2110625 - Data Science Architecture

Big Data File Systems

Asst.Prof. Natawut Nupairoj, Ph.D.
Department of Computer Engineering
Chulalongkorn University
natawut.n@chula.ac.th

CHULA ENGINEERING
Foundation toward Innovation

Typical Big Data File Systems

- Based on scale-out architecture ใช้พื้นฐาน scale out
- Must support structured and unstructured data (หนัง, ภาพ, เสียง)
- Processing historical data
  - Hardly changed เปลี่ยนน้อยมาก เหมาะเป็นข้อมูลอดีต เกิดขึ้นแล้ว
  - Whole-file scanning accessing pattern อ่านไฟล์ตัน ถึง byte สุดท้ายของไฟล์
- Fault tolerance is required ทำอย่างไรถ้าเครื่องตาย ยังมีสำรอง → Reprocess

HDFS : Hadoop Distributed File System

- A distributed file system in Apache Hadoop Core project designed to run on commodity hardware ออกแบบให้ทำงานบน HW ทั่วไป (3 copy)
- Highly fault-tolerant and high throughput access to application data

bandwidth ของ CPU, I/O ไปยัง HDD

- Supports a traditional hierarchical file organization (directories and files) with user quotas and access permissions (does not support hard links or soft links)

มี directory เหมือน file ปกติ แต่ไม่มี hard / soft link (shortcut)

Data Lake

- A system or repository of data stored in its natural/raw format, usually object blobs or files
  - All data is loaded from source systems, no data is turned away
  - Data is stored at the leaf level in an untransformed or nearly untransformed state
  - Data is transformed and schema is applied to fulfill the needs of analysis
- Can be on premise (HDFS is quite popular), or on cloud (Amazon S3, Google Cloudstore, etc.)

Assumptions and Goals

- Hardware failure is the norm, focus on quick failure detection and automatic recovery masking ซ่อนระบบ ไว้ node สำรองแทน
- Applications on HDFS are assumed to need streaming data access, batch (not interactive), high throughput (not low latency) มองเป็น batch ปริมาณที่ส่งไปยัง client สูง แต่ใช้เวลาสูง (รถบรรทุก throughput สูง รถเก๋ง latency สูง)
- HDFS is for large data sets (GBs to TBs) and provides high aggregated bandwidths from hundreds of nodes in a single cluster
- Simple Coherency Model with write-once-read-many access model, a file once created and closed can only append or truncate ไม่ให้ update (เขียน 1 ครั้งจบ) (strong consistency)
- Moving Computation is Cheaper than Moving Data HDFS มองว่าการ move การคำนวณง่ายกว่า move ข้อมูล
- Portability Across Heterogeneous Hardware and Software Platforms HDFS รองรับหลายภาษา

Play with HDFS and Hadoop in Containers

- Note: this still does not work on Apple Silicon
- Use big-data-europe/docker-hive on GitHub

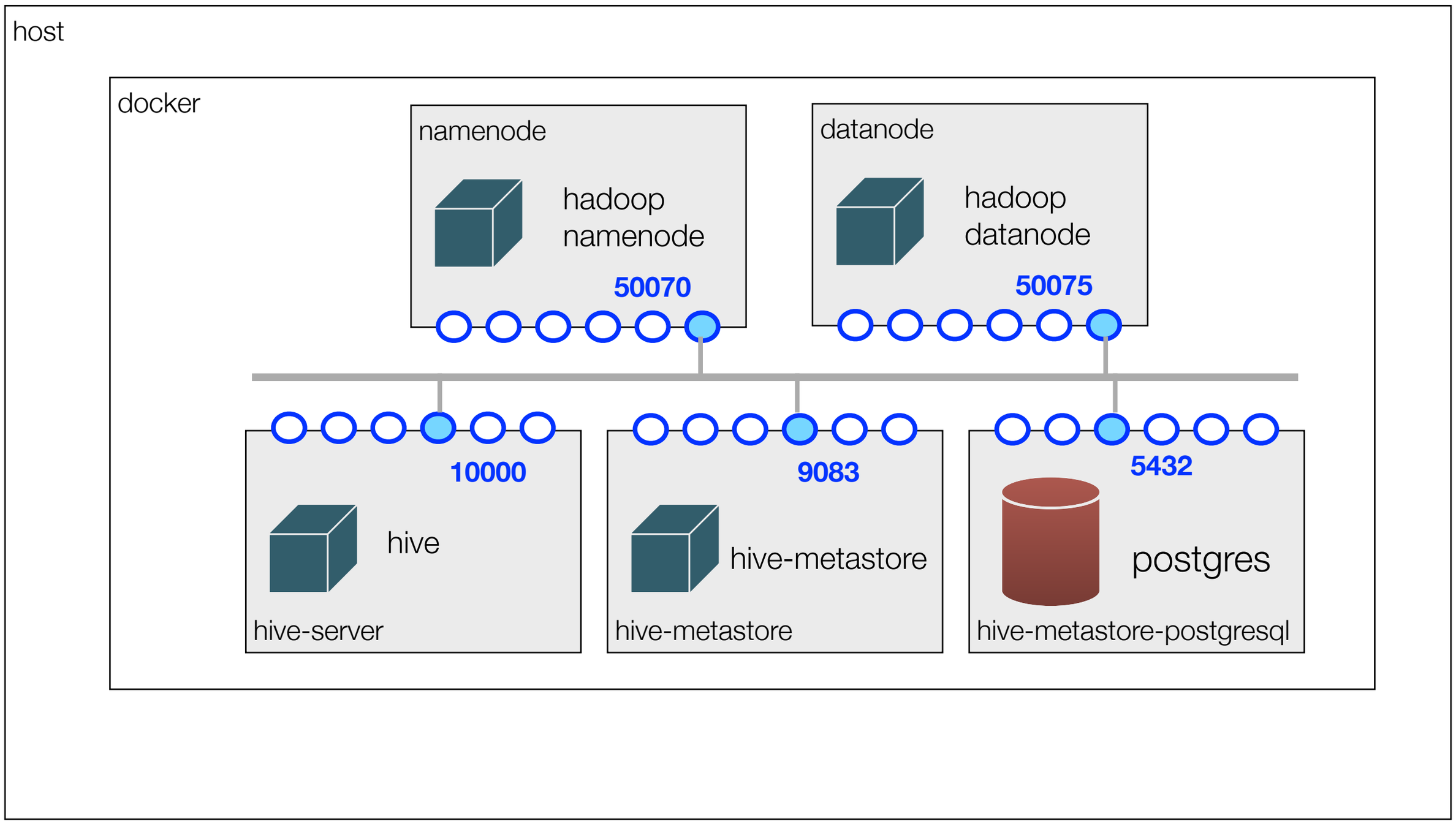
```
git clone https://github.com/big-data-europe/docker-hive.git
```

- Use docker-compose to create all containers

```
sudo docker-compose up
```

- Use docker-compose to create all containers
  - Namenode: http://localhost:50070
  - Datanode: http://localhost:50075
  - Hive-server: http://localhost:10000
  - Hive-metastore: http://localhost:9083

Sample Hadoop Cluster Architecture



Hadoop FS Commands behave similar to Unix Commands

```
sudo docker-compose exec namenode bash

hadoop version

hdfs dfs -mkdir /path/directory_name

hdfs dis -ls /path

hdfs dfs -copyFromLocal <localsrc> <hdfs destination>

hdfs dfs -copyToLocal <hdfs source> <localdst>

hdfs dfs -cat /path_to_file_in_hdfs

hdfs dfs -mv <src> <dest>

hdfs dfs -cp <src> <dest>
```

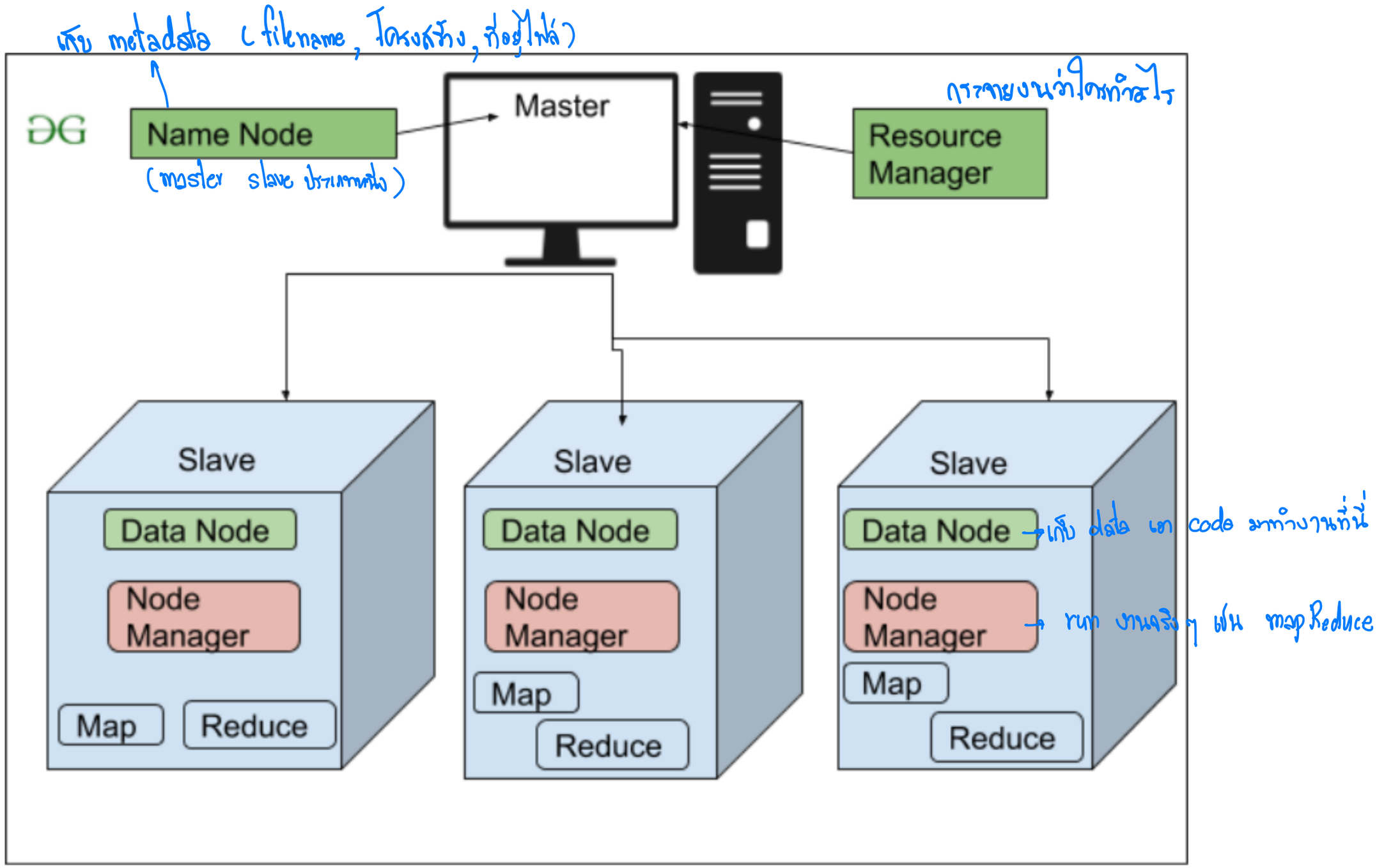
HDFS Sample Commands

```
root@2602201f31b9:~# hdfs dfs -ls /
Found 3 items
drwxr-xr-x   - root supergroup          0 2021-08-17 11:45 /bigdata
drwxrwxr-x   - root supergroup          0 2021-03-20 06:31 /tmp
drwxr-xr-x   - root supergroup          0 2021-03-16 04:05 /user
root@2602201f31b9:~# hdfs dfs -mkdir /myhome
root@2602201f31b9:~# hdfs dfs -copyFromLocal NetflixOriginals.csv /myhome
root@2602201f31b9:~# hdfs dfs -ls /myhome
Found 1 items
-rw-r--r--   3 root supergroup      38678 2021-08-18 02:37 /myhome/NetflixOrigin
als.csv
root@2602201f31b9:~#
```

