

# Google MapReduce

*DS 5110: Big Data Systems (Spring 2023)*

Lecture 3b

Yue Cheng



UNIVERSITY  
*of* VIRGINIA

## Applications

Batch

SQL

ETL

Machine  
learning

Emerging  
apps?

Scalable computing engines

Scalable storage systems



Datacenter infrastructure



# The big picture (motivation)

- Datasets are **too big** to process using a single computer

# The big picture (motivation)

- Datasets are **too big** to process using a single computer
- Good parallel processing engines are **rare (back then in the late 90s)**

# The big picture (motivation)

- Datasets are **too big** to process using a single computer
- Good parallel processing engines are **rare (back then in the late 90s)**
- Want a parallel processing framework that:
  - is **general** (works for many problems)
  - is **easy to use** (no locks, no need to explicitly handle communication, no race conditions)
  - can **automatically parallelize** tasks
  - can **automatically handle machine failures**

# Context (Google circa 2000)

- Starting to deal with **massive** datasets
- But also addicted to cheap, unreliable hardware
  - Young company, expensive hardware not practical
- Only a few expert programmers can write distributed programs to process them
  - Scale so large jobs can complete before failures



# Context (Google circa 2000)

- Starting to deal with **massive** datasets
- But also addicted to cheap, unreliable hardware
  - Young company, expensive hardware not practical
- Only a few expert programmers can write distributed programs to process them
  - Scale so large jobs can complete before failures
- **Key question:** how can every Google engineer be imbued with the ability to write **parallel, scalable, distributed, fault-tolerant** code?
- **Solution:** **abstract out** the redundant parts
- **Restriction:** relies on job semantics, so restricts which problems it works for

# Application: Word Count

```
cat data.txt  
  | tr -s '[:punct:][:space:]' '\n'  
  | sort | uniq -c
```

```
SELECT count(word), word FROM data  
GROUP BY word
```



# Deal with multiple files?

# Deal with multiple files?

1. Compute word counts from individual files

# Deal with multiple files?

1. Compute word counts from individual files
2. Then merge intermediate output

# Deal with multiple files?

1. Compute word counts from individual files
2. Then merge intermediate output
3. Compute word count on merged outputs

# What if the data is too big to fit in one computer?

# What if the data is too big to fit in one computer?

1. In parallel, send to worker:
  - Compute word counts from individual files
  - Collect results, wait until all finished

# What if the data is too big to fit in one computer?

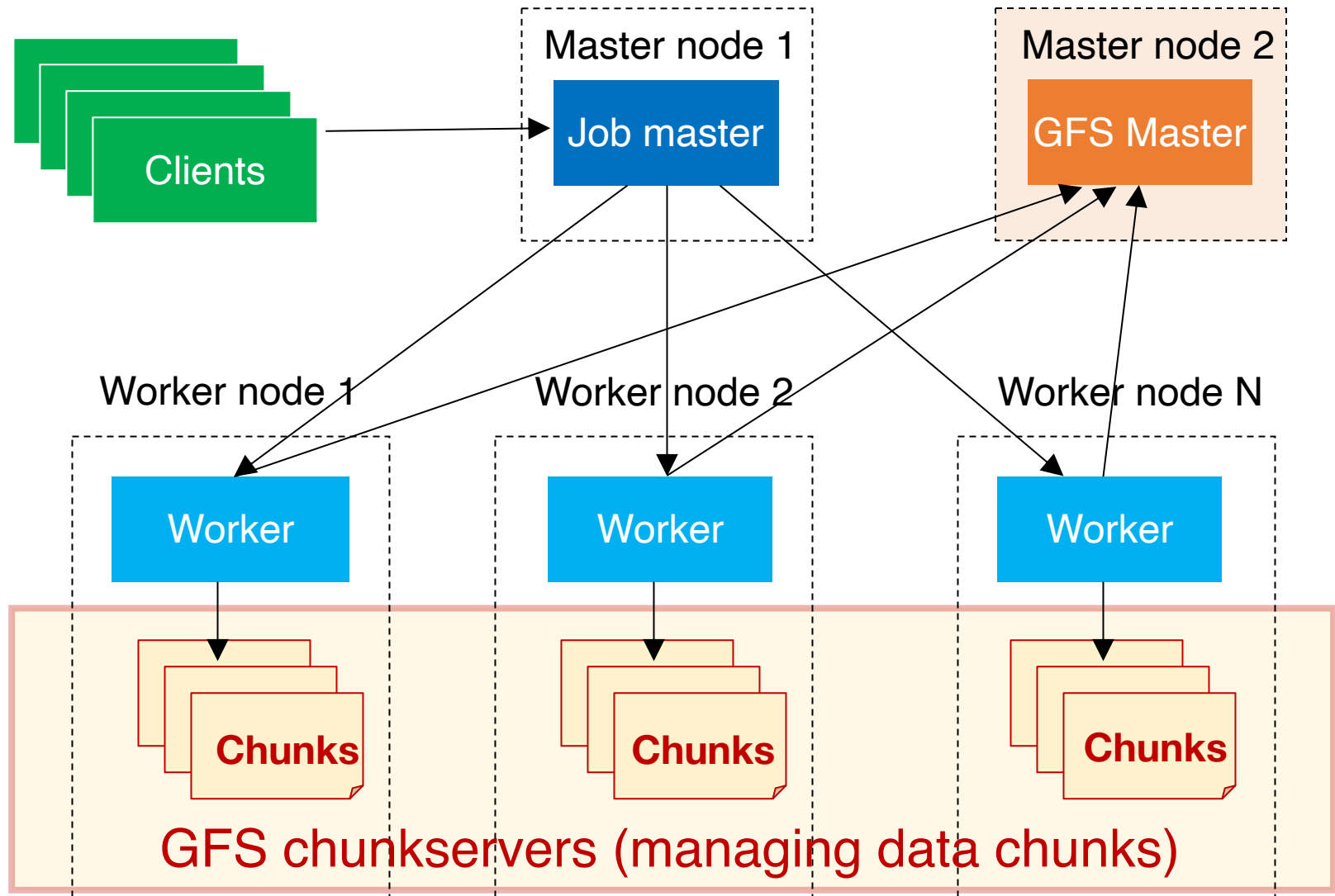
1. In parallel, send to worker:
  - Compute word counts from individual files
  - Collect results, wait until all finished
2. Then merge intermediate output

# What if the data is too big to fit in one computer?

1. In parallel, send to worker:
  - Compute word counts from individual files
  - Collect results, wait until all finished
2. Then merge intermediate output
3. Compute word count on merged intermediates



