Паралелно и дистрибуирано процесирање

Домашна задача 3

Петар Атанасовски - 216052

2.1. Differentiate and exemplify the following terms related to clusters:

a. **Compact versus slack clusters**:

Compact clusters имаат јазли блиску еден до друг, што ја прави комуникацијата побрза со помало доцнење. Од друга страна, slack clusters имаат раширени јазли, што може да резултира со побавна комуникација. Одлуката помеѓу компактни и слаби кластери зависи од потребите на компјутерската средина. Компактните кластери се добри кога физичката блискост е важна, додека слабите кластери се подобри за локации кои се оддалечени една од друга.

b. Centralized versus decentralized clusters:

Centralized clusters имаат една главна контролна точка, која може да го олесни управувањето, но претставува ризик доколку таа точка откаже. decentralized clusters ја шират контролата низ многу точки, правејќи го системот поотпорен и помалку склон на неуспех. Одлуката помеѓу двете зависи од потребите на системот и доверливоста.

c. Homogeneous versus heterogeneous clusters:

Хомогените кластери имаат слични јазли за лесно управување, но помала флексибилност. Хетерогените кластери имаат различни јазли за поголема флексибилност, но бараат покомплексно управување. Изборот зависи од задачите и балансот помеѓу потребната едноставност и приспособливост.

d. Enclosed versus exposed clusters:

Enclosed clusters се затворени, обезбедувајќи контролирана и безбедна средина. Exposed clusters, од друга страна, се отворени и немаат такви ограничувања, што овозможува поголема пристапност, но потенцијално послаба безбедност.

e. Dedicated versus enterprise clusters:

Dedicated clusters се за специфични задачи, додека enterprise clusters се повеќе разновидни, опслужувајќи различни апликации во една организација. Изборот зависи од тоа дали е потребна специјализирана или широка примена.

2.2. This problem refers to the redundancy technique. Assume that when a node fails, it takes 10 seconds to diagnose the fault and another 30 seconds for the workload to be switched over.

a. **What is the availability of the cluster if planned downtime is ignored?**

Ако се игнорира планираното време на застој, availability зависи од распределбата на неуспехот на јазолот, така што ни требаат повеќе податоци за да ја одредиме достапноста.

b. **What is the availability of the cluster if the cluster is taken down one hour per week for maintenance, but one node at a time?**

Претпоставуваме дека кластерот има повеќе од еден јазол, така што bottle neck е време за дијагностицирање и време за префрлување на товарот, така што имаме:

availability = 7(days)/(7 days + 40s) = 604800s/604840s = 0.999

2.4. This problem consists of two parts related to cluster computing:

1. **Define and distinguish among the following terms on scalability:**
   a. **Scalability over machine size**

   Способноста на системот да управува со зголемен број машини или јазли во кластер или мрежа.
   b. **Scalability over problem size**

   Капацитетот на системот да се справи со поголеми датасета или посложени проблеми без намалување на перформансите.
   c. **Resource scalability**

   Ефикасноста на системот во користењето на дополнителни ресурси, како процесорска моќ или меморија, за справување со растечкиот обем на работа.
   d. **Generation scalability**

   Способноста на системот да се прилагоди на различни технолошки генерации и да остане ефективен со развојот на хардверот и софтверот.

2. **Explain the architectural and functional differences among three availability cluster configurations: hot standby, active takeover, and fault-tolerant clusters. Give two example commercial cluster systems in each availability cluster configuration. Comment on their relative strengths and weaknesses in commercial applications.**

Hot Standby Configuration: Во овој кластер, еден сервер активно го опслужува обемот на работа додека друг сервер стои на страна, подготвен да го преземе во случај на пад (дефект). Примерите се Microsoft Windows Server Failover Clustering (WSFC) и Linux High Availability (HA). Јака страна е брзото време на неуспех, но слабост е потребата од редундантни ресурси.

Active Takeover Configuration: Вклучува примарен сервер кој активно го опслужува обемот на работа, а во случај на неуспех, серверот на подготвеност го презема. Примери се Veritas Cluster Server (VCS) и IBM PowerHA. Предностите вклучуваат ефикасно искористување на ресурсите,

но слабостите може да вклучуваат подолго време на откажување во споредба со hot standby configurations.

Fault-Tolerant Cluster Configuration: Обезбедуваат континуирано работење со дистрибуција на обемот на работа низ повеќе сервери кои работат истовремено. Stratus ftServer и HPE NonStop се примери. Јаките страни се континуираното работење и моменталното откажување, но слабост е повисоката цена за одржувањето на синхронизираните системи.

2.7. This problem is related to the use of high-end x86 processors in HPC system construction. Answer the following questions:

a. Referring to the latest Top 500 list of supercomputing systems, list all systems that have used x86 processors. Identify the processor models and key processor characteristics such as number of cores, clock frequency, and projected performance.

На листата на топ 500 суперкомпјутери https://www.top500.org/ од Ноември 2023 нема ниеден со x86 процесор

b. Some have used GPUs to complement the x86 CPUs. Identify those systems that have procured substantial GPUs. Discuss the roles of GPUs to provide peak or sustained flops per dollar.

Овие системи користат GPUs:

- Summit - IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband with 148,600.0 TFlop/s
- Sierra - IBM Power System AC922, IBM POWER9 22C 3.1GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR Infiniband, IBM / NVIDIA with 94,640.0 TFlop/s,
- Selene - NVIDIA DGX A100, AMD EPYC 7742 64C 2.25GHz, NVIDIA A100, Mellanox HDR Infiniband 63,460.0 TFlop/s,
- JUWELS Booster Module - Bull Sequana XH2000 , AMD EPYC 7402 24C 2.8GHz, NVIDIA A100, Mellanox HDR InfiniBand/ParTec ParaStation ClusterSuite 44,120.0 TFlop/s,
- HPC5 - PowerEdge C4140, Xeon Gold 6252 24C 2.1GHz, NVIDIA Tesla V100, Mellanox HDR Infiniband, 35,450.0 TFlop/s.