All Hadoop FS Commands

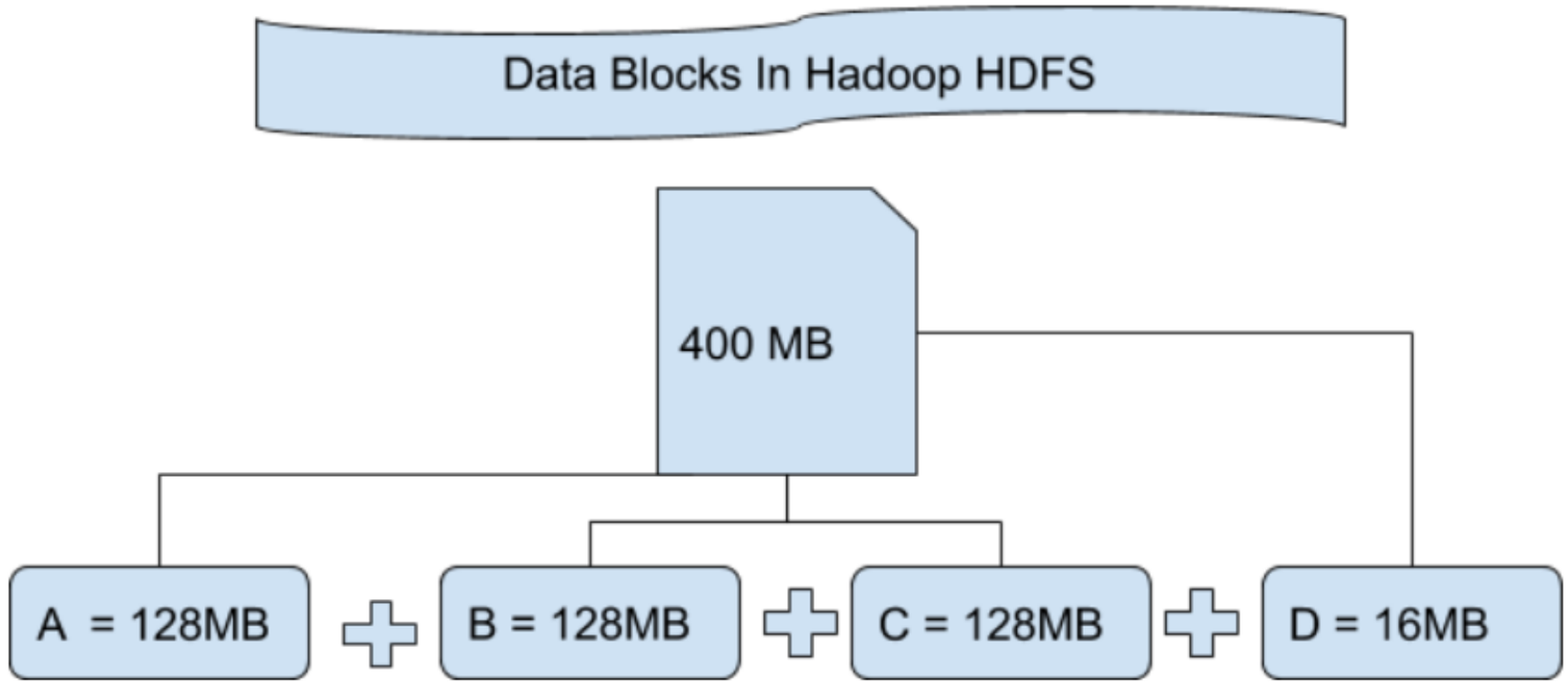
appendToFile	df	lsr	setrep
cat	du	mkdir	stat
checksum	dus	moveFromLocal	tail
chgrp	expunge	moveToLocal	test
chmod	find	mv	text
chown	get	put	touch
copyFromLocal	getfacl	renameSnapshot	touchz
copyToLocal	getfattr	rm	truncate
count	getmerge	rmdir	usage
cp	head	rmr	
createSnapshot	help	setfacl	
deleteSnapshot	ls	setfattr	

High Level Architecture Of Hadoop



Source: https://www.geeksforgeeks.org/hadoop-architecture/

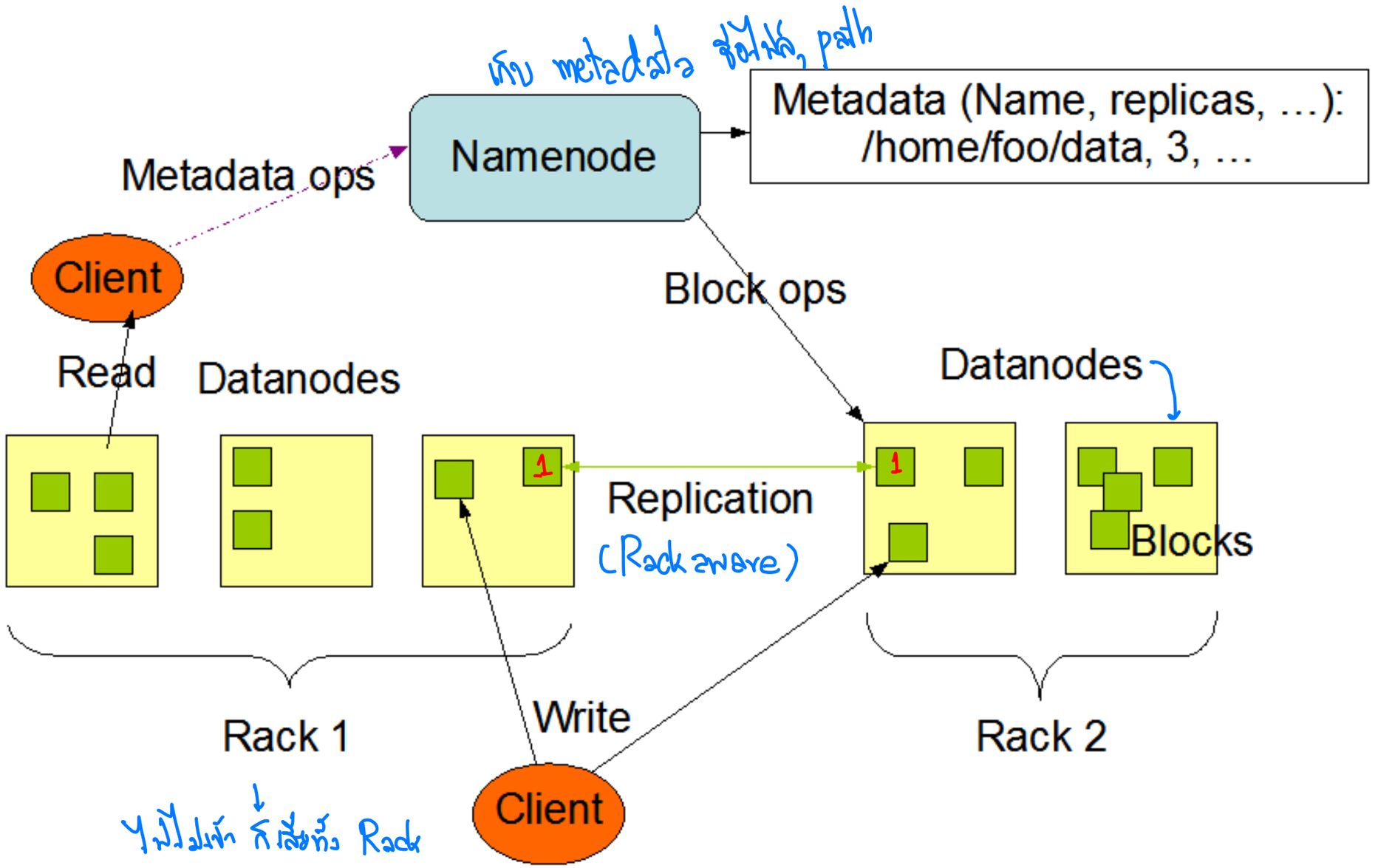
File Block in HDFS



Source: https://www.geeksforgeeks.org/hadoop-architecture/

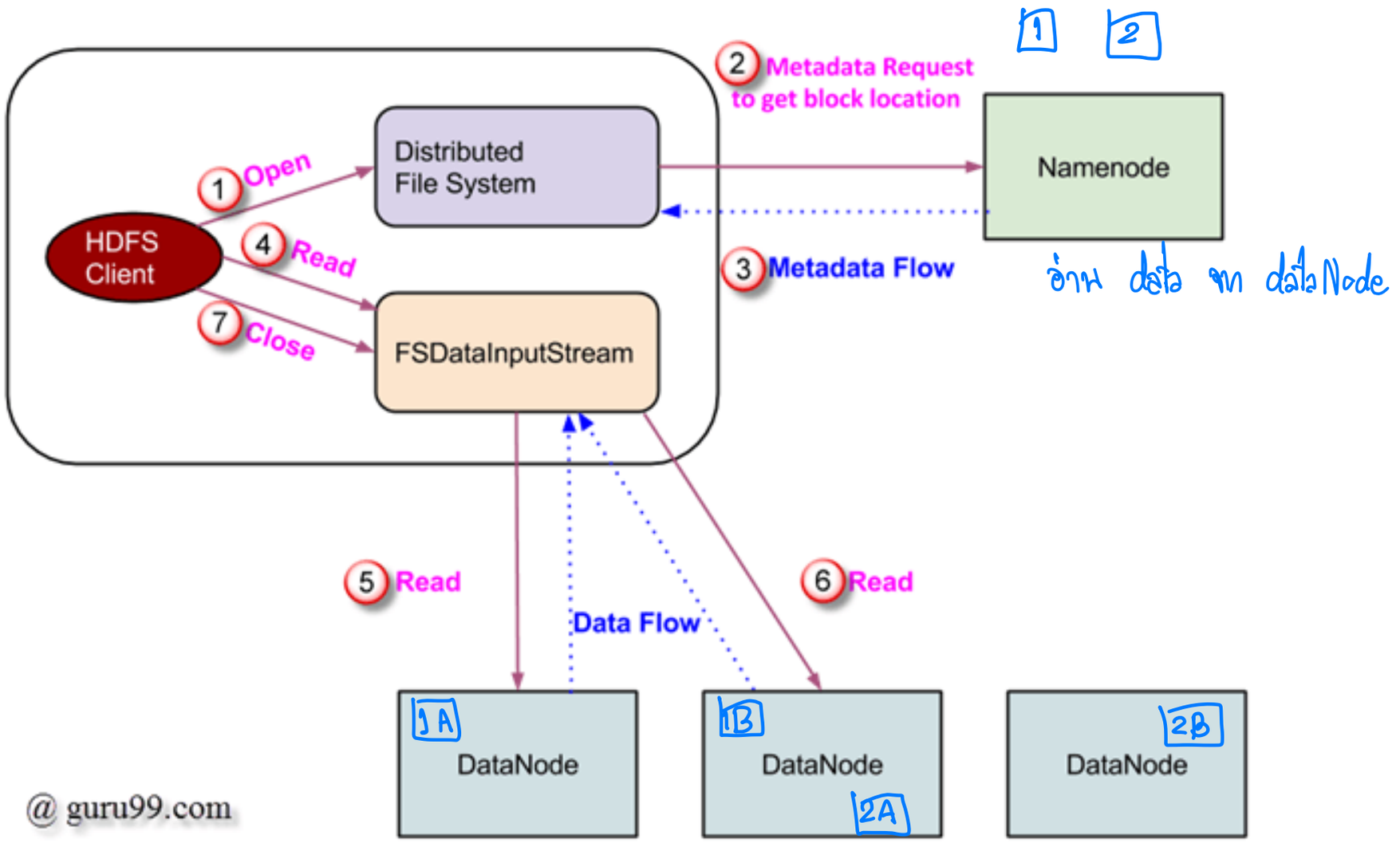
- A data file is splitter into blocks
- Each block is 128 MBs (can be configured), except the last block
- This is much larger comparing to 4 KBs of normal Linux file system

