BIG DATA AND OSFA

Omer Trajman, Cloudera
"...the commercial world will fracture into a collection of independent database engines..."


RESULTING IN...

- EDW Data Marts
- ETL Grid
- Enterprise Storage
- CEP and ESB
- OLTP Logs Sensors
- Data Marts
DATA ON OSFA

Kept for limited periods of time

Used for limited applications

Difficult to integrate with other data

Archived (deleted)

Photo credit/Flickr/bcostin
BIG DATA IS UNFILTERED

Different processing engines for different stages.

Lots of data makes it expensive to move.

Dynamic businesses requires flexibility.

Photo credit/Flickr/dervish
BIG DATA MOVES LESS

CEP and ESB

Flume

OLTP
Logs
Sensors

Flume Sqoop

MR

HDFS

Unified Scale Out Storage and Compute

Sqoop

EDW
MR = ETL?
IT'S NOT (JUST) ABOUT MR

HDFS

Avro  Crunch  Flume  Giraph  Hama  HBase  Hive  Mahout  MaReduce  Pig  Sqoop
OR (JUST) ABOUT OSS

Informatica
MicroStrategy
Tableau
RainStor
Hadoop
Datameer
Karmasphere
NGdata
Platfora
WibiData

HDFS

Avro
Crunch
Flume
Giraph
Hama
HBase
Hive
Mahout
MaReduce
Pig
Sqoop
OSFA AND BIG DATA

- Informatica
- Tableau
- MicroStrategy
- Pig
- HBase
- Hive
- HDFS

Flume

Sqoop
GOOGLE FIGHT?