

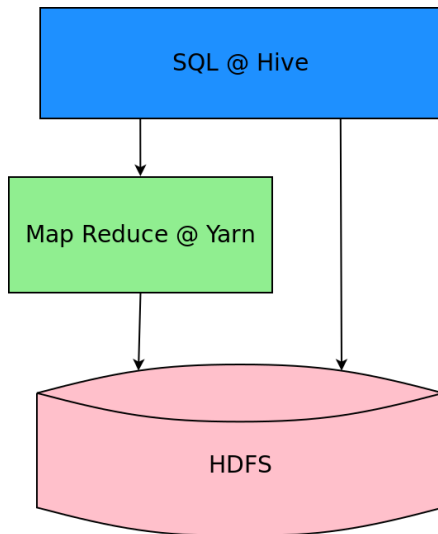
Hadoop

Jakub Podeszwik

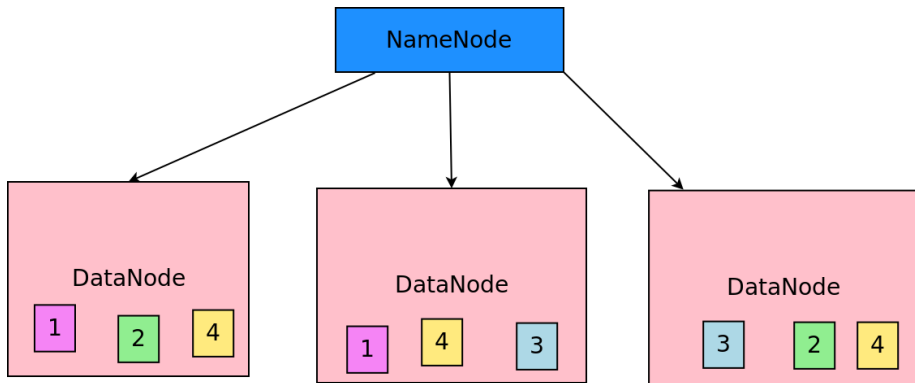
infoShare Academy

20.03.2019

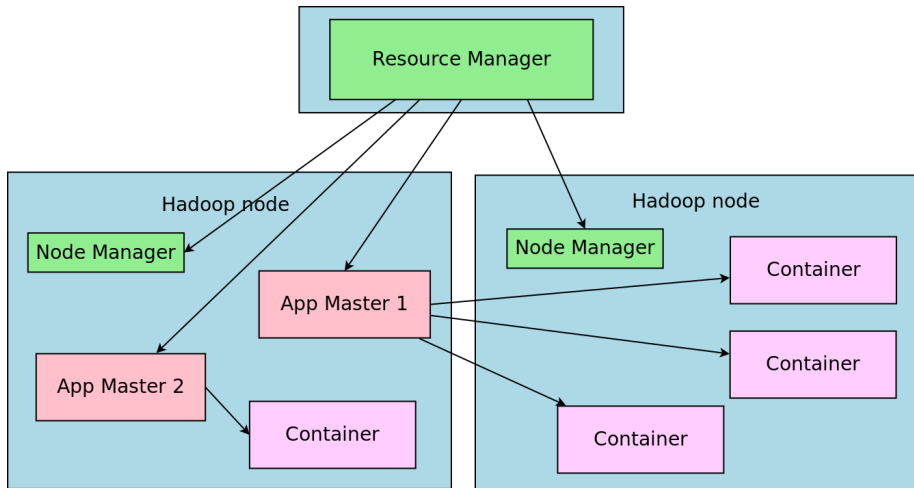
- 2 dni
- 10% prezentacja
- 40% live coding
- 50% zadań głównie programistycznych