HDFS Architecture



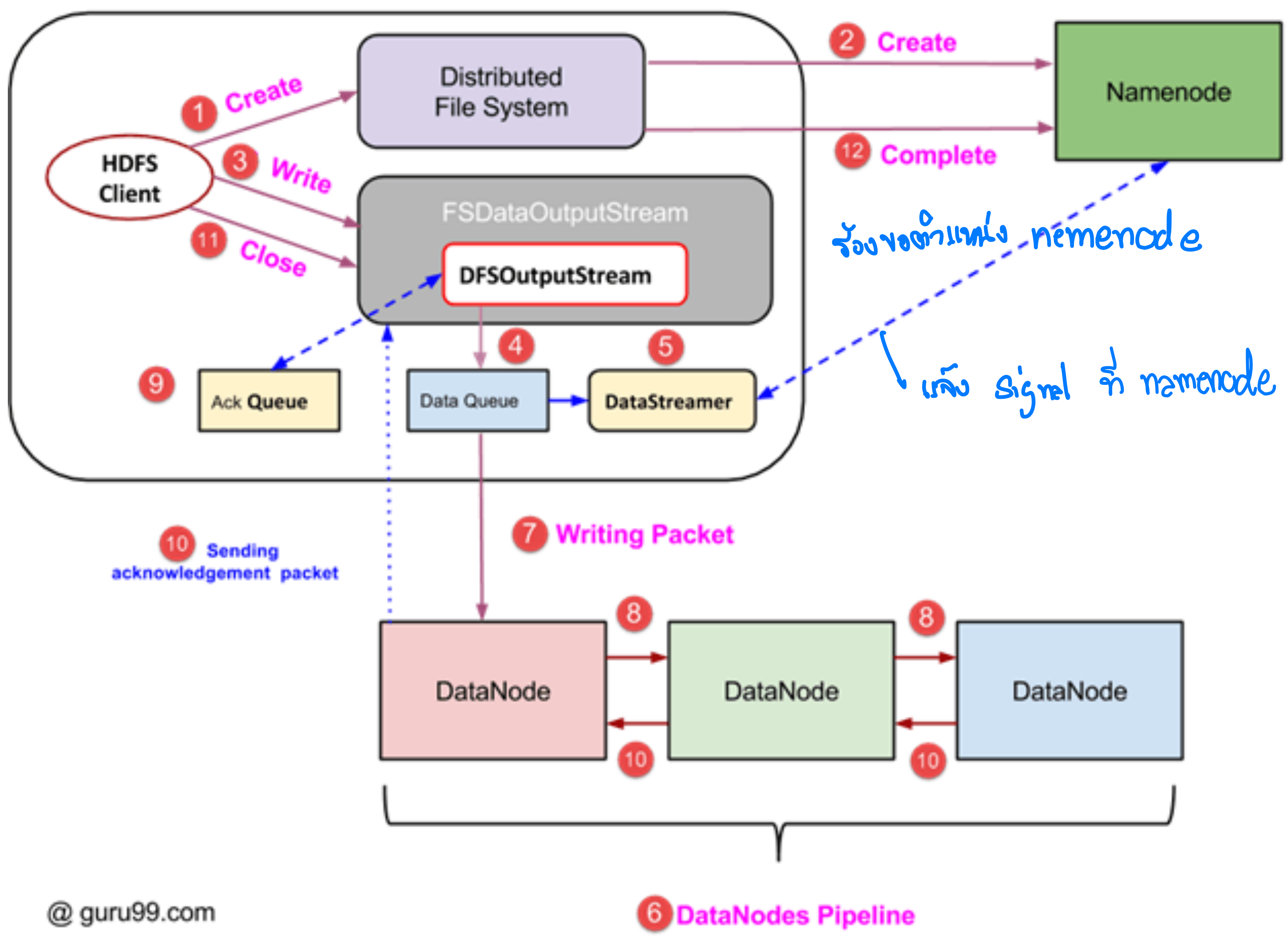
Source: https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

HDFS File Read Operation



Source: https://www.guru99.com/learn-hdfs-a-beginners-guide.html

HDFS File Write Operation



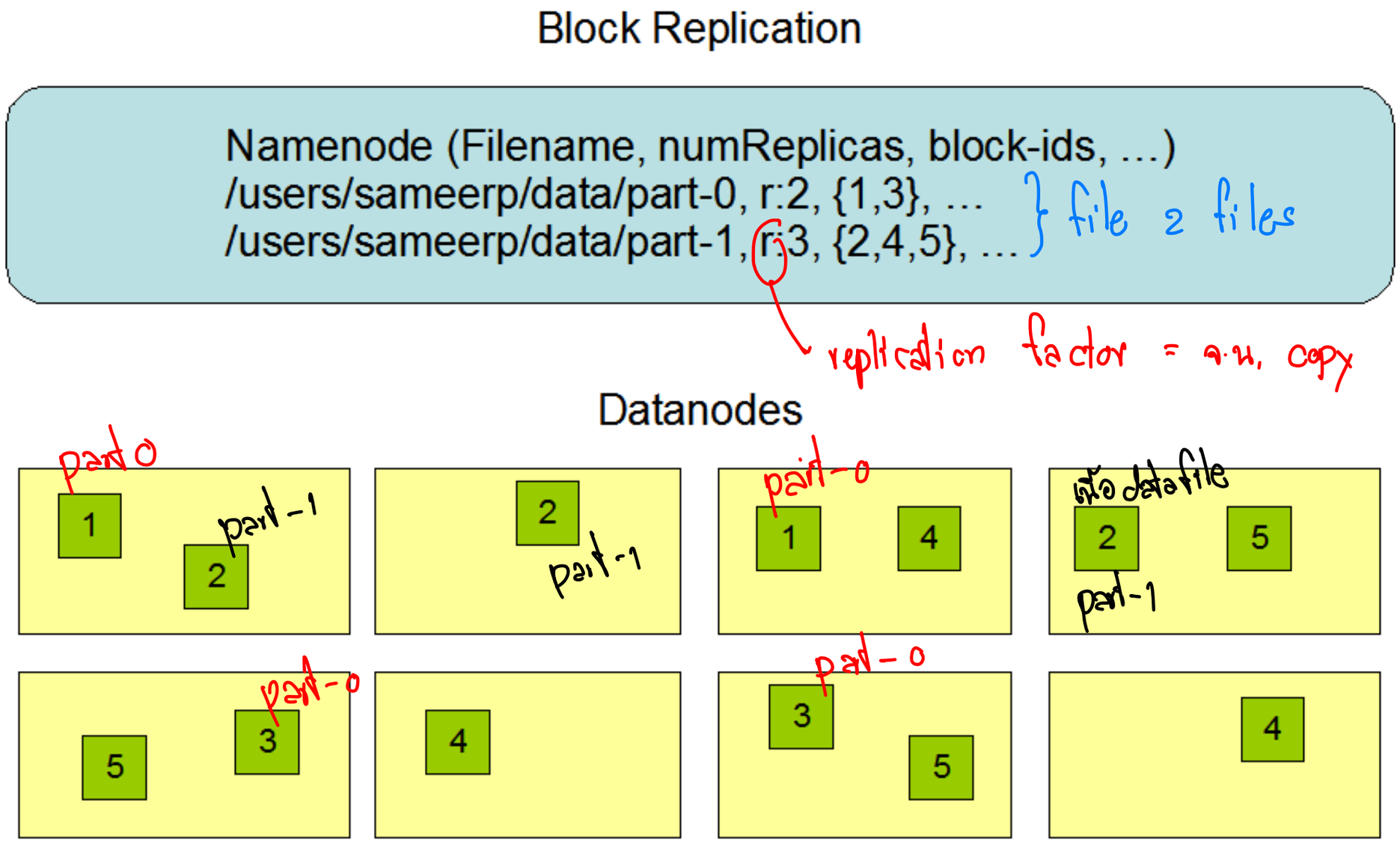
ร้องขอตำแหน่ง namenode

แจ้ง signal ที่ namenode ว่าเขียนแล้ว

file จะปรากฏให้ผู้ใช้งานเห็นเมื่อเขียนเสร็จ (copy ครบ 3 datanode)

Source: https://www.guru99.com/learn-hdfs-a-beginners-guide.html

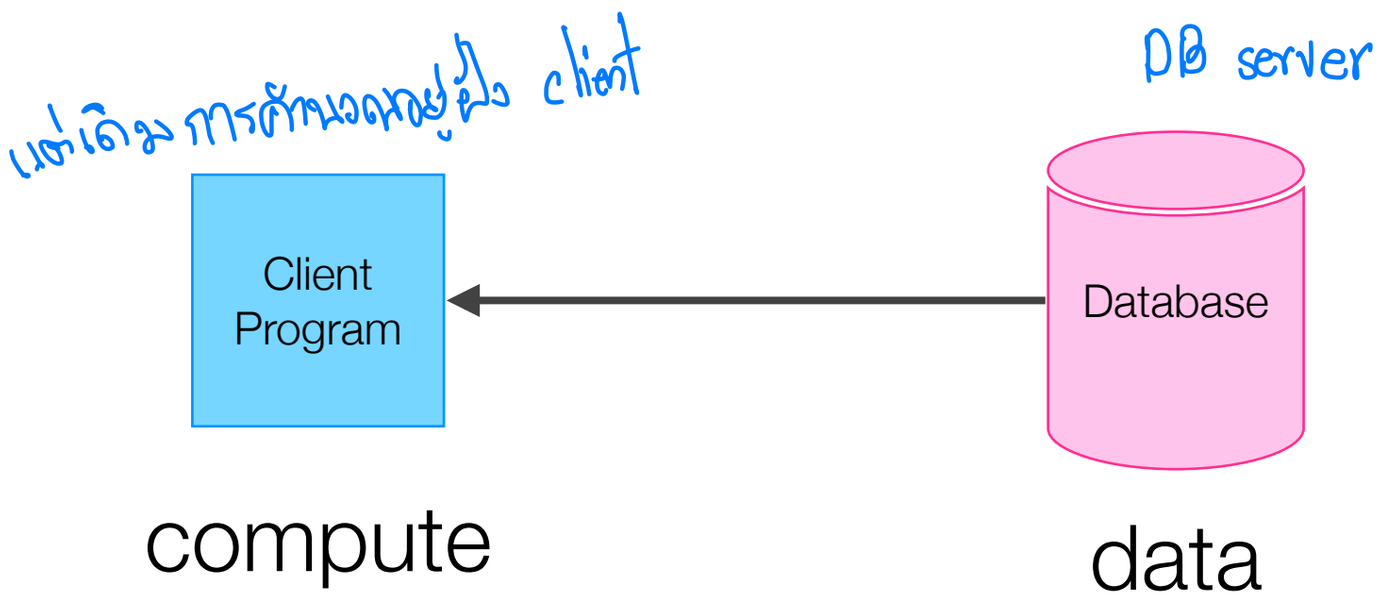
Data Replication



Source: https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html

HDFS and Data Locality in Hadoop

- Data locality in Hadoop in the process of moving computation close to the actual data on the node
  - Faster execution and higher throughput
- Typical transaction processing system move data to the computation e.g. database and client program

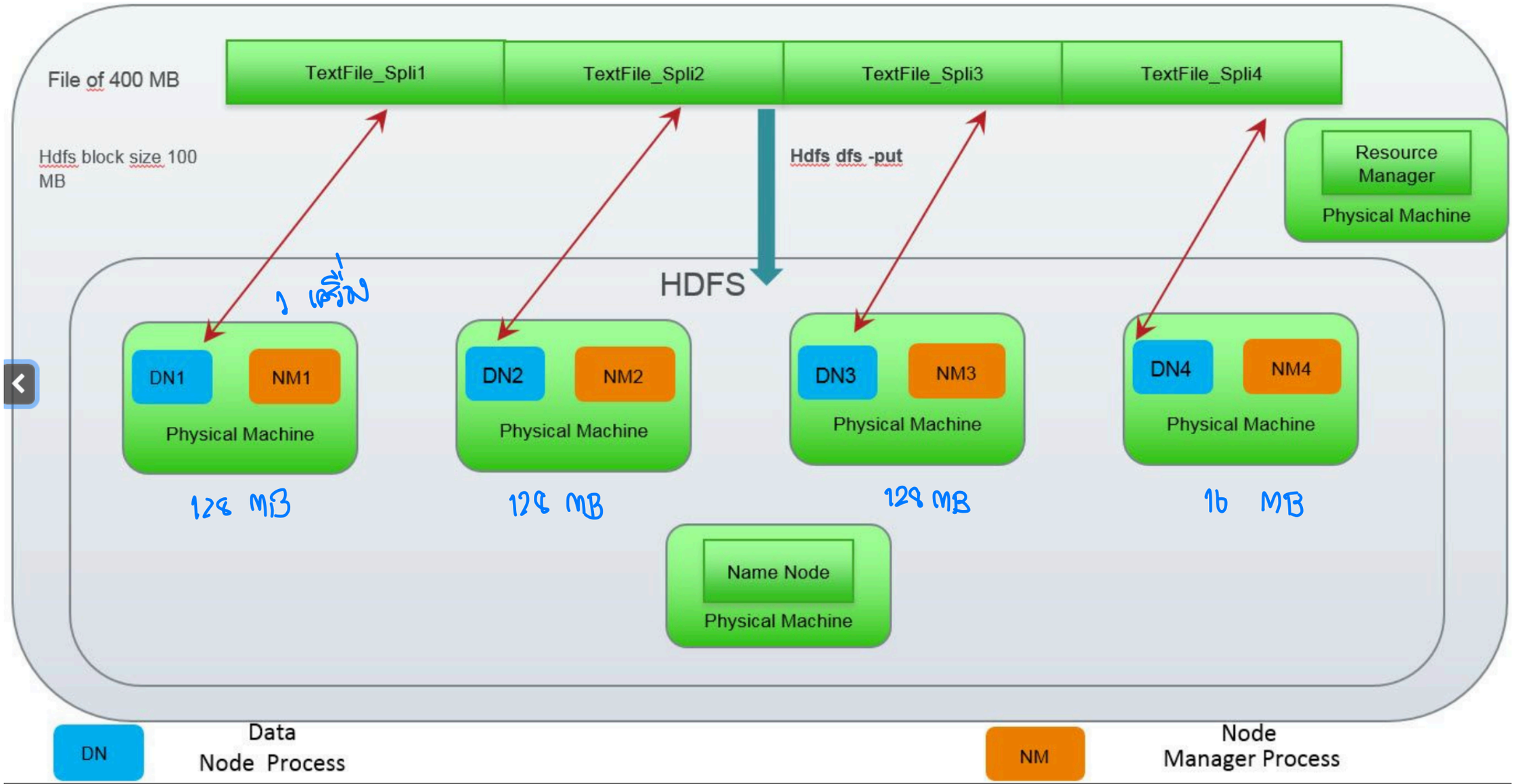


HDFS and Data Locality in Hadoop

- In Hadoop, data is much bigger than code — moving code is much more efficient
  - Data is stored in blocks (with multiple copies) across DataNodes in HDFS
  - NameNode moves code of a MapReduce job to DataNodes to process data blocks (called data local data locality — preferred scenario)
  - If data local data locality is not possible, topology (intra-rack and inter-rack) comes into consideration for code execution placements

