

Паралелно и дистрибуирано процесирање

Домашна задача 5

Петар Атанасовски - 216052

1. Што преставува Big data? Најдете повеќе дефиниции и споредете ги?

Big Data е област која опфаќа начини за анализирање, систематско извлекување информации, справување со збирки на податоци кои се премногу големи или сложени за да можат да се занимаваат со традиционалниот апликативен софтвер за обработка на податоци.

Big Data е однесува на податоци кои се толку големи, брзи или сложени што е тешко или невозможно да се обработат со користење на традиционални методи. Пристап и складирање на големи количини на информации за анализа трае долго време. Со едноставни зборови, Big Data претставува поголеми, посложени сетови на податоци.

Големите податоци се тип на податоци што не можат да се решат со традиционални методи, изворот и популарноста на овие податоци потекнуваат од новитетите во човештвото управувано од податоци.

2. Што преставуваат VVV кај Big data? Дадете пример за секое V.

Трите V се волумен (Volume), разновидност (Variety), брзина (Velocity).

Volume кажува колку податоци има, Variety кажува колку се различни различните типови на податоци, а Velocity кажува со која брзина се генерираат податоците.

- Пример за Volume: Податоците што ги има Амазон се 1000000 TB.
- Пример за Variety: Facebook има податоци за медиуми, разговор, финансиски податоци, податоци за аналитика и многу други.
- Пример за Velocity е: Facebook собира 500+ TB податоци дневно.

3. Што преставува HDFS? Опишете накратко.

Hadoop Distributed File System (HDFS) е дистрибуиран file систем дизајниран да работи на хардвер. Има многу сличности со постоечките дистрибуирани file системи но и некои разлики. HDFS е многу толерантен на грешки и е дизајниран да се користи на хардвер со ниска цена. HDFS обезбедува висока пропусна моќ за пристап до податоците од апликациите и е погоден за апликации кои имаат големи многу податоци.

4. Зошто Hadoop преставува рамка за дистрибуирано процесирање?

Бидејќи Hadoop се состои од многу модули кои обезбедуваат: дистрибуиран file систем, можност за извршување на паралелни програми, за распоред на job-ови, управување со јазли, следење на job-ови и многу повеќе. Hadoop ги обезбедува сите работи што се потребни за дистрибуирано програмирање.

5. Наведете ги моделите за услуга во облак и наведете компании кои нудат услуги за секој модел.

- **IaaS**, обезбедува услуги базирани на cloud и pay as you go модел, услуги за складирање, мрежа, виртуелизација и целосна инфраструктура која поддржува хоризонтално и вертикално скалирање на различните аспекти. Примери се: DigitalOcean, Linode, Rackspace, Amazon Web Services (AWS), Cisco Metapod, Microsoft Azure, Google Compute Engine (GCE).
- **PaaS**, обезбедува софтвер и хардвер што е достапен преку интернет и cloud, тој се потпира на некоја инфраструктура со која може да управува provider-от или може да биде IaaS, крајниот корисник има комплетна платформа, подготвена за користење на апликациите на клиентот. Примери се: AWS Elastic Beanstalk, Windows Azure, Heroku, Force.com, Google App Engine.
- **SaaS**, обезбедува софтвер кој е достапен преку интернет, но со него управуваат трети страни. Некои примери се: Google Workspace, Dropbox, Salesforce, Cisco WebEx, Concur, GoToMeeting.

6. Наведете пример на пресметување во облак кој што користи нуди ресурси за складирање на податоци! Образложете го примерот!

CloudStorage на Firebase е таков пример. Обезбедува практично бесконечна количина на складирање, скалабилна на барање.

7. Наведете неколку предизвици за воведување на пресметување со облак.

Некои од предизвиците околу имплементацијата на cloud computing се: Безбедносни прашања, Управување и ограничување на трошоците, недостаток на ресурси/експертиза, управување/контрола, усогласеност, управување со повеќе облаци, перформанси, градење приватен облак, миграција.

Постојат многу примери на несоодветно управување со трошоците што доведуваат до речиси банкрот.

8. Наведете неколку примери на фирми кои нудат решенија за приватен облак и дадете краток опис.

- HPE - Клучен лидер во приватниот cloud, нуди хардвер и услуги.
- VMware – Еден од најдоброто сервиси за приватен cloud.
- Oracle - Лидер во бази на податоци и складирање за претпријатија.
- IBM / Red Hat OpenShift - понуди на PaaS, работи со други cloud платформи.
- Cisco - Лидер во вмрежувањето.
- AWS – Најголем продавач во областа на cloud компјутерите.