Traccia 1 – Seconda prova

- Parlare delle random forest, della loro implementazione nelle librerie più diffuse di Python ed esemplificare il loro utilizzo pratico.

- Se dovesse calcolare semplici funzioni statistiche, quale applicazione del pacchetto Office userebbe? Quali sono i limiti di tale applicazione nel soddisfare i requisiti di applicazioni di analytics?

- Tradurre in Italiano il seguente paragrafo: "High throughput at low latency: Bigtable is ideal for storing very large amounts of data in a key-value store and supports high read and write throughput at low latency for fast access to large amounts of data. Throughput scales linearly—you can increase QPS (queries per second) by adding Bigtable nodes. Bigtable is built with proven infrastructure that powers Google products used by billions such as Search and Maps. Cluster resizing without downtime: Scale seamlessly from thousands to millions of reads/writes per second. Bigtable throughput can be dynamically adjusted by adding or removing cluster nodes without restarting, meaning you can increase the size of a Bigtable cluster for a few hours to handle a large load, then reduce the cluster's size again—all without any downtime. Flexible, automated replication to optimize any workload: Write data once and automatically replicate where needed with eventual consistency—giving you control for high availability and isolation of read and write workloads. No manual steps needed to ensure consistency, repair data, or synchronize writes and deletes. Benefit from a high availability SLA of 99.99% for instances with multi-cluster routing (99.9% for single-cluster instances)."

Traccia 2 – Seconda prova

- Parlare dei sistemi di analisi dei dati da social media, fornendo almeno un esempio di applicazione e le relative caratteristiche di progetto.

- Se dovesse manipolazione di dati numerici, quale applicazione del pacchetto Office userebbe? Quali sono i limiti di tale applicazione nel soddisfare i requisiti di data preparation e successiva analisi?

- Tradurre in Italiano il seguente paragrafo: "Amazon DynamoDB is a key-value and document database that delivers single-digit millisecond performance at any scale. It's a fully managed, multi-region, multi-active, durable database with built-in security, backup and restore, and in-memory caching for internet-scale applications. DynamoDB can handle more than 10 trillion requests per day and can support peaks of more than 20 million requests per second. Many of the world's fastest growing businesses such as Lyft, Airbnb, and Redfin as well as enterprises such as Samsung, Toyota, and Capital One depend on the scale and performance of DynamoDB to support their mission-critical workloads. Hundreds of thousands of AWS customers have chosen DynamoDB as their key-value and document database for mobile, web, gaming, ad tech, IoT, and other applications that need low-latency data access at any scale. Create a new table for your application and let DynamoDB handle the rest."