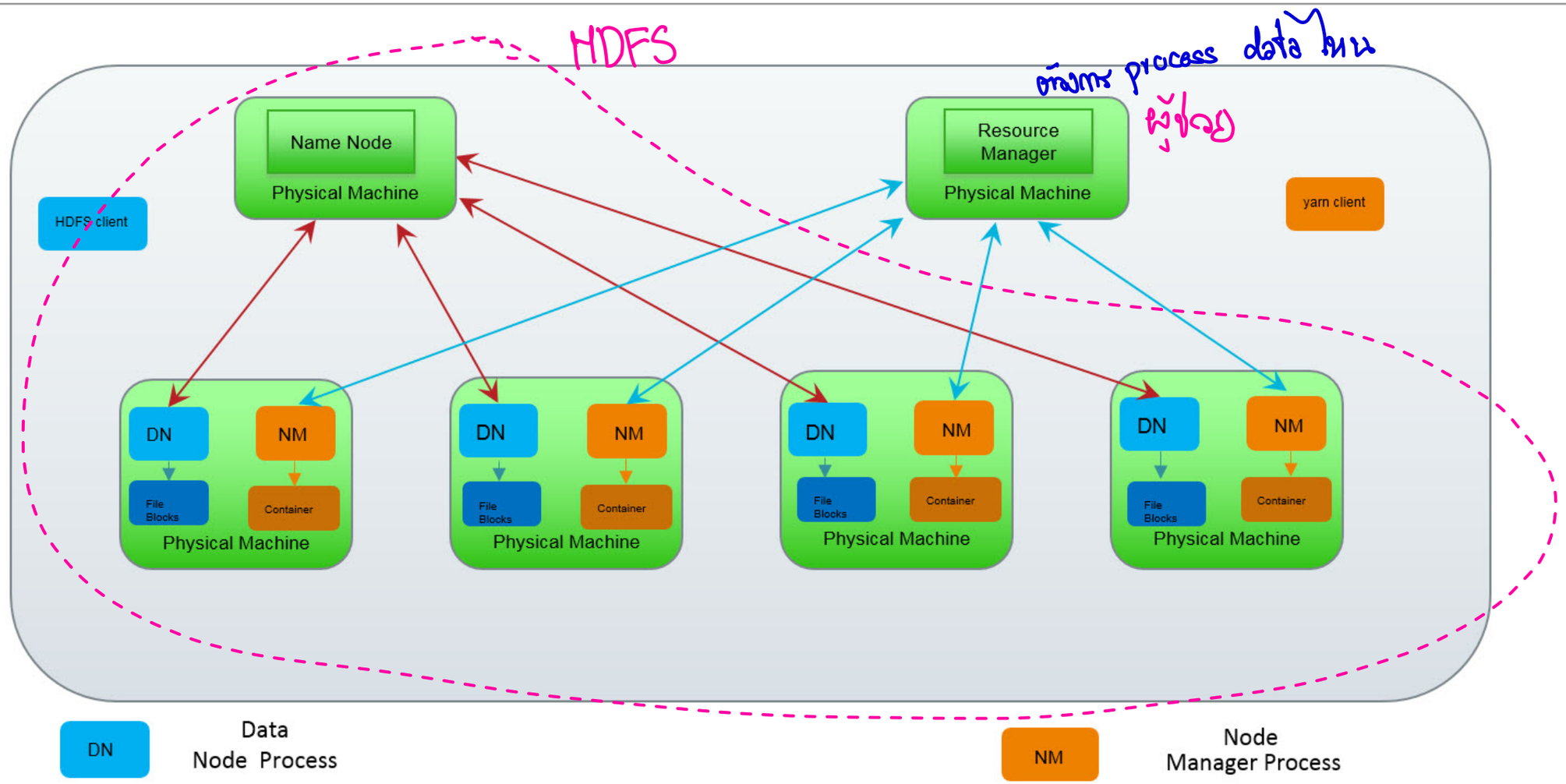
HDFS and Data Locality in Hadoop

Source: https://community.cloudera.com/t5/Community-Articles/Understanding-basics-of-HDFS-and-YARN/ta-p/248860



HDFS and Data Locality in Hadoop

Source: https://community.cloudera.com/t5/Community-Articles/Understanding-basics-of-HDFS-and-YARN/ta-p/248860

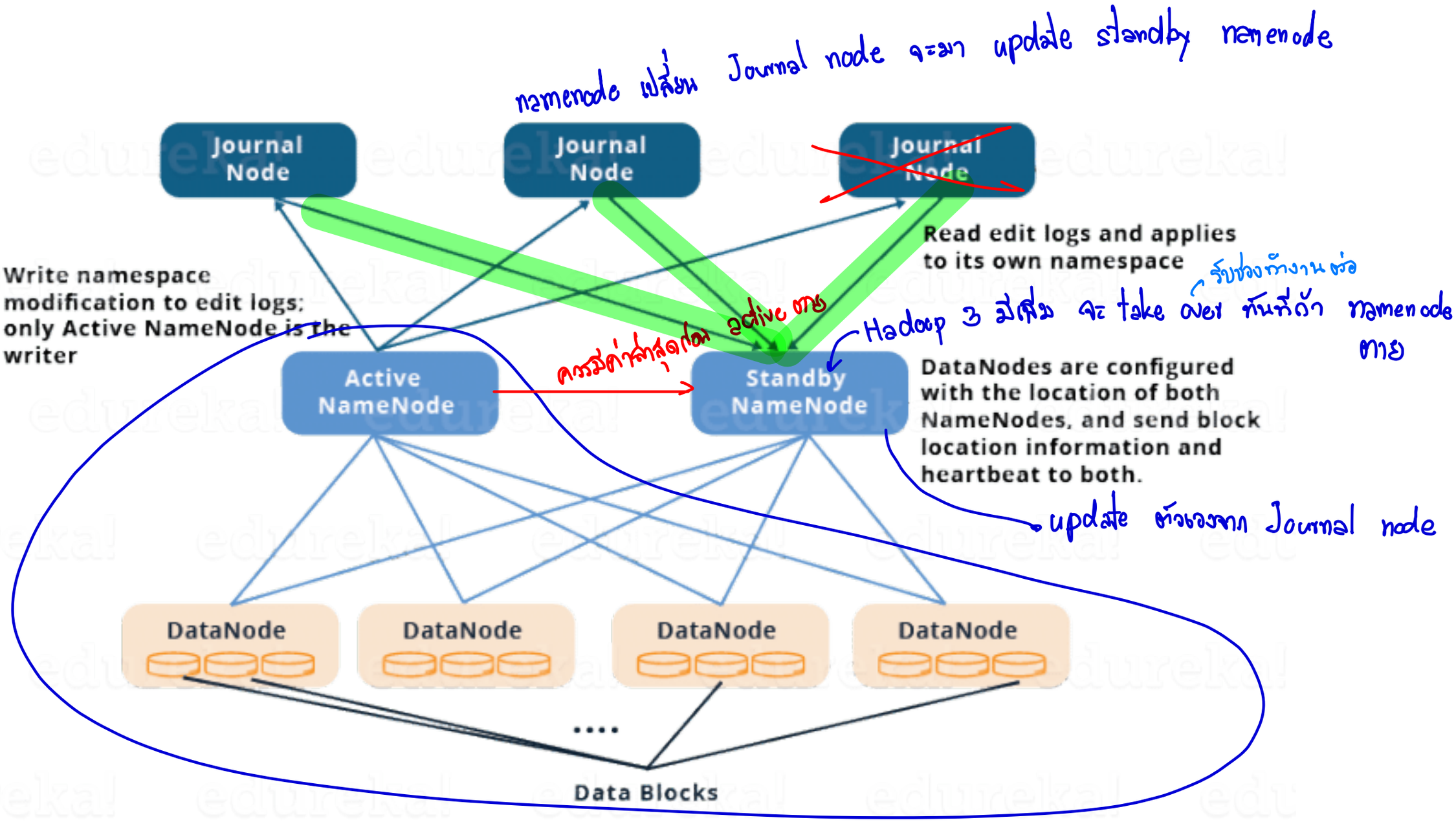


- MapReduce Job Manager is a yarn client that submit codes to resource manager
- Resource manager distributes codes to selected node managers to run in yarn containers
- A Yarn container consists of codes and HDFS client to access data blocks in the same (preferred) or other DataNode

HDFS Pros and Cons

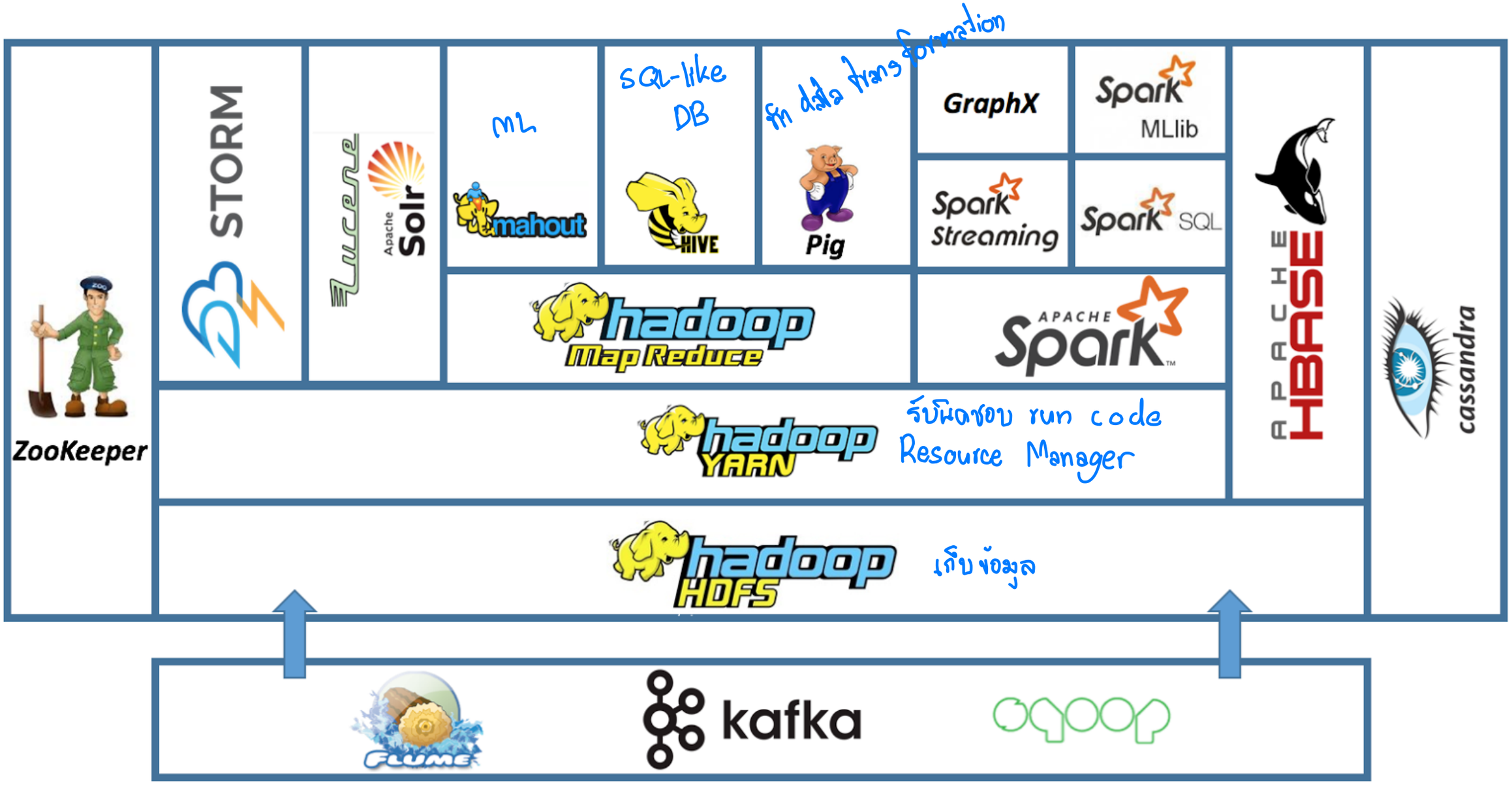
- Pros
  - Cost effective ทำงานบน HW ถูก
  - Highly scalable and high throughput
  - Fault tolerance มี 3 replication (default)
- Cons
  - Problem with small files - NameNode bottleneck, low performance (namenode ช้าลงถ้ามีไฟล์เล็กมากๆ)
  - NameNode is a single point of failure namenode ไม่มี backup
  - Dependency on disks - more overhead for read/write
  - Suitable for only batch processing ไม่เหมาะในการทำงาน interactive