2.9. Compare the latest Top 500 list with the Top 500 Green List of HPC systems. Discuss a few top winners and losers in terms of energy efficiency in power and cooling costs. Reveal the greenenergy winners' stories and report their special design features, packaging, cooling, and management policies that make them the winners. How different are the ranking orders in the two lists? Discuss their causes and implications based on publicly reported data.

Energy-efficient победници се: Supercomputer Fugaku with A64FX 48C 2.2GHz and Tofu interconnect D, Summit featuring IBM POWER9 22C 3.07GHz and NVIDIA Volta GV100, Sierra with IBM POWER9 22C 3.1GHz and NVIDIA Volta GV100, Sunway TaihuLight utilizing Sunway SW26010 260C 1.45GHz, Selene equipped with AMD EPYC 7742 64C 2.25GHz and NVIDIA

A100, and JUWELS Booster Module with AMD EPYC 7402 24C 2.8GHz and NVIDIA A100. Најголемиот победник користи водено ладење за да ги намали трошоците за електрична енергија и користи паметно управување со енергијата за да ја минимизира потрошувачката на јазлите во мирување.

2.10. This problem is related to processor selection and system interconnects used in building the top three clustered systems with commercial interconnects in the latest Top 500 list.

a. **Compare the processors used in these clusters and identify their strengths and weaknesses in terms of potential peak floating-point performance.**
   1. **FRONTIER:** This system leverages 8,699,904 cores from its AMD Optimized 3rd Generation EPYC 64C 2GHz processor.
   2. **AURORA:** This system leverages 4,742,808 cores from its Xeon CPU Max 9470 52C 2.4GHz processor.
   3. **EAGLE:** This system leverages 1,123,200 cores from its Xeon Platinum 8480C 48C 2GHz processor.

b. **Compare the commercial interconnects of these three clusters. Discuss their potential performance in terms of their topological properties, network latency, bisection bandwidth, and hardware used.**

   Се користат следните interconnects Slingshot-11 и NVIDIA Infiniband NDR.

2.16. Study various SSI features and HA support for clusters in Section 2.3 and answer the following questions, providing reasons for your answers. Identify some example cluster systems that are equipped with these features. Comment on their implementation requirements and discuss the operational obstacles to establish each SSI feature in a cluster system.

a. **Single entry point in a cluster environment**

   Овој атрибут се однесува на презентацијата на системот пред крајните корисници. Наместо да се поврзуваат со поединечен јазол, корисниците се поврзуваат со целиот систем, а кластерот ги користи своите механизми за да утврди кои јазли ќе бидат назначени за одредената задача.

b. **Single memory space in a cluster system**

   Оваа карактеристика се фокусира на перспективата на поединечните јазли во однос на меморијата. Секој јазол има пристап до заеднички мемориски простор, поради што се потребни дополнителни механизми за да се гарантира конзистентност.

c. **Single file hierarchy in a cluster system**

   истемот што ја вклучува оваа карактеристика ги вклучува сите јазли кои споделуваат заедничка хиерархија на датотеки, опфаќајќи го и датотечниот систем и вистинската структура на датотеки. Оваа карактеристика наложува дополнителни механизми за одржување на конзистентноста.

d. Single I/O space in a cluster system

Некои системи го користат овој атрибут, овозможувајќи им на сите јазли пристап до уредите за влез/излез на други јазли, вклучувајќи ленти, дискови, сериски линии итн. Привилегиите за пристап може да се регулираат со правила и приоритети. Комерцијалните системи што ја вклучуваат оваа функција вклучуваат HP NSK Guardian, Inferno, LOCUS, OpenSSI, Plan9, Sprite и многу други.

e. Single network space in a cluster system

Целиот систем работи во рамките на единствена LAN или WAN мрежа, во зависност од големината и видот на кластерскиот систем, било да е тој централизиран или децентрализиран.

f. Single networking in a cluster system

Јазлите се меѓусебно поврзани преку заедничка инфраструктура, која сама по себе може да претставува bottleneck за целиот систем.

g. Single point of control in a cluster system

Целиот систем е надгледуван од централна власт, која може да биде еден јазол, група јазли или посебен систем. Обезбедувањето на континуирана достапност на единствената контролна точка е од клучно значење, бидејќи таа служи како bottleneck на системот и секој дефект може да го направи целиот систем нефункционален.

h. Single job management in a cluster system

Некои јазли ја преземаат улогата на управување со работните места, а сите други јазли зависат од нивните услуги за управување со работните места. Јазлите со улога на управување со работни места обично ги имплементираат истите алгоритми за управување со работни места.

i. Single user interface in a cluster system

Целиот систем, заедно со неговите различни модули, користи заеднички кориснички интерфејс. Ова го подобрува искуството на крајниот корисник, ја поедноставува употребливоста и придонесува за целокупната одржливост на системот.

j. Single process space in a cluster system

Одредени системи ја користат функцијата Single System Image (SSI) за да создадат перцепција дека сите процеси се извршуваат на една машина. Алатките за управување со процесите, како што е „ps“ на системи слични на Unix, работат на сите процеси низ целиот кластер. Комерцијалните системи што ја вклучуваат оваа функција вклучуваат Ameba, HP NSK Guardian, Kerrighed, LOCUS, OpenSSI, TidalScale, VMScluster, z/VM, UnixWare NonStop кластери и многу други.