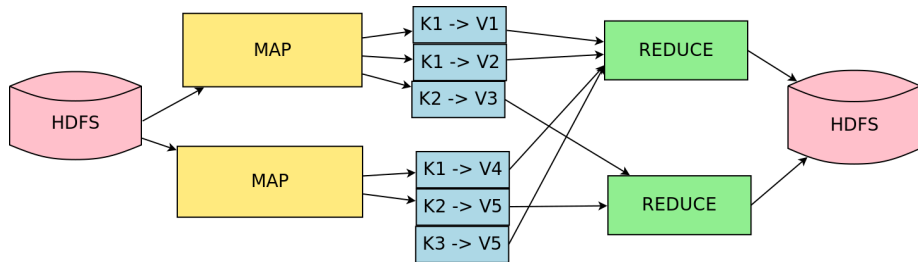
HDFS



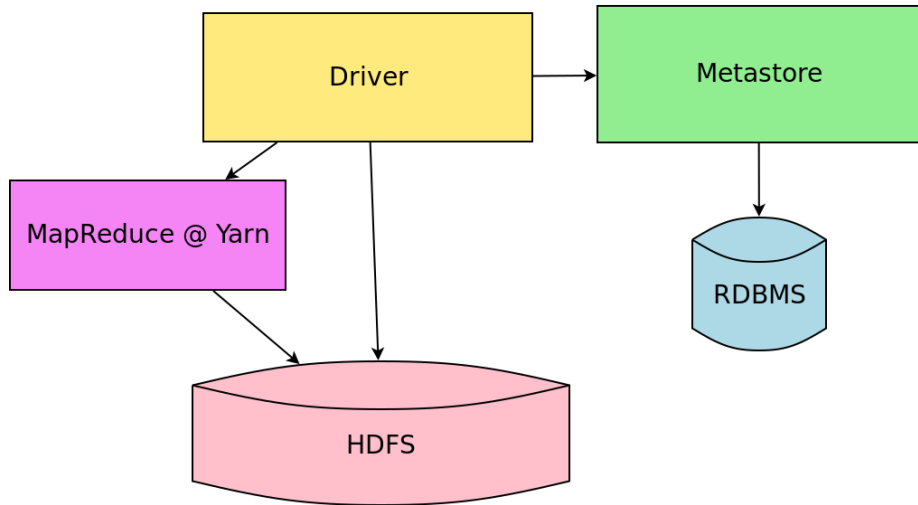
YARN



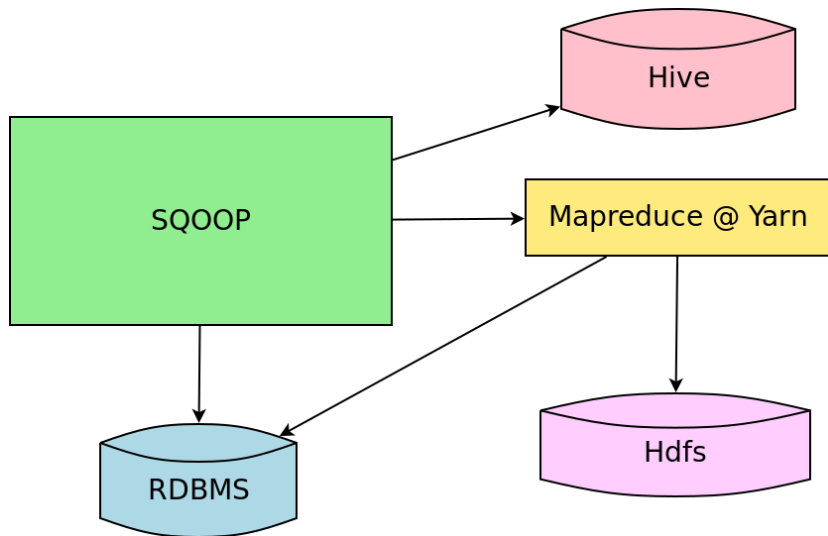
MapReduce



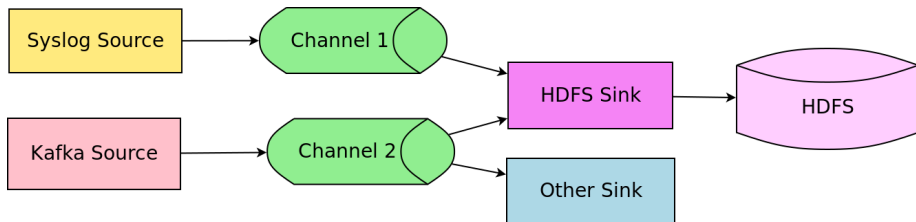
Hive

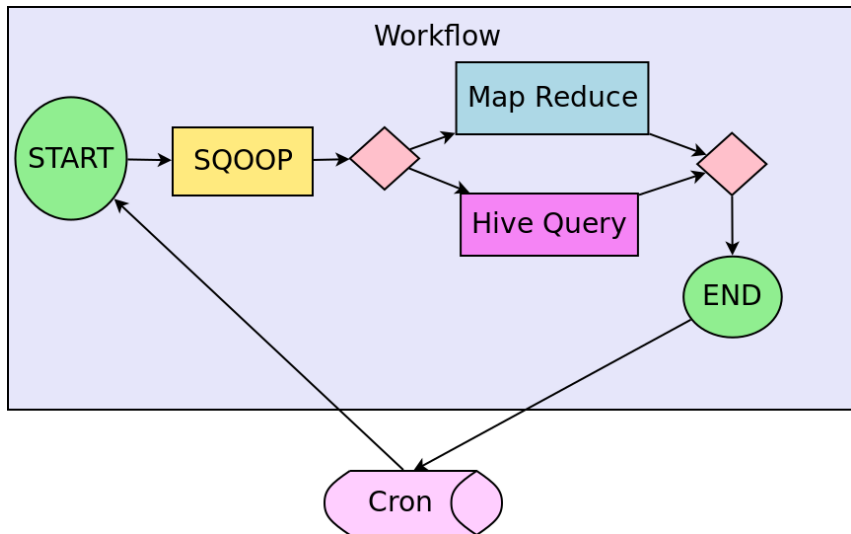


Sqoop



Flume





Query

Search data and saved documents...

Jobs

Impala

Joins

Add a description...

u_rmain

Tables

account_stnew

brinetest1

doc_feedback_aug

free_logs

customerkey (string)

clustername (string)

collectiontimestamp (bigint)

service (string)

role1pest (string)

role (string)

host (string)

filename (string)

ts (bigint)

dt (string)

level (string)

class (string)

message (string)

year (int)

month (int)

free_logs_2016_06

hue_comments

id (string)

author (string)

date (string)

title (string)

jira_c

marktable

query_impala

queryresult

queryresultfromsfdc

sfcdqueryresult

tes_ux

test

testresult

ticket

ticket2

ticket3

ticket4

20

↑

↓

Some high risks were detected.

57

58

59 -- the objective is to find the JIRAs in Hue where there are multiple SPDC tickets linked

60 -- it reveals the soft spots in the product

61

62

63 SELECT sfdc.jira_c.name,

64 sfdc.jira_c.jira_summary_c,

65 count(jira_c.name) AS tickets

66 FROM sfdc.cases, sfdc.jira_c, jira.ticket

67 WHERE sfdc.cases.component_c IN ('Hue')

68 AND sfdc.jira_c.case_c = sfdc.cases.id

69 AND jira.ticket.issuekey = sfdc.jira_c.name

70 AND jira.ticket.statusname NOT IN ('Resolved', 'Closed')

71 AND sfdc.jira_c.name NOT LIKE 'CLRS'

72 GROUP BY jira_c.name, jira_c.jira_summary_c

73 HAVING count(jira_c.name) > 3