Hadoop 3 : HDFS HA Architecture



Source: https://www.edureka.co/blog/how-to-set-up-hadoop-cluster-with-hdfs-high-availability/

Apache Hadoop Ecosystem



Hive

สร้าง DW บน HDFS ใช้คำสั่ง SQL ได้

- An information platform built by a team from Facebook to create a data warehouse framework on top of Hadoop
- A mechanism to impose structure on a variety of data formats stored directly in Apache HDFS or in other data storage systems such as Apache HBase
- Enable analysts with strong SQL skills to run HPL-SQL queries via Apache Tez, Apache Spark, or MapReduce

Hive Architecture

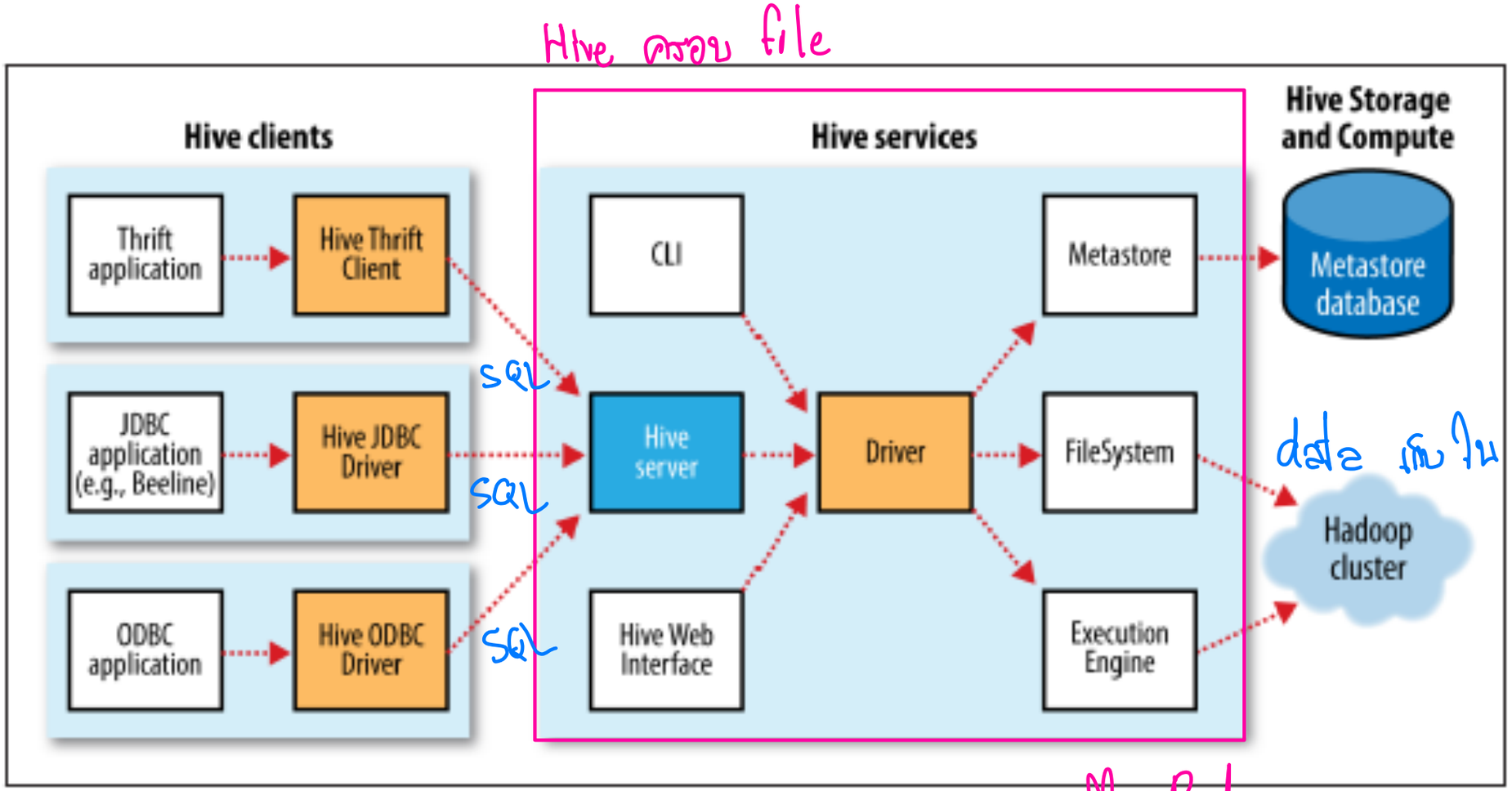
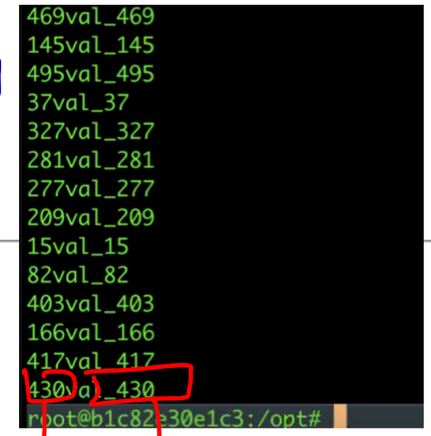


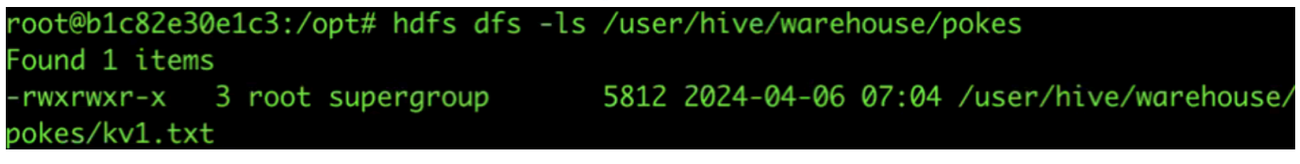
Figure 17-1. Hive architecture

Source: Hadoop: The Definitive Guide, 4th Edition

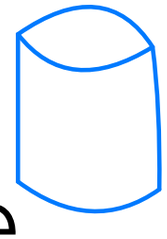
Play with Hive

```
sudo docker-compose exec hive-server bash
more /opt/hive/examples/files/kv1.txt
hive

SHOW TABLES;

CREATE TABLE pokes (foo INT, bar STRING);

DESCRIBE pokes;

LOAD DATA LOCAL INPATH '/opt/hive/examples/files/kv1.txt'
OVERWRITE INTO TABLE pokes;

SELECT * FROM pokes;

hadoop fs -ls /user/hive/warehouse/pokes
```





Traditional Database = Schema on Write

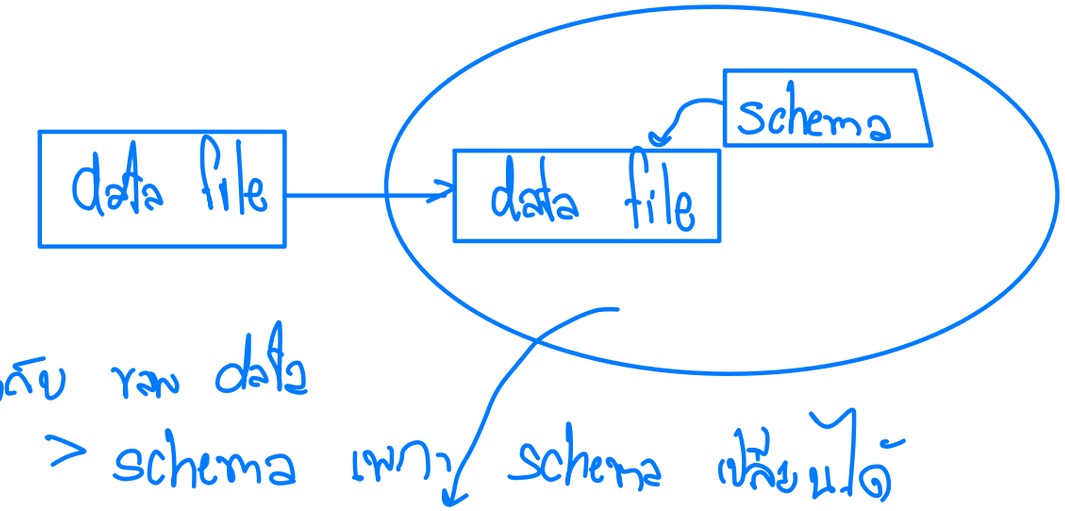


- Table's schema is enforced at data load time
- Stop if data being loaded does not conform to the schema
- Pros
  - Faster query performance
  - Known level of data quality before executing any queries
- Cons
  - Long time to load

Hive = Schema on Read

เปลง schema ตอนนำไปใช้ ข้อมูลไม่หาย

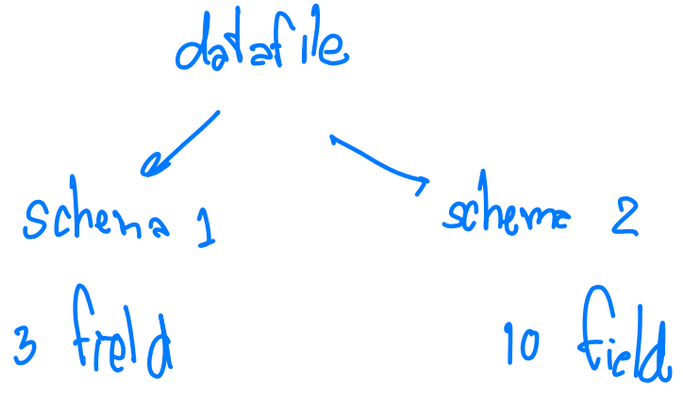
- Do not verify the data at load time, just copy the data file into the data store
- Enforce schema when read
- Pros
  - Faster initial load
  - More flexible for having multiple schemas on the same underlying data
  - Good for scenarios where the schema or index is not known at load time
- Cons
  - Slower query performance



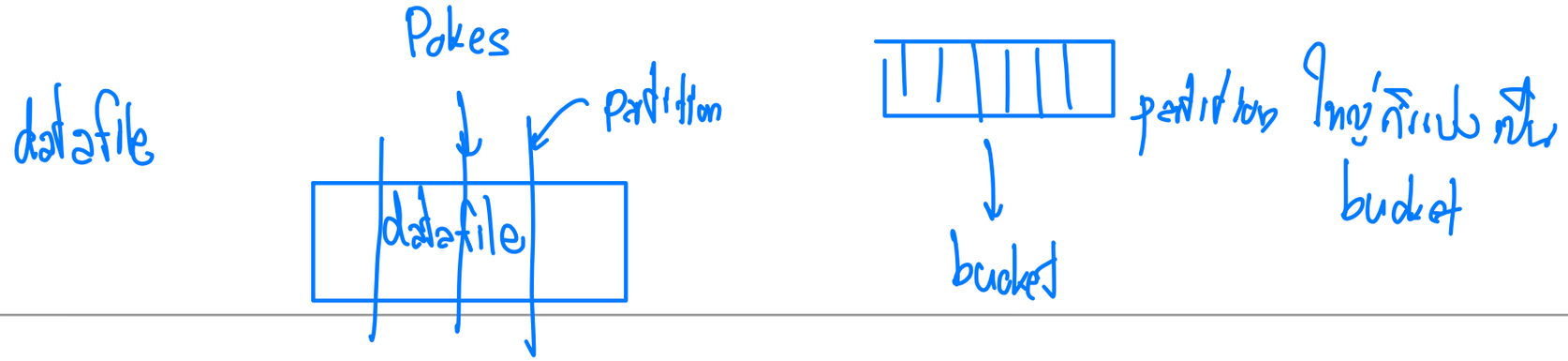
data sci ให้ความสำคัญกับ data > schema เพราะ schema เปลี่ยนได้