# MapReduce+GFS: Put everything together



# MapReduce: Programming interface

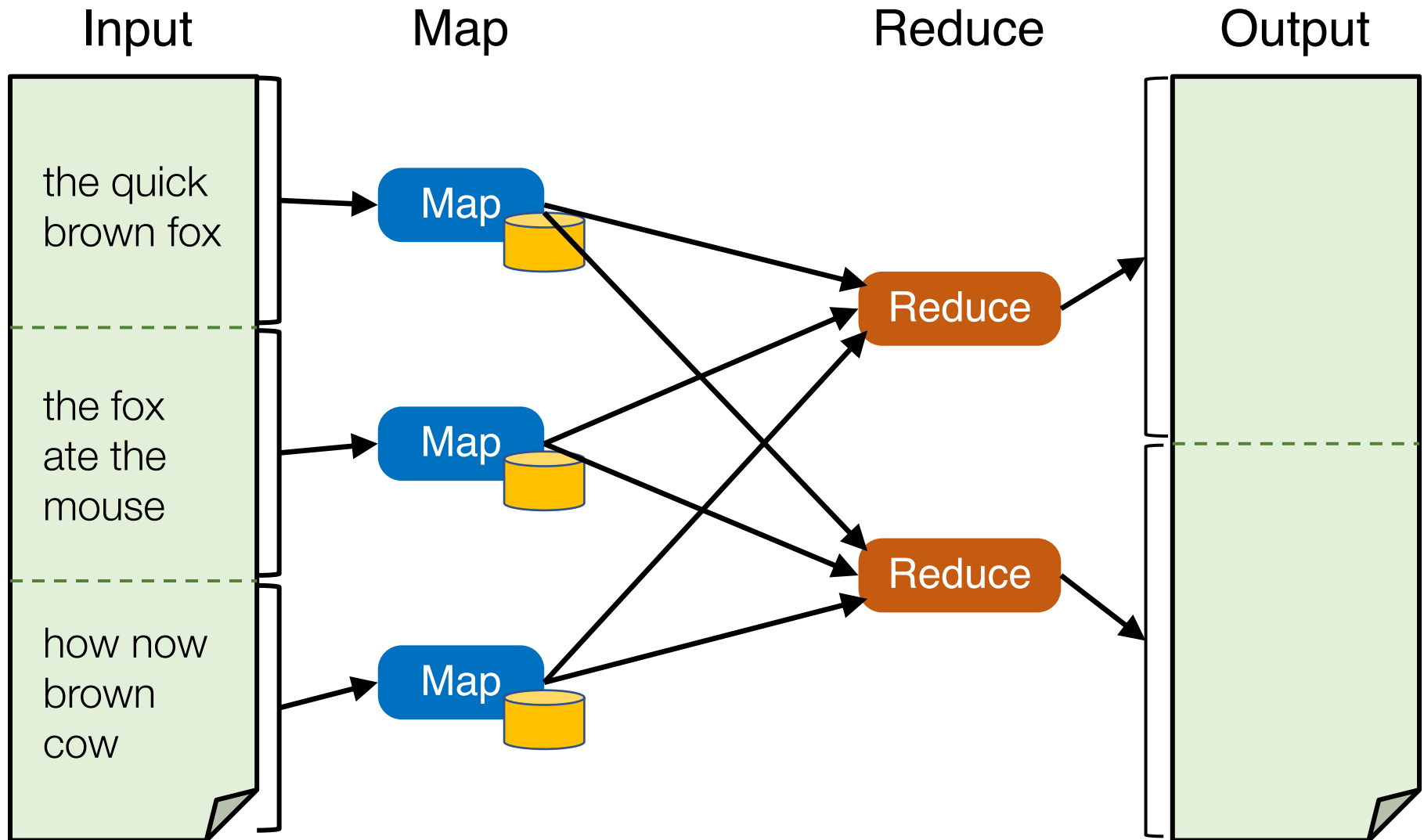
- $\text{map}(k1, v1) \rightarrow \text{list}(k2, v2)$ 
  - Apply function to  $(k1, v1)$  pair and produce set of intermediate pairs  $(k2, v2)$
  
- $\text{reduce}(k2, \text{list}(v2)) \rightarrow \text{list}(k3, v3)$ 
  - Apply aggregation (reduce) function to values
  - Output results

# MapReduce: Word Count

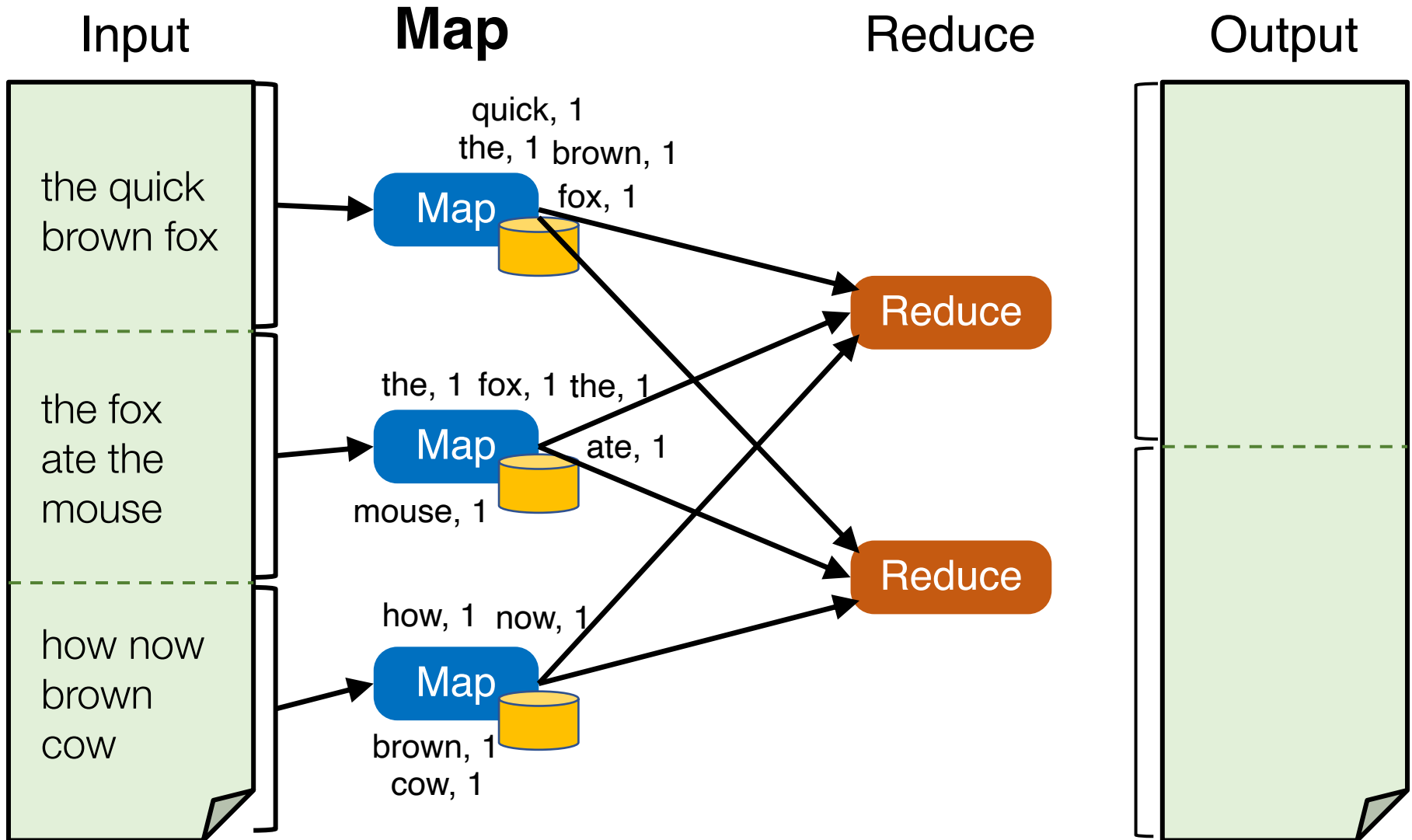
```
map(key, value):  
    for each word w in value:  
        EmitIntermediate(w, "1");
```

```
reduce(key, values):  
    int result = 0;  
    for each v in values:  
        result += ParseInt(v);  
    Emit(AsString(result));
```

