

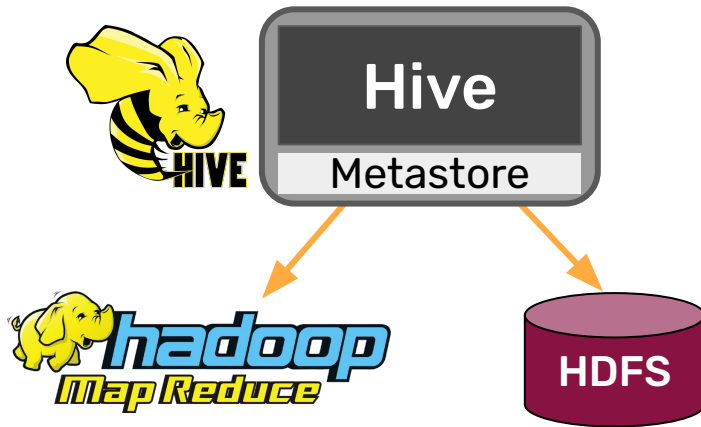


Hive DML

Драль Алексей, study@bigdatateam.org
CEO at BigData Team, <https://bigdatateam.org>
<https://www.facebook.com/bigdatateam>



Hive DDL, Hive DML и HiveQL



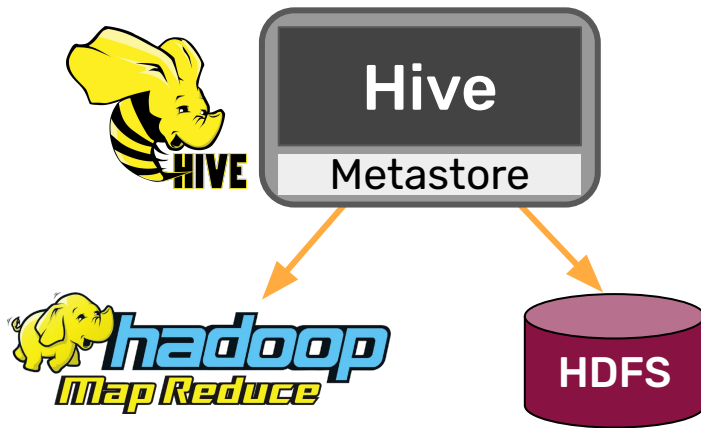


Hive DDL, Hive DML и HiveQL





Q&A: HiveQL, реализация

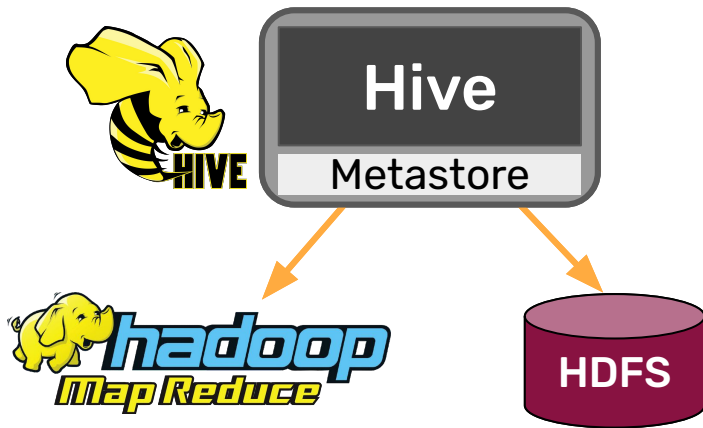


MapReduce (?)

