Wejść na strony: Amazon Web Services (AWS), Microsoft Azure, Google Cloud Platform (GCP)

„Po zapoznaniu się z każdą platformą, przygotuj krótką notatkę z odpowiedzią na pytania:

1. Która platforma wydaje Ci się najbardziej odpowiednia do analizy danych genomowych i

dlaczego?

Najbardziej adekwatną platformą wydaje mi się Google Cloud Platform, po pierwsze uważam że platforma ma największy potencjał rozwojowy biorąc pod uwagę narzędzia AI/Machine Learning. Mają autorski API Google cloud life sciences i Deepvariant (narzędzie do wywoływania wariantów genomowych AI-based) – z racji że sam robię takie projekty doceniam ich jakość, szybkość i poziom automatyzacji. Google z tego co czytałem dobrze się integruje z narzędzia do ML, takimi jak TensorFlow. Microsoft Azure oferuje Azure Genomics integrujący się z Cromwellem i Nextflow, ale jest mniej „gotowy” i wyczuwam problemy z potencjalną optymalizacją wielu narzędzi. Na Google Cloud wiele rzeczy jest po prostu gotowych. Dodatkowo Azure integruje się z narzędziami Microsoft 365, jednak to raczej drugorzędna w kontekście danych genomowych, szczególnie biorąc pod uwagę konieczność posiadania licencji w organizacji. Co może być jednak korzystne, to narzędzia takie jak BWA i GATK dostępne na stronie. Amazon Web Services najwcześniej rozpoczął współdziałać z dziedziną bioinformatyki, ma gotowe narzędzia do do analizy i przechowywania FASTQ/BAM/VCF files. Dodatkowo ma AWS HealthOmics, dobrze zabezpieczony i zautomatyzowany serwis do analizy danych genomowych, także tych poufnych. Biorąc pod uwagę te wszystkie aspekty, wciąż wierzę w Google Cloud, ale wydaje mi się, że AWS również nie jest daleko od niego, szczególnie pod kątem przyszłości.

2. Jakie konkretne narzędzia lub usługi chciałbyś przetestować w swoim projekcie?

W AWS chciałbym sprawdzić jak działa integracja z Nextflow, bo zazwyczaj mam z tym trochę problemów – może byłoby łatwiej ?

W Azure chciałbym sprawdzić modele predykcyjne oparte na AI

W Google CLoud chętnie potestowałbym narzędzie do wywoływania wariantów (Deep Variant)

3. Czy widzisz istotne różnice w cenach lub łatwości użytkowania między platformami?”

AWS wydaje się być dość skomplikowany w konfiguracji, darmowe usługi mają niewielkei zasoby (5 GB), podejrzewam że ceny mogą bardzo szybko rosnąć jeśli przekroczymy limit, a przy danych genomowych na pewno tak będzie (często mają po kilkaset gb w analizach ludzkich genomów, które zazwyczaj robię).

Azure będzie bardziej intuicyjny dla ludzi, którym bliski jest interfejs Microsoftu, darmowy kredyt to 200 dolarów.

Google Cloud wydaje się mi być subiektywnie najbardziej intuicyjny o największej liczbie gotowych narzędzi genomowych, dodatkowo daje 300 dolarów darmowego kredytu, czyli najwięcej z 3 platform. Nadal wybieram Google Cloud.