Data warehouse pertama kali muncul pada tahun akhir 1980-an. Tujuan awalnya adalah untuk membantu proses aliran data dari sistem operasional ke dalam sistem pendukung keputusan atau *decision-support system*(DSS). Seiring berjalannya waktu, data warehouse berkembang menjadi lebih efisien. Ia berevolusi dari penyimpanan informasi pendukung platform *business intelligence*menjadi infrastruktur analitis luas yang mendukung berbagai macam aplikasi.

AI dan machine learning telah mengubah banyak hal dari mulai industri, layanan, aset perusahaan, juga sistem data warehouse. Ekspansi big data dan penerapan teknologi digital mendorong perubahan dalam kapabilitas data warehouse. Untuk mendukung hal tersebut, ada beberapa *tools* yang perlu kita ketahui.

**RDBMS**

Dalam model relasional, sebuah database terdiri dari banyak tabel. Sebuah tabel dibentuk dari kolom dan baris yang memuat nilai tertentu. Konsep Relational Database Management System (RDBMS) sendiri merupakan sistem yang mendukung adanya hubungan atau relasi antar tabel pada suatu database. Setiap tabel dihubungkan dengan tabel lainnya dengan menggunakan primary key dan foreign key. Saat ini sudah banyak jenis database yang menerapkan model RDBMS. Sebut saja MySQL, PostgreSQL, dan Microsoft SQL Server.

**NoSQL**

Sesuai dengan namanya NoSQL adalah jenis basis data yang tidak menggunakan bahasa SQL dalam manipulasi datanya. Dalam penyimpanan datanya, NoSQL memiliki beberapa teknik penyimpanan yaitu: dokumen, graph, key-value, dan column based.

1. Dokumen : Menghubungkan setiap kunci dengan struktur data kompleks yang disebut dokumen.
2. Graph : Menyimpan informasi tentang jaringan data, seperti koneksi sosial.
3. Nilai-kunci : Database NoSQL paling sederhana di mana setiap elemen dalam database disimpan sebagai nilai yang diasosiasikan dengan sebuah kunci.
4. Kolom : Menyimpan data yang memiliki volume besar, di mana setiap elemen data disimpan pada kolom bukan pada baris.

Beberapa database NoSQL terpopuler adalah MongoDB, CouchDB, Cassandra, Redis, Neo4J, dan Riak. Jika ingin mengetahui lebih lanjut tentang NoSQL, kunjungi [tautan](https://www.quora.com/What-are-the-main-differences-between-the-four-types-of-NoSql-databases-KeyValue-Store-Column-Oriented-Store-Document-Oriented-Graph-Database) berikut.

**Firebase Realtime Database**

Sesuai namanya, “Database Realtime” adalah database yang menyimpan data yang berubah seiring waktu. Data jumlah penjualan harian, pengunjung mall setiap jam, arus lalu lintas setiap menit, atau fluktuasi saham setiap detik merupakan beberapa contoh data realtime. Data pada database realtime disimpan dalam format waktu dan nilai pada waktu yang terkait seperti gambar di bawah.

| **Timestamp** | **Metric 1** |
| --- | --- |
| 2019-03-28 00:00:01 | 2356 |
| 2019-03-28 00:00:02 | 6874 |
| 2019-03-28 00:00:03 | 3245 |
| 2019-03-28 00:00:04 | 2340 |

Firebase Realtime Databas*e* (FRD) adalah database berbasis cloud yang didesain khusus untuk mengelola data realtime. FRD dapat menyimpan dan melakukan sinkronisasi data secara realtime di mana setiap kali ada perubahan data terbaru, FRD langsung menyimpannya pada Cloud. FRD juga dilengkapi fitur offline di mana ketika tidak ada koneksi internet, FRD akan menyimpan data secara lokal, kemudian saat *online*, akan melakukan sinkronisasi ke Cloud. Keren, kan?

**Spark**

Apache Spark adalah perangkat lunak untuk pemrosesan dan analisis data berskala besar. Spark dapat digunakan dalam proses ETL (Extract, Transform, Load), data streaming, perhitungan grafik, SQL, dan machine learning. Untuk machine learning, Spark menyediakan MLlib yang berisi implementasi model machine learning seperti klasifikasi, regresi, pengklasteran, penurunan dimensi, dan pemfilteran kolaboratif.

**Big Query**

BigQuery adalah *data warehouse* berbasis cloud untuk perusahaan yang menawarkan penyimpanan data berbasis SQL dan analisis data berukuran besar. Karena berbasis cloud dan tidak ada infrastruktur yang perlu dikelola, pengguna dapat berfokus pada pengolahan data tanpa memerlukan seorang *administrator database*.

Sebagai ML Engineer masa depan, kita harus mampu mengoperasikan berbagai jenis data storage dan data warehouse. Sebabnya, perusahaan tempat kita ingin bekerja sebagai ML Engineer nanti, tidak selalu menggunakan *data warehouse* atau *data storage* yang sama.