- ▶ SELECT .. FROM
- ▶ WHERE
- ▶ GROUP BY + HAVING
- ▶ JOIN
- ▶ ORDER BY / **SORT BY**



Q&A: HiveQL, реализация

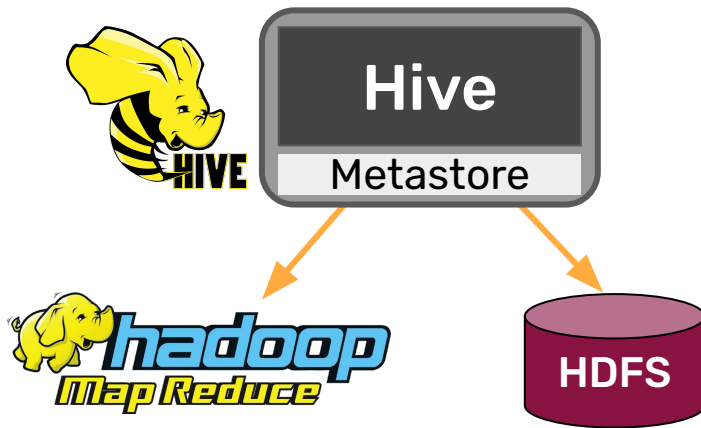


MapReduce (?)

- ▶ SELECT .. FROM [<--Map]
- ▶ WHERE
- ▶ GROUP BY + HAVING
- ▶ JOIN
- ▶ ORDER BY / **SORT BY**



Q&A: HiveQL, реализация

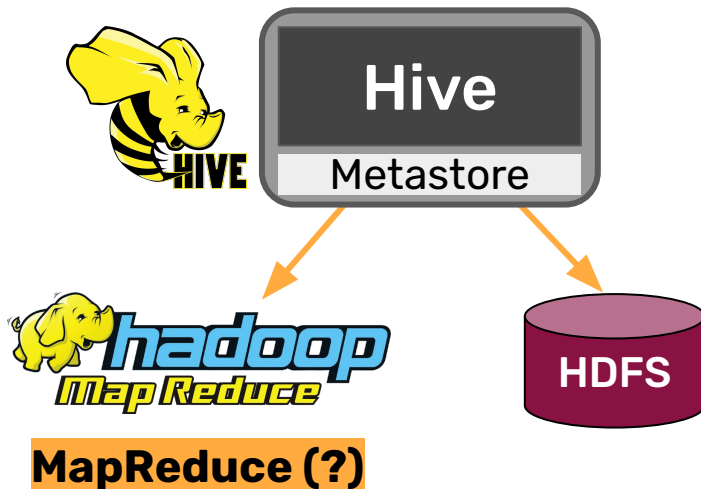


MapReduce (?)

- ▶ SELECT .. FROM [<--Map]
- ▶ WHERE [<--Map]
- ▶ GROUP BY + HAVING
- ▶ JOIN
- ▶ ORDER BY / **SORT BY**



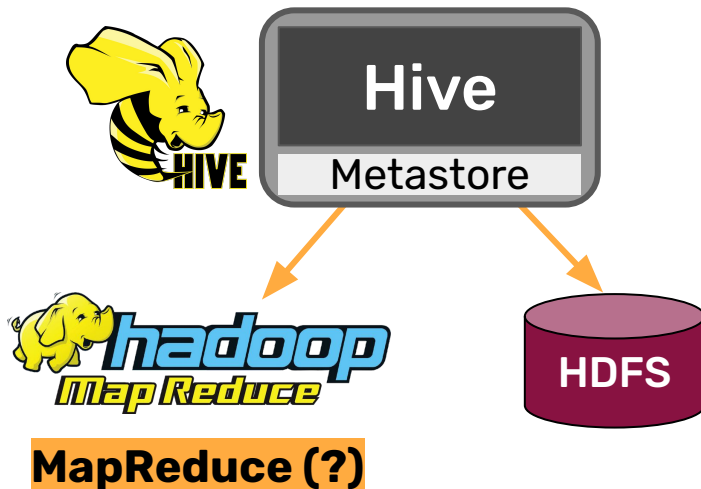
Q&A: HiveQL, реализация



- ▶ SELECT .. FROM [**<--Map**]
- ▶ WHERE [**<--Map**]
- ▶ GROUP BY [**<--Shuffle & Sort**] + HAVING [**<--Reduce**]
- ▶ JOIN
- ▶ ORDER BY / **SORT BY**