schema ยืดหยุ่น

ต้องแลกด้วยเวลา

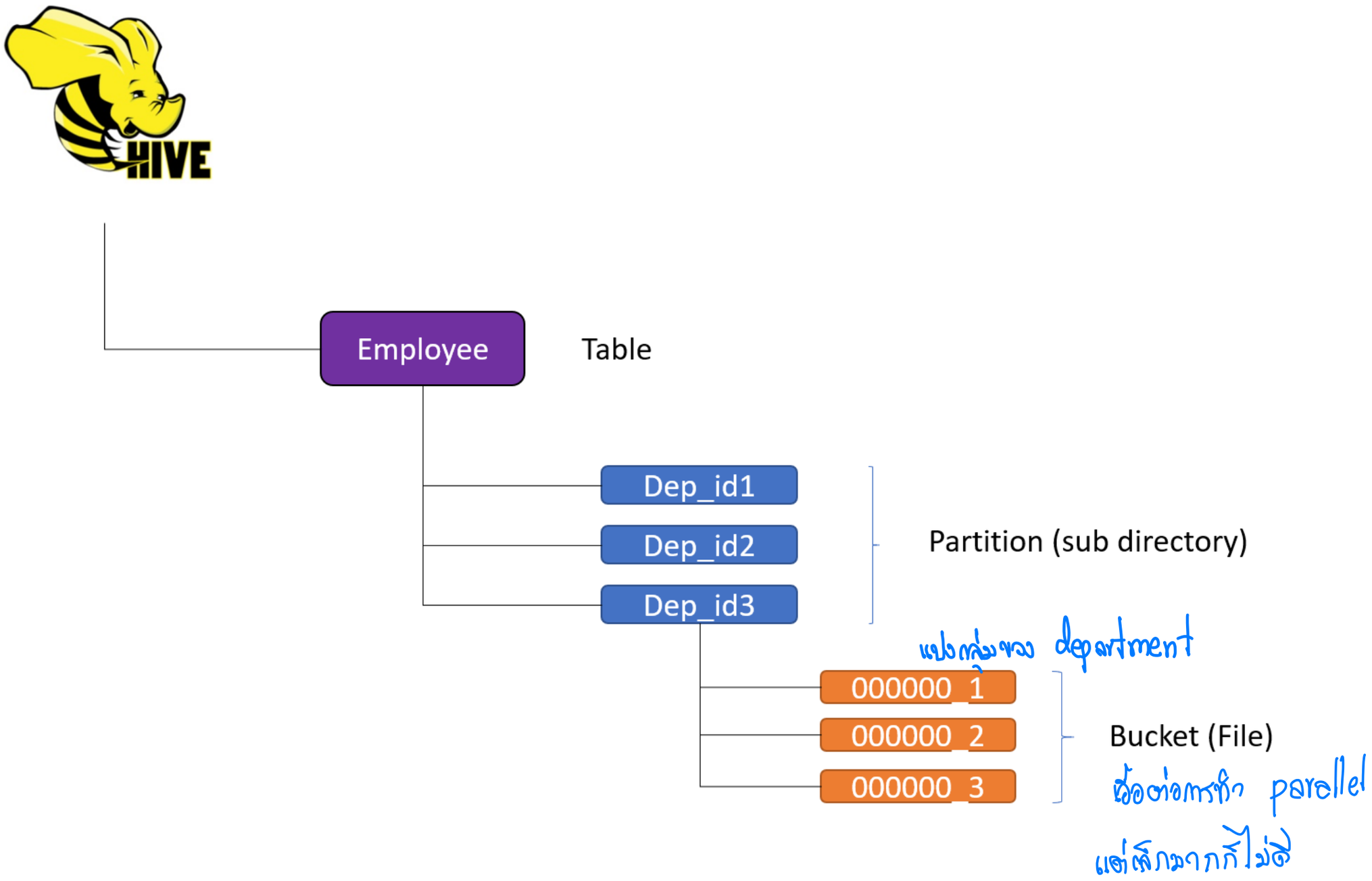


Data Units



- **Databases**: Namespaces function to avoid naming conflicts for tables, views, partitions, columns, and so on
- **Tables**: Homogeneous units of data which have the same schema
- **Partitions**: Each Table can have one or more partition Keys which determines how the data is stored, query only on the relevant partition of the table, thereby speeding up the analysis significantly
- **Buckets (or Clusters)**: Data in each partition may in turn be divided into Buckets based on the value of a hash function of some column of the Table
- Note that it is not necessary for tables to be partitioned or bucketed

Data Units and HDFS Storages เหมาะสม DB ใหญ่



Source: https://medium.com/nerd-for-tech/hive-data-organization-partitioning-clustering-3e14ef6ab121

Hive - Primitive Types

- Integers (TINYINT, SMALLINT, INT, BIGINT)
- Boolean type (BOOLEAN—TRUE/FALSE)
- Floating point numbers (FLOAT, DOUBLE)
- Fixed point numbers (DECIMAL)
- String types (STRING, VARCHAR, CHAR)
- Date and time types (TIMESTAMP, TIMESTAMP WITH LOCAL TIME ZONE, DATE)
- Binary types (BINARY)

Hive - Complex Types

- **Structs**: the elements within the type can be accessed using the DOT (.) notation

- **Maps** (key-value tuples): The elements are accessed using ['element name'] notation

- **Arrays** (indexable lists): The elements in the array can be accessed using the [n] notation where n is an index (zero-based) into the array

Built In Operators and Functions

- Relational Operators
- Arithmetic Operators
- Logical Operators
- Operators on Complex Types
- Built In Basic Functions
- Built In Aggregate Functions
- WHERE, SELECT, JOIN, SORT BY, GROUP BY
- User-Defined Functions

Big Data File Formats

- HDFS treats files as objects, can be any file formats
- Popular file formats
  - Text-based: txt, csv, json, XML
  - Binary-based: avro, parquet, orc ทำ random access ได้ดีกว่า binary base
- Benefits of binary-based file formats
  - Random-access, support parallel processing
  - On-the-fly compression, smaller size but not at the cost of reading speed อ่านข้างในหลัง
  - Embedded schema
  - Support table-oriented data store

คือ data รูปแบบ compress แต่มัน unzip ระหว่างทางเลย ไม่ต้อง unzip ภายหลัง

Table-Oriented Data Store

- Typical data analytics are done on “table” of data
- For big data, a table may not be fit in a single file
- Multiple files can be processed in parallel effectively
- Row-store file format is more efficient for add/update/delete operations
- Column-store file format is more efficient for data analytic operations

Table

	Country	Product	Sales
Row 1	India	Chocolate	1000
Row 2	India	Ice-cream	2000
Row 3	Germany	Chocolate	4000
Row 4	US	Noodle	500

Row Store

Row 1	India
	Chocolate
	1000
Row 2	India
	Ice-cream
	2000
Row 3	Germany
	Chocolate
	4000
Row 4	US
	Noodle
	500

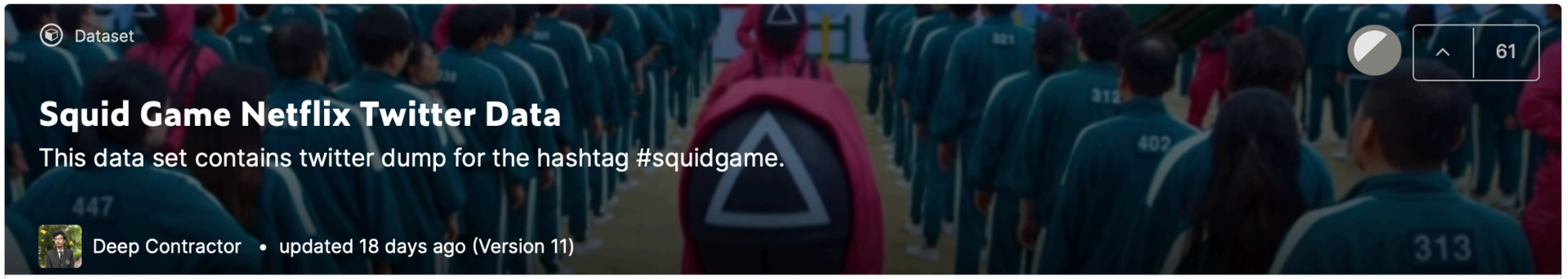
Column Store

Country	India
	India
	Germany
	US
Product	Chocolate
	Ice-cream
	Chocolate
	Noodle
Sales	1000
	2000
	4000
	500

Source: https://www.linkedin.com/pulse/real-time-big-data-analytics-parquet-spark-bonus-garren-staubli/

AVRO

- Row-based schema-oriented data serialization framework
  - Use JSON to define schema กำหนด schema ด้วย JSON ก่อน
  - Serialize data in a compact binary format
  - Similar to Java Object Serialization Pack row เป็น data ก้อนหนึ่ง
  - Support both primitive and complex data types (allows hierarchical data structure) integer, list, hierarchy
- Primary usage in Big Data
  - Apache Hadoop for persistent data (input/output/checkpoint)
  - A wire format for communication (e.g. Kafka, Spark Streaming) ข้อมูลที่ส่งผ่านบน network ส่งผ่าน socket
- Have libraries for many programming languages



Dataset

61

Squid Game Netflix Twitter Data

This data set contains twitter dump for the hashtag #squidgame.

Deep Contractor • updated 18 days ago (Version 11)

Data | Code (2) | Discussion | Activity | Metadata

Download (26 MB) | New Notebook

Usability 10.0 | **License** CC0: Public Domain | **Tags** arts and entertainment, online communities, text data, beginner, exploratory data analysis

Description

- The dataset contains the recent tweets about the record-breaking Netflix show "Squid Game"
- The data is collected using tweepy Python package to access Twitter API.

Data Explorer

261 MB

tweets_v8.csv (26.1 MB)

Detail | Compact | Column

คำอธิบาย (motto)

tweet

user_name	user_locati...	user_descr...	user_creat...	user_follo...	user_friends	user_favou...	user_verified	date	text
the _ûndër-ratèd niggáh🤎		@ManUtd die hard❤️❤️💪💪 YOLO J'ai besoin de quelqu'un qui peut m'aimer au pire🤗 Non, je ne...	2019-09-06 19:24:57+00:00	581	1035	8922	False	2021-10-06 12:05:38+00:00	When life hits and the same time poverty strikes you Gong Yoo : Lets play a game #SquidGame #Netfli...
Best uncle on planet earth			2013-05-08 19:35:26+00:00	741	730	8432	False	2021-10-06 12:05:22+00:00	That marble episode of #SquidGame ruined me. 😭😭😭
marcie		animal crossing. chicken nuggets. baby yoda. smol animals. tv shows. 🏳️‍🌈 pronouns: any	2009-02-21 10:31:30+00:00	562	1197	62732	False	2021-10-06 12:05:22+00:00	#Squidgame time
YoMo.Mdp	Any pronouns	Where the heck is the karma I'm going on my school grave brb #Technosupport	2021-02-14 13:21:22+00:00	3	277	1341	False	2021-10-06 12:05:04+00:00	//Blood on 1st slide I'm joining the squidgame thing, I'm already dead by sugar honeycomb ofc #Squi...
Laura Reactions	France	I talk and I make reactions videos about shows I love #theexpanse #peakyblinders #thelastkingdom #la...	2018-12-19 20:38:28+00:00	330	152	2278	False	2021-10-06 12:05:00+00:00	The two first games, players were killed by the mask guys ; the bloody night and the third game, the...
Peyman 🅚🅐🅘	United Kingdom	Official @KardiaChain $KAI Ambassador Marketing Advisor @kephigallery Graphics and Film Artist https...	2018-01-27 12:07:31+00:00	546	318	6265	False	2021-10-06 12:04:54+00:00	$THG Going to explode to 4B Marketcap very soon. The world first MOBA This game is on another level!...
Aeriaaaa♡		Fujoshi 🙈/ Thai BL-obsessed/Always distracted by poetry 🌸/ CAP 🦙	2021-06-01 14:08:10+00:00	14	110	518	False	2021-10-06 12:04:45+00:00	@B_hundred_Hyun pls use that gun on me. 😩 #BAEKHYUN #EXO #weareoneEXO #SquidGame https://t.co/ksk...