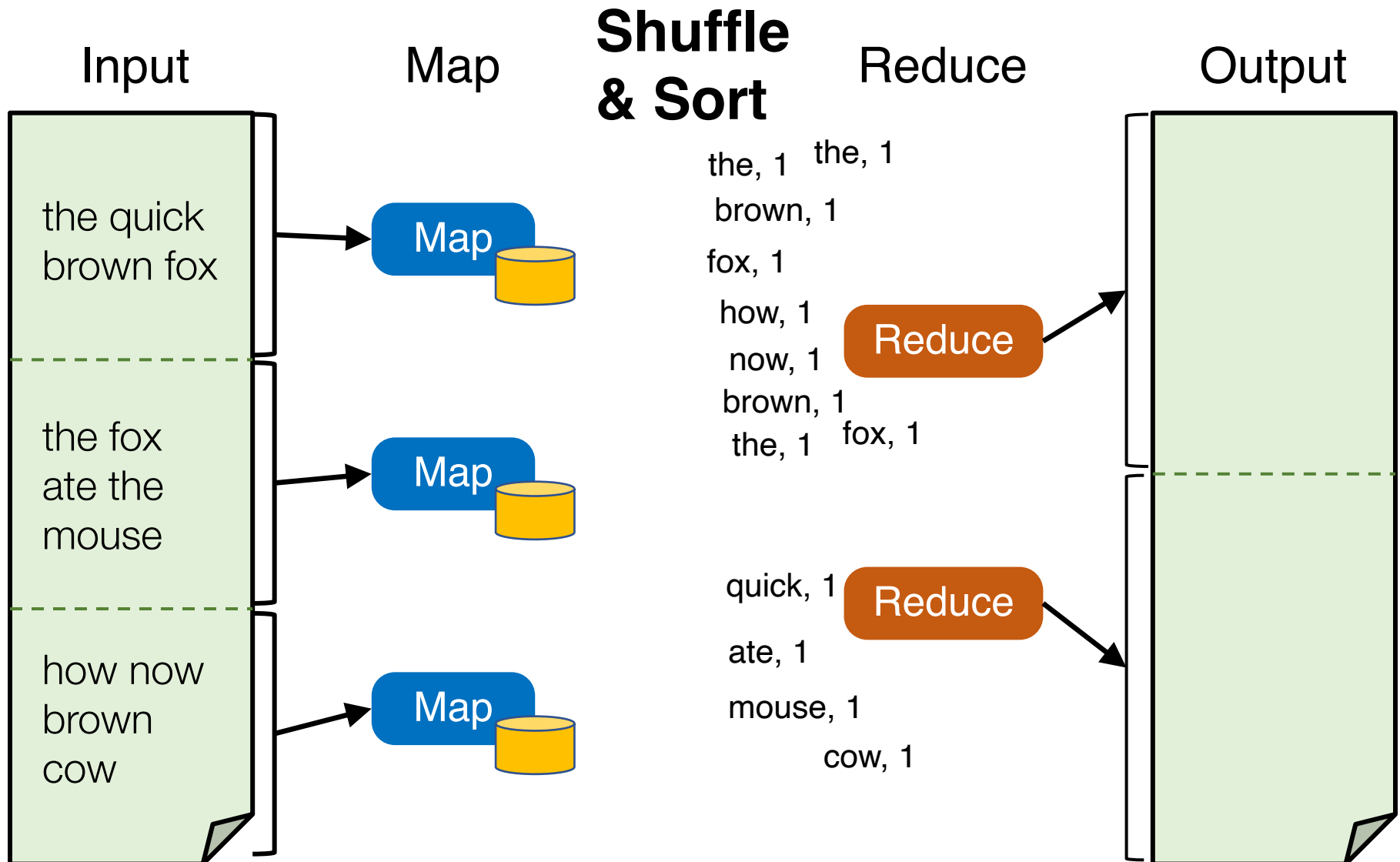
# Word Count execution



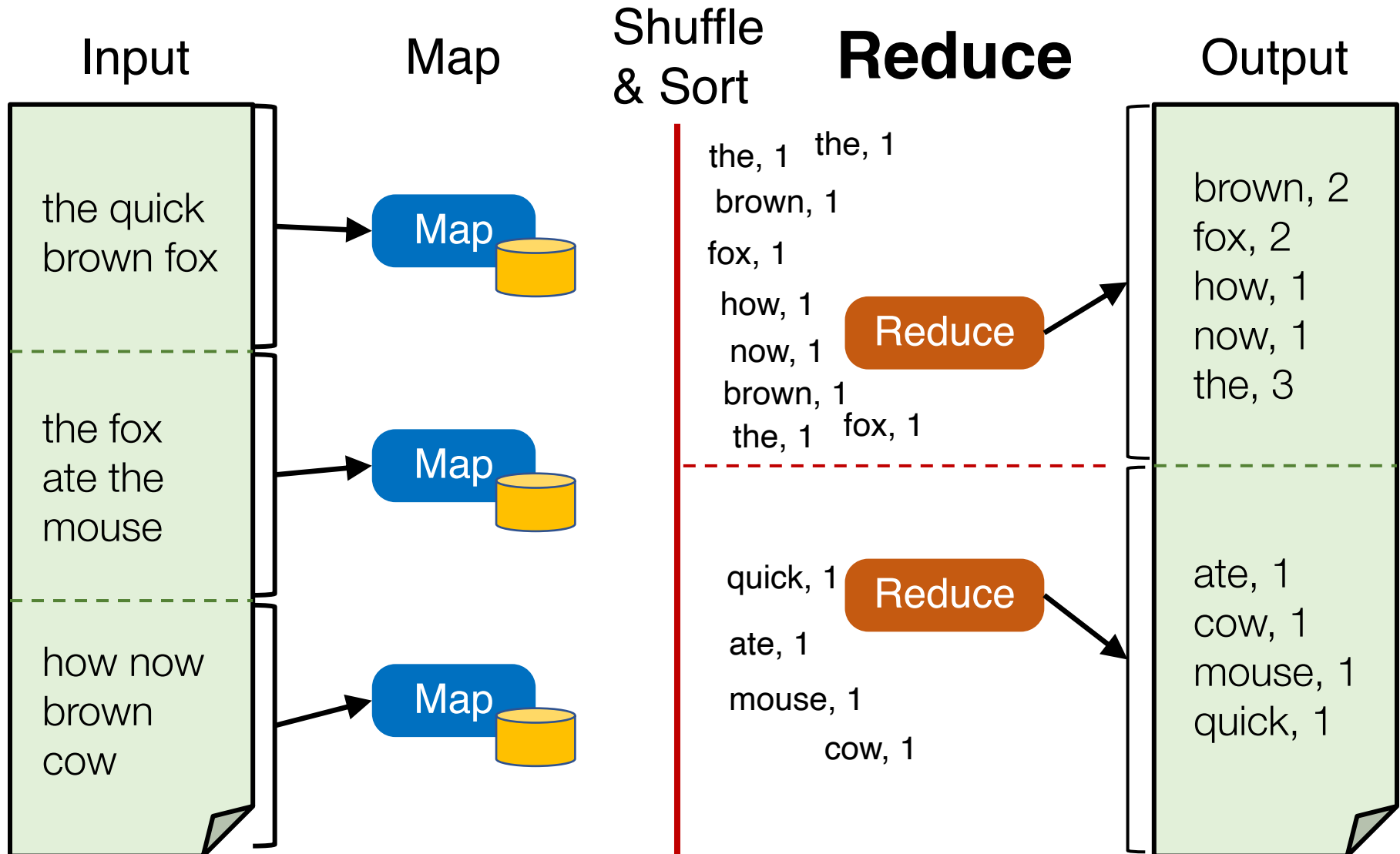
# Word Count execution



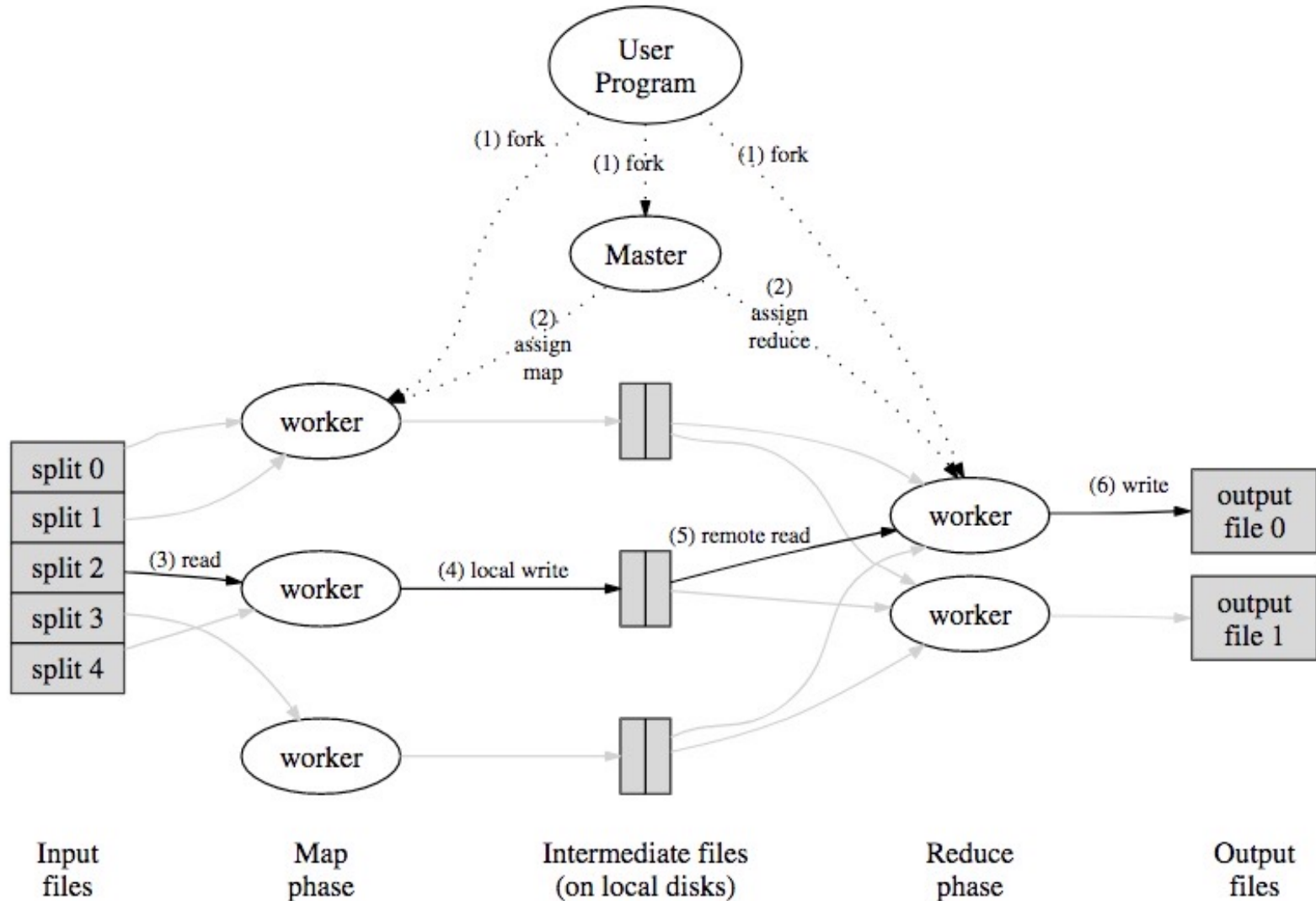
# Word Count execution



# Word Count execution



# MapReduce data flows in paper



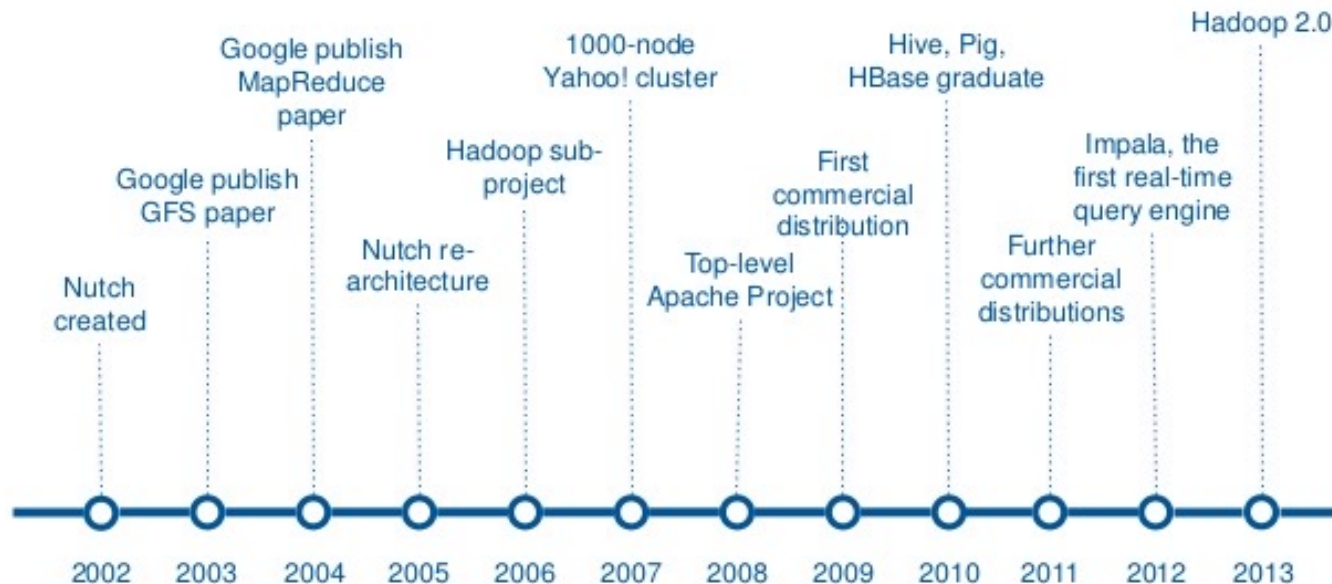


# How it started: Apache Hadoop

- An open-source implementation of Google's MapReduce framework
  - Hadoop MapReduce atop Hadoop Distributed File System (HDFS)



## A Brief History of Hadoop



# How it's going ...

## DATA & AI LANDSCAPE 2019

### INFRASTRUCTURE

**HADOOP ON-PREMISE**  
 cloudera Hortonworks  
 MAPR Pivotal  
 IBM InfoSphere  
 jethro

**HADOOP IN THE CLOUD**  
 AWS Microsoft Azure  
 Google Cloud  
 SAP Cloud Platform  
 IBM InfoSphere BigInsights  
 ORACLE  
 HADOPIE CAZENA

**STREAMING / IN-MEMORY**  
 Amazon Kinesis  
 SAP Cloud Platform  
 ORACLE  
 confluent  
 stream hazelcast GridGain  
 GIGASPACE Wallaroo AWS FASTDATA Ix

### ANALYTICS & MACHINE INTELLIGENCE

**DATA ANALYST PLATFORMS**  
 Microsoft pentaho alteryx  
 Digital Reasoning QUAVUS AYASDI  
 ATTIVO Datameer incorta.  
 interlana MODE ENDOR  
 sisu switchboard Starburst

**DATA SCIENCE PLATFORMS**  
 IBM databricks dataiku  
 DOMINO rapidminer TIBCO  
