HIVE TUTORIAL

simplilearn

What's in it for you?

1. History of Hive

What's in it for you?

1. History of Hive
2. What is Hive?

What's in it for you?

1. History of Hive
2. What is Hive?
3. Architecture of Hive

What's in it for you?

1. History of Hive
2. What is Hive?
3. Architecture of Hive
4. Data flow in Hive

What's in it for you?

1. History of Hive
2. What is Hive?
3. Architecture of Hive
4. Data flow in Hive
5. Hive Data Modeling

What's in it for you?

1. History of Hive
2. What is Hive?
3. Architecture of Hive
4. Data flow in Hive
5. Hive Data Modeling
6. Hive Data types

What's in it for you?

1. History of Hive
2. What is Hive?
3. Architecture of Hive
4. Data flow in Hive
5. Hive Data Modeling
6. Hive Data types
7. Different modes of Hive

What's in it for you?

1. History of Hive
2. What is Hive?
3. Architecture of Hive
4. Data flow in Hive
5. Hive Data Modeling
6. Hive Data types
7. Different modes of Hive
8. Difference between Hive and RDBMS

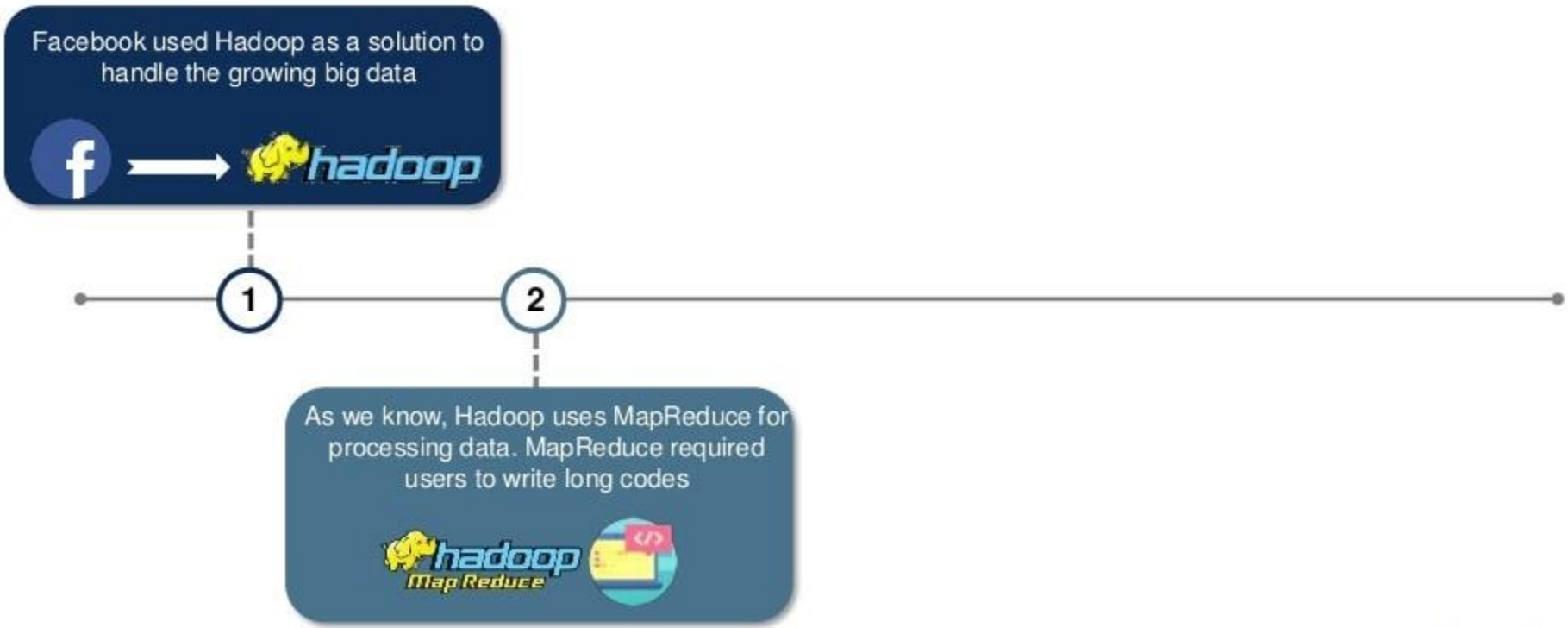
What's in it for you?

1. History of Hive
2. What is Hive?
3. Architecture of Hive
4. Data flow in Hive
5. Hive Data Modeling
6. Hive Data types
7. Different modes of Hive
8. Difference between Hive and RDBMS
9. Features of Hive

History of Hive

History of Hive

History of Hive



simplilearn

History of Hive

1. Facebook used Hadoop as a solution to handle the growing big data

2. As we know, Hadoop uses MapReduce for processing data. MapReduce required users to write long codes

3. Not all users were well versed with Java and other coding languages. This proved to be a disadvantage for them

simplilearn

History of Hive

1. Facebook used Hadoop as a solution to handle the growing big data

2. As we know, Hadoop uses MapReduce for processing data. MapReduce required users to write long codes

3. Not all users were well versed with Java and other coding languages. This proved to be a disadvantage for them

4. Users were comfortable with writing queries in SQL

simplilearn

History of Hive

1. Facebook used Hadoop as a solution to handle the growing big data

2. As we know, Hadoop uses MapReduce for processing data. MapReduce required users to write long codes

3. Not all users were well versed with Java and other coding languages. This proved to be a disadvantage for them

4. Users were comfortable with writing queries in SQL

5. Hive was developed with a vision to incorporate the concepts of tables, columns just like SQL

simplilearn

Why Hive?

Problem

For processing and analyzing data, users found it difficult to code as not all of them were well versed with the coding languages

Processing

Analyzing

Solution

Users required a language similar to SQL which was well known to all the users

Solution

HiveQL

simplilearn

What is Hive?

What is Hive?

Hive is a data warehouse system which is used for querying and analyzing large datasets stored in HDFS

What is Hive?

Hive is a data warehouse system which is used for querying and analyzing large datasets stored in HDFS

Hive uses a query language call HiveQL which is similar to SQL

What is Hive?

Hive is a data warehouse system which is used for querying and analyzing large datasets stored in HDFS

Hive uses a query language call HiveQL which is similar to SQL

simplilearn

What is Hive?

Hive is a data warehouse system which is used for querying and analyzing large datasets stored in HDFS

Hive uses a query language call HiveQL which is similar to SQL

simplilearn

What is Hive?

Hive is a data warehouse system which is used for querying and analyzing large datasets stored in HDFS

Hive uses a query language call HiveQL which is similar to SQL

simplilearn

Architecture of Hive

Architecture of Hive

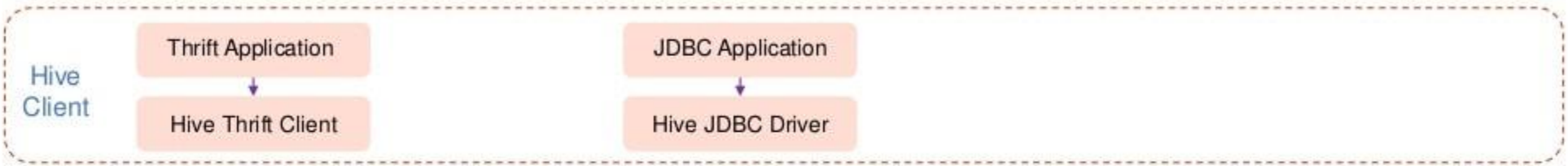
Architecture of Hive

Hive Client supports different types of client applications in different languages for performing queries

Architecture of Hive

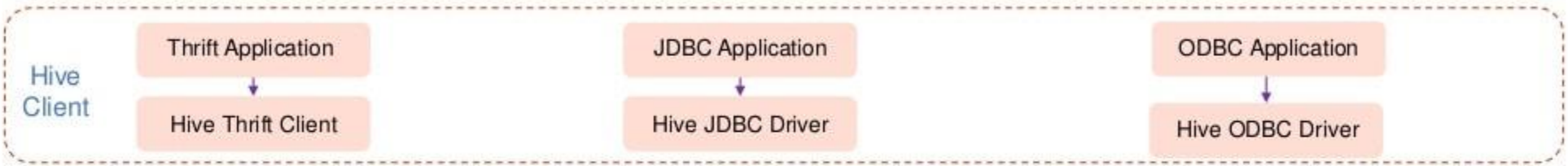
Thrift is a software framework. Hive server is based on thrift, so it can serve the request from all the programming languages that supports thrift

Architecture of Hive



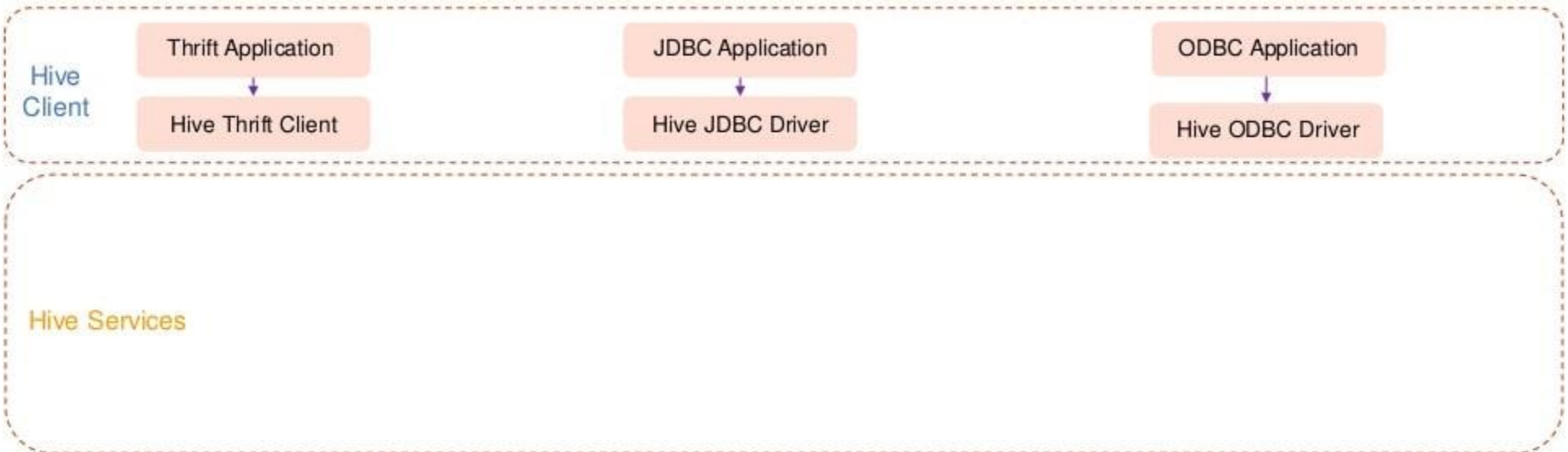
JDBC - Java Database Connectivity
JDBC application is connected through JDBC Driver

Architecture of Hive

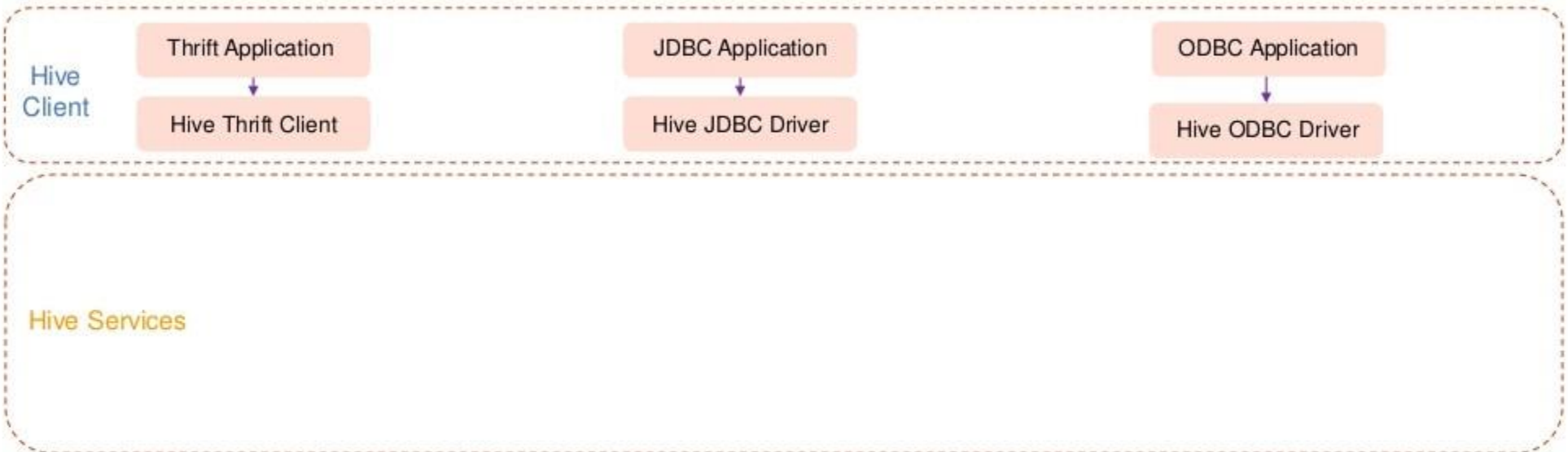


ODBC - Open Database Connectivity
ODBC application is connected through ODBC Driver

