



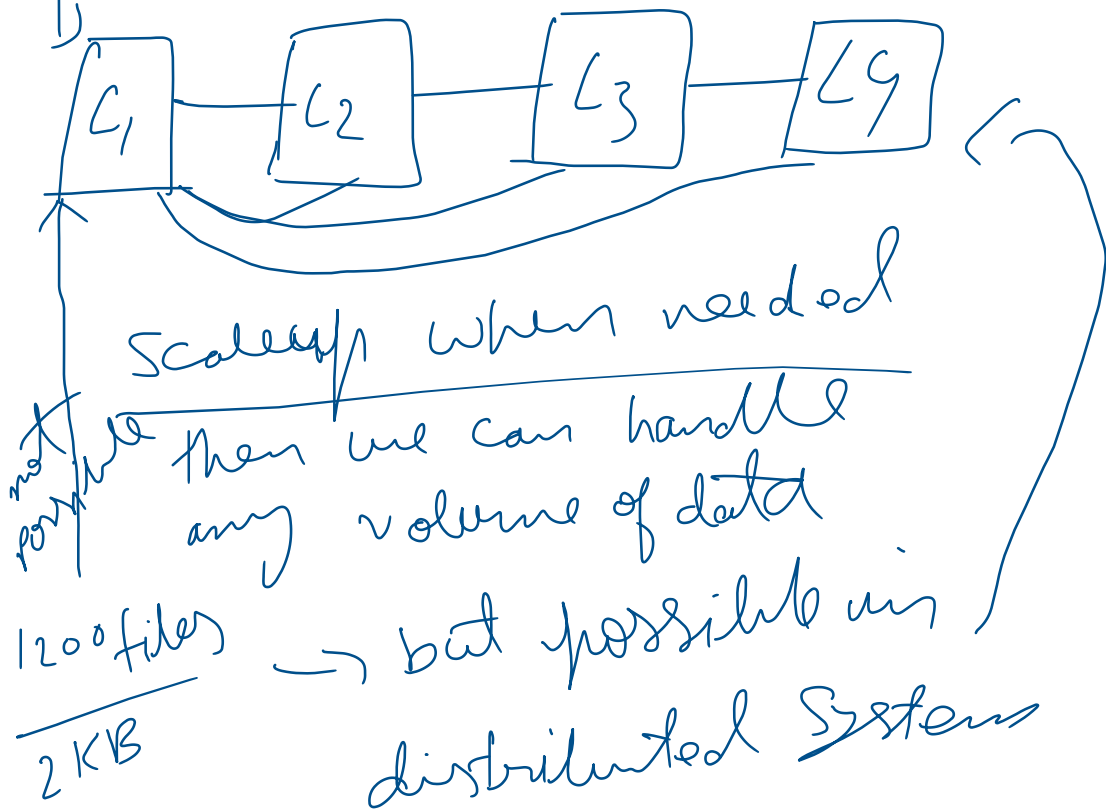
GFS → HDFS

## 5 V's System

- (1) Volume → Size of data → 100GB / day
- (2) Velocity → Speed of data → Streaming, Batch, mini-Batch
- (3) Variety → type of data
- (4) Veracity → trustworthiness of data in terms of accuracy
- (5) Value → Final insight

This 5 problem HDFS  
trying to Solve.

$\frac{120 \text{ files}}{\text{per file } 2 \text{ KB}}$



HDFS [1x, 2x, 3x]

↳ Hadoop Distributed file System

next class