 ANA CONDA SAS Altair  
 KNIME MathWorks

### APPLICATIONS - ENTERPRISE

**SALES**  
 CHORUS INSIDESALES.COM people.ai  
 conversica clarifai avso tactai  
 fusejmachines Clearbit

**MARKETING - B2B**  
 RADIUS EVERSTING Lattice  
 MINTIGO osense  
 tubular JENGGAGIO  
 KNOTCH mpro

**MARKETING - B2C**  
 bloom SendGrid  
 braze ACTIONIQ BLUECORE  
 CONTENTSQUARE TEALUMIN Zampartico  
 Amplitude ampensity QUANTIGO  
 Simon Lyfika PERSADO  
 remesh

**CUSTOMER EXPERIENCE / SERVICE**  
 MEDALLIA SurveyMonkey  
 CLARABRIDGE zendesk Kustomer  
 mycom DMN LUMIVISION Gainsight  
 HEAP Amplitude Watson Assistant  
 DigitalGenius ASAPP ada AUTOMAT ahni  
 CarDesk metorm

**ENTERPRISE PRODUCTIVITY**  
 slack ORACLE  
 GURU lumiatla  
 DIFFBOT clarifai  
 talla Kasisto

**NoSQL DATABASES**  
 Google Cloud AWS ORACLE Microsoft Azure  
 mongoDB MarkLogic  
 Couchbase DATASTAX  
 redislabs REEROSPIKE  
 AlangoDB SCYLLA

**NewSQL DATABASES**  
 SAP Clustrix PLUGIX  
 Pivotal  
 MEMSQL influxdata  
 Couchroach Labs  
 VOLTRB splice  
 paradigms

**GRAPH DBs**  
 Amazon Neptune  
 IBM ORACLE  
 GraphDB  
 OrientDB  
 Neo4j  
 GraphSense

**MPP DBs**  
 TERADATA VERTICA  
 IBM Data Warehouse Systems  
 Exasol  
 Exor  
 dremio Yellowbrick

**CLOUD EDW**  
 AWS Google Cloud  
