



ARQUITECTURA LAMBDA

Neftalí Valdez Mtz
Alejandra Siqueiros Tarazón

Arquitectura Lambda

- Es una arquitectura de procesamiento de datos diseñada para administrar grandes cantidades de datos tomando ventaja de métodos de procesamiento batch y streaming.
- Este enfoque busca considerar:
 - Latencia
 - Thourghput
 - Tolerancia a fallos

Arquitectura Lambda



BATCH

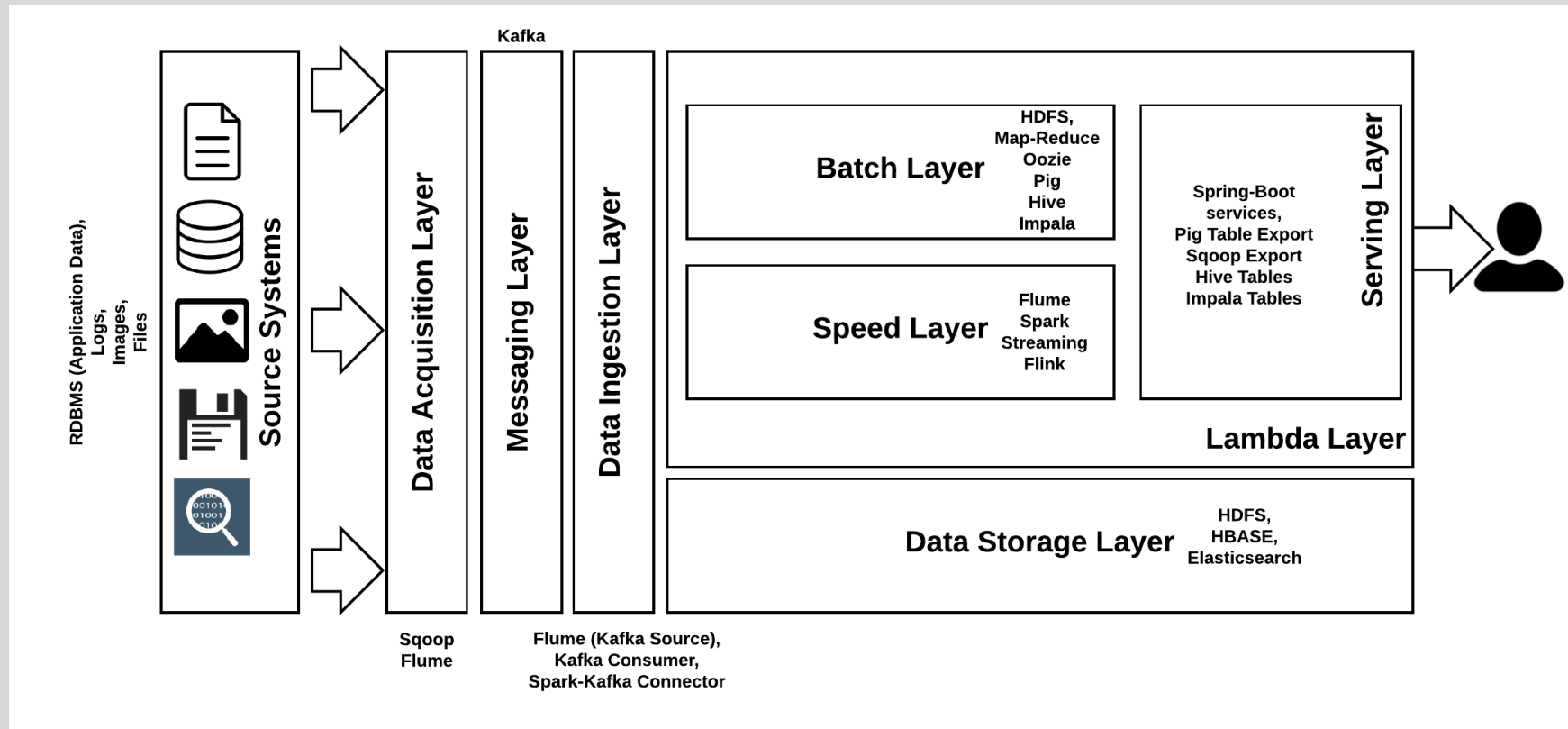


SPEED



SERVING

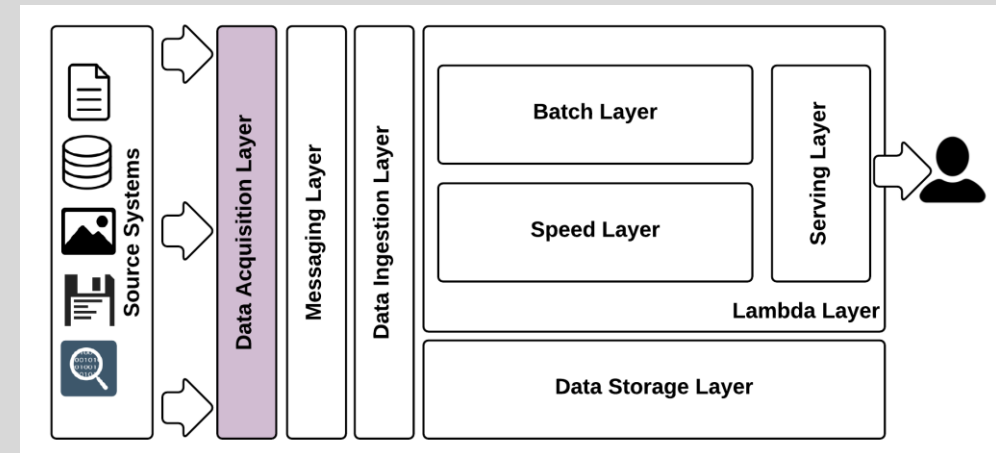
Arquitectura Lambda



Fuente: Data Lake for Enterprises. PACKT

Sqoop

- Principal herramienta para la adquisición de datos de bases de datos relacionales.
- Permite la transferencia en bruto de datos.
- Es fácil integrarlo con Hive, Hbase y Oozie.



SQOOP

- Ventajas
- Se comunica con Postgres, MySQL, Oracle, Teradata...
- Permite algún procesamiento ETL
- Ejecuta data transfer en paralelo.
- JDBC y conectores directos
- Formatos como Avro, Text, Parquet, Sequential.

Sqoop

- Desventajas de Sqoop
- Utiliza JDBC y en ocasiones esto puede ser ineficiente.
- En bulk transfer puede estresar las fuentes de datos.