74 ORDER BY count(jira_c.name) DESC

75

component

Hue

type

Escalation

date

2016-10-01

Query y f1e94325573f8fda:cof9a10e00000000 100% Complete (388 out of 389)

Query f1e94325573f8fda:cof9a10e00000000 90% Complete (386 out of 389)

Query f1e94325573f8fda:cof9a10e00000000 90% Complete (386 out of 389)

Query f1e94325573f8fda:cof9a10e00000000 90% Complete (386 out of 389)

Query f1e94325573f8fda:cof9a10e00000000 100% Complete (389 out of 389)

Query History

Saved Queries

Results (19)

Q

u

name

jira_summary_c

tickets

1 CDH-45011 Improve interaction between Hue and Impala 66

2 CDH-51313 Tracking jira document2 upgrade 5.7 and below to 5.8 and above 47

3 OPSAPS-25666 Offer option in add service wizard to automatically set as a dependency for another service 16

4 CDH-46194 Security analysis for Hue security Jiras 15

5 CDH-46197 Improve integration between Hue and HiveServer2 14

6 OPSAPS-27028 Ease of Embedded DB Causing Frustration and Database Migration Asks 11

7 OPSAPS-39656 Hue needs Load Balancer parameter for SPNEGO auth 10

8 OPSAPS-28330 We should automatically add the value of ldap_username in Hue to hive and impala proxy users 8

9 OPSAPS-24974 Add LDAP Properties for HST and Impala to Hue

Assistant

Functions

Tables

jira_c

cases

ticket

Statement 5/5

Suggestions

Query on partitioned table is missing filters on partitioning columns.

Rewrite query to add filtering conditions.

Improve Analysis

(<http://gethue.com/>)



Dzień 1

- ① HDFS
- ② MapReduce
- ③ Hadoop Streaming
- ④ Sqoop

- 1 Flume
- 2 Hive
- 3 Hive UDFs
- 4 Oozie

Pytania?