 Pivotal  
 snowflake  
 Infoworks

**SERVERLESS**  
 AWS Google Cloud  
 Pivotal  
 snowflake  
 Infoworks

**BI PLATFORMS**  
 looker databricks  
 Domo ARCADIA DATA ThoughtSpot  
 ATSCALE BUSINESS Qlik  
 GoodData Informatica  
 MicroStrategy Keen IO

**VISUALIZATION**  
 tableau Power BI  
 SAP Google Cloud  
 celonis  
 zepl Periscope Data  
 CHARTIO

**MACHINE LEARNING**  
 Amazon SageMaker  
 H2O  
 DataRobot gamalon  
 VIZENIZE ELEMENT  
 deeppersonal

**HUMAN CAPITAL**  
 pymentics  
 hiit  
 Ailyo Textio  
 Workday Stella  
 entelo  
 comcast  
 boomy

**LEGAL**  
 RAVEL  
 DISCO Qtrio  
 JUDICATA  
 BROEVIA  
 IRONCLAD  
 PRESENTATION  
 ROSS

**REGTECH & COMPLIANCE**  
 Regi  
 text IQ  
 PARTNERSHIPS  
 CROSSBRAIN  
 DATA REPUBLIC

**FINANCE - AN/AP**  
 ZUORO  
 SAHANA  
 TRADESHIFF  
 SCALE FACTOR  
 bukeeper  
 pilot

**BACK OFFICE AUTOMATION & RPA**  
 UiPath  
 blueprism  
 VIDAO  
 Workfusion Workato  
 RPAWORKS  
 ALKYRI

**SECURITY**  
 TANIUM CYLANCE  
 CODE42 CyberCloud  
 VECTRA  
 pindrop  
 CodeSecure  
 bitbitgloss  
 sponcongration  
 BLUEHEXAGON Semble  
 Zscaler StackPath  
 ANOMALI  
 DATAVISOR  
 feedzai  
 CyberSec  
 SecurityScorecard  
 SECURE  
 BITSIGHT  
 J.A.S.K  
 FORTER  
 riskrecon  
 J.A.S.K  
 AEMOTIC

