

Projektowanie i wdrażanie systemów w chmurze

Lista zadań na ćwiczenia 19.01.18

1. Przedstaw znaczenie i omów działanie wszystkich głównych etapów obliczeń w modelu MapReduce.
2. Przedstaw prosty program napisany w modelu MapReduce. Możesz użyć pseudokodu, lub dowolnego języka, ale upewnij się, że interfejs którego używasz, jest ściśle zgodny z jakimś prawdziwym narzędziem obsługującym MapReduce (np. Hadoop).
3. Porównaj w szczegółach Sparka i Hadoopa. Do jakich obliczeń się nadają, a kiedy ich wybór jest złym pomysłem?
4. Przedstaw poglądowo organizację obliczeń związanych z:
 - a. [1 pkt] Indeksowaniem
 - b. [1 pkt] Ustalaniem wyników wyszukiwaniadla dużej wyszukiwarki. Inspirację możesz znaleźć tutaj:
<https://www.google.com/search/howsearchworks/algorithms/>
Zaproponuj kroki i sposób ich zrównoleglenia. Przykładowo, dla wyników wyszukiwania, całe zapytanie może być dzielone na słowa, które mogą mieć jakieś własne indeksy, a te indeksy mogą być trzymane w rozproszonym systemie plików o określonej hierarchii. Znaczenie może mieć kontekst geograficzny, informacja o zalogowanym użytkowniku i wiele innych.
W przypadku budowania indeksu wyszukiwarki, konieczne może być spojrzenie na linki na stronie, słowa kluczowe i ich rozmieszczenie, czas ładowania i wiele innych.
Odpowiednio przetworzone i zapisane informacje pozwolą na sprawne określanie wyników wyszukiwania.
5. Przedstaw kilka zastosowań dla usługi AWS Batch. Opowiedz jak dokładnie zostanie ona wykorzystana w każdym z nich (jakie instancje zostaną użyte i dlaczego, jakie będą reguły opisujące wykonanie zadań itp.).
6. Omów szczegółowo usługi EMR (AWS) i Dataproc (GCP). Zaprezentuj integracje z innymi usługami w ramach platformy.
7. Wiele firm chwali się w Internecie w jaki sposób użycie chmury pomogło im rozwiązać wyzwania związane z dużymi obliczeniami. Znajdź przykład takiego sukcesu i omów co sprawiło, że rozwiązanie chmurowe było właściwe dla takiego scenariusza.