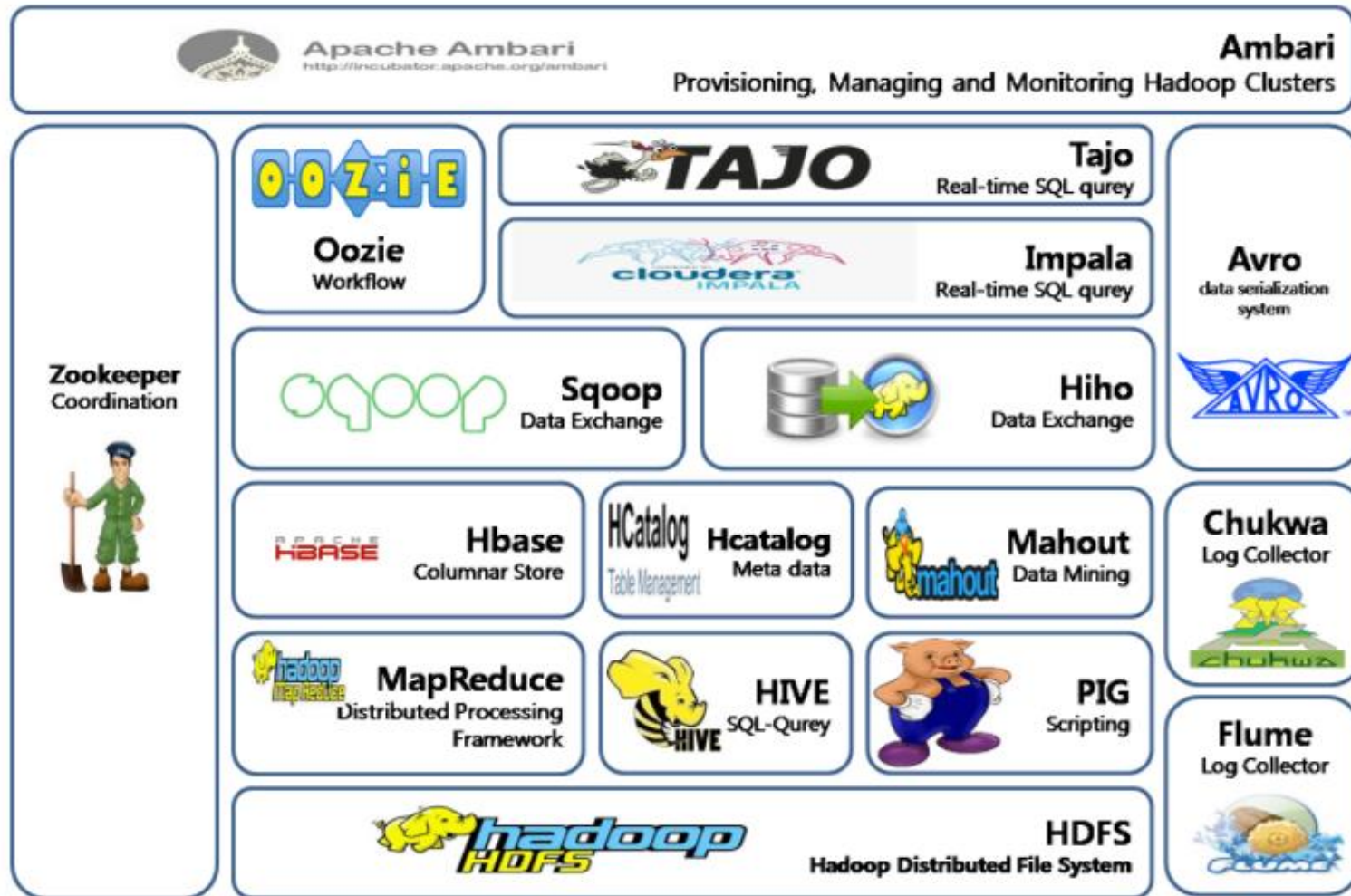

Hadoop Ecosystems

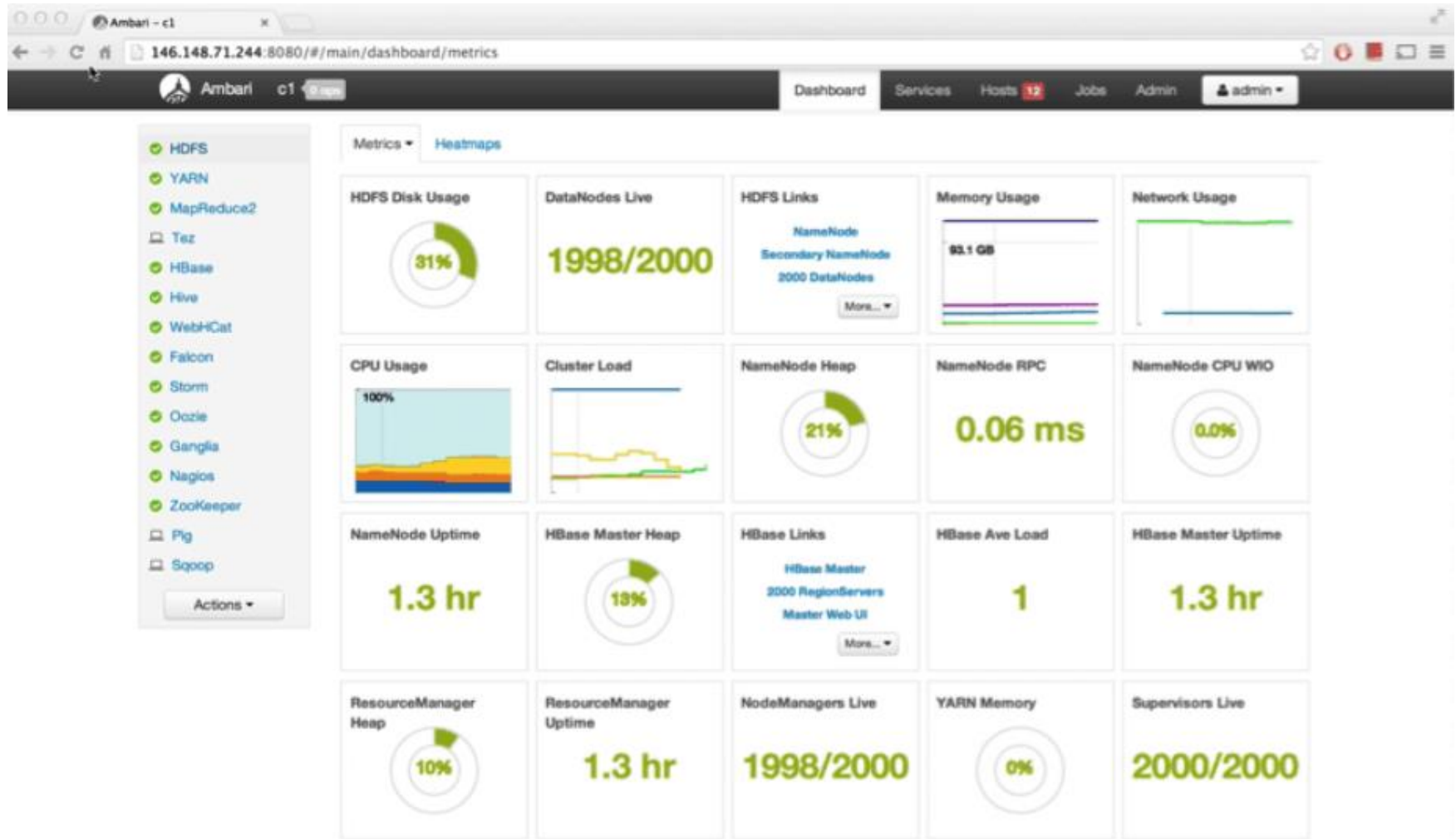
Table of Contents

- 1 Apache Ambari
- 2 Apache Zepplin
- 3 HDP
- 4 CDH
- 5 MAPR

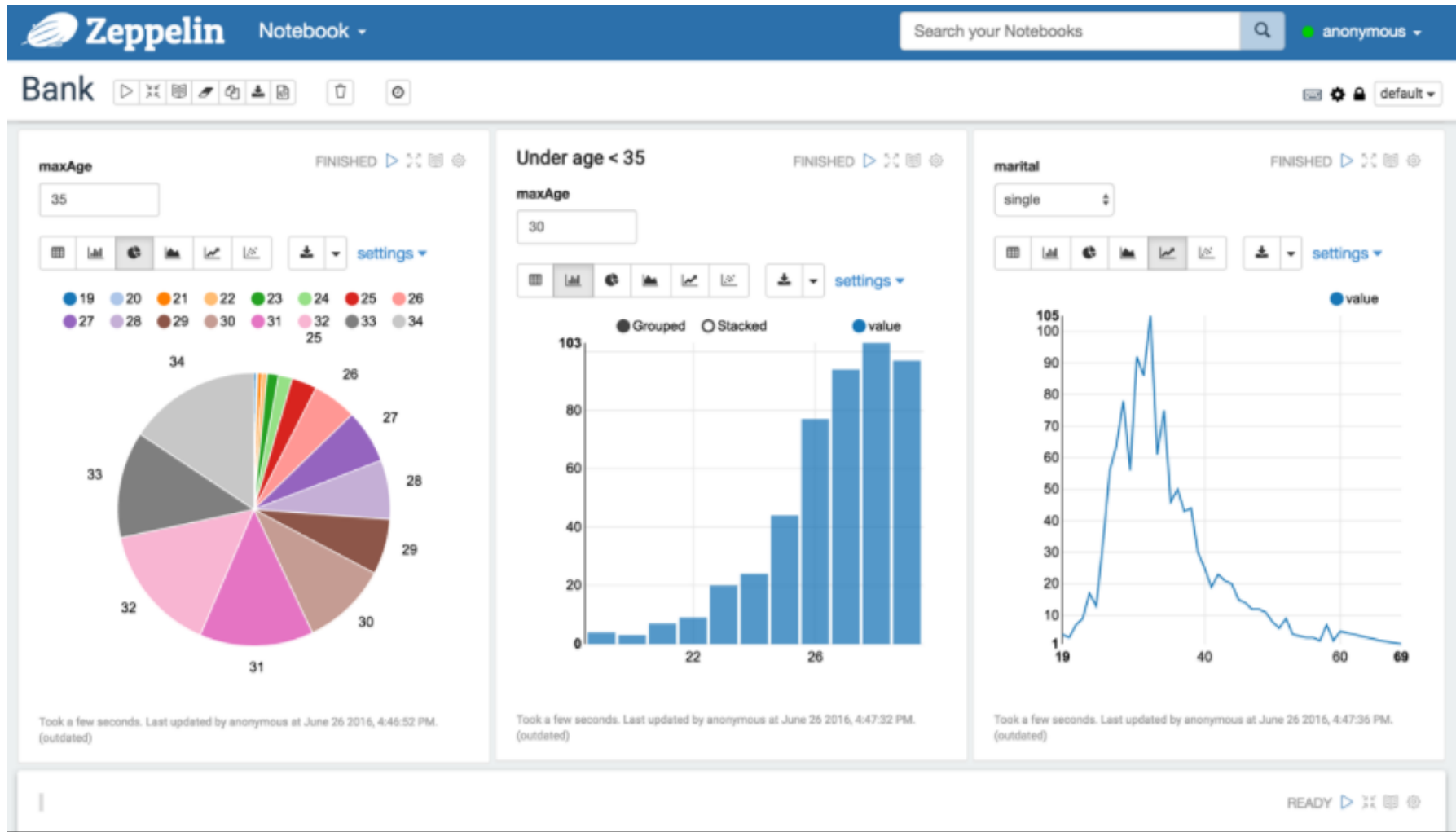
Apache Ambari




Apache Ambari



Apache Zeppline



Apache Zeppelin

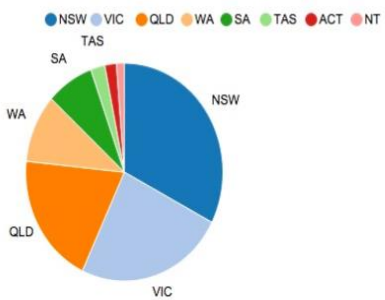
 Notebook - Interpreter

Connected

import sys.process._
// sc is an existing SparkContext.
