Agenda

- Map|Reduce Wrap up
- Math behind replica conflict resolution

Map Reduce Recap

Map Phase :

map: input K-V pair → set of intermediate K-V pairs

<Doc1, [the, quick]> → {<the, Doc1>, <quick, Doc1>}

Shuffle Phase :

Data from map workers is sent to reduce workers, according to some data partitioning function.

Eg: hash(key) % R
↓
No. of reduce workers

Reduce Phase

reduce: (key, set of values) → set of output

values

<dog, {Doc3, Doc5, Doc6}> → <dog, {Doc3, Doc5, Doc6}>

If reduce function is associative, map worker can do local computation using combiners

(a+b)+c = (a+b)+c Associative Property.

"Flume Java" ⇒ 2010 Paper