Architecture of Hive

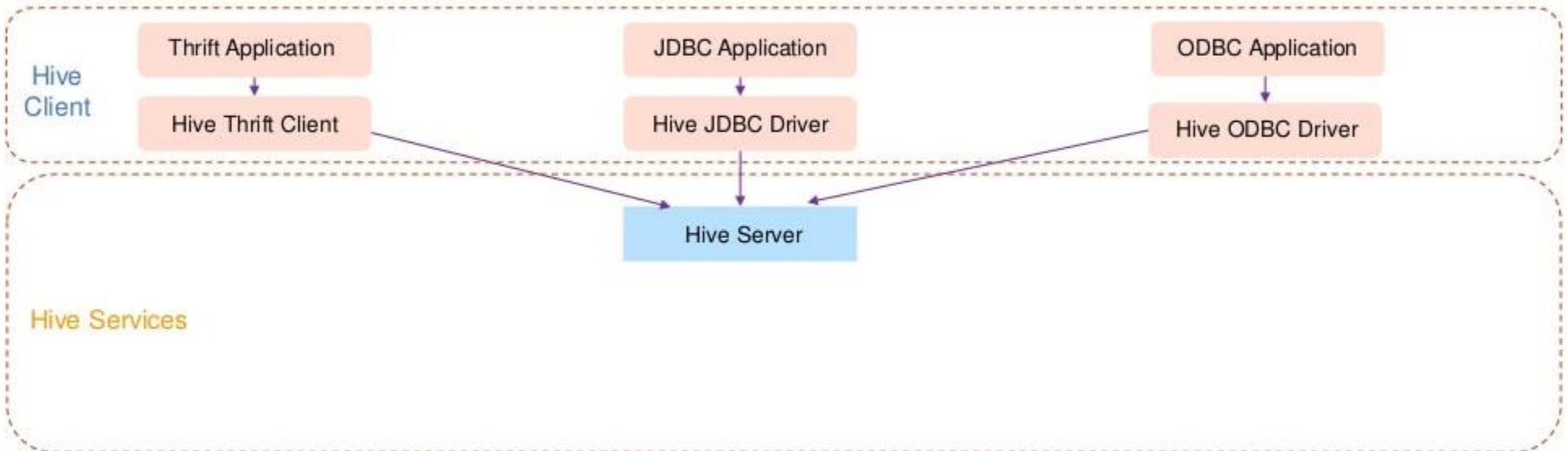


Architecture of Hive



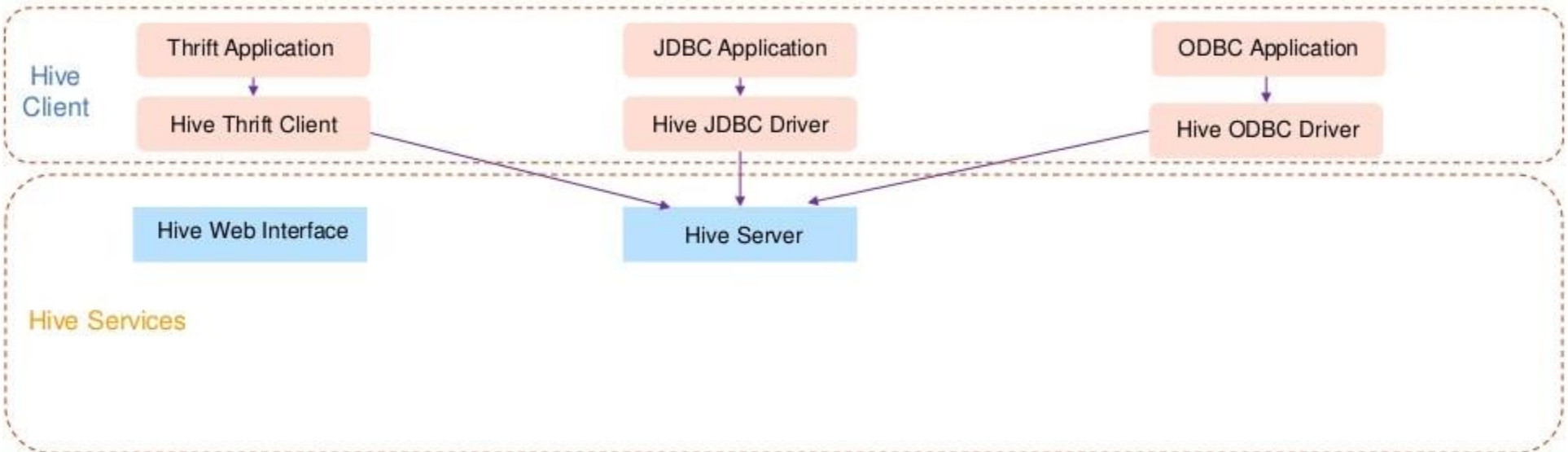
Hive supports various **services**

Architecture of Hive



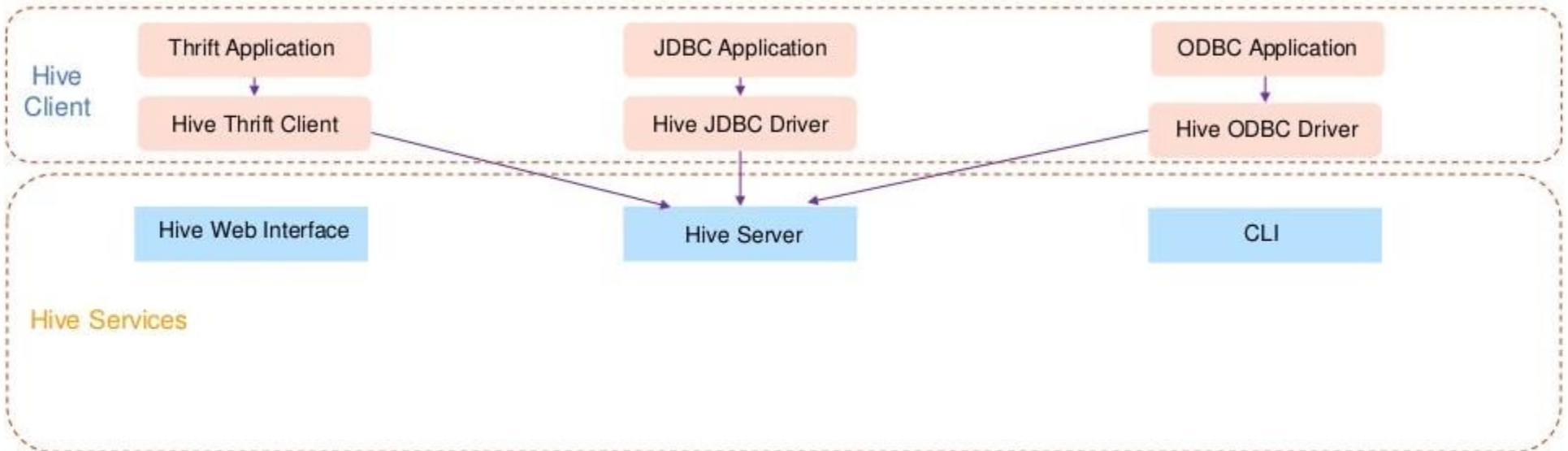
All the client requests are submitted to the **Hive server**

Architecture of Hive



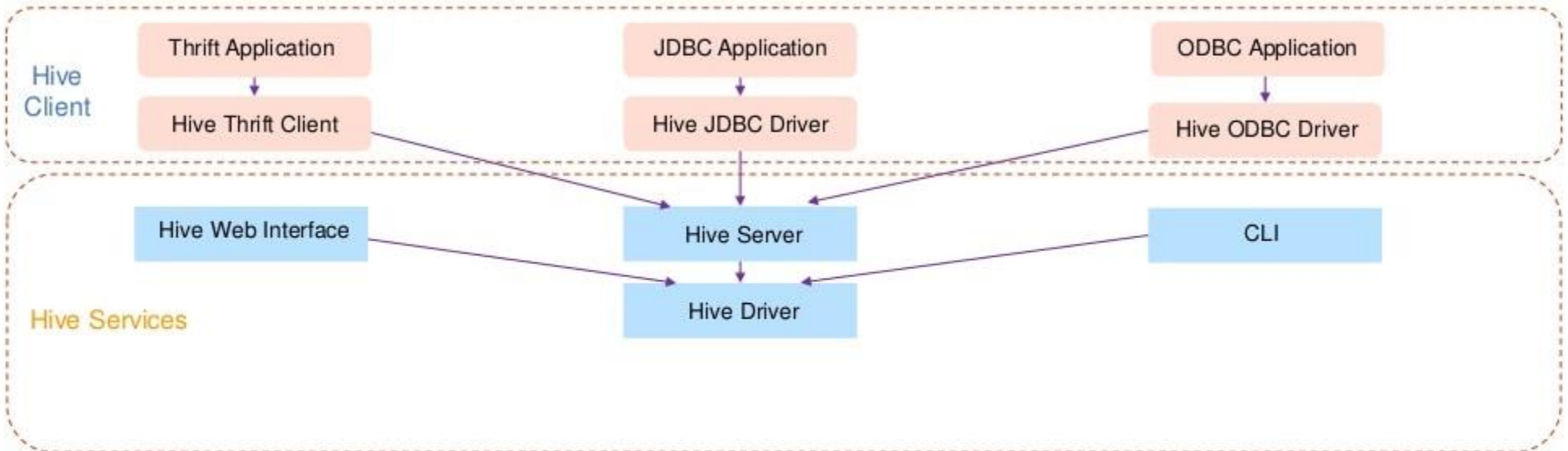
GUI is provided to execute Hive queries

Architecture of Hive



Commands are executed directly in **CLI**

Architecture of Hive

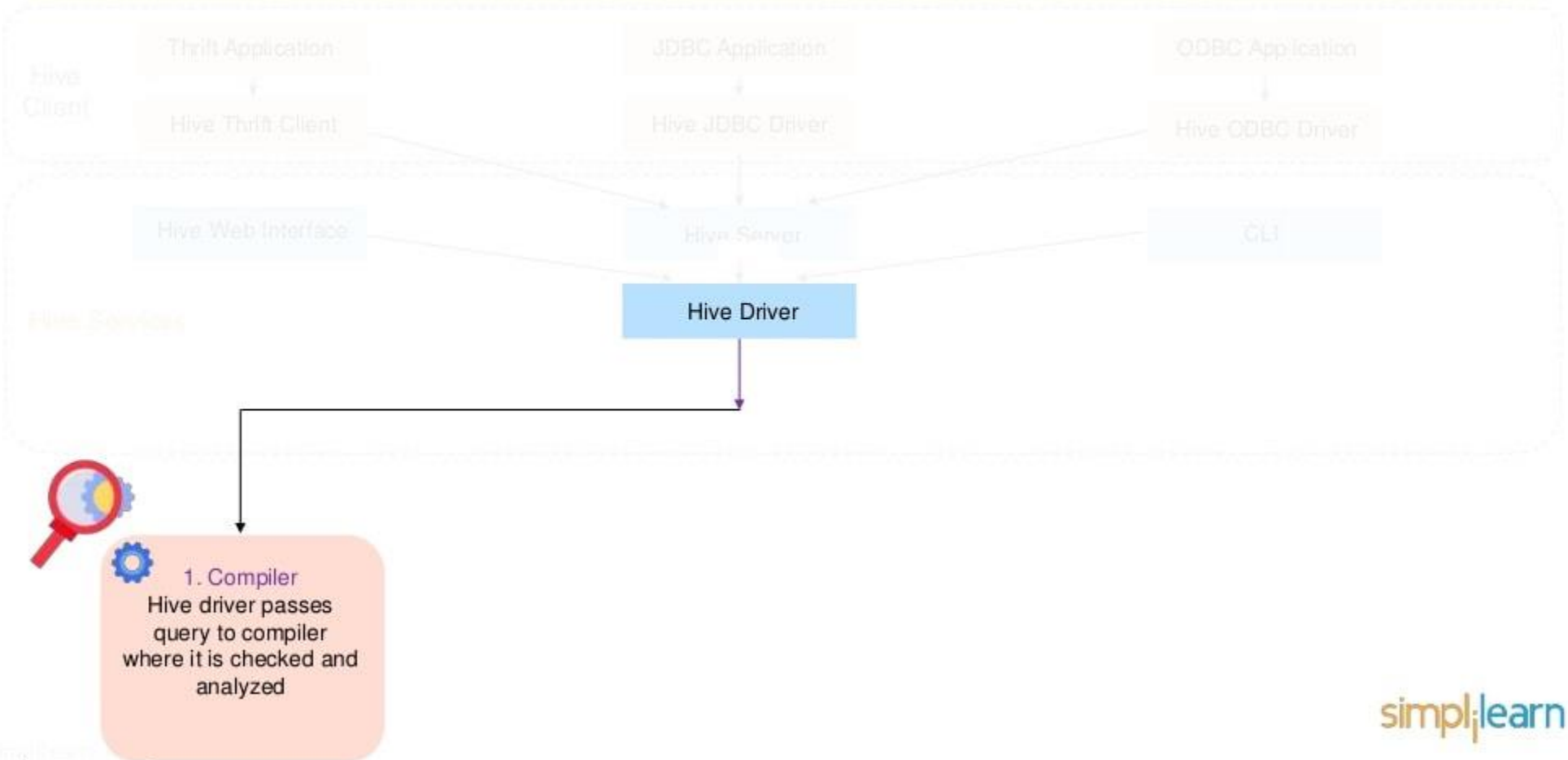


Hive driver is responsible for all the queries submitted

Architecture of Hive

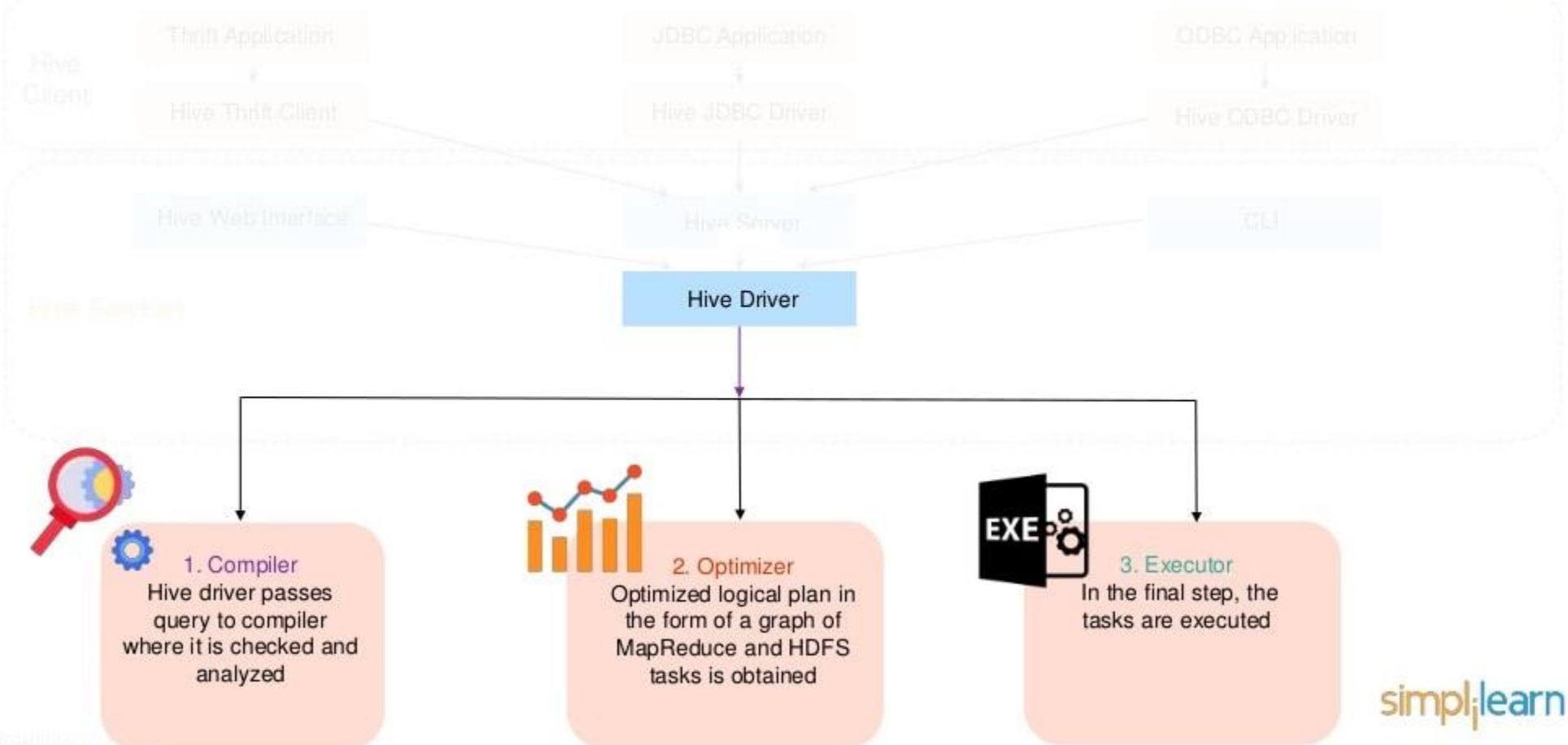
The Hive Driver now performs 3 steps internally

