

Un caso de transferencia de tecnología de HPC entre CCAD-UNC, FAMAf-UNC y Tarjeta Naranja S.A.

Carlos S. Bederián¹, Gustavo Gaudiano², Julio Larrechart², Marcós Mambrín²,
Martín Onetti³, Alejandro I. Silva³, Nicolás Wolovick³

¹ IFEG-CONICET, Córdoba, Argentina

carlos.bederian@unc.edu.ar

<http://ifeg.famaf.unc.edu.ar>

² Tarjeta Naranja S.A., Córdoba, Argentina

{gustavo.gaudiano,julio.larrechart,marcos.mambrin}@naranja.com

<https://www.naranja.com/>

³ FAMAf-Universidad Nacional de Córdoba, Córdoba, Argentina

{oc,asilva591,nicolasw}@famaf.unc.edu.ar

<http://www.famaf.unc.edu.ar>

Resumen. Luego de algunas reuniones durante el 2018 sobre “Sinergia Universidad-Industria” mediadas por Carlos Gil y Álvaro Ruiz de Mendarozqueta entre Naranja y profesionales de FAMAf-UNC y CCAD-UNC, se plantea la necesidad de trabajar conjuntamente entre la empresa y la FAMAf para abordar distintas problemáticas específicas de la empresa crediticia. El primer problema abordado se enmarca dentro del área de Computación de Alto Desempeño (HPC) a fin de optimizar un proceso en lotes que produce los resúmenes de cuenta en PDF a partir de una salida en formato PCL. El objetivo es procesar 4 millones de resúmenes en 7hs, lo que implica acelerar 3.73x sin cambiar el hardware ni el sistema operativo. Se logra a través de GNU Parallel y la nueva herramienta PDFastix bajar el tiempo a 4.7 hs. El desarrollo, luego de una etapa de adaptación interna de Naranja de aproximadamente 5 meses se encuentra actualmente en producción.

1 Caracterización General del Proyecto

1.1 Instituciones y Empresas Participantes

CCAD-UNC: El Centro de Computación de Alto Desempeño de la Universidad Nacional de Córdoba se crea en 2010 para dar marco a la llegada de un equipo de HPC de gran porte conocido como Cluster Cristina. Actualmente es un centro de servicios de toda la UNC, posee tres supercomputadoras operativas y brinda servicios de cómputo, formación y asesoramiento a toda la comunidad científica de la UNC.

FAMAf-UNC: Creada en 1956 como Instituto de Matemática, Astronomía y Física (IMAF), en la llegada a la democracia en diciembre de 1983 se convierte en Facultad. En 1993 se crea la carrera Lic. en Cs. de la

Computación y en 2018 pasa a llamarse Facultad de Matemática, Astronomía, Física y Computación. La FAMAFA es una facultad pequeña en alumnado dentro de la UNC, pero fuerte en impacto respecto a investigación y vinculación con el medio al medio. Tiene aproximadamente 300 docentes y 1000 alumnos.

Naranja: es la empresa líder en emisión de tarjetas de crédito del país. Actualmente está evolucionando para consolidarse como un ecosistema de servicios basados en la tecnología, con la misión de facilitar la vida de sus clientes a través de experiencias únicas.

Nació en Córdoba, Argentina en 1985, como medio de pago para las cuentas corrientes de la casa de deportes Salto 96 fundada por dos profesores de Educación Física, David Ruda y Gerardo Asrin. En 1995, Banco Galicia se convirtió en su accionista y favoreció su expansión hacia todo el país. Hoy, con 9 millones de plásticos, emite Tarjeta Naranja, Tarjeta Naranja Visa, Tarjeta Naranja Mastercard y Tarjeta Naranja American Express.

Su filosofía de trabajo basada en la amistad, el respeto y la comunicación la posicionan entre las compañías con mejor clima laboral en Argentina.

1.2 Descripción del proyecto:

El proyecto se centra en un problema de desempeño a la hora de generar los resúmenes de cuenta mensuales. El proceso por lotes tarda demasiado, lo cual de alguna manera pone un freno al crecimiento de la cartera de clientes. El objetivo del proyecto es claro y sencillo, además tiene una interfaz limpia con el resto del sistema de generación de resúmenes de cuenta. A partir de archivos PCL que describen el resumen de cuenta, hay que acelerar la transformación a PDF que se hacía con un software privativo específico para la tarea. Dicho software era comandado por un shell script para aprovechar las características multicore del equipo de producción. Se pide reemplazar esa solución por una alternativa que manteniendo el formato de salida, aproveche mejor el equipo de producción que es un servidor enterprise-grade de última generación.

La factibilidad técnica del proyecto está garantizada desde su inicio, ya que la generación de los cuatro millones de resúmenes son procesos independientes, y por lo tanto masivamente paralelizables. Con el hardware de última generación que la empresa posee, no hay ningún cuello de botella en procesador, memoria o almacenamiento que impida bajar dramáticamente los tiempos.

La interfaz se realiza entre un encargado técnico de Naranja y el equipo de FAMAFA/CCAD. Previo al inicio del proyecto se entregan ejemplos de archivos PCL, y el resultado obtenido en PDF. También se realizan dos reuniones presenciales donde se discuten propuestas de mejora que se motorizaron desde la empresa que no lograron solucionar el problema. Hay contacto con dos líderes técnicos de nivel superior pero sólo para cuestiones generales del proyecto.

A través de la Oficina de Vinculación Tecnológica de FAMAFA-UNC,

Naranja pide un servicio que se presupuesta y es aceptado por la misma. Luego de este punto se inicia el proyecto. Desde el primer contacto por el proyecto en julio de 2018, se comienza a trabajar en septiembre de ese año, cuando finalmente podemos acceder a los equipos de desarrollo que contienen el hardware y el sistema operativo similares a los equipos de producción. A fines de octubre se entrega el informe final con la solución que procesa 4.000.000 de resúmenes en 4.7 hs.