**DATA TRANSFORMATION**  
 talend pentaho  
 alteryx TRIFACTA  
 tmnr Paxata  
 StreamSets UNIFI

**DATA INTEGRATION**  
 SAP Data Services Informatica  
 Mirth Connect  
 Inmagic  
 Segment  
 ZALONI  
 Import.io  
 Infovorks  
 Pinesai  
 SNOWFLOW  
 MATILLION

**DATA GOVERNANCE**  
 IBM ORACLE  
 collibra  
 Alation  
 OKERA  
 MANTA  
 data.world

**MGMT / MONITORING**  
 AWS New Relic octio  
 rubrik APPDYNAMICS  
 dynatrace WAVEFOOT  
 Signalix  
 splunk  
 Moogsoft pagerduty  
 zentral  
 Numeracy  
 OKERA  
 MANTA  
 data.world

**COMPUTER VISION**  
 Microsoft Azure  
 Amazon Rekognition  
 clarifai  
 EVER AI  
 neurologix  
 xUBIQUITY  
 YRUBIA  
 YITU  
 trac  
 synthesis  
 Qubica

**HORIZONTAL AI**  
 IBM Watson Cortana  
 Affective  
 sentiment  
 HUMANITA  
 PETUM  
 CIRIUS AI  
 OSARO  
 VISIUM  
 Ustream  
 PolyAI

**SPEECH & NLP**  
 Google Cloud  
 Amazon Transcribe  
 narrative science  
 Movell  
 Soundhound Inc.  
 voicebox  
 cogito slips  
 SHARTELE  
 UN  
 PolyAI

**ADVERTISING**  
 AppNexus  
 Oracle MOAT  
 InMobi  
 TAPAL  
 Opip

**EDUCATION**  
 Knewton  
 Clever  
 edclara  
 kidaptive  
 Panorama  
 Knowme  
 gradscope

**REAL ESTATE**  
 Redfin  
 Opener  
 VTS  
 CREDIFI  
 GEOPHY  
 COMPSTAK  
 STACEMAKER  
 QUANTUM

**GOVT**  
 OPENOVY  
 mark43  
 LiveStories  
 Passport  
 SmartProcure  
 STREETIMAGINE  
 QUANTUM

**INTELLIGENCE**  
 Palantir  
 Dataminr  
 Quid  
 PRIMER  
 FORGE

**FINANCE - INVESTING**  
 KENSH  
 ADAPTOR  
 MINDEN  
 ISENTIUM  
 AOROR  
 ProsperPack  
 PAGAYA

**FINANCE - LENDING**  
 ondeck  
 Affirm  
 KREDITCHEX  
 AVANT  
 TALIA  
 LEASBANC  
 Upgrade  
 100CREDIT  
 WeLab  
 WEBOSS  
 ZESTY AI  
 CIRE  
 CIGNA

**INSURANCE**  
 MONOMIE  
 ANOMALI  
 CYENCE  
 CLOPO  
 SHIFT Technology  
 ROOT  
 ZESTY AI  
 CIRE  
 CAPE

**STORAGE**  
 AWS Google Cloud  
 Microsoft Azure  
 PURE STORAGE  
 ALLUXIO  
 HUBBOX  
 COHEITY

**CLUSTER SVCS**  
 Amazon EMR  
 IBM Cloudfoundry  
 Google Cloud  
 Pivotal  
 SAP  
 SAP HANA

**DATA GENERATION & LABELLING**  
 Amazon Mechanical Turk  
 workwork  
 oppen  
 HIVE  
 Labelbox  
 Lionbridge

**AI OPS**  
 ALGORITHMIA  
 Vertaai  
 datom  
 datation  
 Weights & Biases  
 DETERMINED AI  
 fidder

**GPU DBs & CLOUD**  
 kinetica  
 ORACLE  
 PG-Stream  
 FLOYDHUB

**HARDWARE**  
 Google TPU  
 ARM  
 intel  
 MYTHIC  
 Ahabox  
 WAVE  
 CELEBRITY  
 DEFINI

**SEARCH**  
 elastic search  
 algolia  
 Ludwicks  
 swifttype  
 alpha sense  
 omni-us  
 SINEQUA

**LOG ANALYTICS**  
 splunk  
 sumologic  
 solaredge  
 TIMBER  
 HObono  
 logz.io

**SOCIAL ANALYTICS**  
 Hootsuite  
 Netbase  
 synthesize  
 simple reach  
 bitly  
 SimilarWeb

**WEB / MOBILE / COMMERCE ANALYTICS**  
 Google Analytics  
 mixpanel  
 AIRAMPLITUDE  
 Airtable  
 RESCI  
 SIGOPT  
 granify  
 CUSTORA

**HEALTHCARE**  
 Flatiron  
 Cliver  
 YRUBIA  
 METABOTIA  
 Zingber  
 GLOW  
 bivation  
 3DMED  
 GIBERIA  
 PAPA  
 ovia  
 TEMPOS  
 patersilume  
 AICURE  
 Inestro  
 UNICOM  
 CITIZEN I  
 PROGNOS  
 notable  
 entelic  
 mpro  
 ARTERYS  
 IMAGEN  
 PAIGE  
 DATAVANT

**LIFE SCIENCES**  
 Benchling  
 color  
 Benevolent AI  
 Verily  
 WUOLNECO  
 TIBCO  
 CLIC LABS  
 FINANXUS  
 DIVANEXUS  
 PHAROS  
 CITRINE  
 AEMOTIC  
 AEMOTIC  
 AEMOTIC  
 AEMOTIC

**TRANSPORTATION**  
 UBER  
 TESLA  
 CLEARPATH  
 CRUISE  
 NURO  
 ZIGO  
 DRIVE AI  
 CAMBRIDGE  
 AMOTIVE  
 G7  
 PILOT AI  
 AURORA  
 NIO  
 AMOTIVE  
 MOOVIT  
 IIA  
 NEXAR  
 KODIAK  
 COMM.A.I  
 netradyne  
 GENMA  
 CIVIL MAPS  
 COPILOT

**AGRICULTURE**  
 FARMERS EDGE  
 Granular  
 JOHN DEERE  
 BLUEBERIDGE  
 FARMERS EDGE  
 FarmLogs  
 TARMANIS  
 GAMAYA  
 Electric  
 ZINER

**COMMERCE**  
 Instacart  
 FAIR F  
 STITCH FIX  
 HURLOCK  
 HEURLOCK  
 FARMERS EDGE  
 TARMANIS  
 GAMAYA  
 Electric  
 ZINER