Architecture of Hive

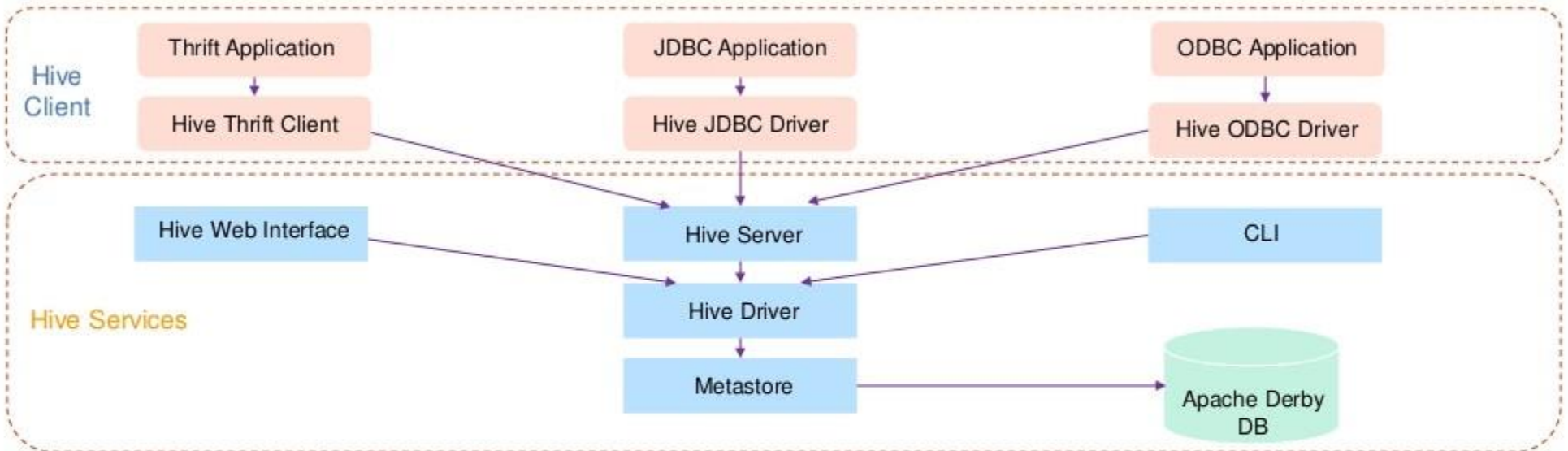


Architecture of Hive

Architecture of Hive

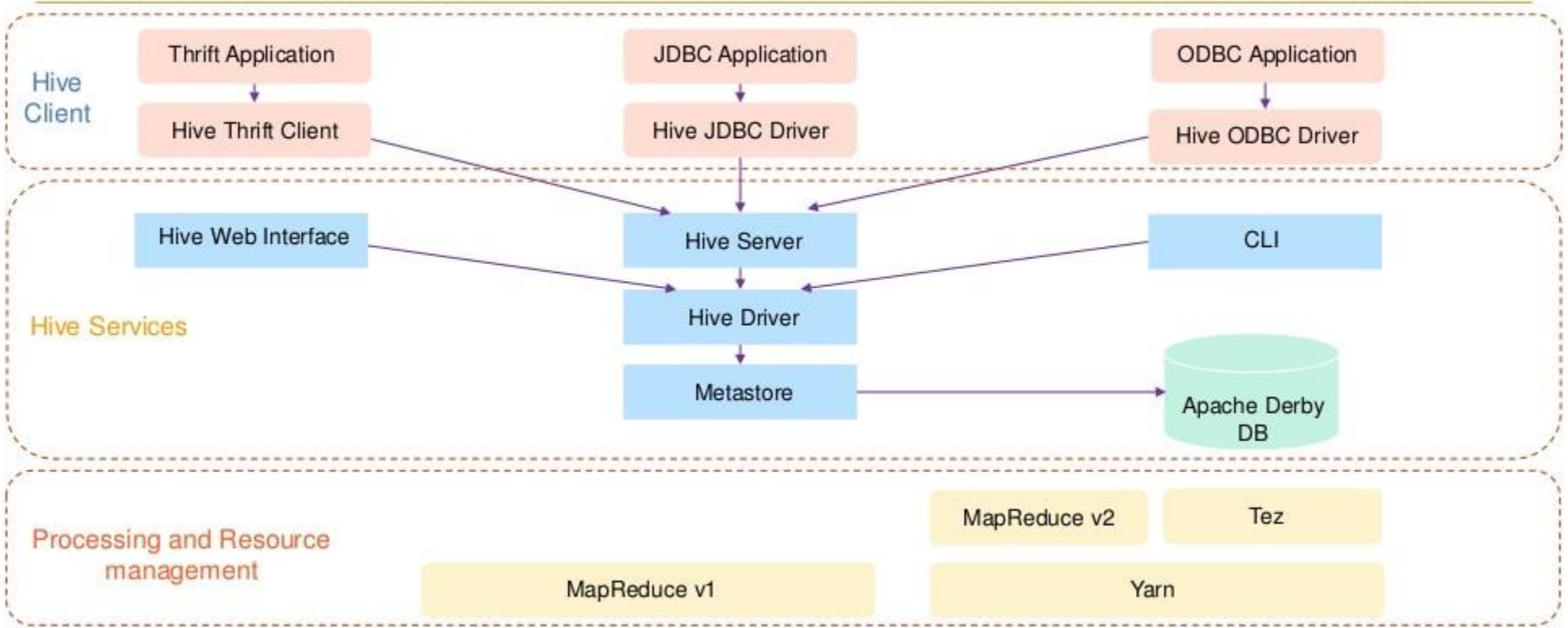


Architecture of Hive

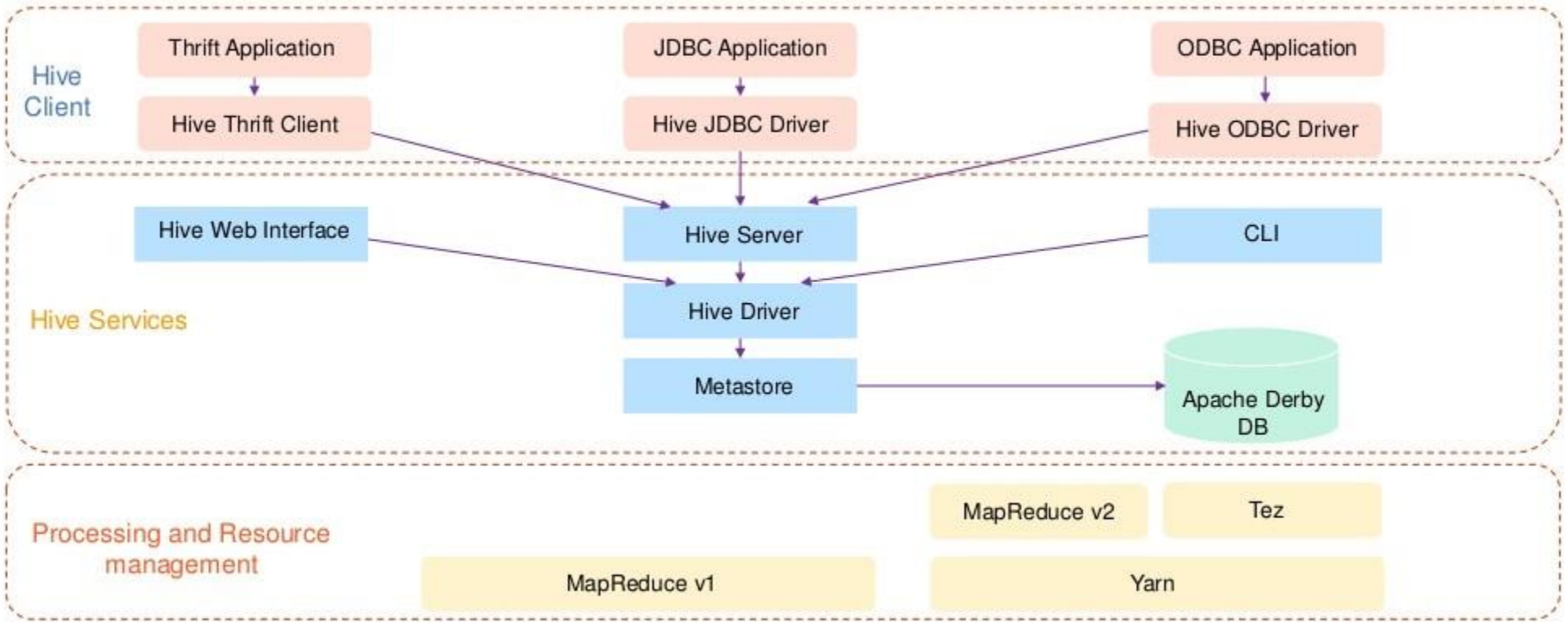


Metastore is a repository for Hive metadata. Stores metadata for Hive tables

Architecture of Hive

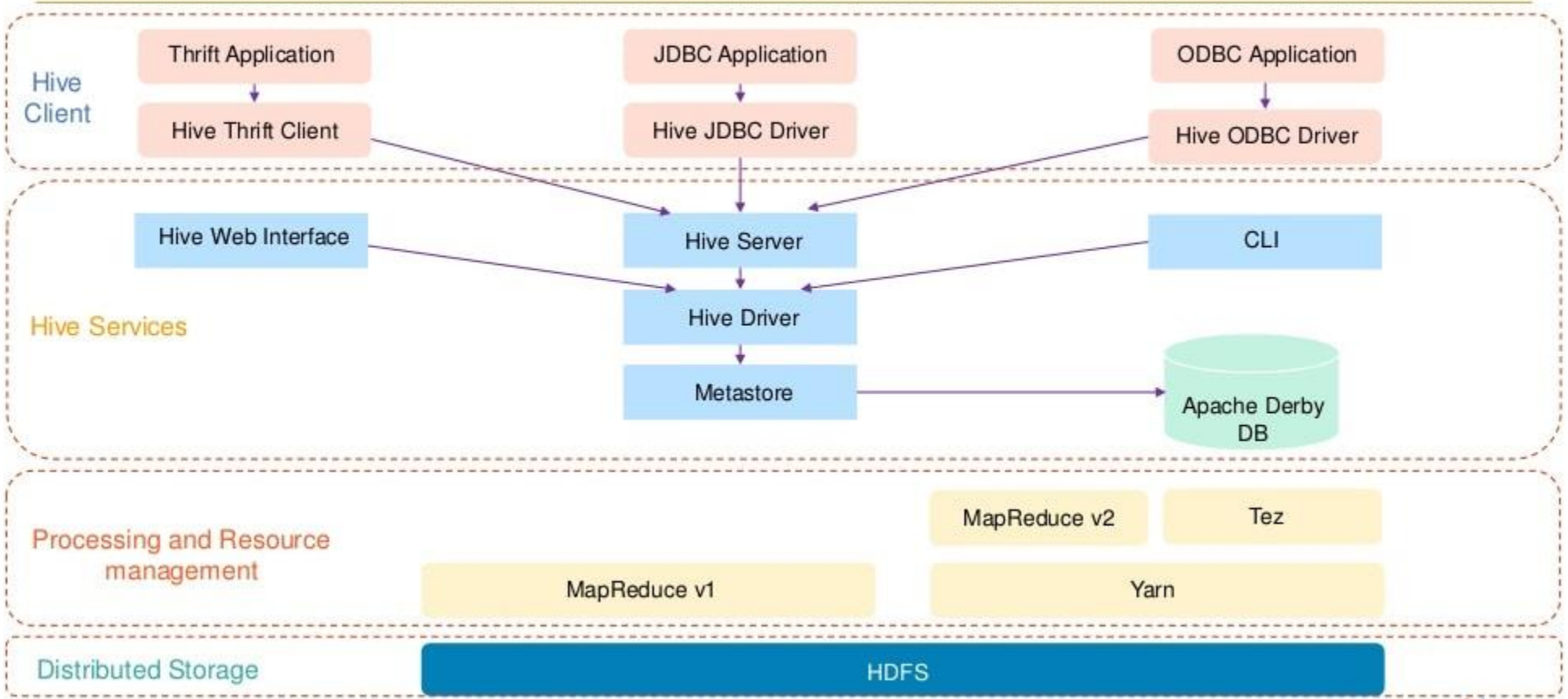


Architecture of Hive



Hive uses MapReduce framework to process queries

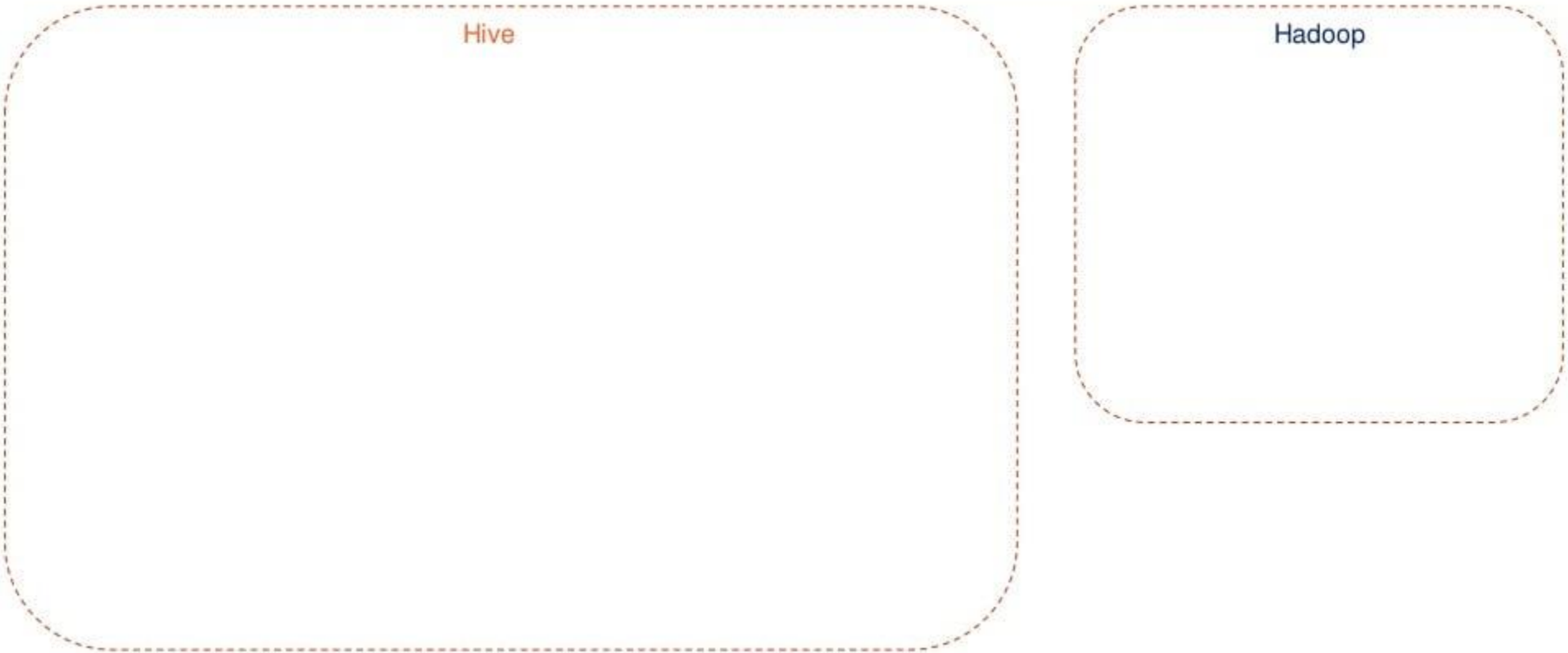
Architecture of Hive



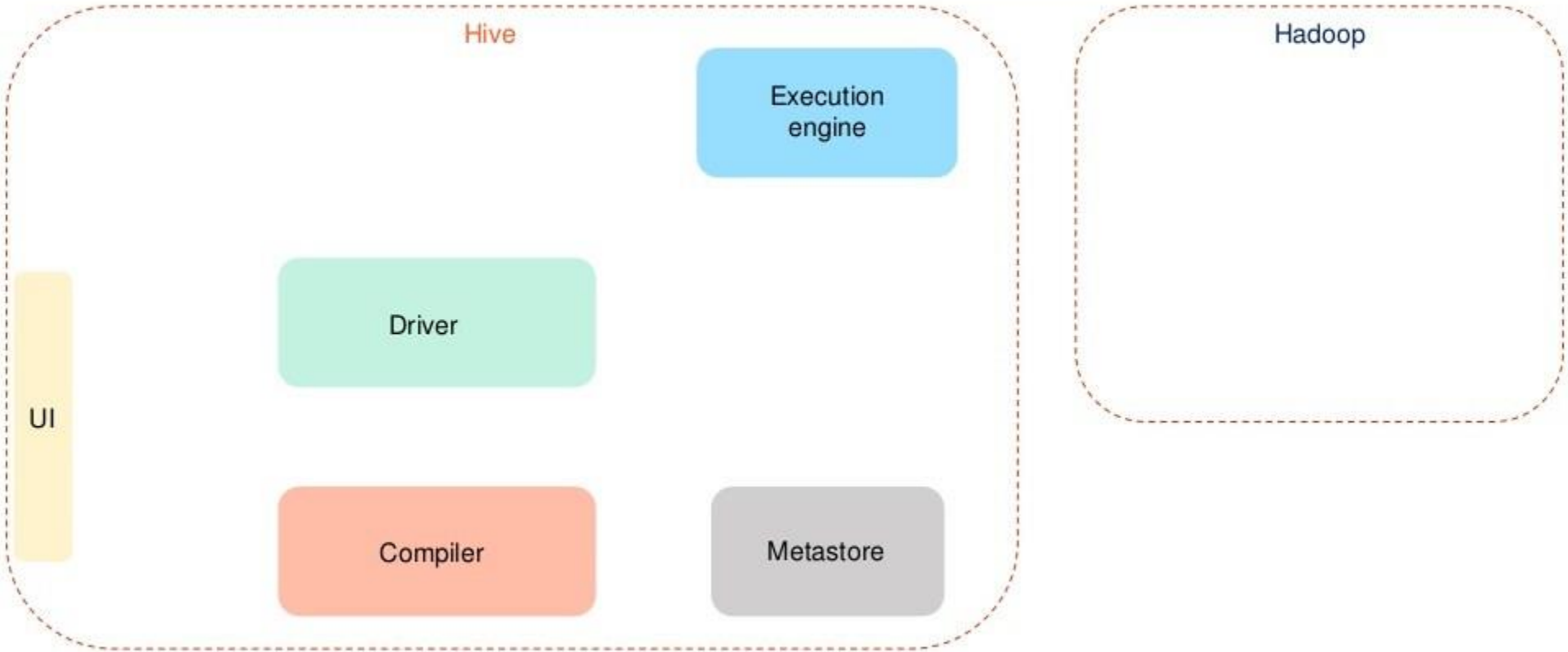
simplilearn

Data flow in Hive

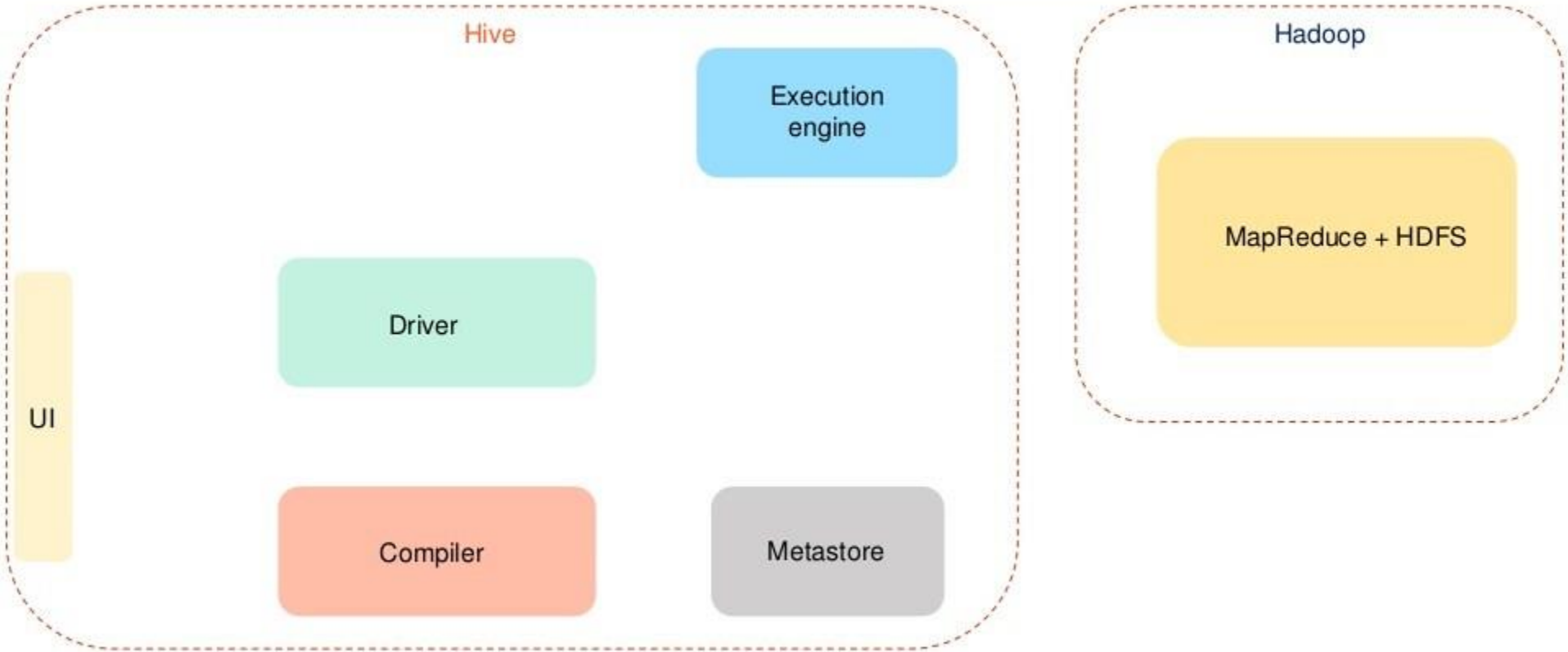
Data flow in Hive



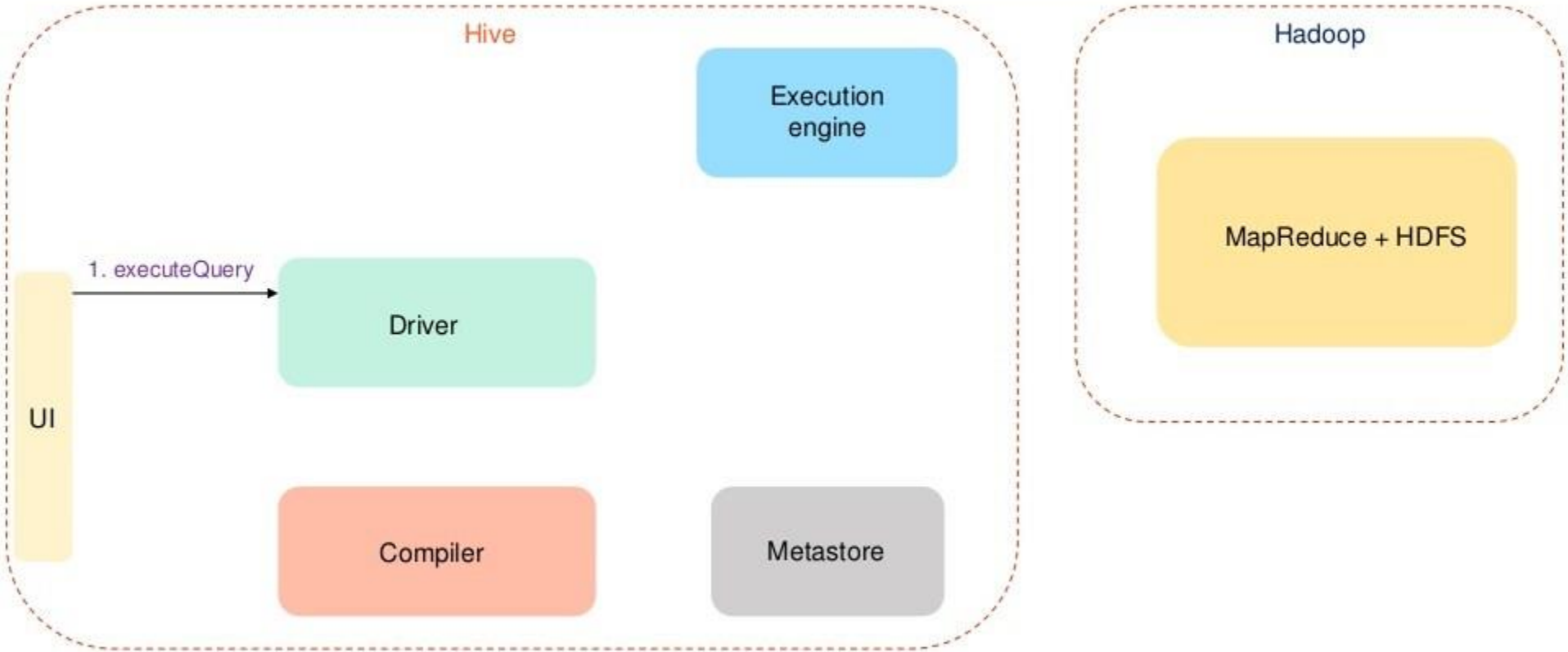
Data flow in Hive



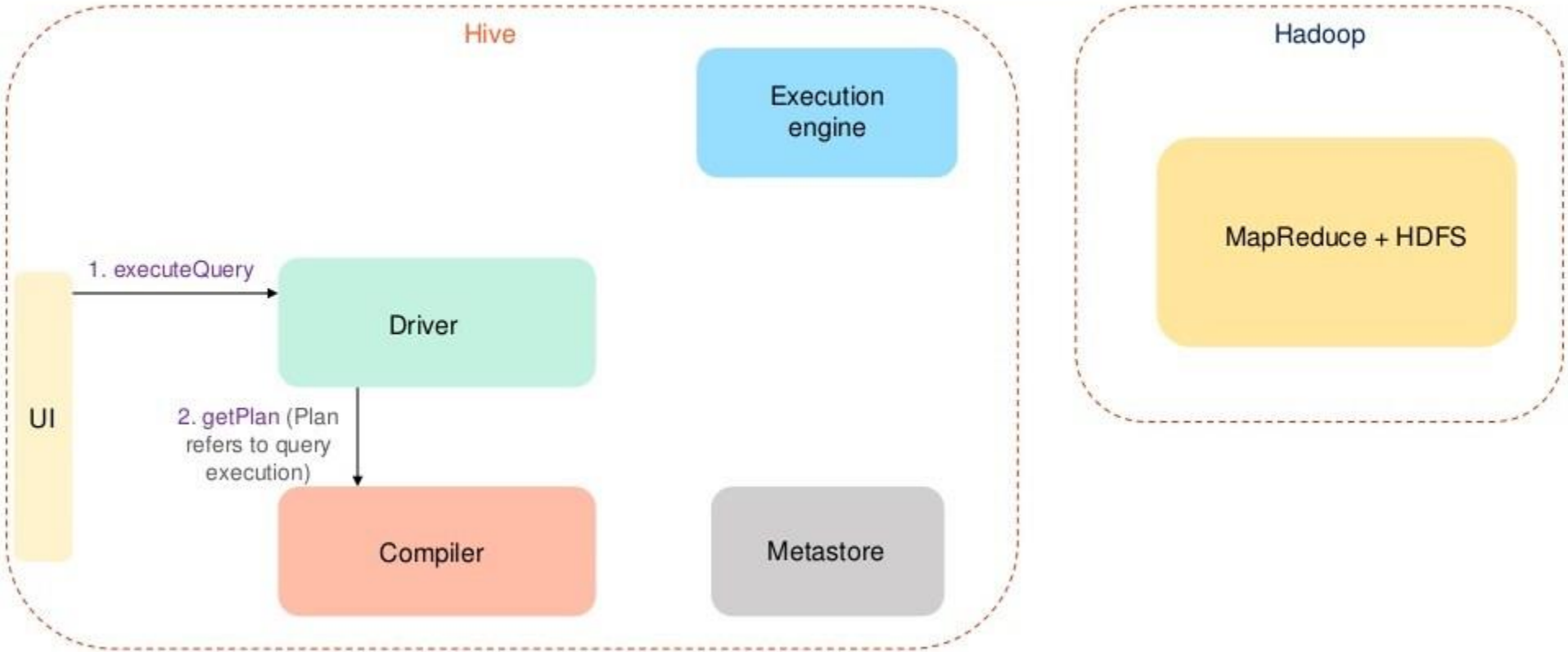
Data flow in Hive



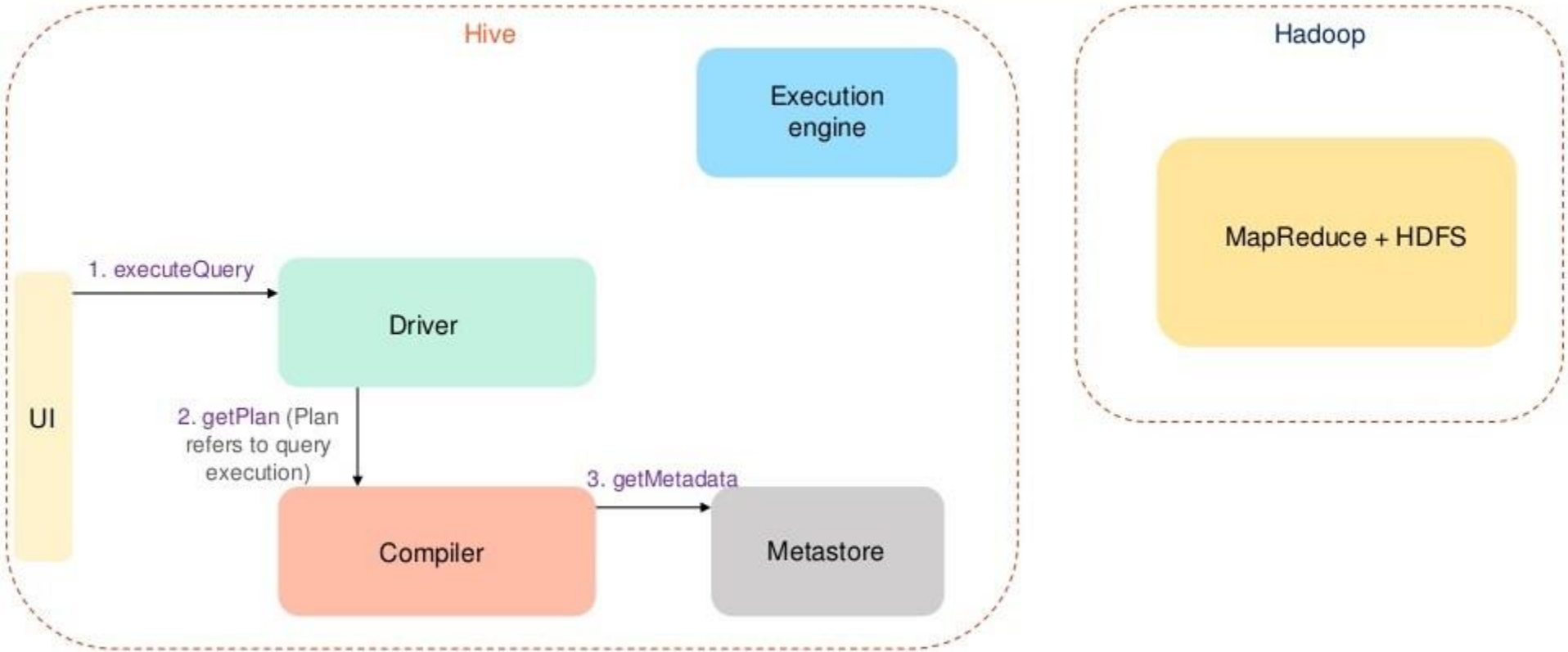
Data flow in Hive



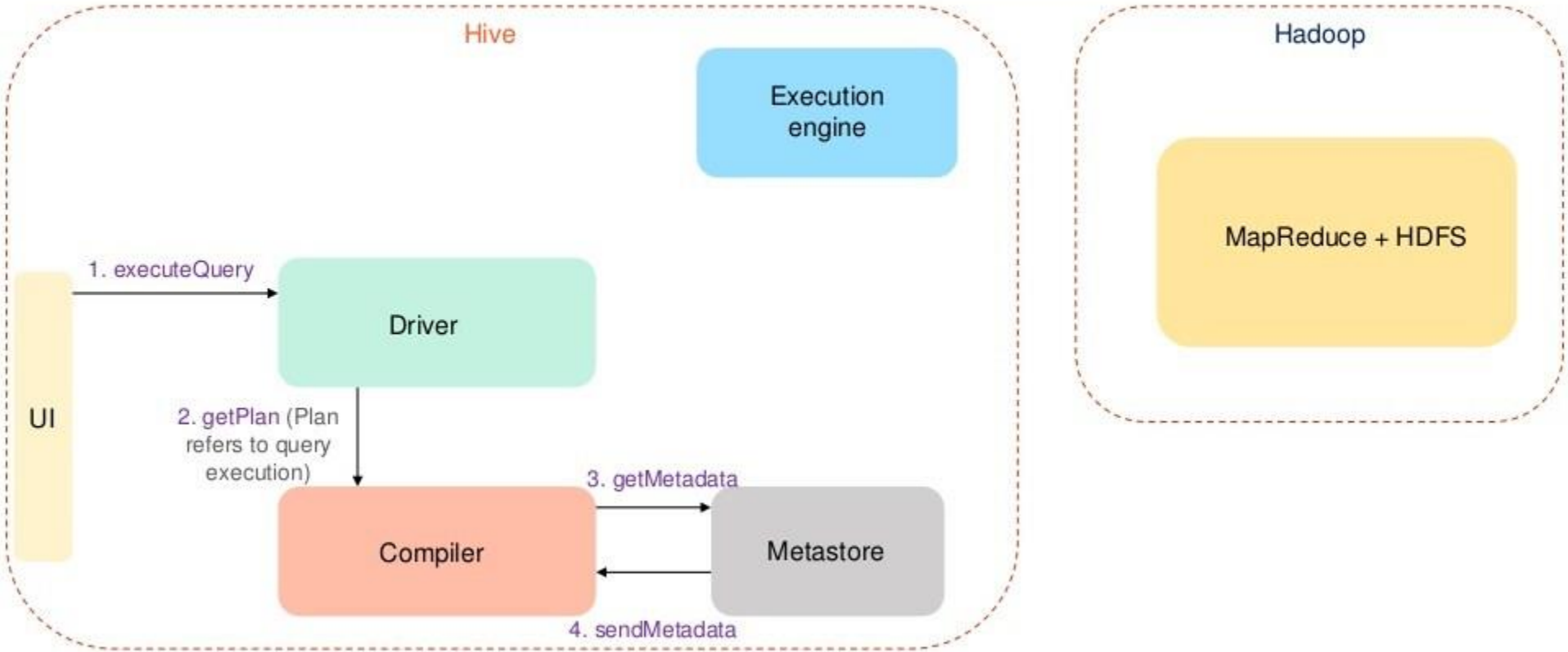