1.3 Tipo de Interacción

Colaboración en I+D	X
Asistencia técnica de investigadores a empresas	X
Comercialización de resultados de I+D	
Desarrollo de currícula y clases en conjunto	
Emprendedorismo (start-up, spin-off)	
Otro Especificar:	

Tabla 1: Tipo de interacción universidad-empresa

2 Detalles de Ejecución del Proyecto

2.1 Actividades Realizadas

Dos reuniones presenciales, entrega de scripts UNIX y archivos de ejemplo. Intercambio periódico de correos electrónicos entre todo el equipo tanto de FAMAF/CCAD como de Naranja. Reunión final de entrega de informe, programa y scripts.

2.2 Hitos y Cumplimiento de los Objetivos

El proyecto se plantea como un sprint de 4 semanas con un equipo de 3 personas. Un project manager 10 hs/semana, un technical leader 15 hs/semana y un programador semi-senior 20 hs/semana. El esfuerzo total es de 180 hs/hombre para lograr los objetivos. Dos de los integrantes del equipo son docentes en FAMAF-UNC y también forman parte del CCAD-UNC. El programador semi-senior es un ex-alumno de la materia Computación Paralela que dictan los otros dos integrantes año de por medio.

Al ser un sprint de sólo cuatro semanas, no hubo mojones intermedios. Se trabajó sobre el problema, se obtuvo una solución, se elaboró el informe y se entregó junto al código fuente y los scripts necesarios.

2.3 Principales Desafíos

Gracias a las tres reuniones previas (Enero, Abril y Junio) dentro de la actividad “Sinergia Universidad-Empresa” las partes ya se tenían mutua confianza. Además los líderes técnicos de la empresa y de la institución se conocían previamente del ámbito de las BBS cordobesas de principios de la década de los 90.

Los objetivos del trabajo eran claros, la interfaz con el resto del sistema era muy delgada y con un input/output específico.

Tal vez el mayor problema resultó en el acceso a la plataforma de desarrollo y posteriormente a la de pruebas. La combinación de un hardware y sistema operativo muy diferente al que el CCAD o FAMAFA poseen, hizo muy trabajoso el proceso de medición de desempeño de las posibles soluciones. La exploración fue lenta y se consumió mucho tiempo en esto. Mientras que en nuestros equipos Linux/amd64 podíamos explorar libre y rápidamente, en el entorno de desarrollo teníamos una máquina ocupada, con poco espacio de disco y con procesos de aprobación largos entre un pedido de reconfiguración de sistema y que esto se pusiera en práctica.

Toda el prototipado y pruebas se realizaron en un entorno Linux/amd64 y desde allí se portó al sistema operativo/hardware destino confirmado las hipótesis obtenidas en las pruebas informales de los servidores de HPC del CCAD.

La falta de flexibilidad en la plataforma respecto a compiladores instalados en el sistema de desarrollo y producción también fue un problema a solucionar, donde se tuvo que compilar varios compiladores de manera incremental en versiones hasta obtener el compilador con una versión suficiente que incorporaba optimizaciones necesarias para la plataforma destino. En una plataforma Linux/amd64, esto es un proceso que demora un par de minutos.

3 Resultados del Proyecto

3.1 Resultados de cada Actividad

Se encontró el cuello de botella del problema, se obtuvo una solución que superaba ampliamente el objetivo inicial y que además brindaba ciertas flexibilidades respecto a la generación de resúmenes en PDF que la solución anterior no disponía, como por ejemplo componer el logo de la empresa a colores. El software privativo de conversión de PCL a PDF se siguió utilizando, porque cuando se acota su tarea sólo a la parte que varía entre distintos resúmenes, resulta muchísimo más eficiente que cualquier solución de código abierto para la transformación, como GhostPDL. La solución propuesta además escala de manera lineal con la cantidad de núcleos, con lo cual se puede mejorar el tiempo “tirándole más hardware” al proceso. En pruebas con los nodos de Mendieta¹ segunda generación corriendo Linux/amd64 logramos tiempos

¹ Una de las supercomputadoras administrada por FAMAFA/CCAD

de 1.5 hs para los 4 millones de resúmenes y una eficiencia del 80% respecto a la paralelización.

3.2 Evaluación de los Resultados y Lecciones Aprendidas

El resultado fue mucho mejor de lo esperado, el rápido análisis de factibilidad resultó correcto y la solución planteada es más flexible que lo que se tenía anteriormente en varios aspectos.

Creemos que el acceso a equipos de desarrollo con la plataforma destino lo antes posible y sin tanta carga, hubiera acortado los tiempos y sobre todo hecho menos dificultoso el desarrollo.

La confianza fue clave y darnos acceso a equipos de respaldo de producción, fue una muestra de ello.

4 Balance y Conclusiones

A raíz del éxito que tuvo este proceso de trabajo, ambas instituciones en 2019 firmaron un convenio de colaboración para la asesoría en el desarrollo/mejora de algunos problemas que la empresa planea resolver a través de los servicios que se brindan en FAMAF.

En este marco, actualmente estamos trabajando sobre una nueva problemática planteada por la empresa acerca de autorizaciones crediticias. Hay varios alumnos de Naranja en la “Diplomatura en Ciencia de Datos, Aprendizaje Automático y sus Aplicaciones”. La relación se muestra fructífera y creemos que en estas épocas donde el Presupuesto Nacional en CyT está bajando, una relación buena y madura con empresas del medio local, permitirá mantener vivo el ecosistema del cual se nutren tanto las empresas como las universidades.