UNIVERSIDAD DE LAS PALMAS DE GRAN CANARIA
Instituto Universitario de Microelectrónica Aplicada

Máster de Tecnologías de Telecomunicación

Trabajo Fin de Máster

Diseño e implementación de una arquitectura MapReduce para aplicaciones Big Data mediante Síntesis de Alto Nivel

Julian Spahr

Pedro Pérez Carballo, Antonio Núñez Ordóñez

Julio de 2017

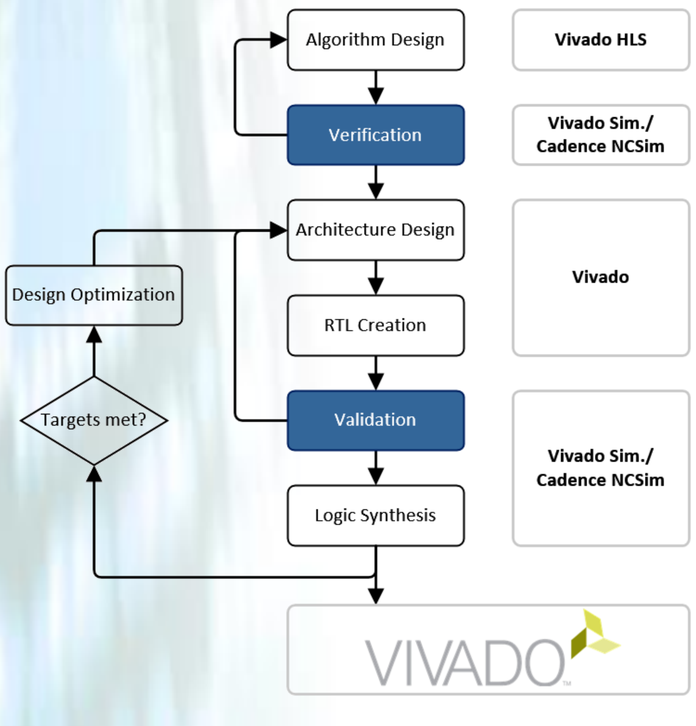
Resumen:

- Este proyecto abarca el modelado, síntesis e implementación de una plataforma hardware que integra algoritmos críticos orientados a aplicaciones Big Data.
- Se integra un *framework* MapReduce sobre una placa de prototipado Xilinx Zynq ZC706, empleando una metodología de Síntesis de Alto Nivel.
- La aplicación empleada para evaluar el *worker* MapReduce es del tipo Word Count, con descripciones algorítmicas en C/C++.
- El *worker* MapReduce segmenta el dataset en múltiples fracciones (Split), traduce el contenido a parejas Key-Values (Map), combina los datos acorde a un criterio de combinación (Reduce) y une todas las fracciones para retornar el resultado total (Merge).
- Segmentación de los datos sobre múltiples pares de bloques IP dedicados a las funcionalidades Map y Reduce, permiten explotar eficientemente la capacidad de paralelismo de la FPGA.

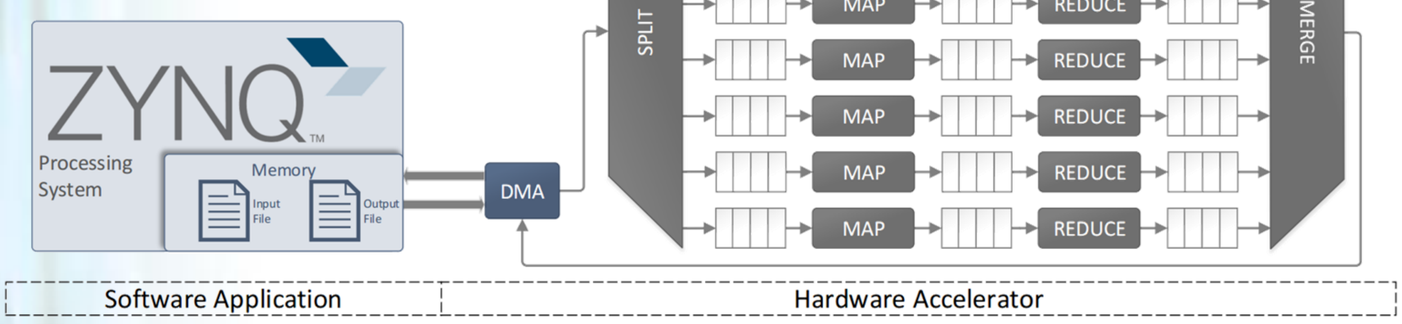
Big Data y MapReduce:

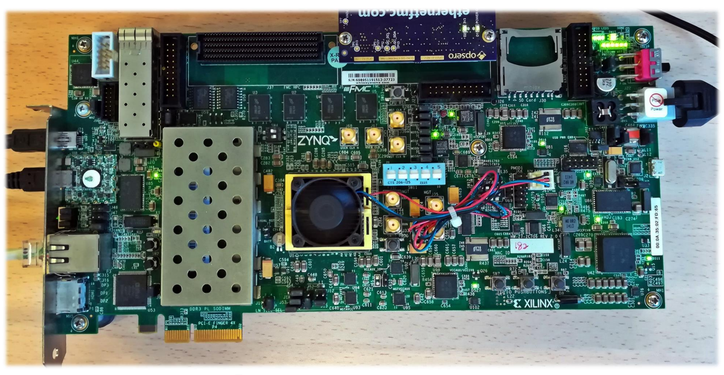
Big Data se define como el tráfico de datos heterogéneo y creciente que fluye en incalculables cantidades entre fuentes. Se entiende como "**demasiado grande, demasiado rápido o demasiado difícil para ser procesado por herramientas ordinarias**". Para realizar un análisis eficiente de estos datos y extraer información válida, se presenta el *framework* MapReduce, desarrollado por Google para facilitar procesado de datos en Big Data. Comprende dos fases principales: **Map** (traducción) y **Reduce** (combinación).

Flujo de diseño:



Arquitectura:





Características:

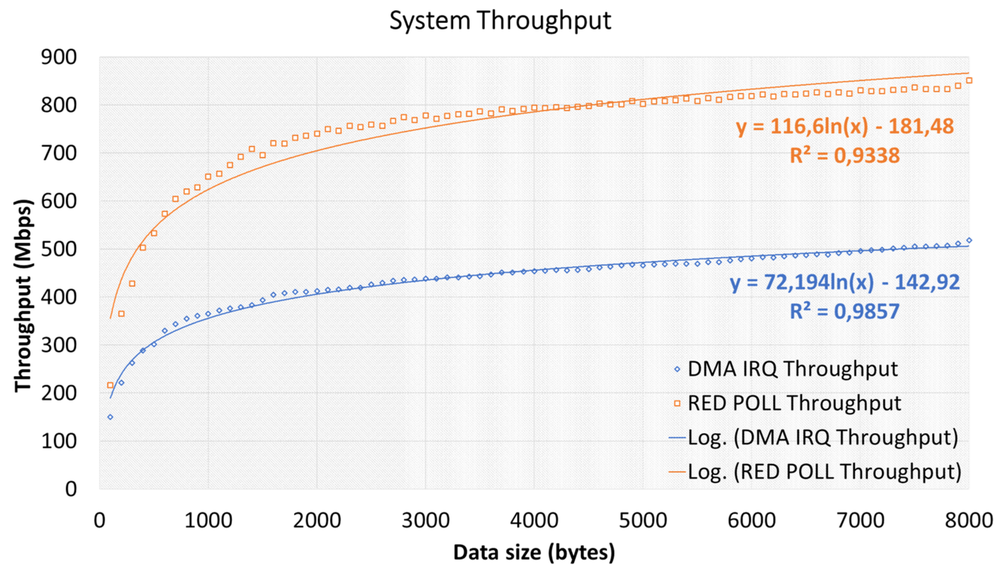
El dispositivo empleado es el Xilinx Zynq XC7Z045.La Lógica Programable (PL) y el Sistema de Procesamiento (PS) trabajan a **las frecuencias máximas de 250 y 800 MHz** respectivamente. La plataforma permite **key's de 8 KB** a la entrada, produciendo un **value de 16 KB a la salida**.