Data flow in Hive



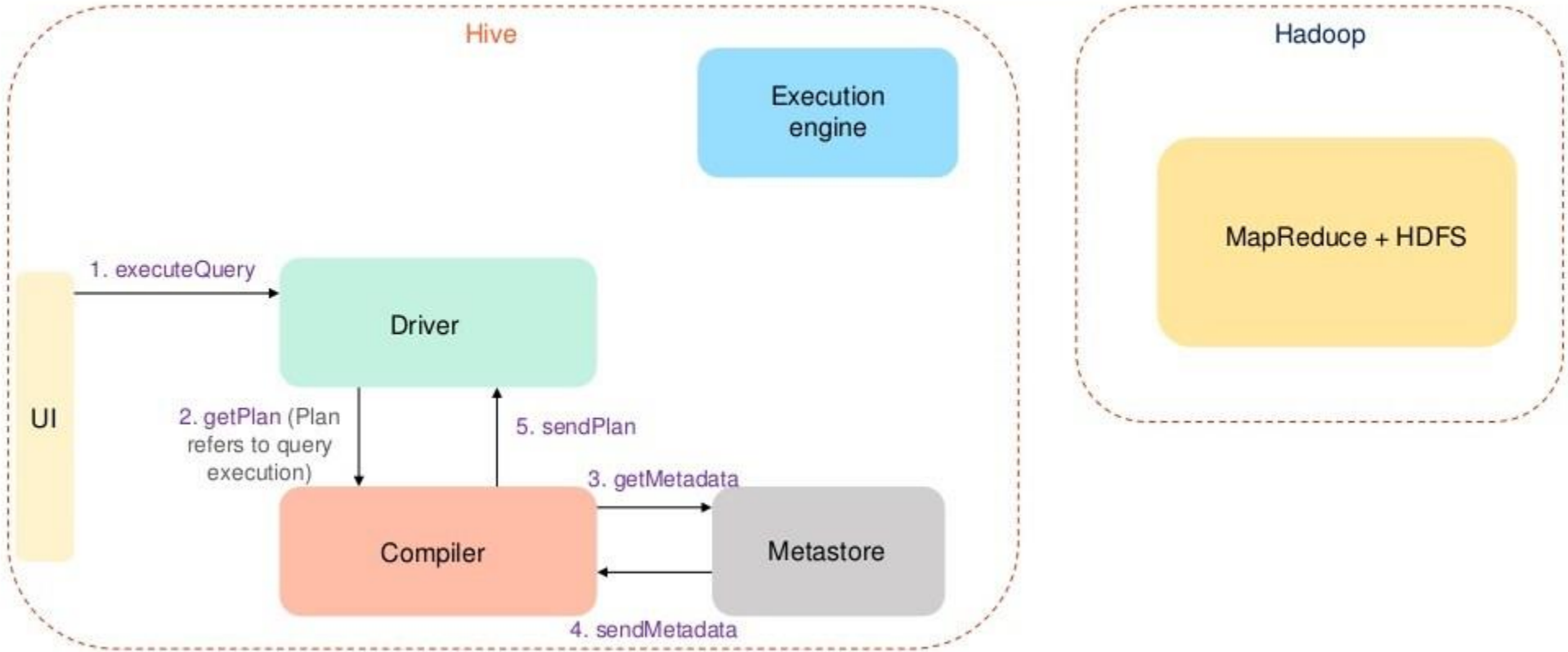
Data flow in Hive



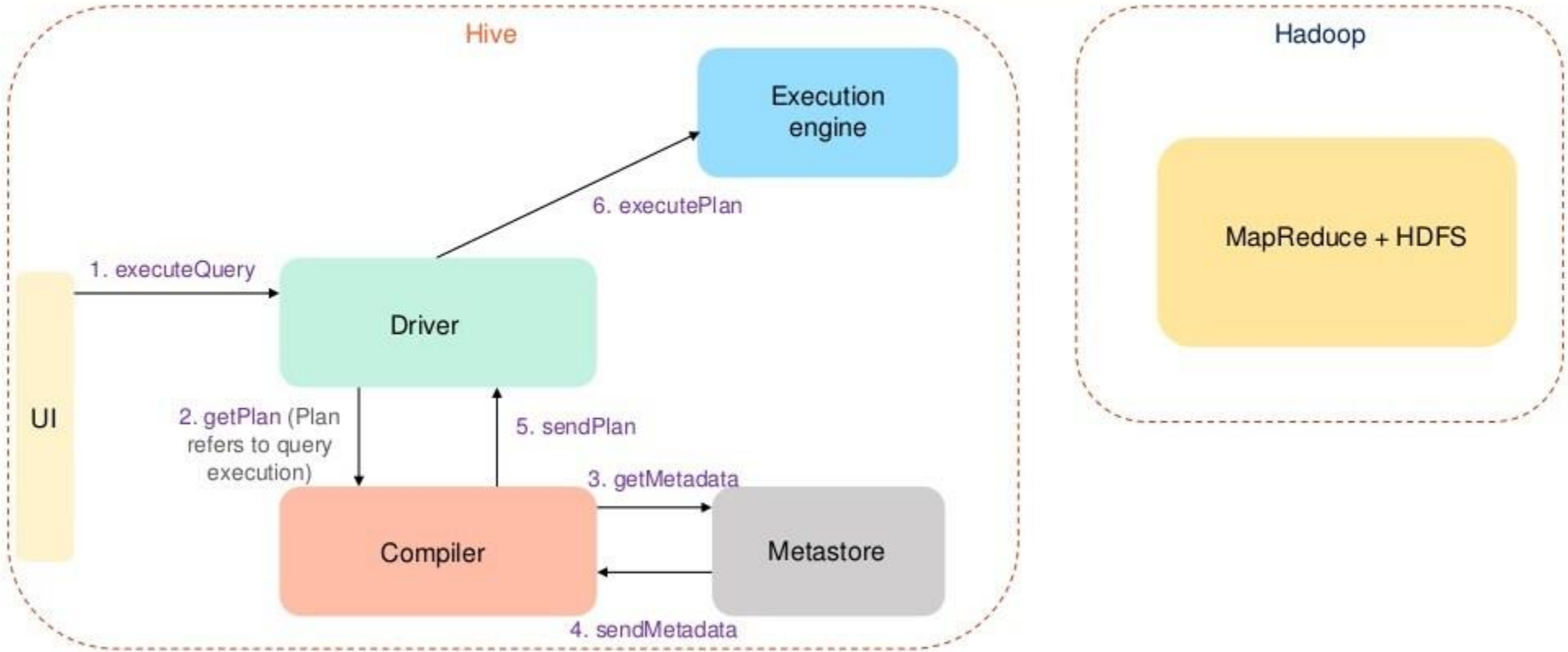
Data flow in Hive



Data flow in Hive

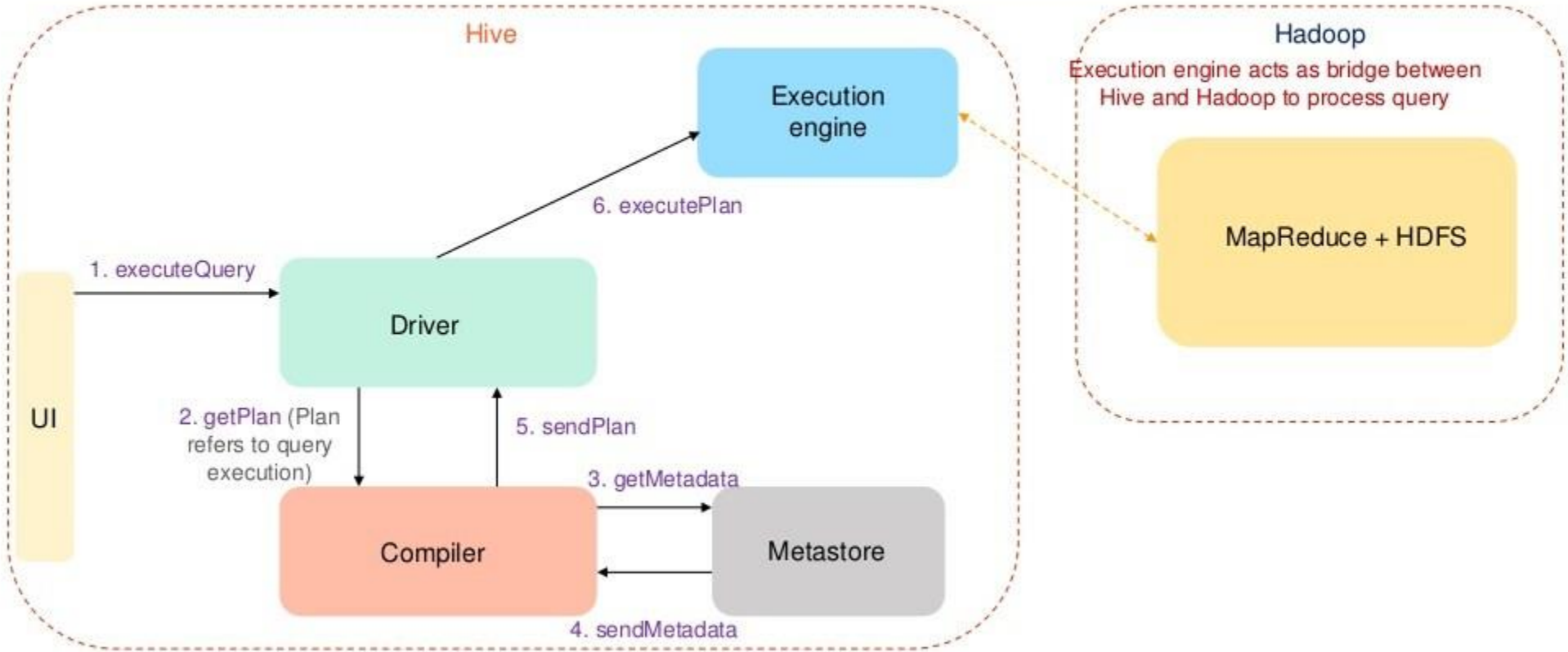


Data flow in Hive



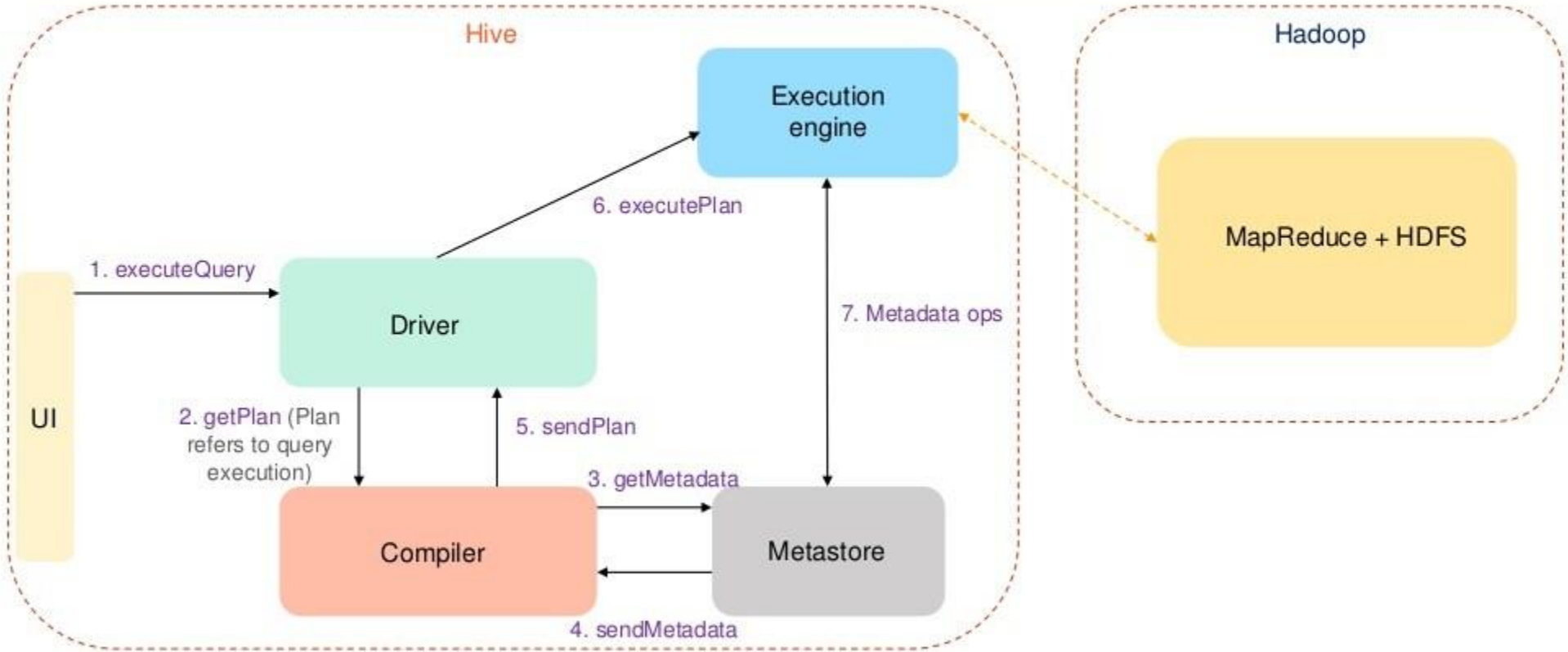
simplilearn

Data flow in Hive



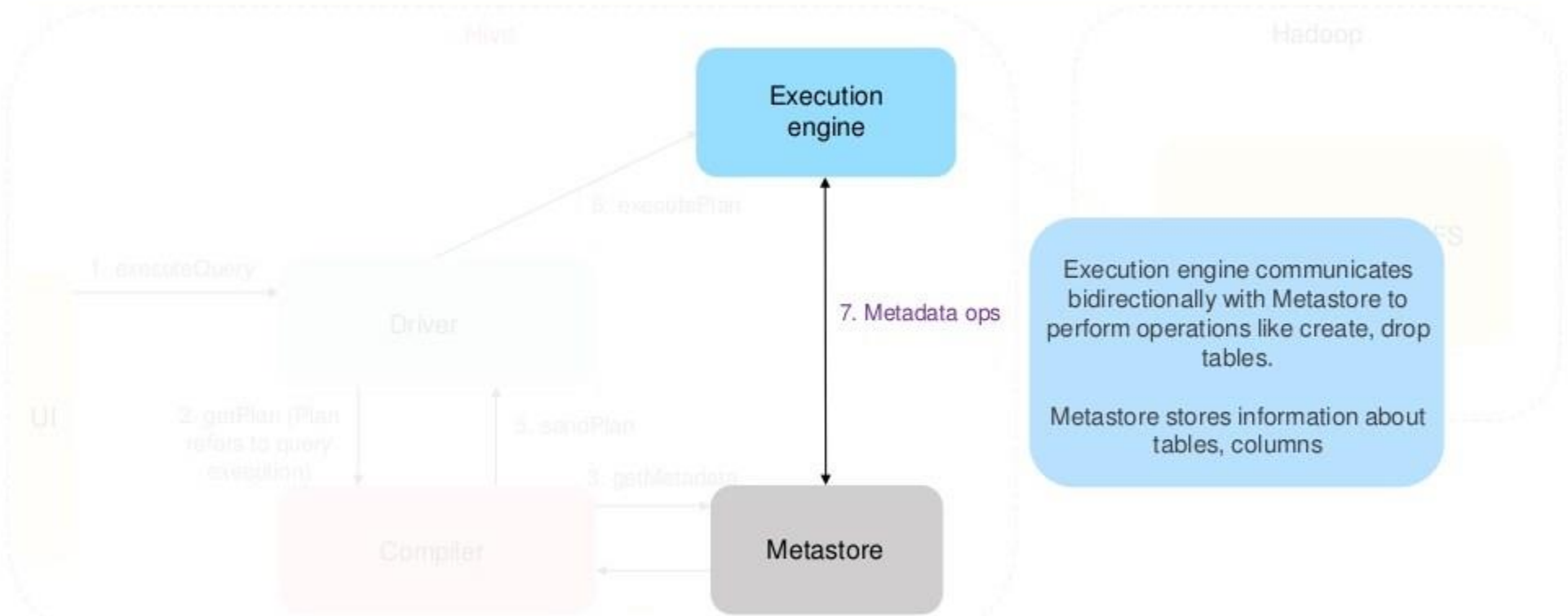
simplilearn

Data flow in Hive

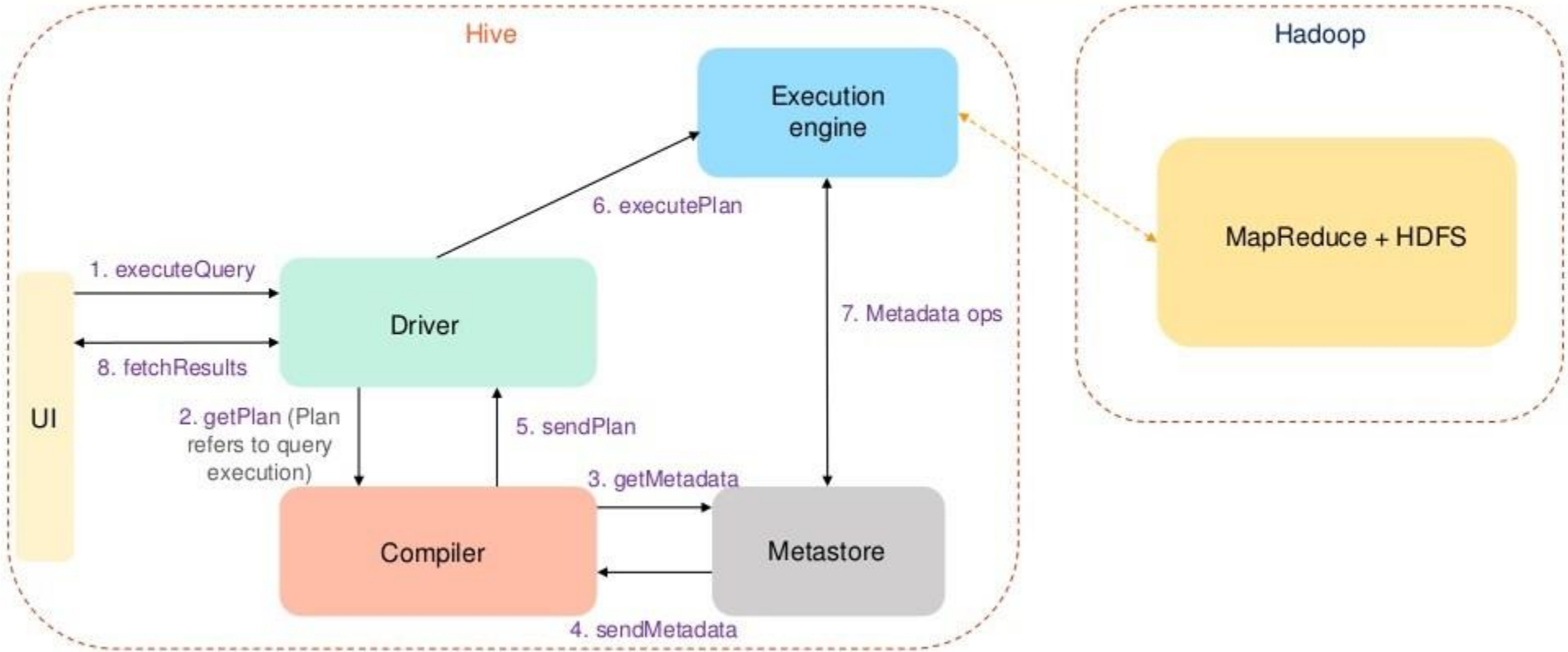


simplilearn

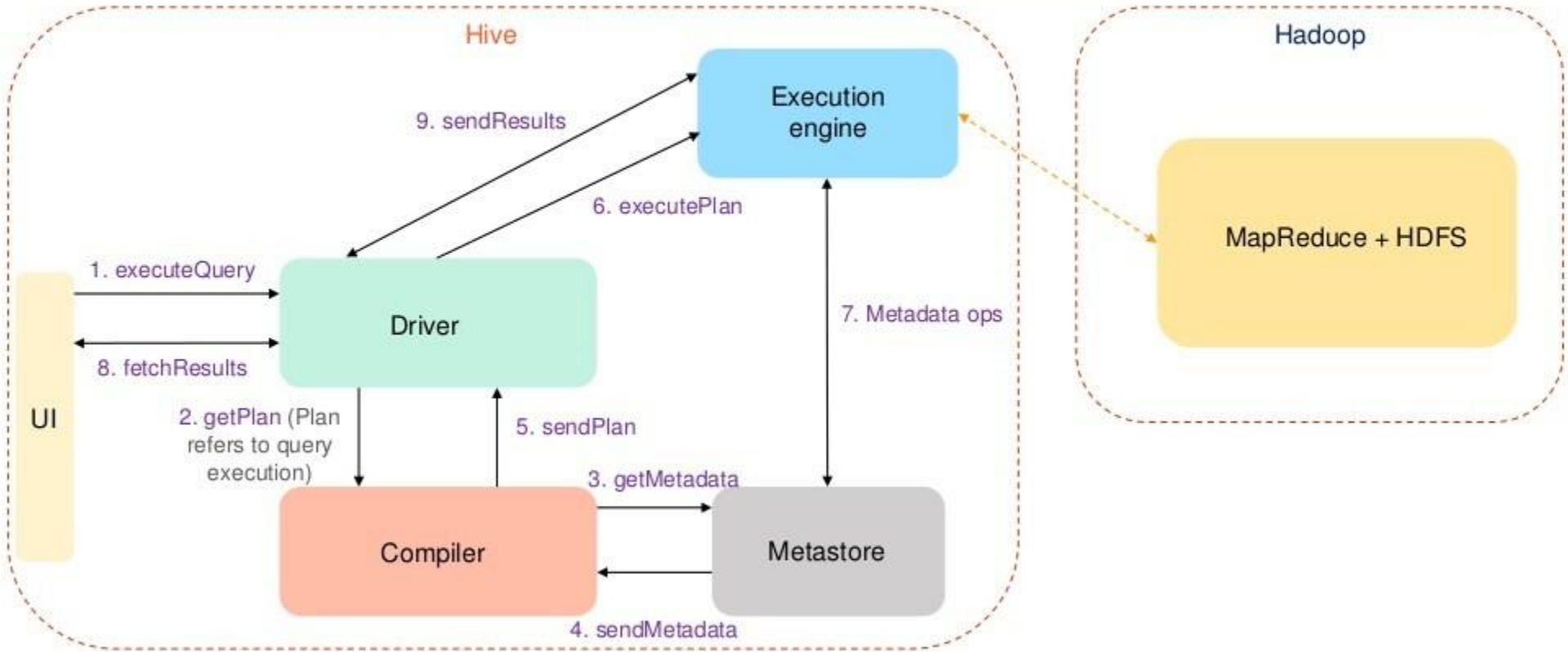
Data flow in Hive



Data flow in Hive



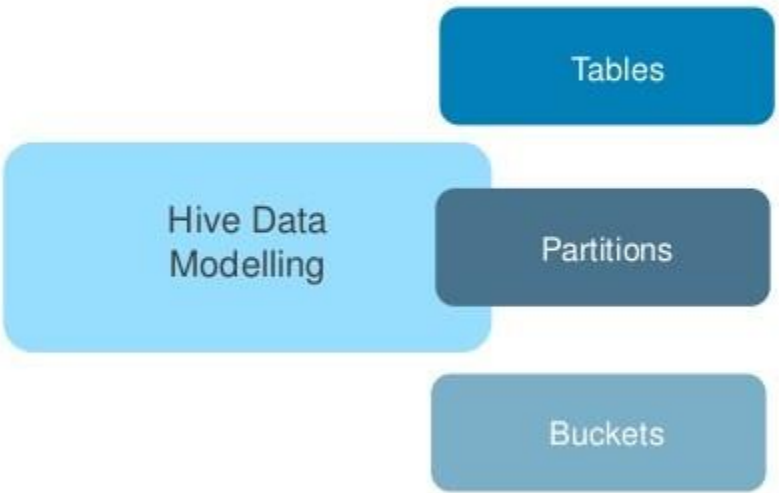
Data flow in Hive



simplilearn

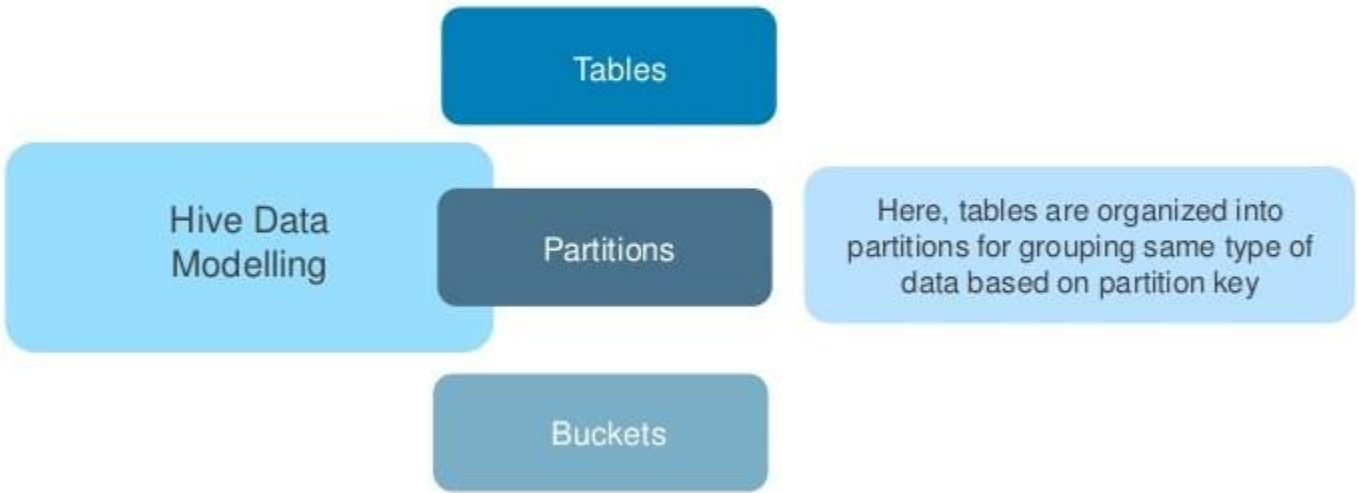
Hive Data Modeling

Hive Data Modeling



Hive Data Modeling

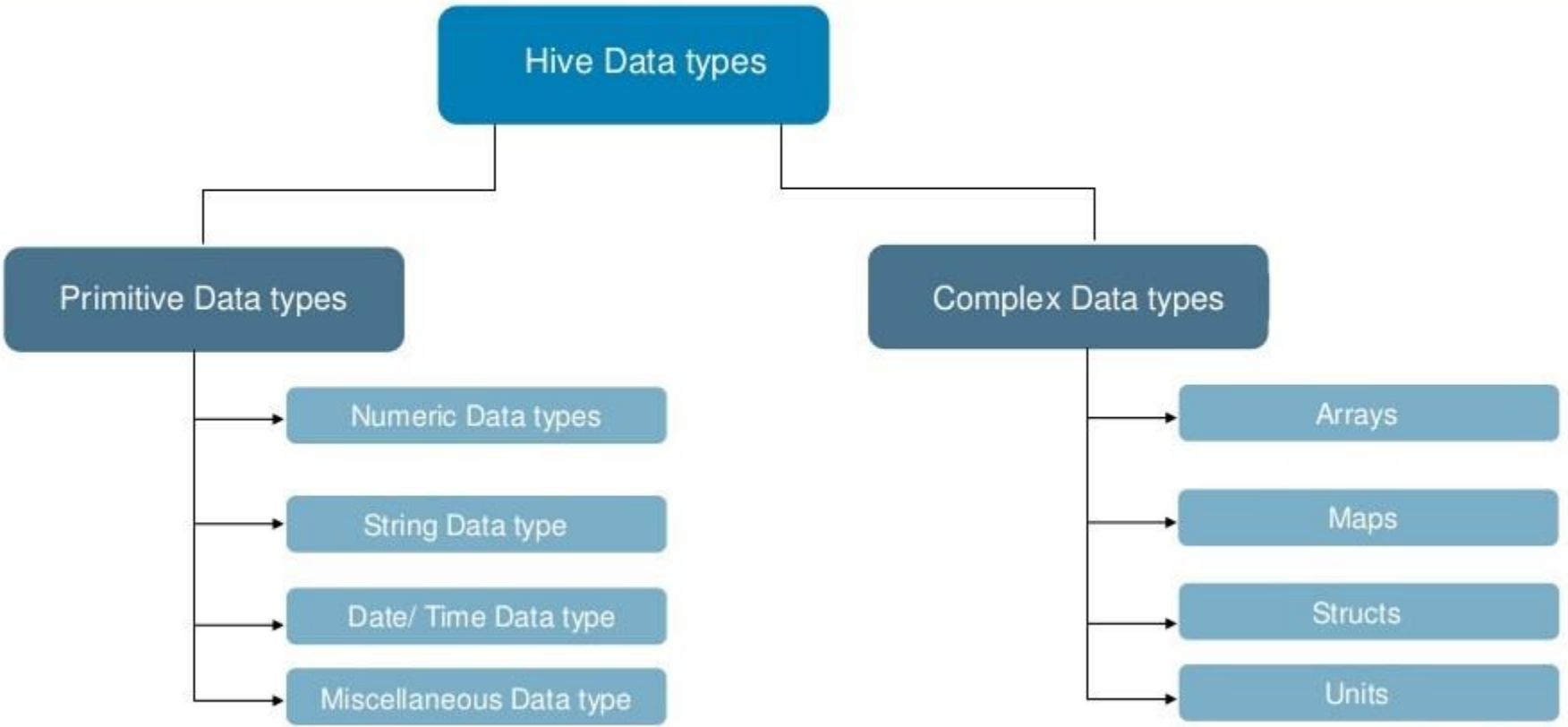
Hive Data Modeling



Hive Data Modeling

Hive Data types

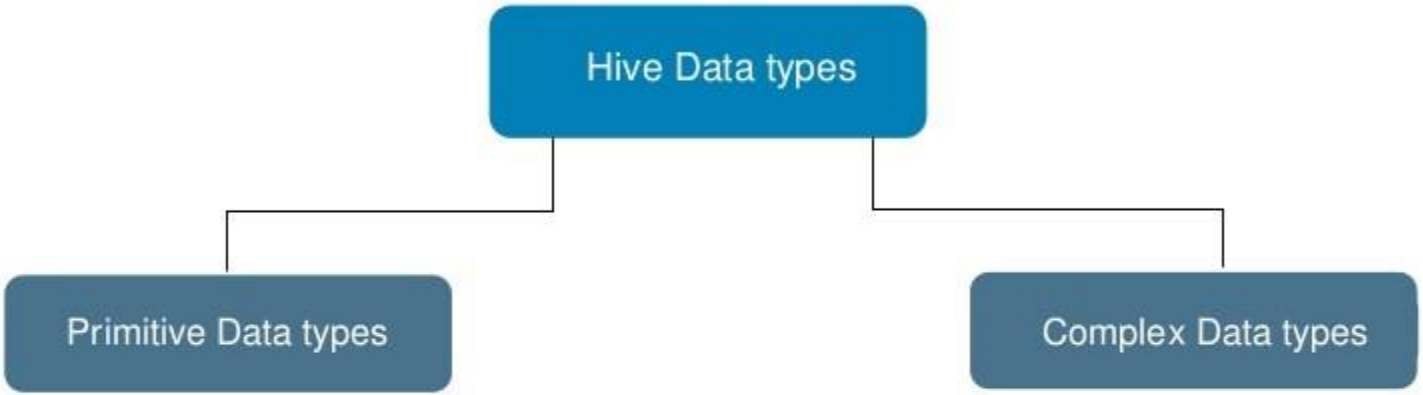
Hive Data types



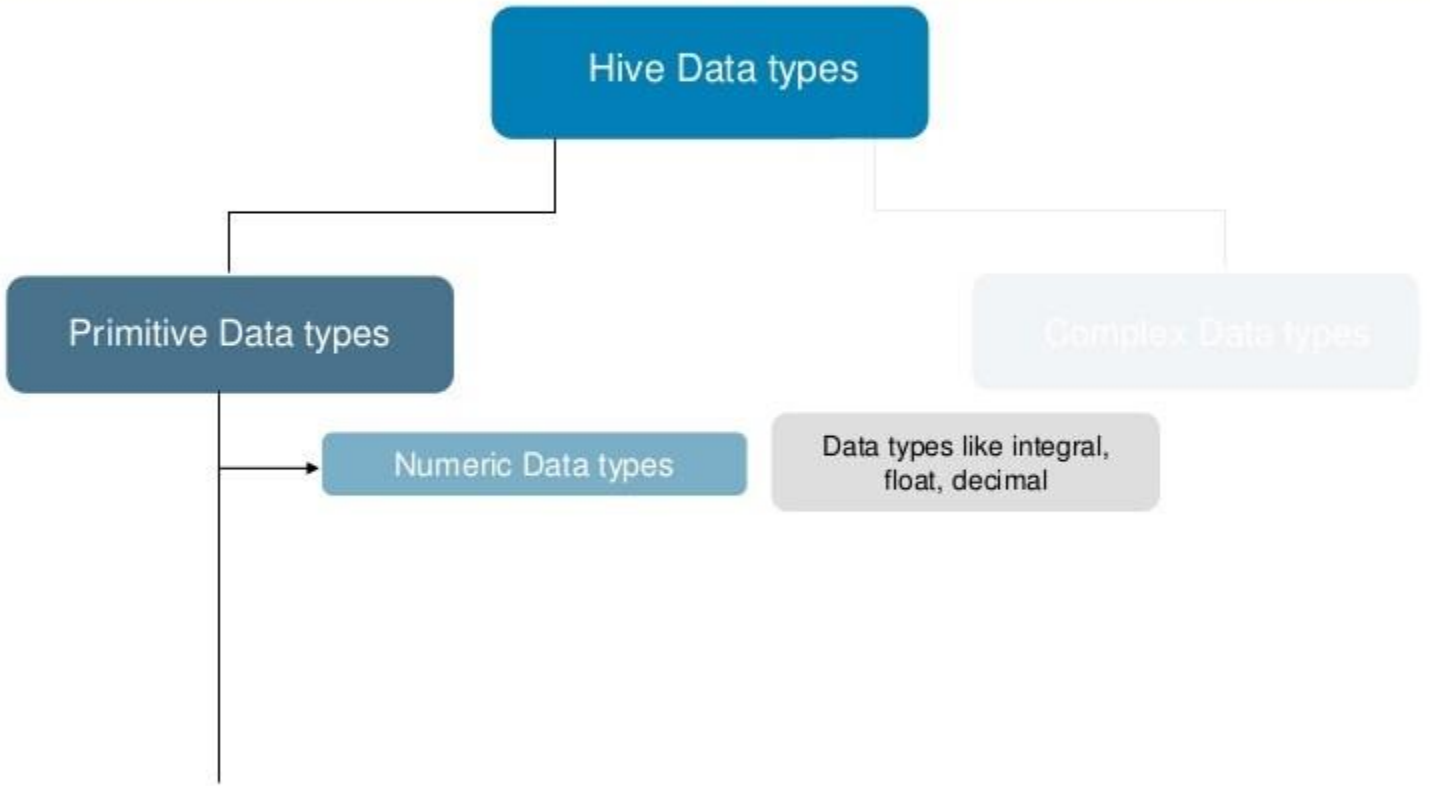