val sqlContext = new org.apache.spark.sql.SQLContext(sc)
val health_dataset = sc.textFile("/Users/nshawa/Downloads/health_data_expenses.csv")
case class Health(year: String, state: String, category: String, funding_src1: String, funding_src2: String, spending: Integer)
val health = health_dataset.map(k=>k.split(",")).map(
 k => Health(k(0),
 k(1),
 k(2),
 k(3),
 k(4),
 k(5).toInt
)
).toDF()
health.registerTempTable("health_table")

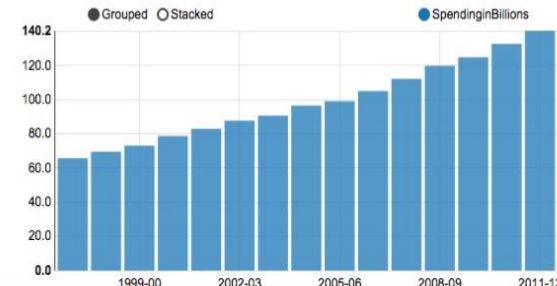
import sys.process._
sqlContext: org.apache.spark.sql.SQLContext = org.apache.spark.sql.SQLContext@7e76cc70
health_dataset: org.apache.spark.rdd.RDD[String] = /Users/nshawa/Downloads/health_data_expenses.csv MapPartitionsRDD[566] at textFile at <console>:182
defined class Health
health: org.apache.spark.sql.DataFrame = [year: string, state: string, category: string, funding_src1: string, funding_src2: string, spending: int]
Took 4 seconds

