**SparkSQL**

[SparkSQL](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://spark.apache.org/sql/&usg=ALkJrhh7hl44d7fk2ZTFooDb-WwosiuYfQ) adalah komponen Spark yang mendukung data query baik melalui SQL atau melalui [Hive Query Language](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://cwiki.apache.org/confluence/display/Hive/LanguageManual&usg=ALkJrhgQq9I68K4rtkAOByGgtpkgRzvJAQ) . Ini berasal sebagai port Apache Hive untuk berjalan di atas Spark (di tempat MapReduce) dan sekarang terintegrasi dengan Spark stack. Selain menyediakan dukungan untuk berbagai sumber data, memungkinkan untuk menenun query SQL dengan transformasi kode yang menghasilkan alat yang sangat hebat. Berikut adalah contoh permintaan kompatibel dengan Hive:

// sc is an existing SparkContext. val sqlContext = new org.apache.spark.sql.hive.HiveContext(sc) sqlContext.sql("CREATE TABLE IF NOT EXISTS src (key INT, value STRING)") sqlContext.sql("LOAD DATA LOCAL INPATH 'examples/src/main/resources/kv1.txt' INTO TABLE src") // Queries are expressed in HiveQL sqlContext.sql("FROM src SELECT key, value").collect().foreach(println)

**Spark SQL** adalah modul Apache Spark untuk bekerja dengan data terstruktur.

Terpadu

Murni mencampur query SQL dengan program Spark.

Spark SQL memungkinkan Anda [mengurutkan](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://spark.apache.org/docs/latest/sql-programming-guide.html&usg=ALkJrhhMmnSJE1HSxYQyB1q5DTVh72fsAA) data terstruktur di dalam program Spark, menggunakan SQL atau [API DataFrame yang](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://spark.apache.org/docs/latest/sql-programming-guide.html&usg=ALkJrhhMmnSJE1HSxYQyB1q5DTVh72fsAA) sudah [dikenal](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://spark.apache.org/docs/latest/sql-programming-guide.html&usg=ALkJrhhMmnSJE1HSxYQyB1q5DTVh72fsAA). Digunakan di Jawa, Scala, Python dan R.

konteks = HiveContext (sc)   
hasil = konteks. sql   
"SELECT \* FROM people" )   
nama = hasil peta ( lambda p: p.name )

Menerapkan fungsi untuk hasil query SQL.

Uniform Data Access

Hubungkan ke sumber data dengan cara yang sama.

DataFrames dan SQL menyediakan cara yang umum untuk mengakses berbagai sumber data, termasuk Hive, Avro, Parket, ORC, JSON, dan JDBC.Anda bahkan dapat menggabungkan data ke sumber ini.

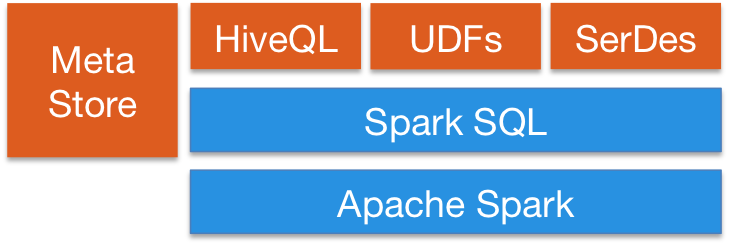
konteks. jsonFile ( "s3n: // ..." )   
.registerTempTable ("json")   
hasil = konteks. sql   
"" "PILIH \*   
DARI orang   
JOIN json ... "" " )

Permintaan dan gabungkan sumber data yang berbeda.

Integrasi sarang

Jalankan query SQL atau HiveQL pada gudang yang ada.

Spark SQL mendukung sintaks HiveQL serta Serous Serigala dan UDF, yang memungkinkan Anda mengakses gudang sarang yang ada.

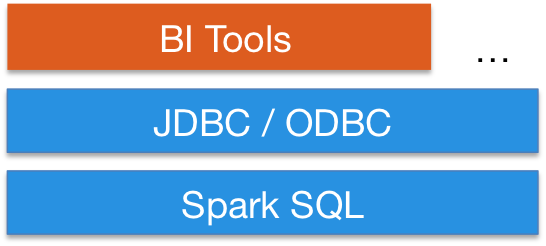


Spark SQL dapat menggunakan metastora Hive yang ada, SerDes, dan UDFs.

Konektivitas Standar

Hubungkan melalui JDBC atau ODBC.

Modus server menyediakan konektivitas JDBC standar dan konektivitas standar untuk alat intelijen bisnis.



Gunakan alat BI yang ada untuk mengajukan permintaan data yang besar.

Kinerja & Skalabilitas

Spark SQL mencakup pengoptimalan berbasis biaya, penyimpanan kolumnar dan pembuatan kode untuk membuat kueri dengan cepat. Pada saat yang sama, skala pada ribuan node dan query multi jam menggunakan mesin Spark, yang memberikan toleransi kesalahan mid-query penuh. Jangan khawatir menggunakan mesin yang berbeda untuk data historis.

Masyarakat

Spark SQL dikembangkan sebagai bagian dari Apache Spark. Dengan demikian akan diuji dan diperbarui dengan setiap rilis Spark.

Jika Anda memiliki pertanyaan tentang sistem ini, tanyakan pada [milis Spark](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://spark.apache.org/community.html&usg=ALkJrhhTA3hJFnFMci4OiqEu2XFgk92MjA" \l "mailing-lists) .

Para pengembang Spark SQL menyambut baik kontribusi. Jika Anda ingin membantu, baca [bagaimana berkontribusi pada Spark](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://spark.apache.org/contributing.html&usg=ALkJrhh4h7YmiWzi3ojDxn_tSI29y_-mJA) , dan kirimkan patch!

Mulai

Untuk memulai dengan Spark SQL:

* [Download Spark](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://spark.apache.org/downloads.html&usg=ALkJrhjOlfLTwIivw9wiXCPk8tecA_i6SQ) . Ini termasuk Spark SQL sebagai modul.
* Baca panduan [Spark SQL dan DataFrame](https://translate.googleusercontent.com/translate_c?depth=1&hl=id&prev=search&rurl=translate.google.co.id&sl=en&sp=nmt4&u=https://spark.apache.org/docs/latest/sql-programming-guide.html&usg=ALkJrhhMmnSJE1HSxYQyB1q5DTVh72fsAA) untuk mempelajari API.