simplilearn

Hive Data types

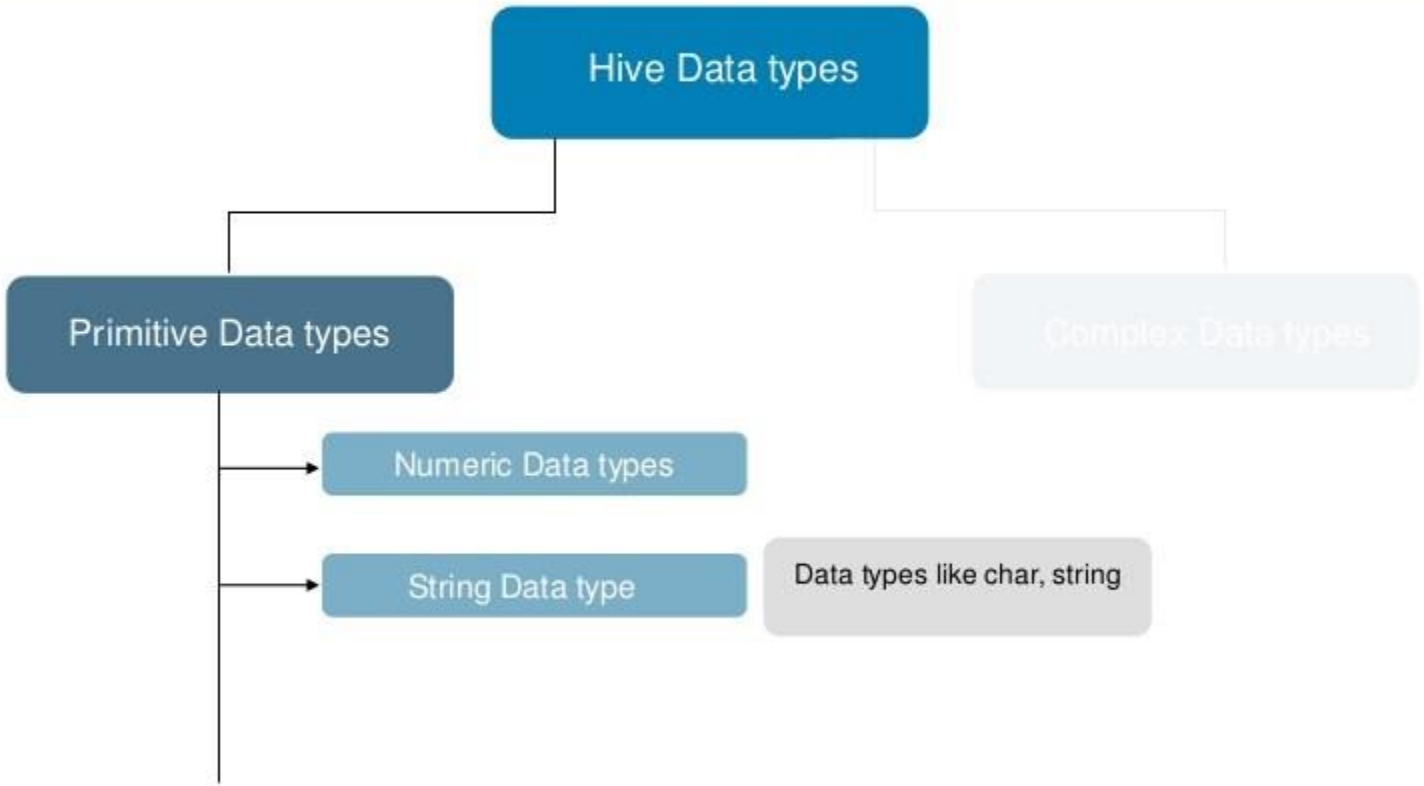
Hive Data types



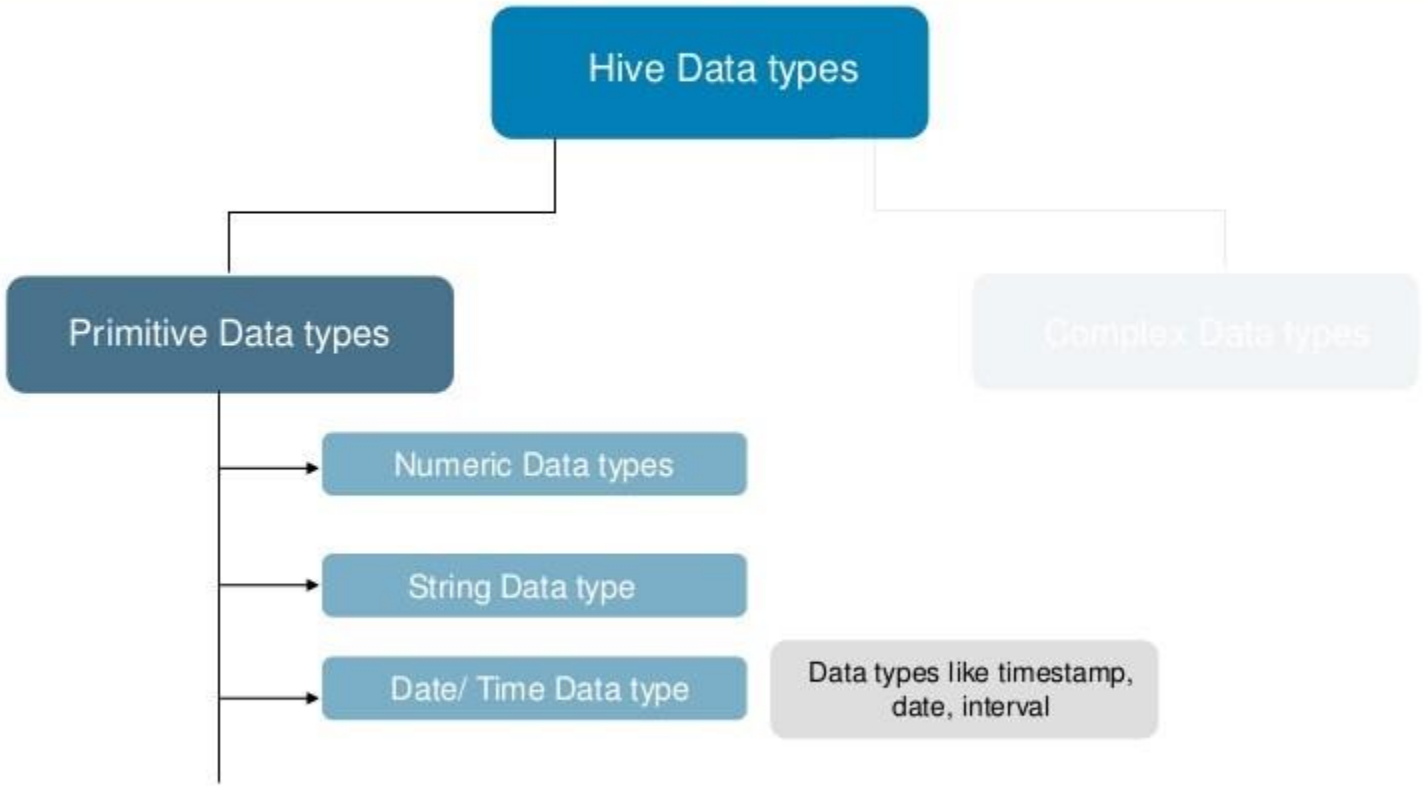
Hive Data types



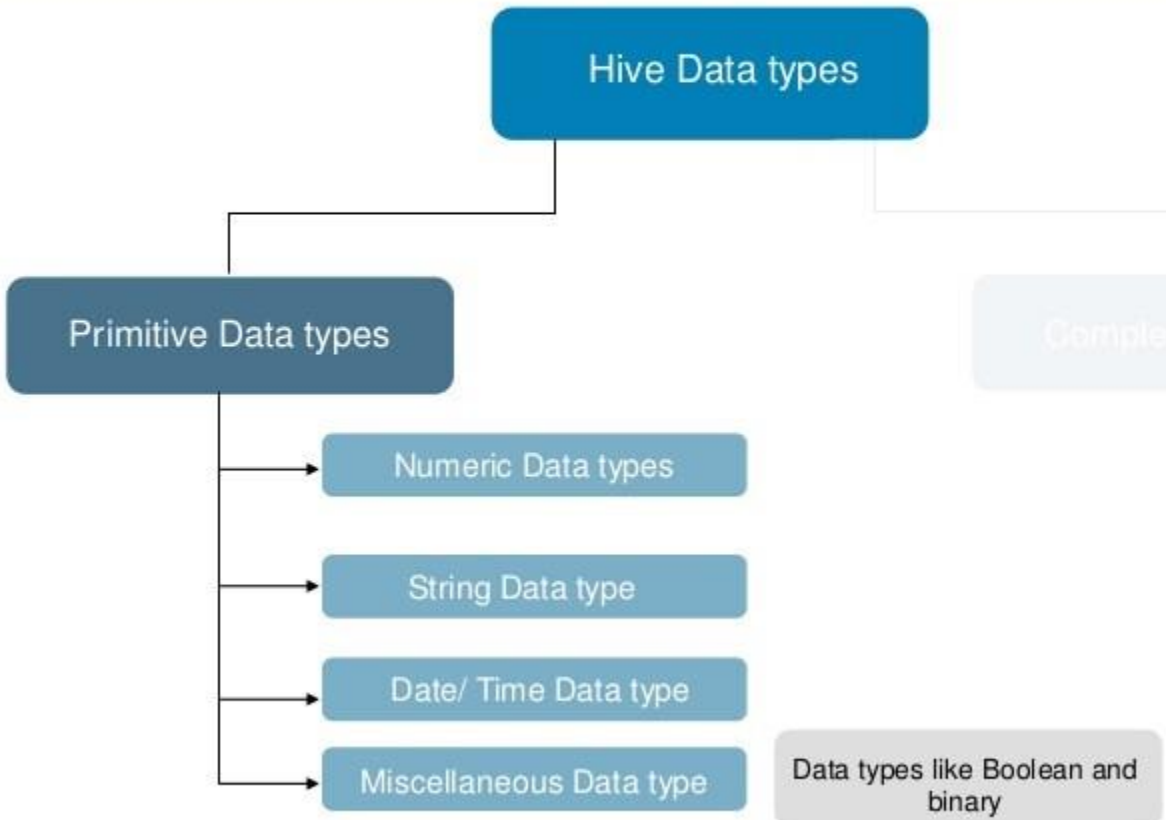
Hive Data types



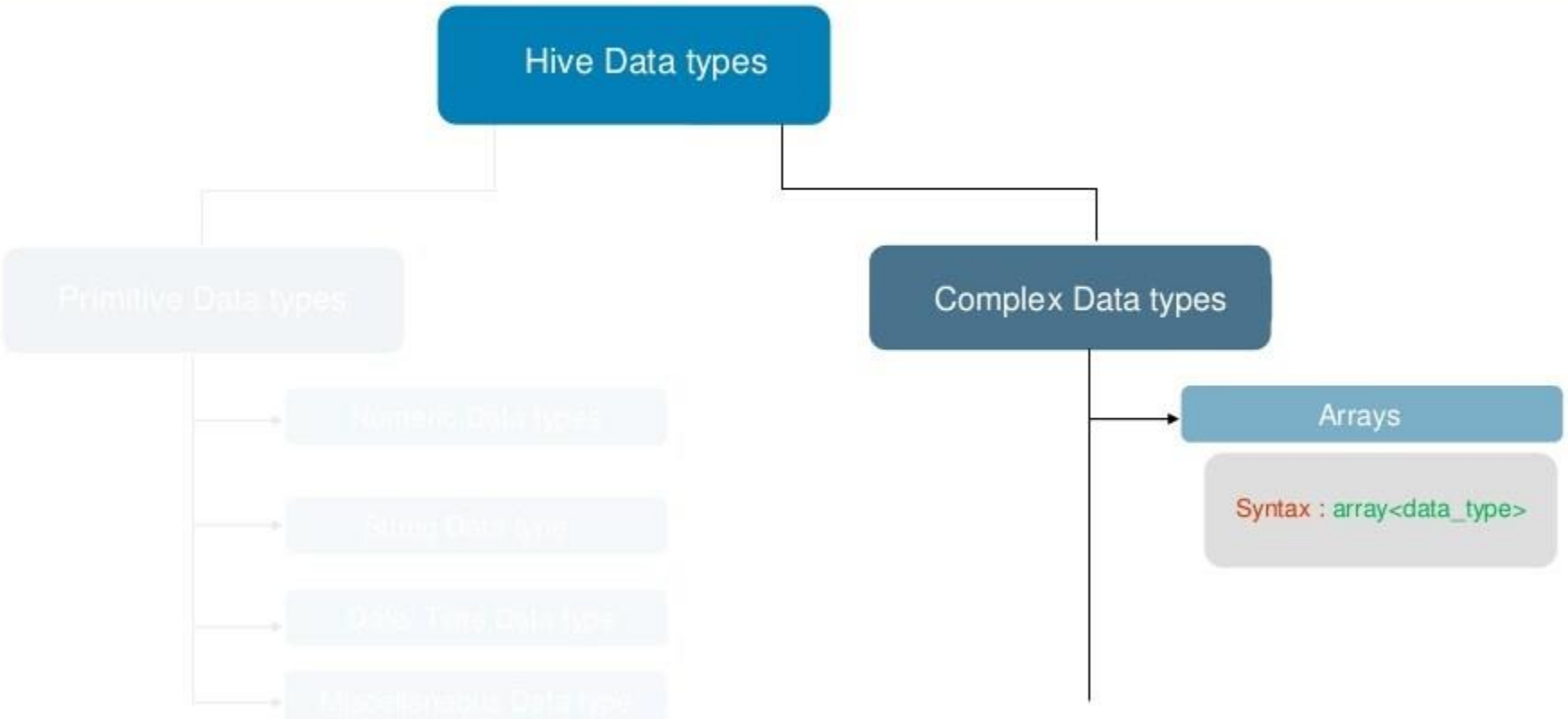
Hive Data types



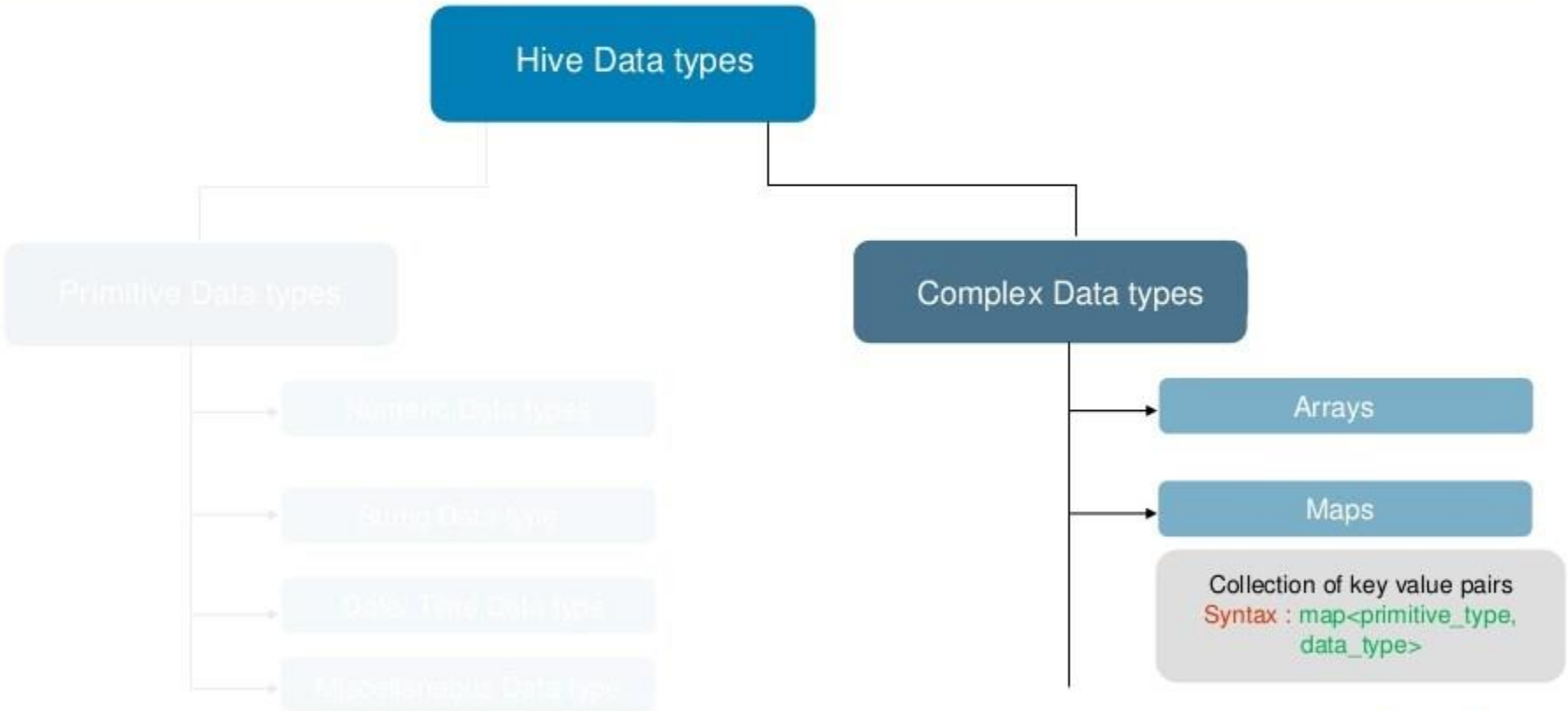
Hive Data types



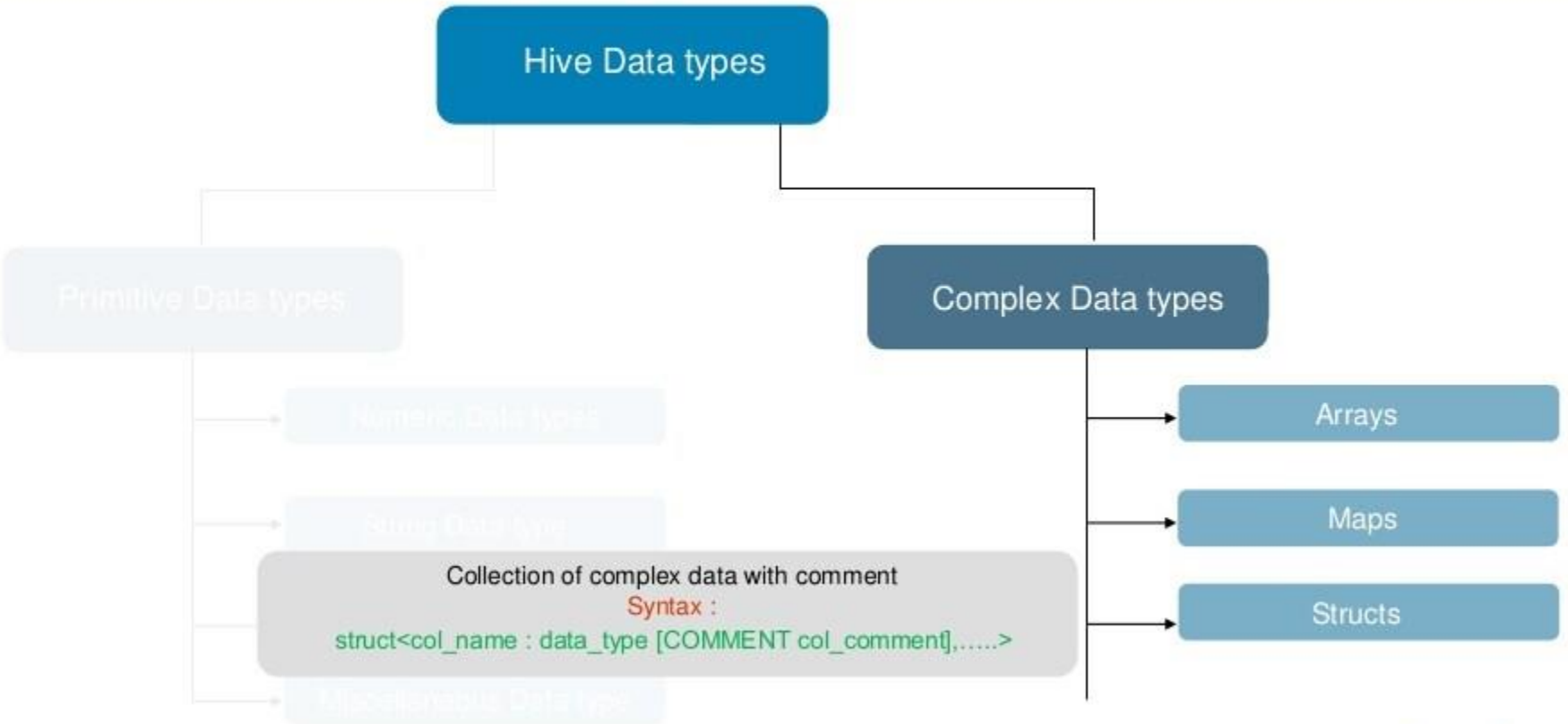
Hive Data types



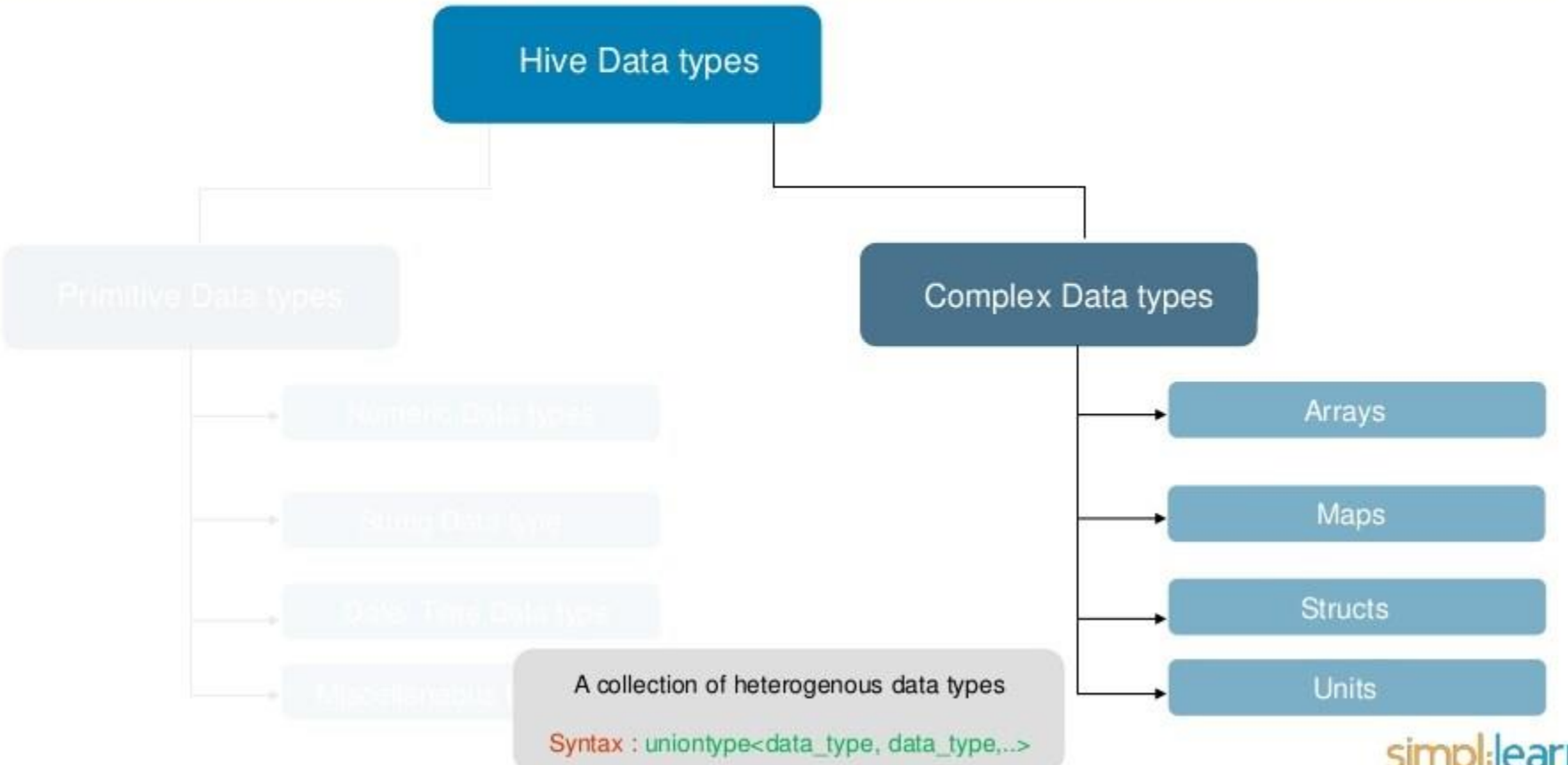
Hive Data types



Hive Data types



Hive Data types



Different modes of Hive

Different modes of Hive

Hive operates in two modes depending on the number and size of data nodes

Different modes of Hive

Hive operates in two modes depending on the number and size of data nodes

Local Mode

MapReduce Mode

Different modes of Hive

Hive operates in two modes depending on the number and size of data nodes

