What is a parallel database?
Parallel v.s. distributed databases

- **Parallel database system:**
  - Improve performance through parallel implementation

- **Distributed database system:**
  - Data is stored across several sites, each site managed by a DBMS capable of running independently
Parallel DBMSs

- **Goal**
  - Improve performance by executing multiple operations in parallel

- **Key benefit**
  - Cheaper to scale than relying on a single increasingly more powerful processor

- **Key challenge**
  - Ensure overhead and contention do not kill performance
Performance Metrics for Parallel DBMSs

- **Speedup**
  - More processors $\rightarrow$ higher speed
  - Individual queries should run faster
  - Should do more transactions per second (TPS)
  - Fixed problem size *overall*, vary # of processors ("strong scaling")

- **Scaleup**
  - More processors $\rightarrow$ can process more data
  - Fixed problem size *per processor*, vary # of processors ("weak scaling")
  - **Batch scaleup**
    - Same query on larger input data should take the same time
  - **Transaction scaleup**
    - N-times as many TPS on N-times larger database
    - But each transaction typically remains small
Linear v.s. Non-linear **Speedup**

![Graph showing linear and non-linear speedup with respect to the number of processors](graph.png)
Linear v.s. Non-linear Scaleup

Batch Scaleup

# processors (=P) AND data size

$\times 1$  $\times 5$  $\times 10$  $\times 15$
Challenges to Linear Speedup and Scaleup

- **Startup cost**
  - Cost of starting an operation on many processors

- **Interference**
  - Contention for resources between processors

- **Skew**
  - Slowest processor becomes the bottleneck
Architectures for Parallel Databases

- Shared memory
- Shared disk
- Shared nothing
Shared memory

Interconnection Network

Global Shared Memory

P

P

P

D

D

D
Shared disk

Interconnection Network
Shared nothing

Interconnection Network

P
M
D

P
M
D

P
M
D
Shared nothing

- Most scalable architecture
  - Minimizes interference by minimizing resource sharing
  - Can use commodity hardware
- Also most difficult to program and manage

We will focus on shared nothing

Important question: what exactly can we actually parallelize in a parallel database?
Taxonomy for parallel query evaluation

- **Inter-query parallelism**
  - Each query runs on one processor

- **Inter-operator parallelism**
  - A query runs on multiple processors
  - An operator runs on one processor

- **Intra-operator parallelism**
  - An operator runs on multiple processors
Data partitioning schemes

Partitioning a table:

Range

Hash

Round Robin
Data partitioning

What are the pros and cons?

- Round robin
  - Good load balance but always needs to read all the data

- Hash based partitioning
  - Good load balance but works only for equality predicates and full scans

- Range based partitioning
  - Works well for range predicates but can suffer from data skew
Parallel join: R join S on attribute x

R

Hash on x

Join each hash bucket

S
MapReduce
MapReduce

- Next: quick overview and comparison of MapReduce
Abridged Declaration of Independence

A Declaration By the Representatives of the United States of America, in General Congress Assembled. When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident; that all men are created equal and independent; that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying it's foundation on such principles and organizing it's power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will dictate that governments long established should not be changed for light and transient causes: and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies; and such is now the necessity which constrains them to expunge their former systems of government. the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unsullied by falsehood.

How many “big”, “medium”, and “small” words are used?
Example: Word length histogram

Big = Yellow = 10+ letters
Medium = Red = 5..9 letters
Small = Blue = 2..4 letters
Tiny = Pink = 1 letter

A Declaration By the Representatives of the United States of America. In General Congress Assembled,

When in the course of human events, it becomes necessary for a people to alter their体制, and to assume among the powers of the earth the equal and independent station to which the laws of nature and of nature's God entitle them; a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident: that all men are created equal and independent; that from their equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute a new government, laying its foundation on such principles as organizing it's power to effect the ends for which they shall seem most likely to effect their safety and happiness. Prudence indeed will dictate that governments long established should not be changed for light and transient causes; and accordingly it experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun in a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies; and such is now the necessity which constrains them to expunge their former systems of government. The history of the present majesty is the history of unremitting injuries and usurpations, among which we must account single or solitary as contradict the uniform tenor of the rest. All of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith as well unsullied as falsehood.
A Declaration By the Representatives of the United States of America, in General Congress Assembled.

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident; that all men are created equal and independent; that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying it's foundation on such principles and organizing it's power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will dictate that governments long established should not be changed for light and transient causes; and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies; and such is now the necessity which constrains them to expunge their former systems of government. the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unsullied by falsehood.
Example: Word length histogram
A Declaration By the Representatives of the United States of America, in General Congress Assembled.

When in the course of human events it becomes necessary for a people to advance from that subordination in which they have hitherto remained, and to assume among powers of the earth the equal and independent station to which the laws of nature and of nature's god entitle them, a decent respect to the opinions of mankind requires that they should declare the causes which impel them to the change.

We hold these truths to be self-evident; that all men are created equal and independent; that from that equal creation they derive rights inherent and inalienable, among which are the preservation of life, and liberty, and the pursuit of happiness; that to secure these ends, governments are instituted among men, deriving their just power from the consent of the governed; that whenever any form of government shall become destructive of these ends, it is the right of the people to alter or to abolish it, and to institute new government, laying it's foundation on such principles and organizing it's power in such form, as to them shall seem most likely to effect their safety and happiness. Prudence indeed will
dictate that governments long established should not be changed for light and transient causes: and accordingly all experience hath shewn that mankind are more disposed to suffer while evils are sufferable, than to right themselves by abolishing the forms to which they are accustomed. But when a long train of abuses and usurpations, begun at a distinguished period, and pursuing invariably the same object, evinces a design to reduce them to arbitrary power, it is their right, it is their duty, to throw off such government and to provide new guards for future security. Such has been the patient sufferings of the colonies; and such is now the necessity which constrains them to expunge their former systems of government. the history of his present majesty is a history of unremitting injuries and usurpations, among which no one fact stands single or solitary to contradict the uniform tenor of the rest, all of which have in direct object the establishment of an absolute tyranny over these states. To prove this, let facts be submitted to a candid world, for the truth of which we pledge a faith yet unsullied by falsehood.
MapReduce Programming Model

- Input & Output: each a set of key/value pairs
- Programmer specifies two functions:

  map (in_key, in_value) -> list(out_key, intermediate_value)

  - Processes input key/value pair
  - Produces set of intermediate pairs

  reduce (out_key, list(intermediate_value)) -> list(out_value)

  - Combines all intermediate values for a particular key
  - Produces a set of merged output values (usually just one)

Inspired by primitives from functional programming languages such as Lisp, Scheme, and Haskell
Map Reduce

- Google: [Dean 2004]
- Open source implementation: Hadoop

- Map-reduce = high-level programming model and implementation for large-scale parallel data processing
Implementation

- There is one master node
- Master partitions input file into $M$ splits, by key
- Master assigns workers (=servers) to the $M$ map tasks, keeps track of their progress
- Workers write their output to local disk, partition into $R$ regions
- Master assigns workers to the $R$ reduce tasks
- Reduce workers read regions from the map workers’ local disks