AVRO Schema File

JSON Schema

complex types (record, enum, array, map, union, and fixed)

```
{
    "namespace": "example.avro.chula",
    "type": "record",
    "name": "squid_tweets",
    "fields": [
        {"name": "user_name", "type": "string"},
        {"name": "user_location", "type": ["null", "string"]},
        {"name": "user_description", "type": "string"},
        {"name": "user_created", "type": {"type": "long", "logicalType": "local-timestamp-millis"}},
        {"name": "user_followers", "type": "int"},
        {"name": "user_friends", "type": "int"},
        {"name": "user_favourites", "type": "int"},
        {"name": "user_verified", "type": "boolean"},
        {"name": "date", "type": {"type": "long", "logicalType": "local-timestamp-millis"}},
        {"name": "text", "type": "string"},
        {"name": "source", "type": "string"},
        {"name": "is_retweet", "type": "boolean"}
    ]
}
```

optional

primitive types (null, boolean, int, long, float, double, bytes, and string)

AVRO Example in Python

In [1]:
```
import pandas as pd
import fastavro as favro
import json, os
```

In [2]:
```
schema = json.load(open('tweets_v8.avsc'))
parsed_schema = favro.parse_schema(schema)
```

เตรียม AVRO schema JSON file

In [3]:
```
parse_dates = ['user_created', 'date']
df = pd.read_csv('tweets_v8.csv', parse_dates=parse_dates, keep_default_na=False)
df.shape
```

Out[3]: (80019, 12)

In [4]:
```
records = df.to_dict('records')
```

In [5]:
```
with open('tweets_v8.avro', 'wb') as out:
    favro.writer(out, parsed_schema, records)
```

} เขียน file

In [6]:
```
with open('tweets_v8_deflate.avro', 'wb') as out:
    favro.writer(out, parsed_schema, records, codec='deflate')
```

} compressed

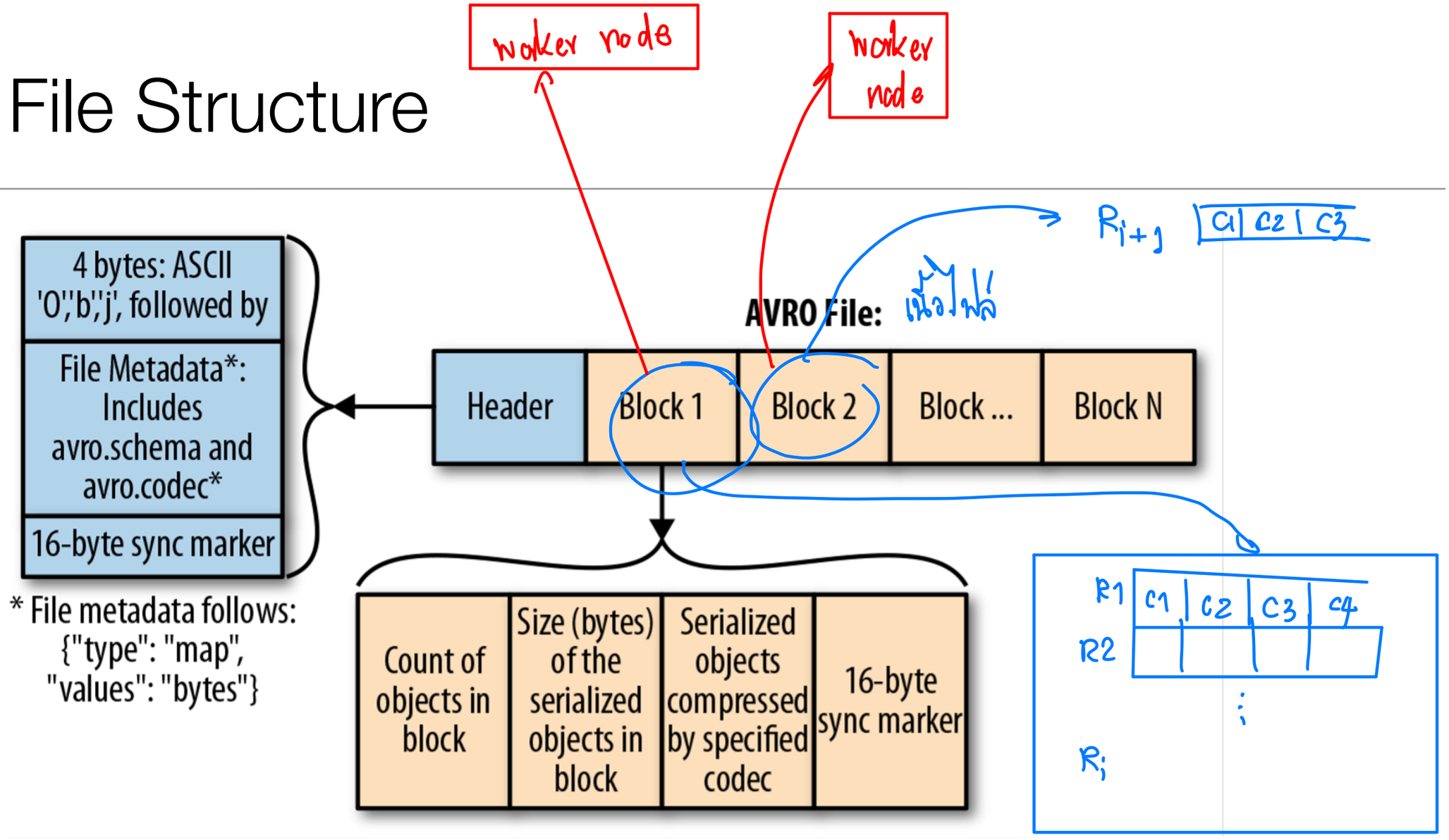
In [7]:
```
print('tweets_v8.csv = {:5.2f}MB'.format(os.path.getsize('tweets_v8.csv')/(1024*1024)))
print('tweets_v8.avro = {:5.2f}MB'.format(os.path.getsize('tweets_v8.avro')/(1024*1024)))
print('tweets_v8_deflate.avro = {:5.2f}MB'.format(os.path.getsize('tweets_v8_deflate.avro')/(1024*1024)))
```

```
tweets_v8.csv = 24.89MB
tweets_v8.avro = 20.49MB
tweets_v8_deflate.avro = 11.51MB
```

avro file

compressed

AVRO File Structure



- Each row is “serialized”, multiple rows store in each block
- Data compressed per block
- Block's sync marker is for MapReduce file splitting

```
00000000  4F 62 6A 01 04 14 61 76 72 6F 2E 63 6F 64 65 63  Obj...avro.codec
00000010  08 6E 75 6C 6C 16 61 76 72 6F 2E 73 63 68 65 6D  .null.avro.schem
00000020  61 DC 0A 7B 22 74 79 70 65 22 3A 20 22 72 65 63  a▄.{"type": "rec
00000030  6F 72 64 22 2C 20 22 6E 61 6D 65 22 3A 20 22 65  ord", "name": "e
00000040  78 61 6D 70 6C 65 2E 61 76 72 6F 2E 63 68 75 6C  xample.avro.chul
00000050  61 2E 73 71 75 69 64 5F 74 77 65 65 74 73 22 2C  a.squid_tweets",
00000060  20 22 66 69 65 6C 64 73 22 3A 20 5B 7B 22 6E 61   "fields": [{"na
00000070  6D 65 22 3A 20 22 75 73 65 72 5F 6E 61 6D 65 22  me": "user_name"
00000080  2C 20 22 74 79 70 65 22 3A 20 22 73 74 72 69 6E  , "type": "strin
00000090  67 22 7D 2C 20 7B 22 6E 61 6D 65 22 3A 20 22 75  g"}, {"name": "u
000000A0  73 65 72 5F 6C 6F 63 61 74 69 6F 6E 22 2C 20 22  ser_location", "
000000B0  74 79 70 65 22 3A 20 5B 22 6E 75 6C 6C 22 2C 20  type": ["null",
000000C0  22 73 74 72 69 6E 67 22 5D 7D 2C 20 7B 22 6E 61  "string"]}, {"na
000000D0  6D 65 22 3A 20 22 75 73 65 72 5F 64 65 73 63 72  me": "user_descr
000000E0  69 70 74 69 6F 6E 22 2C 20 22 74 79 70 65 22 3A  iption", "type":
000000F0  20 22 73 74 72 69 6E 67 22 7D 2C 20 7B 22 6E 61   "string"}, {"na
00000100  6D 65 22 3A 20 22 75 73 65 72 5F 63 72 65 61 74  me": "user_creat
00000110  65 64 22 2C 20 22 74 79 70 65 22 3A 20 7B 22 6C  ed", "type": {"l
00000120  6F 67 69 63 61 6C 54 79 70 65 22 3A 20 22 6C 6F  ogicalType": "lo
```

```
#SquidGame… https://t.co/N4UGv9hxx8",Twitter Web App,False
Laura Reactions,France,I talk and I make reactions videos about shows I love #theexpanse #peakyblinders #thelastkingdom
#lacasadepapel #atla #theboys #thewitcher,2018-12-19 20:38:28+00:00,330,152,2278,False,2021-10-06 12:05:00+00:00,"The two
first games, players were killed by the mask guys ; the bloody night and the third game, they killed each o…
https://t.co/Qf057XDJ7C",Twitter Web App,False
Peyman 🅚🅐🅘,United Kingdom,"Official @KardiaChain $KAI Ambassador
Marketing Advisor @kephigallery
Graphics and Film Artist https://t.co/OsTOSEKrHt",2018-01-27 12:07:31+00:00,546,318,6265,False,2021-10-06
12:04:54+00:00,"$THG
```

row เดียวกัน เก็บสีกัน

00000690	74	65	72	20	57	65	62	20	41	70	70	00	1E	4C	61	75	ter Web App..Lau
000006A0	72	61	20	52	65	61	63	74	69	6F	6E	73	02	0C	46	72	ra Reactions..Fr
000006B0	61	6E	63	65	94	02	49	20	74	61	6C	6B	20	61	6E	64	anceö.I talk and
000006C0	20	49	20	6D	61	6B	65	20	72	65	61	63	74	69	6F	6E	I make reaction
000006D0	73	20	76	69	64	65	6F	73	20	61	62	6F	75	74	20	73	s videos about s
000006E0	68	6F	77	73	20	49	20	6C	6F	76	65	20	23	74	68	65	hows I love #the
000006F0	65	78	70	61	6E	73	65	20	23	70	65	61	6B	79	62	6C	expanse #peakybl
00000700	69	6E	64	65	72	73	20	23	74	68	65	6C	61	73	74	6B	inders #thelastk
00000710	69	6E	67	64	6F	6D	20	23	6C	61	63	61	73	61	64	65	ingdom #lacasade
00000720	70	61	70	65	6C	20	23	61	74	6C	61	20	23	74	68	65	papel #atla #the
00000730	62	6F	79	73	20	23	74	68	65	77	69	74	63	68	65	72	boys #thewitcher
00000740	C0	A6	87	83	F9	59	94	05	B0	02	CC	23	00	C0	E7	F0	└ªçâ·Yö.\.╠#.└τ≡
00000750	D7	8A	5F	9C	02	54	68	65	20	74	77	6F	20	66	69	72	╫è_£.The two fir
00000760	73	74	20	67	61	6D	65	73	2C	20	70	6C	61	79	65	72	st games, player
00000770	73	20	77	65	72	65	20	6B	69	6C	6C	65	64	20	62	79	s were killed by
00000780	20	74	68	65	20	6D	61	73	6B	20	67	75	79	73	20	3B	the mask guys ;
00000790	20	74	68	65	20	62	6C	6F	6F	64	79	20	6E	69	67	68	the bloody nigh
000007A0	74	20	61	6E	64	20	74	68	65	20	74	68	69	72	64	20	t and the third
000007B0	67	61	6D	65	2C	20	74	68	65	79	20	6B	69	6C	6C	65	game, they kille
000007C0	64	20	65	61	63	68	20	6F	E2	80	A6	20	68	74	74	70	d each oΓÇª http
000007D0	73	3A	2F	2F	74	2E	63	6F	2F	51	66	30	35	37	58	44	s://t.co/Qf057XD
000007E0	4A	37	43	1E	54	77	69	74	74	65	72	20	57	65	62	20	J7C.Twitter Web
000007F0	41	70	70	00	26	50	65	79	6D	61	6E	20	F0	9F	85	9A	App.&Peyman ≡ƒàÜ

Parquet

- Columnar schema-oriented data store
- Analytic operations are usually column-based pandas ถนัด per column โอกาสใช้ for loop ใน pandas น้อยมาก
- Schema is embedded in the file
- Support nested schema using record shredding and assembly algorithm (Google Dremel) to map hierarchical data into column-based data store
- Support column-wise data compression compress ตาม column
  - More efficient as each column shares the same datatypes
  - Run length encoding, dictionary encoding, delta encoding
- Support in multiple languages, as well as, pandas, MapReduced, and spark

Parquet Example in Python

In [1]:
```
import pandas as pd
import os
```

In [2]:
```
parse_dates = ['user_created', 'date']
df = pd.read_csv('tweets_v8.csv', parse_dates=parse_dates, keep_default_na=False)
df.shape
```