- Is used when Hadoop is having one data node and the data is small
- Processing will be very fast on smaller datasets which are present in local machine

Local Mode

MapReduce Mode

Different modes of Hive

Hive operates in two modes depending on the number and size of data nodes

Local Mode

MapReduce Mode

- Is used when Hadoop is having multiple data nodes and the data is spread across various data notes
- Processing large datasets can be more efficient using this mode

simplilearn

Difference between Hive and RDBMS

Difference between Hive and RDBMS

Hive	RDBMS
• Hive enforces schema on read	• RDBMS enforces schema on write

Difference between Hive and RDBMS

Hive	RDBMS
• Hive enforces schema on read	• RDBMS enforces schema on write
• Hive data size is in petabytes	• Data size is in terabytes

Difference between Hive and RDBMS

Hive	RDBMS
• Hive enforces schema on read	• RDBMS enforces schema on write
• Hive data size is in petabytes	• Data size is in terabytes
• Hive is based on the notion of write once and read many times	• RDBMS is based on the notion of read and write many times

Difference between Hive and RDBMS

Hive	RDBMS
• Hive enforces schema on read	• RDBMS enforces schema on write
• Hive data size is in petabytes	• Data size is in terabytes
• Hive is based on the notion of write once and read many times	• RDBMS is based on the notion of read and write many times
• Hive resembles a traditional database by supporting SQL but it is not a database. It is a data warehouse	• RDBMS is a type of database management system which is based on the relational model of data

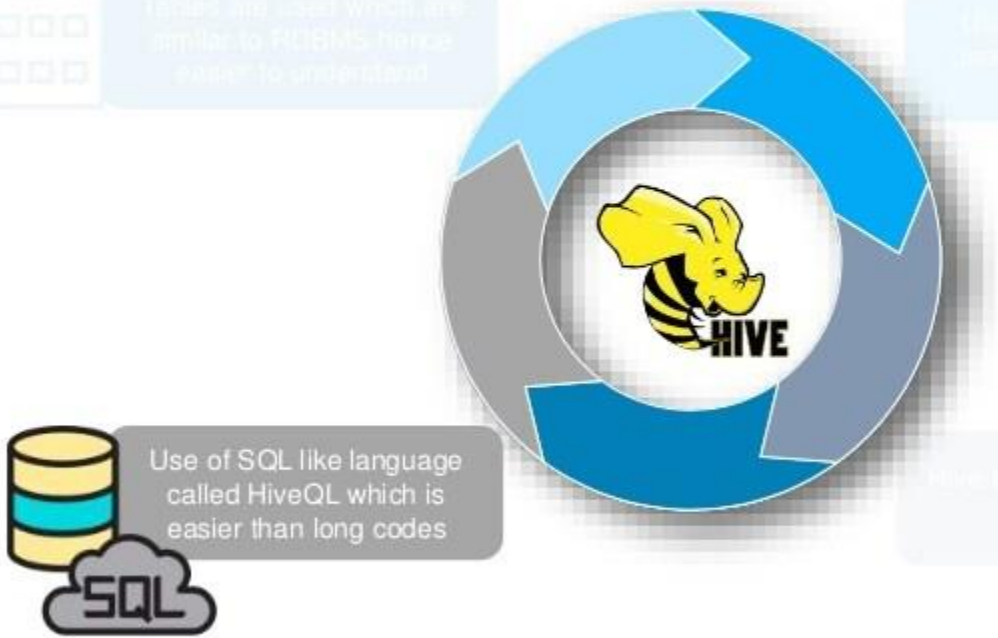
Difference between Hive and RDBMS

Hive	RDBMS
• Hive enforces schema on read	• RDBMS enforces schema on write