%sql
select state,sum(spending)/1000 SpendinginBillions from health_table group by
state order by SpendinginBillions desc



State	SpendinginBillions
NSW	445.845
VIC	272.507
QLD	121.022
WA	104.221
TAS	90.786
ACT	75.765
NT	72.698

%sql
select year,sum(spending)/1000 SpendinginBillions from health_table group by
year order by SpendinginBillions

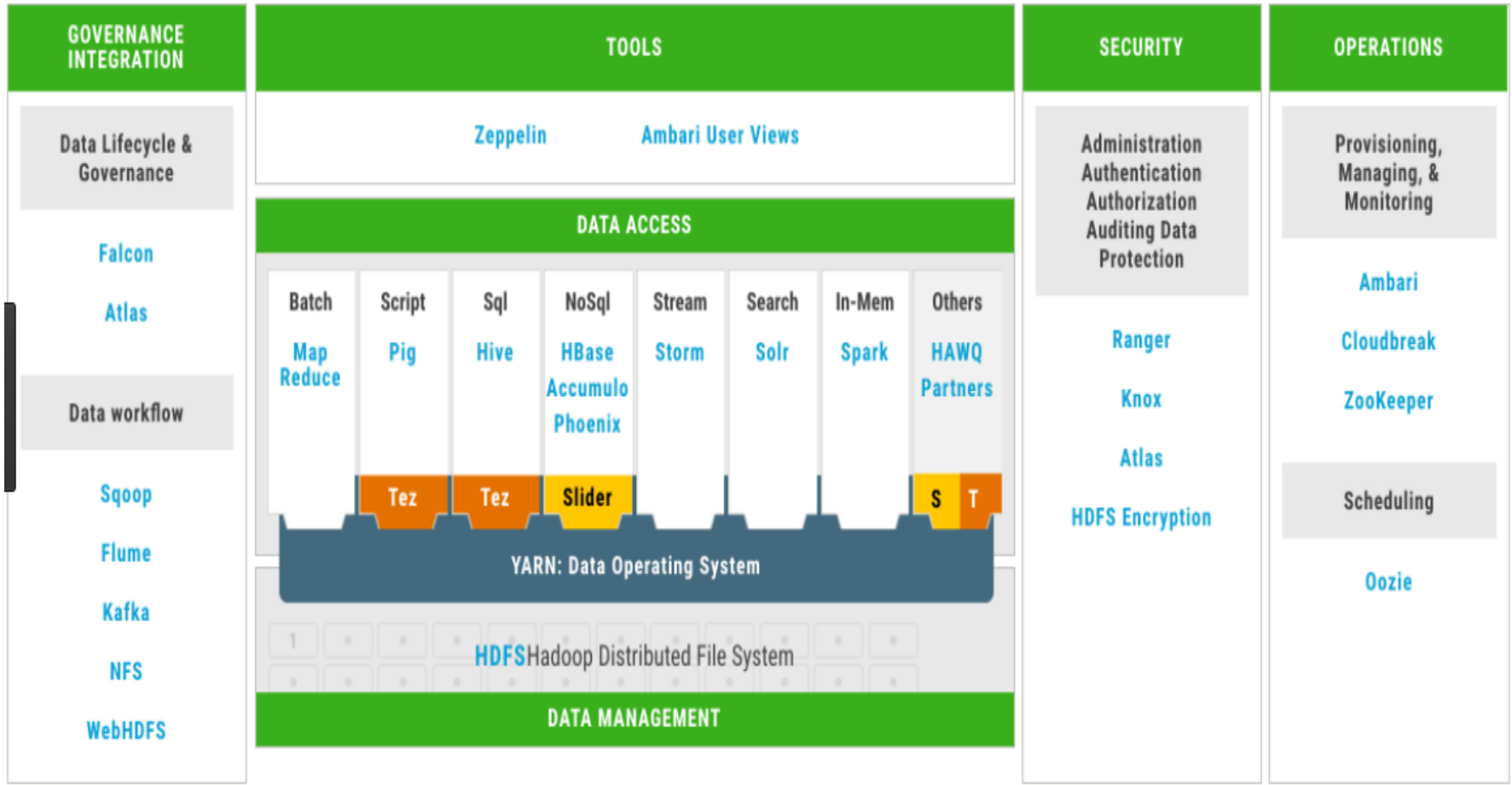


Year	SpendinginBillions
1999-00	65.0
2000-01	70.0
2001-02	75.0
2002-03	80.0
2003-04	85.0
2004-05	90.0
2005-06	95.0
2006-07	100.0
2007-08	105.0
2008-09	110.0
2009-10	115.0
2010-11	120.0
2011-12	125.0

