

DATA ZOO

EVEN
PART 2: MORE
ZOO!

distributed/redundant file stores
glusterFS, minio, S3 (aws), ABS (Azure), GCS (Google)

hive, parquet, orc, avro
columnar compression schema

Impala, Dremio, Drill
ANALYTICS DBs (SQL-based)
BigQuery (Google)
Redshift (AWS)
- columnar
- distributed
Speed!!!
Druid, Clickhouse, Pinot
- distributed
- columnar
- Speed

DATA BASES
SQL Relational DBs (Op)
- Postgres
- MySQL
- Oracle
- MariaDB
- SQLServer

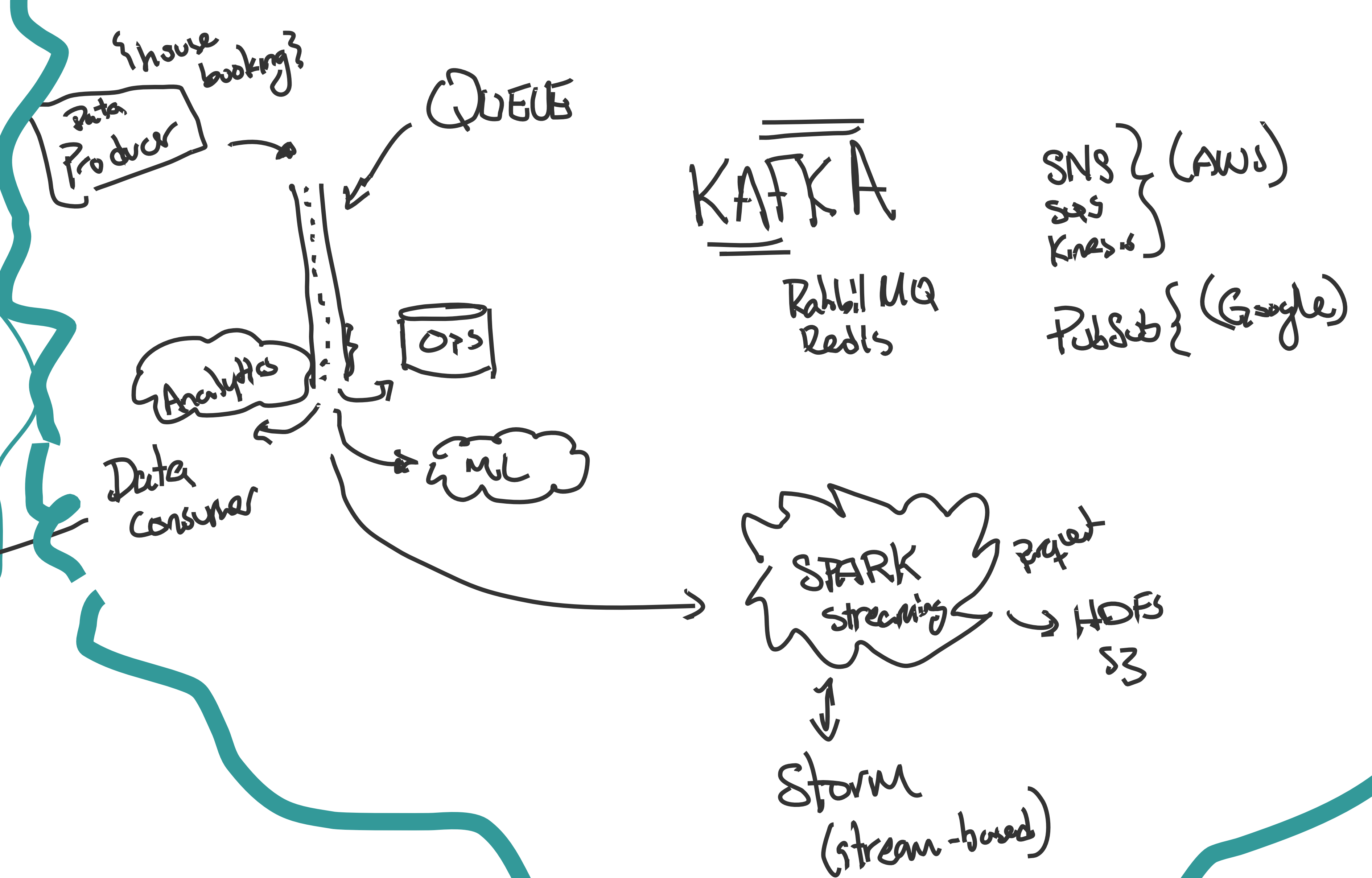
NoSQL DB (Ops)
- MongoDB
- CouchDB
- Cassandra
document
{"foo": "bar", "baz": "foo"}
"wide table", "nested table"
DISTRIBUTED

NoSQL DB (specialized)
SEARCH
- Elasticsearch
KV
in-memory
- Redis
- Memcached
foo: "bar"
: 53
"caching"

STORING DATA

HORIZONTALLY
SCALABLE
100gb

MOVING DATA



SHOWING DATA