• Hive data size is in petabytes	• Data size is in terabytes
• Hive is based on the notion of write once and read many times	• RDBMS is based on the notion of read and write many times
• Hive resembles a traditional database by supporting SQL but it is not a database. It is a data warehouse	• RDBMS is a type of database management system which is based on the relational model of data
• Easily scalable at low cost	• Not scalable at low cost

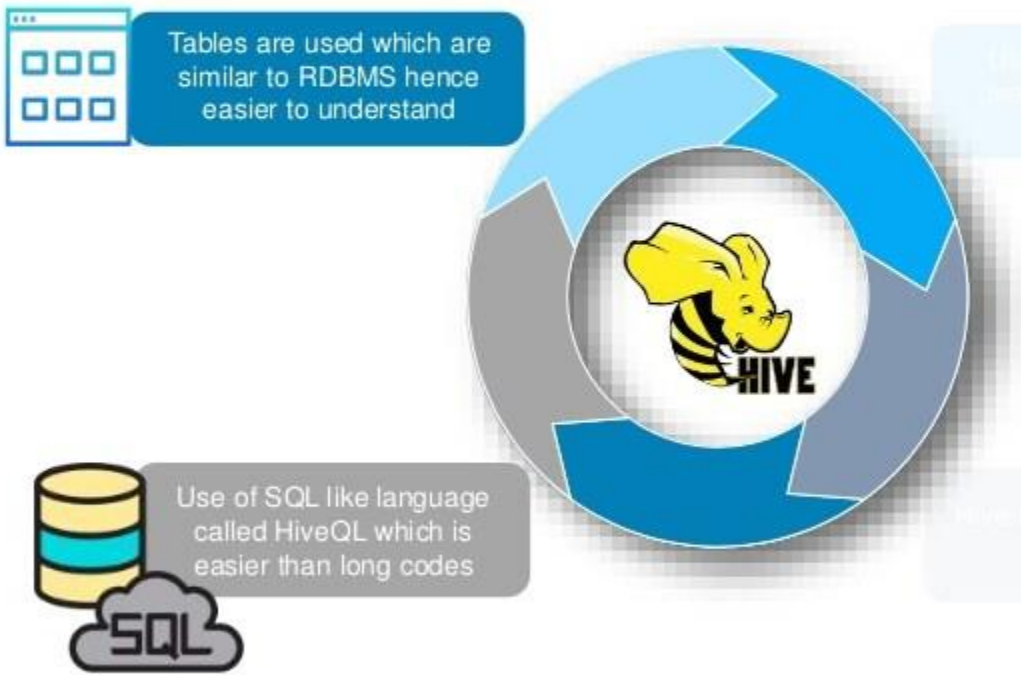
Features of Hive

Features of Hive

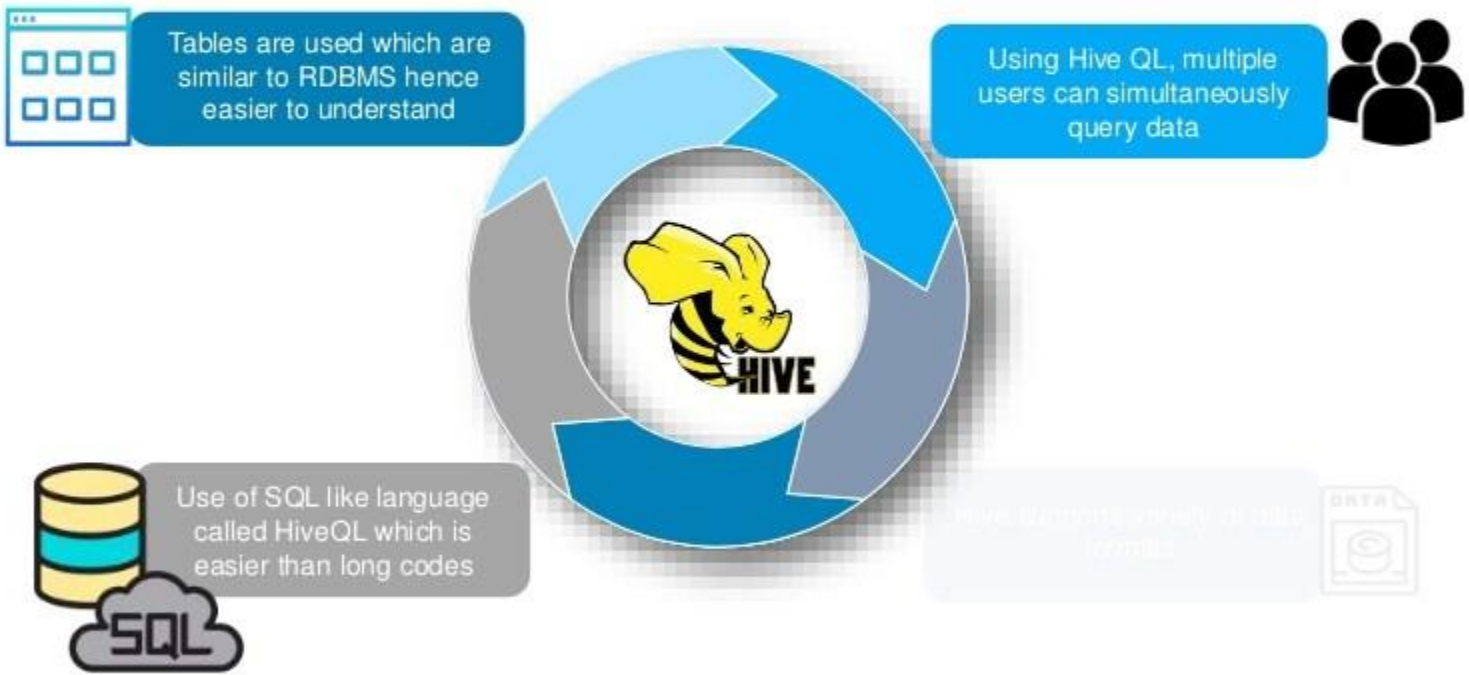
Features of Hive



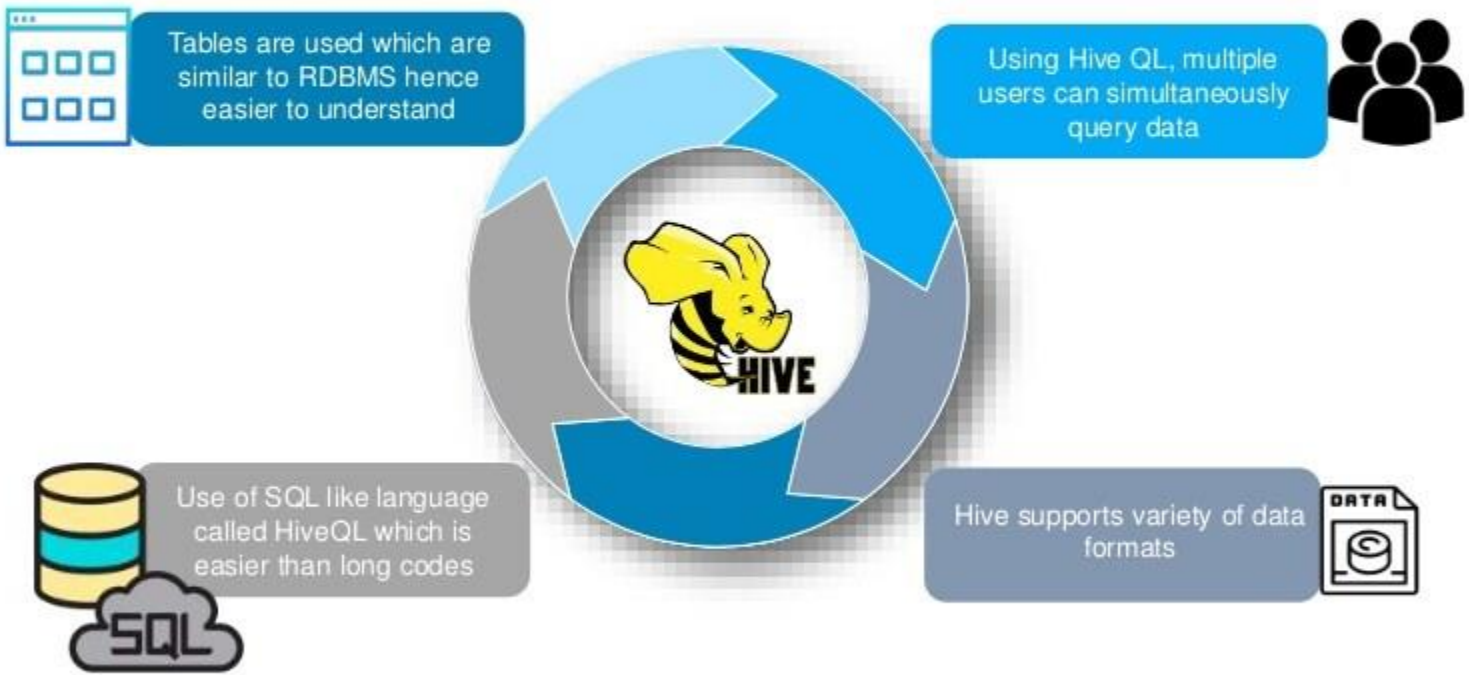
Features of Hive



Features of Hive



Features of Hive



Demo on HiveQL

References

1. https://www.tutorialspoint.com/hive

2. https://www.youtube.com/watch?v=rr17cbPGWGA


THANK YOU
For more information, visit
www.simplilearn.com
simplilearn