**INDUSTRIAL**  
 AVEVA  
 SIEMENS  
 PREDIX  
 UPTAKE  
 SCORTEX  
 KINEX  
 TACHYUS  
 ByteDance  
 AMPER  
 SOIERN  
 BOXEVER  
 ELECTRIC  
 ZINER

### CROSS-INFRASTRUCTURE/ANALYTICS

aws Google Cloud Microsoft IBM SAP Hewlett Packard Enterprise SAS IO10DATA vmware TIBCO TERADATA ORACLE NetApp syncsort MAPR cloudera

### OPEN SOURCE

**FRAMEWORKS**  
 Spark  
 Flink  
 YARN  
 TEZ  
 MESOS  
 DOCKER  
 HELIX

**QUERY / DATA FLOW**  
 Spark SQL  
 presto  
 SLMDATA  
 GRAPHQL  
 Flink

**DATA ACCESS & DATABASES**  
 Cassandra  
 mongoDB  
 redis  
 Cockroach Labs  
 druid  
 ScioDB  
 etcd  
 Kong

**ORCHESTRATION & MGMT**  
 talend  
 Apache Airflow  
 Apache Ambari

**STREAMING & MESSAGING**  
 Spark  
 nifi  
 Flink  
 beam  
 kafka  
 storm  
 Apache RocketMQ

**STAT TOOLS & LANGUAGES**  
 Databricks  
 Scala  
 R  
 Python  
 Julia  
 SPOly

**AI OPS & INFRA**  
 mflow  
 DataDog  
 D.C  
 SELDON  
 Polymon

**AI / MACHINE LEARNING / DEEP LEARNING**  
 TensorFlow  
 Keras  
 PyTorch  
 OpenAI  
 DM TK  
 theano  
 Apache SENGSA  
 DIMSUM  
 FeatureFu  
 VELES  
 MONSTRON  
 GNNX  
 LUDWIG  
 PyForch  
 DSSINE  
 milib  
 MAHOTU  
 Aerosolve  
 fast.ai  
 mir

**SEARCH**  
 elasticsearch  
 Solr  
 Lucene

**LOGGING & MONITORING**  
 elasticsearch  
 kibana  
 logstash  
 fluentbit  
 fluentd  
 Grafana

**VISUALIZATION**  
 matplotlib  
 TIBO  
 TensorBoard  
 seaborn  
 Bokeh

**COLLABORATION**  
 BeakerX  
 Jupyter  
 ANA CONDA

**SECURITY**  
 Apache Ranger  
 KNOX  
 Sentry  
 SCURTYLO

### DATA SOURCES & APIs

**HEALTH**  
 Apple  
 VALIDIC  
 practicefusion  
 fitbit GARMIN  
 HUMAN APP  
 kinsio

**IOT**  
 GE Digital  
 UPTAKE  
 thingworx  
 helium  
 samsara

**FINANCIAL & ECONOMIC DATA**  
 Bloomberg  
 THOMSON REUTERS  
 DOW JONES  
 S&P CAPITAL IQ  
 CBINSIGHTS  
 PLAID  
 GEstimize  
 PREMISE  
 Quantil  
 Stockworks  
 xignite  
 Thinknum  
 earnest  
 predata

**AIR / SPACE / SEA**  
 Orbital Insight  
 planet  
 AIRBOTICS  
 SPIRE  
 UNDERSTANDY  
 WINDWARD  
 MarineTraffic

**PEOPLE / ENTITIES**  
 acxiom  
 experian  
 EPSILON  
 InsideView  
 Crimson Hexagon  
 Quantcast  
 BASIS  
 SAFEGRAPH

**LOCATION INTELLIGENCE**  
 FOURSQUARE  
 sense360  
 PlaceIQ  
 CARX  
 BASIS  
 SAFEGRAPH  
 Mapbox  
 Mapbox  
 Hexagon  
 esri  
 factual  
 Mapillary  
 StreetView  
 cuebio  
 Radar  
 StreetView

**OTHER**  
 DATA GOV  
 IMAGENET  
 CRUX  
 inddpxus

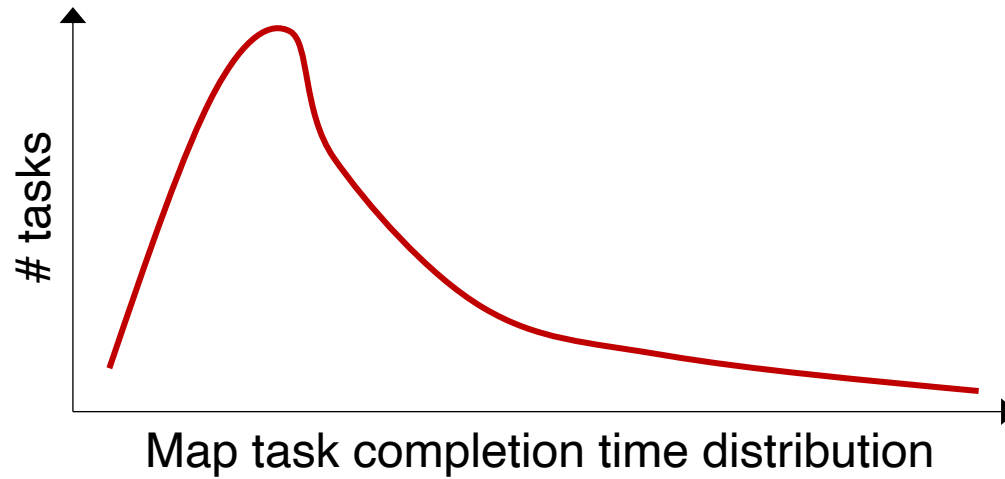
### DATA RESOURCES

**DATA SERVICES**  
 OPERA  
 DATA SCIENCE  
 fractal  
 kaggle  
 DataKind  
 innoplcxus

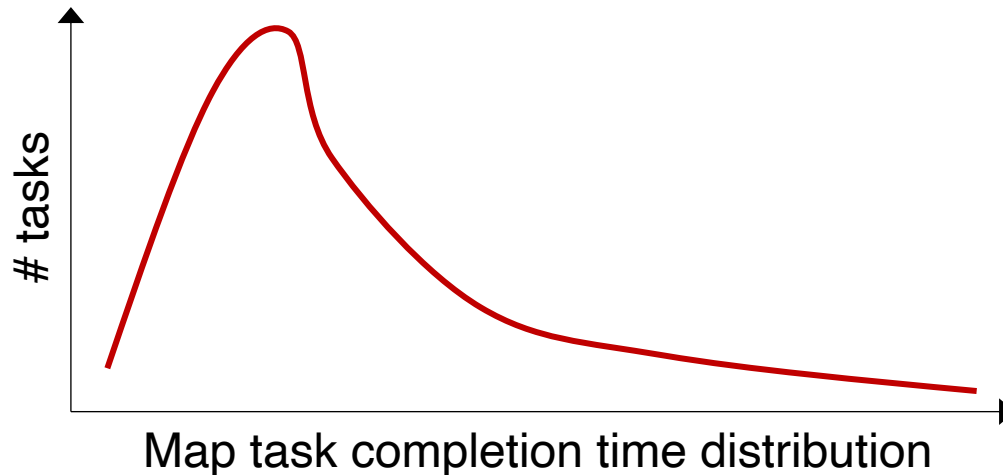
**INCUBATORS & SCHOOLS**  
 PLURALSIGHT  
 DataCamp  
 DataElite  
 INSIGHT  
 THE DATA INCUBATOR  
 METIS