Resultados:

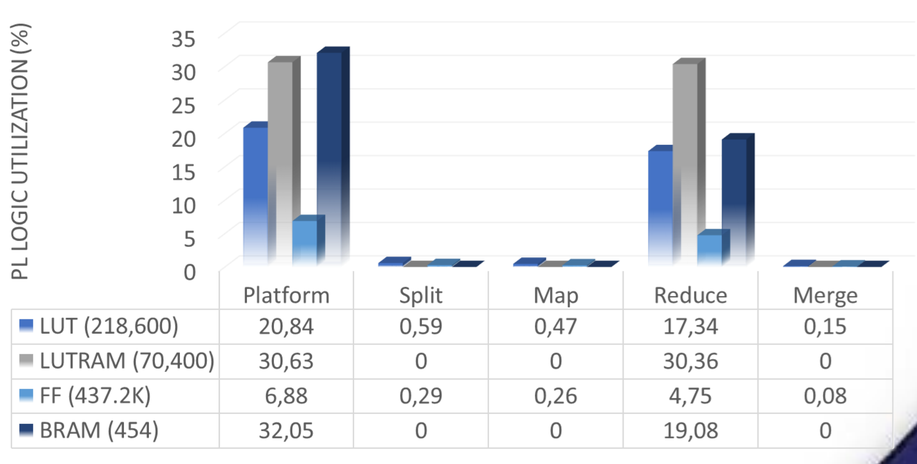
Los parámetros de latencia y throughput se miden considerando dos datapath:

- Comienzo del bloque **Split hasta** el final del bloque **Merge**. Medido empleando un DMA configurado por interrupciones.
- Comienzo del **primer** bloque **Map hasta** el final del **último** bloque **Reduce**. Medido por *polling* a la columna de bloques Reduce.

La validación final del sistema se realizó empleando *samples* de múltiples novelas y publicaciones, tomando su media para el cálculo de la latencia y el throughput final.



Utilización:



	Platform	Split	Map	Reduce	Merge
LUT (218,600)	20,84	0,59	0,47	17,34	0,15
LUTRAM (70,400)	30,63	0	0	30,36	0
FF (437.2K)	6,88	0,29	0,26	4,75	0,08
BRAM (454)	32,05	0	0	19,08	0

Conclusiones:

- Con todas las funcionalidades principales integradas en el PL, se obtienen *throughputs* de ~860 Mbps en las etapas Map-Reduce, y ~500 Mbps incluyendo las funcionalidades Split y Merge.
- El factor de ocupación es de un 32.05 % sobre el PL para un *worker* MapReduce de capacidad de 8 KB y 8 pares Map-Reduce.
- El throughput del sistema posee comportamiento logarítmico, ya que la latencia mínima de transferencia es de ~8 µs, de los cuales ~3.2 µs pertenecen a la etapa Map-Reduce.
- El MapReduce *worker* supone una solución óptima para aplicaciones de altas prestaciones, alta flexibilidad y bajo consumo.
- Para reemplazar la aplicación actual, se han de modificar los cuerpos principales de los bloques Map y Reduce, y ajustar el criterio de la segmentación del bloque Split.

IUMA
Universidad de Las Palmas de Gran Canaria
Campus Universitario de Tafira
Las Palmas de Gran Canaria, SPAIN
tel.: +34 928451086
+34928451250
Fax: +34 928451083
URL: www.iuma.ulpgc.es