Sqoop

- Al invocar Sqoop entran en juego dos componentes:
- Driver
- Conectores
- Tipos de conectores:
 - Built In
 - Pluggable

Sqoop

- Sqoop y HDFS
- Sqoop está pensado para trabajar sobre HDFS.
- También soporta ingesta incremental.

Bibliografía

- Holden Karau, Andy Konwinski, Patrick Wendell, y Matei Zaharia. "Learning SPARK" OREILLY 2015.
- Guller Mohammed. "Big Data Analytics with Spark". Apress 2015.
- TURKINGTON GARRY "Hadoop Beginner's Guide" Packt Publishing 2013.
- SUDHI SESHACHALA; "Bigdata – Understanding Hadoop and Its Ecosystem" JUNE 1, 2015.
<http://devops.com/2015/06/01/bigdata-understanding-hadoop-ecosystem/>
- "NOSQL DATA MODELING TECHNIQUES" <https://highlyscalable.wordpress.com/2012/03/01/nosql-data-modeling-techniques/>
- Ayende Rahien; "That NOSQL Thing Column (Family) Databases". <https://ayende.com/blog/4500/that-no-sql-thing-column-family-databases>
- "Comparing document-oriented and relational data"; <http://docs.couchbase.com/developer/dev-guide-3.0/compare-docs-vs-relational.html>
- IAN ROBINSON, JIM WEBER & EMIL EIFREM "Graph Databases" O'Reilly 2013