**RESEARCH**  
 OpenAI  
 facebook research  
 MIRI  
 VECTOR INSTITUTE  
 A2  
 ALLEN INSTITUTE  
 EARLY STAGE VENTURE CAPITAL

# Stragglers



# Stragglers



- **Tail execution time** means some workers (always) finish late
- Q: How can MR work around this?
  - Hint: its approach to **fault-tolerance** provides the right tool

# Resilience against stragglers

- If a task is going slowly (i.e., **straggler**):
  - Launch second copy of task on another node
  - Take the output of whichever finishes first

# More design

- Master failure
- Locality
- Task granularity

# GFS usage at Google

- 200+ clusters
- Many clusters of 1000s of machines
- Pools of 1000s of clients
- 4+ PB filesystems
- 40 GB/s read/write load
  - In the presence of frequent hardware failures

\* Jeff Dean, LADIS 2009

# MapReduce usage statistics over time

	Aug, '04	Mar, '06	Sep, '07	Sep, '09
Number of jobs	29K	171K	2,217K	3,467K
Average completion time (secs)	634	874	395	475
Machine years used	217	2,002	11,081	25,562
Input data read (TB)	3,288	52,254	403,152	544,130
Intermediate data (TB)	758	6,743	34,774	90,120
Output data written (TB)	193	2,970	14,018	57,520
Average worker machines	157	268	394	488

\* Jeff Dean, LADIS 2009



# MapReduce discussion

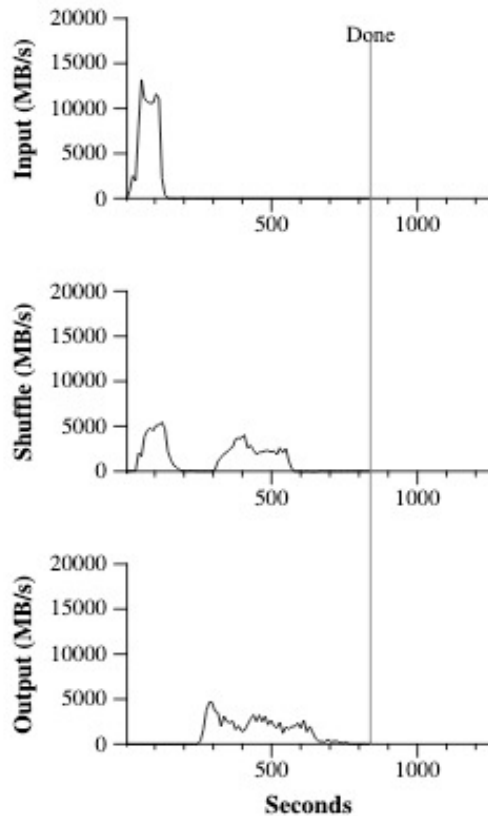
What will likely serve as a performance bottleneck for Google's MapReduce used back in 2004 (or even earlier)? CPU? Memory? Disk? Network? Anything else?

# MapReduce discussion

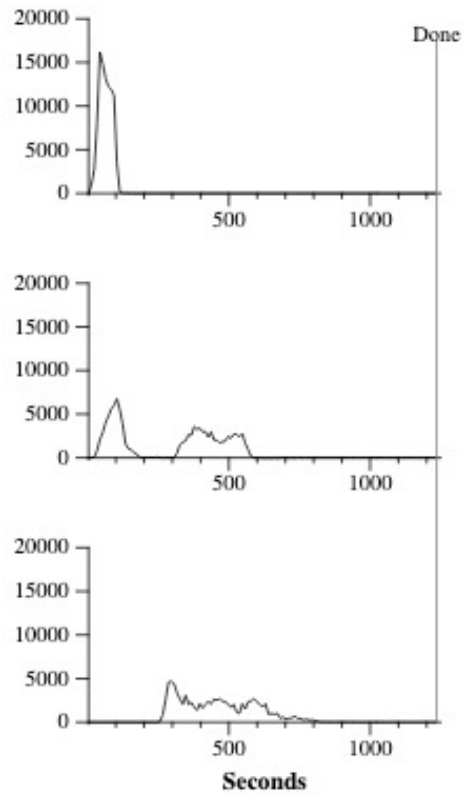
What will likely serve as a performance bottleneck for Google's MapReduce used back in 2004 (or even earlier)? CPU? Memory? Disk? Network? Anything else?

How does MapReduce reduce the effect of slow network?

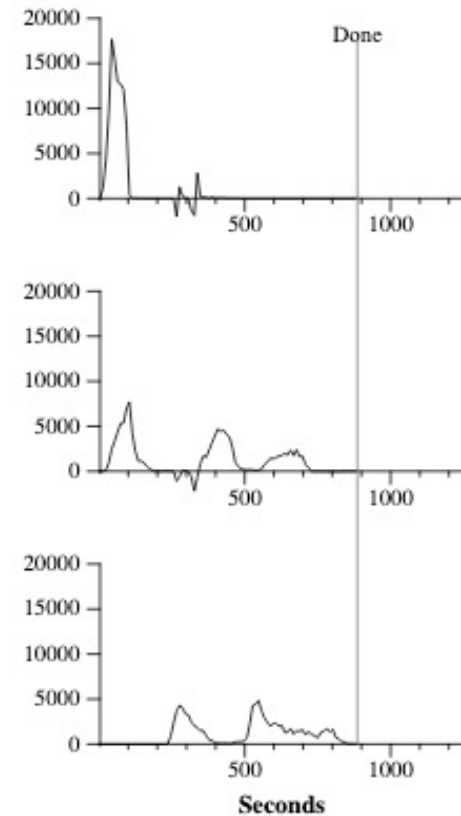
# MapReduce discussion



(a) Normal execution



(b) No backup tasks



(c) 200 tasks killed

# MapReduce discussion

Consider a log analytics job where you perform log-based debugging. You want to extract the timestamp info of all entries that match a keyword and then calculate the count of all matched entries:

1. Filter the entries with the keyword;
2. Calculate the count of all matched entries

What are the main shortcomings of using MapReduce to support such pipeline-like applications?

# Next step

- Look out for
  - Project suggestion doc
  - Fill the team composition form
  - Project bid and team composition due by Feb 24
  
- Next week: Apache Spark