the overall initial cost and maintenance cost.

N-tier

The N-tier architecture is generalization of the three-tier architecture and the advantages and disadvantages are the same according to the number of tiers in the given implementation.

Peer-to-peer

Advantages

- Resistance to failure: The failing of one node won't disrupt the whole system.
- Low maintenance cost: A P2P system doesn't require server(s) so the initial and maintenance costs are practically zero.
- P2P systems don't require network operating system. Each peer can use different operating system which enables more compatibility between peers using different operating systems and file systems.
- Reduction of the technical staff, because special maintenance is not required.

Disadvantages

- Lack of security is present because each peer is independent and the negligence of one peer can be used by malicious third sides to compromise the system.
- Central backup is challenging, because the files are distributed across all the peers.
- In some distributed system the 51% attack is a threat.

The other aspect is on lower level and concerns about the physical implementation of the system, and contains the following architectures: Shared memory architectures (tightly coupled) and Distributed memory architectures (loosely coupled)

The advantages and disadvantages of the mentioned architectures are the following.

Shared memory architectures [4]

Advantages

- Simplicity: Since the memory is shared, no special mechanisms are required for achieving consistency, since directory and control information (e.g., lock tables) are shared by all processors, writing database software is not very different than for centralized systems.
- Load Balancing: Since the memory is treated as central unit, the load balancing can be done more easier.

Disadvantages

- Limited scalability: Since all cores are using the same memory there is limit to the scalability which can be achieved, also the memory bus is bottleneck
- The shared memory is single point of failure of the whole system.

Distributed memory architectures

Advantages

- The price performance ratio is superior compared to shared memory architectures
- The amount of scalability which can be achieved is practically unlimited.

Disadvantages

- Difficult to program: Since the memory is distributed special effort is needed to provide consistency.
- Special effort is needed to process large amount of data, which needs to be divided across the instances of the memory.

3 Describe the workstation-server model. Are there known implementations of this model?

The workstation-server model is distributed system composed of a network of workstations where each workstation provides local processing capability and interface to the whole network.

Some workstations are assigned dedicated roles, i.e. time servers, file servers, web servers and they are usually equipped with a disk and operating system and are named as mini-computers. On the other hand the other workstations are usually disk-less and use the network to send requests to the specialized computers for various tasks and services. Diskless workstations have many advantages, one of them is their plugplay capability. Some workstation-server implementations advocate the usage of idle workstations for providing services to the whole system.

4 Explore the processor-pool model. Try to find real systems where this model is used and list those systems.

The Hadoop's JobTracker System distributes all the map and reduce tasks among a pool of processors. [1]

One early real system was the Cambridge Distributed Computing System[2], which used the term processor bank instead of pool.

Every organisation where it's participants needs a lot of processing power but in small time intervals, for example a researcher needs to run a simulation which could be done in 10 minutes with the processing power available at the pool, and uses the other time to analyze those results should implement processor-pool system.

5 What are the characteristic of Distributed Operating System (DOS) and list implementations.

The characteristic of the Distributed Operating Systems are: [5]

- Resource sharing: All the resources are shared on multiple nodes, i.e. printers, file storage, etc...
- Openess: There is higher degree to which new resource sharing services can be made available to the users
- Concurrency: Many tasks can be achived in the same time.
- Scalability: As demand rises new resources can be added without any special intervention and maintenance required.
- Fault Tolerance: If one node in distributed system malfunction or corrupts then other nodes will replace it easily without affecting the overall performance.
- Transparency is achieved in: access, location, concurrency, replication, failure, migration, performance, scaling

Implementations of distributed operating systems are:

- The DYSEAC
- Lincoln TX-2
- AMOEBA
- Cronus
- DEMOS/MP
- DISTOS
- DISTRIX
- Eden
- Galaxy
- MDX
- MICROS
- MOS
- MOSIX
- Newark
- NSMOS
- Plan9
- REPOS
- RIG
- Roscoe
- Saguaro
- SODA
- SODS/OS
- Spring
- Uniflex
- Windows server 2003
- Windows server 2008
- Windows server 2012
- AEGIS
- Arachne
- Charlotte
- CHOICES
- Clouds
- CMDS
- CONDOR

6 For the implementations in Question 4, state which approach to kernel they use (ex. monolitic kernel, microkernel, hybrid kernel)?

The Cambridge Distributed Computing System uses hybrid kernel.

The Hadoop's JobTracker system can be used on both monolitic and microkernel, so there is a debate. Apache advocates the usage of microkernel, on the other hand Linus Torvalds advocates the usage of monolitic kernel.[3]

7 State implementations of operating systems which use microkernel and what are the their use cases?

Operating systems which implement microkernels are the following:

- V (OS)
- Versatile Real-Time Executive
- Verve (OS)
- VSTa
- Harmony (operating system)
- Harmony OS
- HelenOS
- HeliOS
- Hydra (OS)
- Mac OS 8
- MachTen
- MacMach
- Meiko Scientific
- Micro-Controller Operating Systems
- Midori (OS)
- MINIX
- MkLinux
- MorphOS
- MQX
- Multi-Environment Real-Time

Microkernels are used primarily in servers and in systems with limited resources, they were very popular among the systems in the 80's because of their memory and storage limitations.

8 What are the biggest challenges in creating Distributed Operating System (DOS)?

The challenges are:

- Achieving transparency,
- Achieving security,
- Using global and domain knowledge to achieve process synchronization
- Using global and domain knowledge to achieve process resource management
- Maximizing the fault tolerance and error recovery

References

[1] Multiple Authors. *Hadoop JobTracker*. URL: https://cwiki.apache.org/confluence/display/HADOOP2/JobTracker.

[2] MultipleAuthors. *Cambridge Distributed Computing System*. URL: https://en.wikipedia.org/wiki/Cambridge_Distributed_Computing_System.

[3] Francois Orsinis. *Apache Derby Case Study: Benefits of a Microkernel architecture*. URL: https://francoisorsini.wordpress.com/2008/08/11/apache-derby-case-study-benefits-of-a-microkernel-architecture/.

[4] DM Pressel, M Behr, and S Thompson. *The True Limitations of Shared Memory Programming*. Tech. rep. ARMY RESEARCH LAB ABERDEEN PROVING GROUND MD, 2000.

[5] Junaid Rehman. *What are the advantages and disadvantages of distributed operating systems?* URL: https://www.itrelease.com/2015/09/what-are-advantages-and-disadvantages-of-distributed-operating-systems/.

[6] Majd F. Sakr. *Introduction to Cloud Computing*. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.702.2212&rep=rep1&type=pdf.