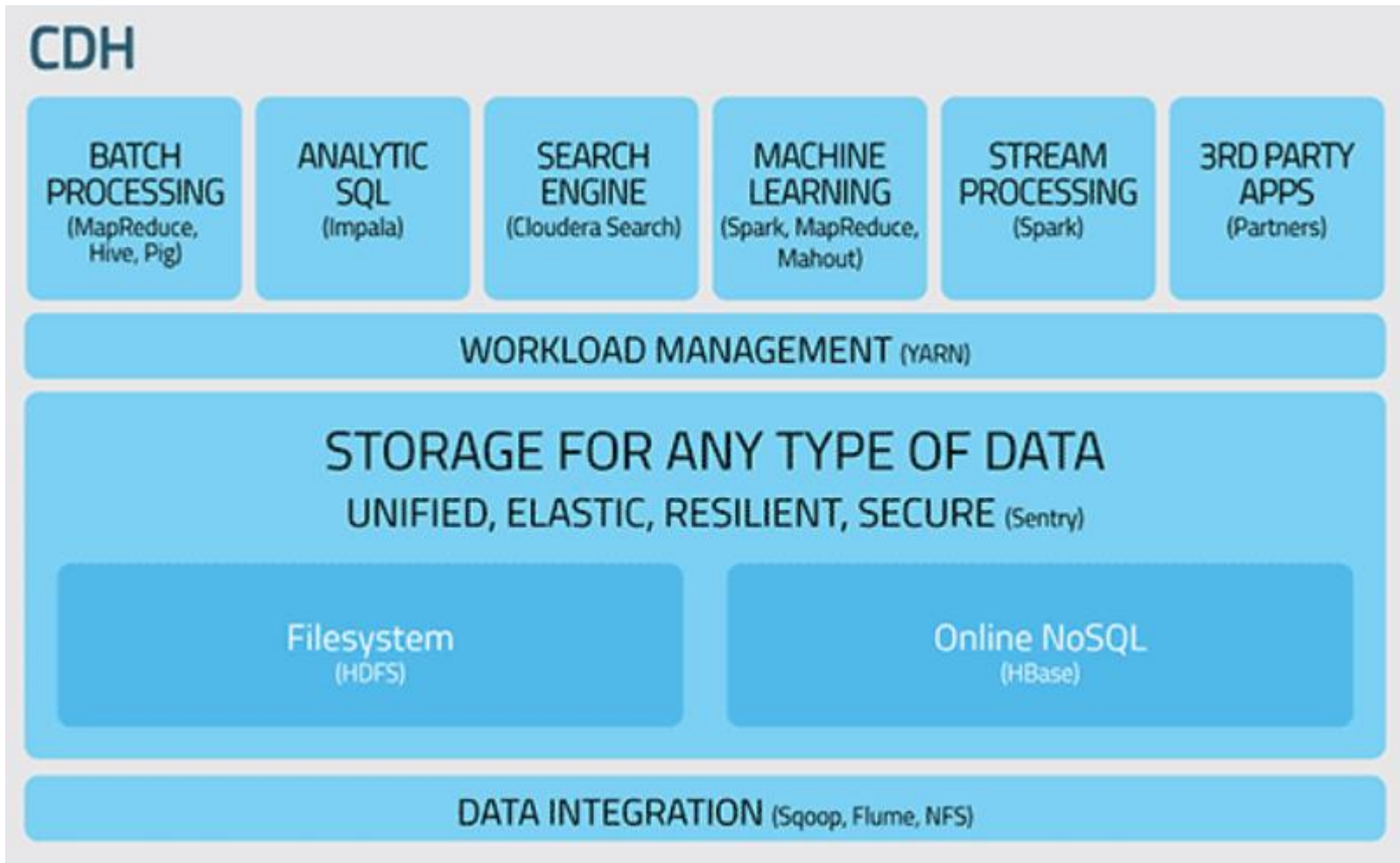
%sql
select category,sum(spending)/1000 SpendinginBillions from health_table group by
category order by SpendinginBillions desc

category	SpendinginBillions
Public hospitals	445.845
Medical services	272.507
Private hospitals	121.022
Benefit-paid pharmaceuticals	104.221
Dental services	90.786
Community health	75.765
Capital expenditure	72.698
All other medications	70.508
Other health practitioners	51.382

HortonWorks HDP



Cloudera CDH



MapR