Out[2]: (80019, 12)

pandas parquet

In [3]:
```
df.to_parquet('tweets_v8.par', compression=None)
```

In [4]:
```
df.to_parquet('tweets_v8_snappy.par')
```

In [5]:
```
print('tweets_v8.csv = {:5.2f}MB'.format(os.path.getsize('tweets_v8.csv')/(1024*1024)))
print('tweets_v8.par = {:5.2f}MB'.format(os.path.getsize('tweets_v8.par')/(1024*1024)))
print('tweets_v8_snappy.par = {:5.2f}MB'.format(os.path.getsize('tweets_v8_snappy.par')/(1024*1024)))
```

```
tweets_v8.csv = 24.89MB
tweets_v8.par = 22.98MB
tweets_v8_snappy.par = 14.19MB
```

```
I'm joining the squidgame thing, I'm already dead by sugar honeycomb ofc

#SquidGame… https://t.co/N4UGv9hxx8",Twitter Web App,False
Laura Reactions,France,I talk and I make reactions videos about shows I love #theexpanse #peakyblinders #thelastkingdom
#lacasadepapel #atla #theboys #thewitcher,2018-12-19 20:38:28+00:00,330,152,2278,False,2021-10-06 12:05:00+00:00,"The two
first games, players were killed by the mask guys ; the bloody night and the third game, they killed each o…
https://t.co/Qf057XDJ7C",Twitter Web App,False
Peyman 🅚🅐🅘,United Kingdom,"Official @KardiaChain $KAI Ambassador
Marketing Advisor @kephigallery
Graphics and Film Artist https://t.co/OsTOSEKrHt",2018-01-27 12:07:31+00:00,546,318,6265,False,2021-10-06
12:04:54+00:00,"$THG
```

```
00000000  50 41 52 31 15 00 15 92 CE BC 01 15 92 CE BC 01  PAR1...Æ╠..Æ╠.
00000010  2C 15 A6 E2 09 15 00 15 06 15 08 00 00 04 00 00  ,.ª┌............
00000020  00 A6 E2 09 01 23 00 00 00 74 68 65 20 5F C3 BB  .ª┌..#...the _├╗
00000030  6E 64 C3 AB 72 2D 72 61 74 C3 A8 64 20 6E 69 67  nd├½r-rat├¿d nig
00000040  67 C3 A1 68 F0 9F 91 8A F0 9F 8F BE 1A 00 00 00  g├íh≡ƒæè≡ƒÅ╛....
00000050  42 65 73 74 20 75 6E 63 6C 65 20 6F 6E 20 70 6C  Best uncle on pl
00000060  61 6E 65 74 20 65 61 72 74 68 06 00 00 00 6D 61  anet earth....ma
00000070  72 63 69 65 08 00 00 00 59 6F 4D 6F 2E 4D 64 70  rcie....YoMo.Mdp
00000080  0F 00 00 00 4C 61 75 72 61 20 52 65 61 63 74 69  ....Laura Reacti
00000090  6F 6E 73 13 00 00 00 50 65 79 6D 61 6E 20 F0 9F  ons....Peyman ≡ƒ
000000A0  85 9A F0 9F 85 90 F0 9F 85 98 0B 00 00 00 41 65  àÜ≡ƒàÉ≡ƒàÿ....Ae
000000B0  72 69 61 61 61 61 E2 99 A1 19 00 00 00 42 61 72  riaaaa┌Öí....Bar
000000C0  42 69 45 20 46 C3 A1 62 72 65 67 61 73 20 F0 9F  BiE F├íbregas ≡ƒ
000000D0  87 BF F0 9F 87 A6 0E 00 00 00 4A 6F 65 6C 20 44  ç┐≡ƒçª....Joel D
000000E0  2E 20 50 61 72 6B 65 72 0C 00 00 00 4B 65 76 69  . Parker....Kevi
000000F0  6E 20 46 72 61 6E 63 6F 0D 00 00 00 54 72 65 61  n Franco....Trea
00000100  73 75 72 65 54 72 65 73 73 09 00 00 00 73 6F 70  sureTress....sop
00000110  68 20 F0 9F 8D 81 0B 00 00 00 41 6C 69 79 61 68  h ≡ƒìü....Aliyah
00000120  20 F0 9F 8C A4 0C 00 00 00 4D 65 67 61 20 56 69  ≡ƒîñ....Mega Vi
00000130  73 69 6F 6E 73 06 00 00 00 59 65 6D 7A 78 79 13  sions....Yemzxy.
00000140  00 00 00 6B 69 6D 20 6A 6F 6F 20 72 79 6F 75 6E  ...kim joo ryoun
00000150  67 20 70 69 63 73 0A 00 00 00 47 65 72 69 61 74  g pics....Geriat
00000160  72 69 78 78 1A 00 00 00 4D 6F 76 69 65 6C 6F 76  rixx....Movielov
```

Column เดียวกัน

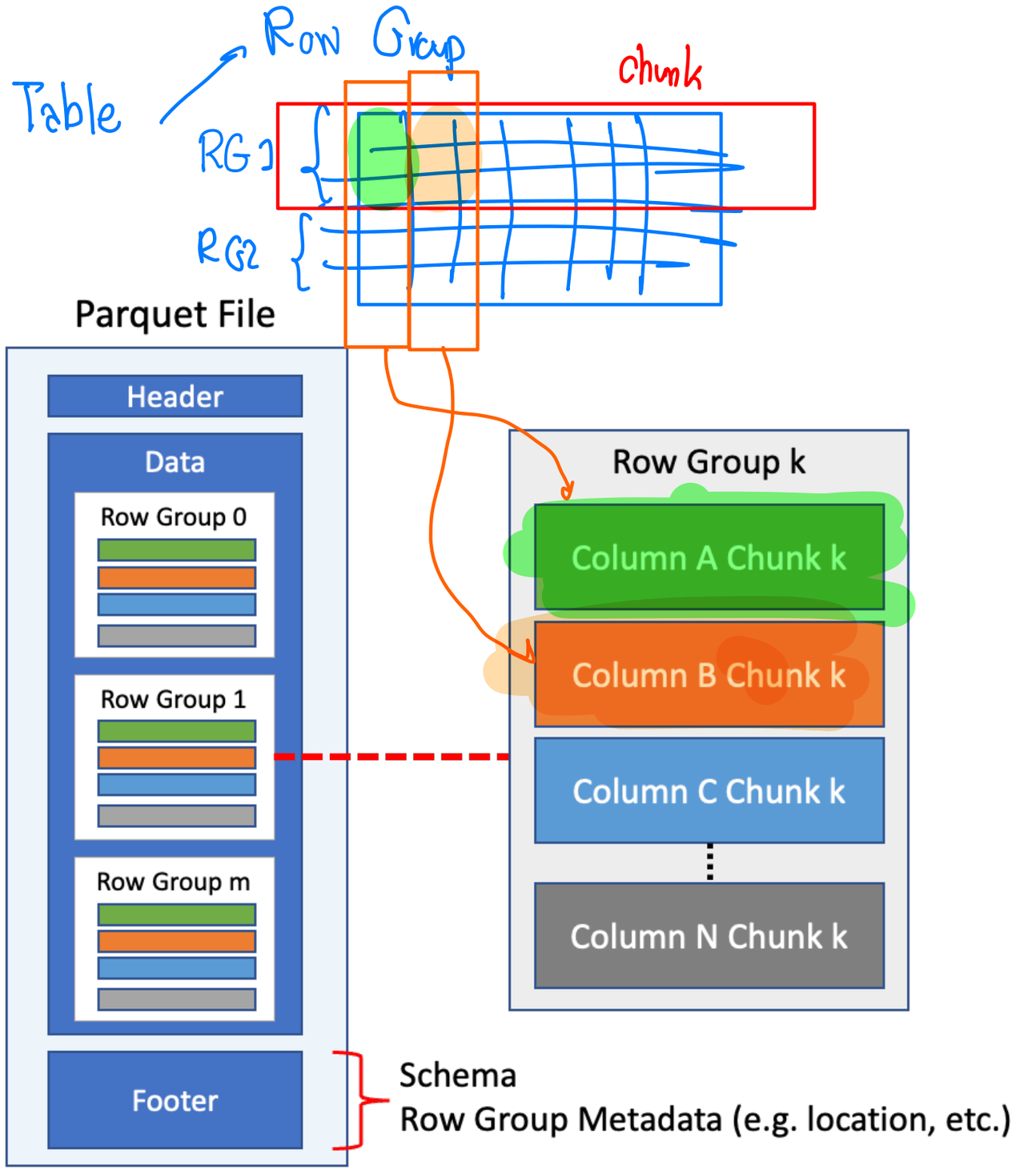
```
I'm joining the squidgame thing, I'm already dead by sugar honeycomb ofc

#SquidGame… https://t.co/N4UGv9hxx8",Twitter Web App,False
Laura Reactions,France,I talk and I make reactions videos about shows I love #theexpanse #peakyblinders #thelastkingdom
#lacasadepapel #atla #theboys #thewitcher,2018-12-19 20:38:28+00:00,330,152,2278,False,2021-10-06 12:05:00+00:00,"The two
first games, players were killed by the mask guys ; the bloody night and the third game, they killed each o…
https://t.co/Qf057XDJ7C",Twitter Web App,False
Peyman 🅚🅐🅘,United Kingdom,"Official @KardiaChain $KAI Ambassador
Marketing Advisor @kephigallery
Graphics and Film Artist https://t.co/OsTOSEKrHt",2018-01-27 12:07:31+00:00,546,318,6265,False,2021-10-06
12:04:54+00:00,"$THG
```

```
00292600  6E 6F 73 75 70 70 6F 72 74 8A 00 00 00 49 20 74  nosupportè...I t
00292610  61 6C 6B 20 61 6E 64 20 49 20 6D 61 6B 65 20 72  alk and I make r
00292620  65 61 63 74 69 6F 6E 73 20 76 69 64 65 6F 73 20  eactions videos
00292630  61 62 6F 75 74 20 73 68 6F 77 73 20 49 20 6C 6F  about shows I lo
00292640  76 65 20 23 74 68 65 65 78 70 61 6E 73 65 20 23  ve #theexpanse #
00292650  70 65 61 6B 79 62 6C 69 6E 64 65 72 73 20 23 74  peakyblinders #t
00292660  68 65 6C 61 73 74 6B 69 6E 67 64 6F 6D 20 23 6C  helastkingdom #l
00292670  61 63 61 73 61 64 65 70 61 70 65 6C 20 23 61 74  acasadepapel #at
00292680  6C 61 20 23 74 68 65 62 6F 79 73 20 23 74 68 65  la #theboys #the
00292690  77 69 74 63 68 65 72 76 00 00 00 4F 66 66 69 63  witcherv...Offic
002926A0  69 61 6C 20 40 4B 61 72 64 69 61 43 68 61 69 6E  ial @KardiaChain
002926B0  20 24 4B 41 49 20 41 6D 62 61 73 73 61 64 6F 72   $KAI Ambassador
002926C0  0A 4D 61 72 6B 65 74 69 6E 67 20 41 64 76 69 73  .Marketing Advis
002926D0  6F 72 20 40 6B 65 70 68 69 67 61 6C 6C 65 72 79  or @kephigallery
002926E0  0A 47 72 61 70 68 69 63 73 20 61 6E 64 20 46 69  .Graphics and Fi
002926F0  6C 6D 20 41 72 74 69 73 74 20 68 74 74 70 73 3A  lm Artist https:
00292700  2F 2F 74 2E 63 6F 2F 4F 73 54 4F 53 45 4B 72 48  //t.co/OsTOSEKrH
00292710  74 4C 00 00 00 46 75 6A 6F 73 68 69 20 F0 9F 99  tL...Fujoshi ≡ƒÖ
00292720  88 2F 20 54 68 61 69 20 42 4C 2D 6F 62 73 65 73  ê/ Thai BL-obses
00292730  73 65 64 2F 41 6C 77 61 79 73 20 64 69 73 74 72  sed/Always distr
00292740  61 63 74 65 64 20 62 79 20 70 6F 65 74 72 79 20  acted by poetry
00292750  F0 9F 8C B8 2F 20 43 41 50 20 E2 99 91 F0 9F 90  ≡ƒîŋ/ CAP ΓÖæ≡ƒÉ
```

ไม่ใช้อ่าน column ต้นฉบับ

Parquet File Structure

- Header (4 bytes) ‘PAR1’ — indicate Parquet file format
- Data contains set of row groups
  - Partition data into **row groups** (m row groups in the example) — for MapReduce parallelization
  - Each row group contains all columns (from A to N in the example) and metadata of the group
  - Column data in the same row group is stored in **a column chunk** with metadata of the chunk — for I/O parallelization
- Footer contains schema and metadata of the file (e.g. location of metadata of each row group and each column in row group)



Assignment Readings

- K. Shvachko, et al. "The hadoop distributed file system." 2010 IEEE 26th symposium on mass storage systems and technologies (MSST), 2010.

- A. Thusoo, et al. "Hive-a petabyte scale data warehouse using hadoop", 2010 IEEE 26th international conference on data engineering (ICDE 2010), 2010.

References

- Hadoop: The Definitive Guide, 4th Edition
- https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html
- https://www.geeksforgeeks.org/anatomy-of-file-read-and-write-in-hdfs/
- https://data-flair.training/blogs/top-hadoop-hdfs-commands-tutorial/
- https://data-flair.training/blogs/data-locality-in-hadoop-mapreduce/
- https://community.cloudera.com/t5/Community-Articles/Understanding-basics-of-HDFS-and-YARN/ta-p/248860
- https://cwiki.apache.org/confluence/display/Hive/Tutorial
- https://www.javatpoint.com/dynamic-partitioning-in-hive