

Projektowanie i wdrażanie systemów w chmurze

Lista zadań na ćwiczenia 2018.01.28

1. 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).
2. Omów szczegółowo usługi EMR (AWS) i Dataproc (GCP). Jak są zintegrowane z innymi usługami platformy chmurowej?
3. 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.
4. Wybierz i omów jakiś framework do tworzenia aplikacji w paradygmacie serverless. Jak się go używa w praktyce? Co czyni go wygodniejszym od samodzielnego pisania kodu np. do AWS Lambda?
5. Czym są procesy ETL i gdzie mają zastosowanie? Jakie proste usługi chmurowe można wykorzystać implementując takie procesy? Jakie narzędzia dedykowane pod ETL oferuje AWS?
6. [0.5 pkt] Podaj przykład (inny niż na wykładzie) procesu obsługi zasobów chmurowych, który jest trudny do skonfigurowania, ale użycie prostej Lambdy (omów pobieżnie jej kod) rozwiązuje problem.
7. Przedstaw jakiś **niebanalny** problem obliczeniowy, który da się wyrazić poprzez MapReduce. Omów w jaki sposób wykorzystanie paradygmatu MapReduce upraszcza myślenie i rozwiązywanie tego problemu.
8. Jakie konsekwencje przy dużych obliczeniach niesie ze sobą wybór typów instancji lub modelu ich rozliczania? Jak podejmować decyzje w tych tematach? Jakie inne charakterystyki instancji mogą wpływać na to czy będą się dobrze nadawały do określonych obliczeń? Jakie wąskie gardła mogą się pojawić?
9. Do jakich problemów lepiej jest wykorzystywać hurtownie danych (np. AWS Redshift) zamiast klasycznej bazy danych? Czym różni się *data warehouse* od *data lake*, i kiedy stosować jedno, a kiedy drugie?