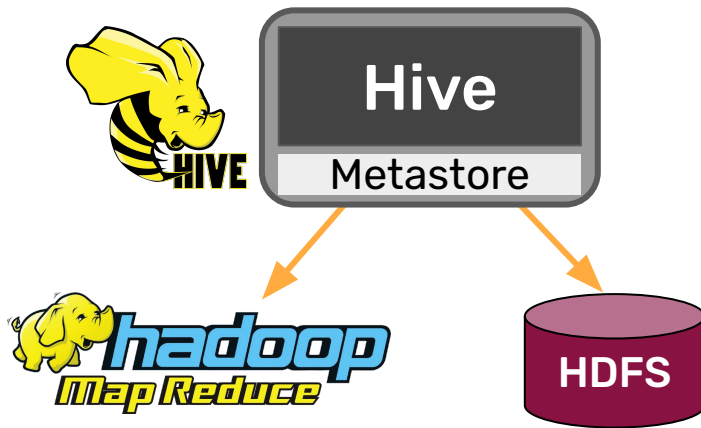
Q&A: HiveQL, реализация



- ▶ SELECT .. FROM [**<--Map**]
- ▶ WHERE [**<--Map**]
- ▶ GROUP BY [**<--Shuffle & Sort**] + HAVING [**<--Reduce**]
- ▶ JOIN [**<--Map / Reduce "-side"**]
- ▶ ORDER BY / **SORT BY**

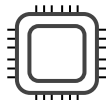


Q&A: HiveQL, реализация



MapReduce (?)

- ▶ SELECT .. FROM [**<--Map**]
- ▶ WHERE [**<--Map**]
- ▶ GROUP BY [**<--Shuffle & Sort**] + HAVING [**<--Reduce**]
- ▶ JOIN [**<--Map / Reduce "-side"**]
- ▶ ORDER BY / **SORT BY**





Explain B HiveQL

EXPLAIN

```
FROM src
INSERT OVERWRITE TABLE dest_g1
SELECT src.key, sum(substr(src.value,4))
GROUP BY src.key;
```



Explain B HiveQL

EXPLAIN

```
FROM src
INSERT OVERWRITE TABLE dest_g1
SELECT src.key, sum(substr(src.value,4))
GROUP BY src.key;
```

(1) The Abstract Syntax Tree

ABSTRACT SYNTAX TREE:

```
(TOK_QUERY (TOK_FROM (TOK_TABREF src))
...)
```



Explain B HiveQL

EXPLAIN

```
FROM src
INSERT OVERWRITE TABLE dest_g1
SELECT src.key, sum(substr(src.value,4))
GROUP BY src.key;
```

(1) The Abstract Syntax Tree

ABSTRACT SYNTAX TREE:

```
(TOK_QUERY (TOK_FROM (TOK_TABREF src))
...)
```

(2) The Dependency Graph

STAGE DEPENDENCIES:

Stage-1 is a root stage

Stage-2 depends on stages: Stage-1

Stage-0 depends on stages: Stage-2



Explain B HiveQL

EXPLAIN

```
FROM src
INSERT OVERWRITE TABLE dest_g1
SELECT src.key, sum(substr(src.value,4))
GROUP BY src.key;
```

(2) The Dependency Graph

STAGE DEPENDENCIES:

Stage-1 is a root stage

Stage-2 depends on stages: Stage-1

Stage-0 depends on stages: Stage-2

(1) The Abstract Syntax Tree

ABSTRACT SYNTAX TREE:

```
(TOK_QUERY (TOK_FROM (TOK_TABREF src))
...)
```

(3) The plans of each Stage

STAGE PLANS:

Stage: Stage-1

Map Reduce

Alias -> Map Operator Tree:

src

Reduce Output Operator

key expressions:

expr: key

type: string

sort order: +