



**“ESTUDIO PARA LA MEJOR EXPLOTACION DE LA BASE HISTÓRICA DE
CERTIFICADOS PARA UNA PRONTA Y EXPEDITA IDENTIFICACIÓN Y
MODIFICACIÓN DE DATOS, ASÍ COMO MEJORES NIVELES DE CONTROL Y
SEGURIDAD”**

**ADENDUM - INDEXACIÓN DE DATOS & REPORTE CON ESTADÍSTICAS DE
REGISTROS HISTÓRICOS**

INSTITUTO NACIONAL PARA LA EDUCACIÓN DE LOS ADULTOS



Confidencial

Durante el tiempo que los trabajos asociados a este documento no sean concluidos, la información aquí contenida es propiedad de SID Solutions S.A. de C.V. en lo subsecuente será SID; su uso y/o divulgación está protegido por derechos reservados por lo que al recibir este documento se establece un acuerdo de confidencialidad absoluta.

Al término y entrega de la totalidad de dichos trabajos toda la información generada pasa a ser parte del INEA para que le dé el mejor uso que le convenga.

Control de Cambios

Fecha	Versión	Autor de Modificación	Organización
31 de Octubre de 2017	1.0	Alejandro Castaño Bajo Aprobación de la Dirección de Proyectos de Sid Solutions.	Sid Solutions

Contenido

CONFIDENCIAL	2
CONTROL DE CAMBIOS	3
TRABAJO EN EQUIPO	5
INTRODUCCIÓN	6
ACTIVIDAD	7
CONCLUSIONES	17

Trabajo en Equipo

La metodología de trabajo para desarrollar los proyectos y servicios de consultoría se basan en un trabajo de equipo en conjunto con nuestro cliente, al brindarle apoyo y mejores soluciones a través del conocimiento que nuestros consultores ofrecen basado en la experiencia y en las certificaciones.

SID ofrece servicios de consultoría que aportan valor a nuestros clientes y nuestro objetivo es establecer una relación de confianza y de trabajo que les permita alcanzar sus objetivos, mejorar la operación y enfocar sus esfuerzos en el análisis para lograr ser una empresa más competitiva.

Introducción

De la misma forma en que los primeros cuatro entregables definidos en el Anexo Técnico asociado al "ESTUDIO PARA LA MEJOR EXPLOTACION DE LA BASE HISTÓRICA DE CERTIFICADOS PARA UNA PRONTA Y EXPEDITA IDENTIFICACIÓN Y MODIFICACIÓN DE DATOS, ASÍ COMO MEJORES NIVELES DE CONTROL Y SEGURIDAD", desde ahora "el estudio", este entregable es respuesta al adendum a dicho Anexo y es parte del mismo trabajo hecho a este respecto.

Esta **quinta entrega** tiene fecha de recepción por el INEA el 31 de octubre del 2017, con el siguiente contenido:

1. Indexación de Datos:
 - 1.1. Que espera como entregable una descripción del valor que agrega el estudio a través de desempeño en la generación de métricas respecto a la explotación de información histórica.
2. Reporte con estadísticas de registros históricos:
 - 2.1. Esto debe permitir saber cómo determinar un reporte dinámico de accesos a la información de registros históricos a través de estadísticos en diferentes aspectos de la base de registros.

Actividad

Con el objeto de cumplir con los puntos indicados en la sección de objetivos de este entregable, procedemos a indicar lo siguiente por cada uno de los rubros:

1. Indexación de Datos

Antes de hablar de la Indexación de Datos y sus beneficios, es importante entender que todo proceso de cómputo es reflejo fiel de un proceso humano, es decir, todo programa de cómputo es la automatización de aquello que el ser humano repite bajo ciertas reglas que definen al proceso en sí, y dada esa identificación de repetición de actividades donde el ser humano "ya no aporta valor a esa parte del proceso", es en donde las computadoras entran en acción para repetir y repetir esa acción asegurando su resultado.

Uno de los procesos que el ser humano siempre ha hecho, es el de encontrar respuestas a sus dudas y para ello en principio hay que recabar datos y luego encontrar reglas que describen la correlación entre ellos y describen así un evento o resultado de un proceso, sin olvidar que a lo largo del tiempo la observación y resguardo de los datos observados dieron paso a lo que hoy conocemos como método científico.

Para este tiempo que ahora vivimos, la tecnología nos permite automatizar la manera en que buscamos respuestas a partir de los datos que almacenemos en una base de datos que es una herramienta para la administración de éstos y no es definida sólo como una herramienta computarizada, ya que la historia del hombre nos da evidencia del manejo de datos, como ejemplo de ello tenemos el Código de Hammurabi, primera base de datos de datos no estructurados escrito en la historia del hombre, los códigos prehispánicos con registro de las temporadas, un registro de censo realizado por los romanos en las épocas bíblicas, una agenda telefónica que históricamente se venía manejando en papel.

Ahora bien, con la masificación de la tecnología de cómputo a partir de los años 80's de nuestra era, y la evolución del poder de cómputo ha venido a revolucionar el mundo del manejo de datos a tal grado que hoy un dispositivo llamado "Smart phone" tiene tanto o más poder de cómputo que un equipo llamado "Main frame" de los años 80; esto ha venido a revolucionar la generación de información a partir de relacionar altos volúmenes de datos a través de algoritmos cada vez más complejos dando por resultado información que antes no podía obtenerse tan fácil y tan

rápidamente, más sin embargo tener algoritmos o programas que automatizan procesos con errores y operamos datos sin previa validación o limpieza de ellos, lo único que nos agregará a nuestra vida diaria es tener errores de proceso e información más rápidamente que en años atrás, lo que nos llevará a tener y tomar malas decisiones para nuestro entorno donde usamos estos datos.

Por lo descrito anteriormente, podemos entender que a partir de contar con una computadora donde opere un manejador de bases de datos puedo realizar de manera más rápida la búsqueda de información, pero bajo el entendido que una computadora automatiza procesos y los procesos son factibles de mejorarse a partir de su mejor entendimiento, entonces podemos concluir que un proceso de búsqueda de datos se puede mejorar ¿Cómo? A partir del uso de una estructura de datos definida como Índice de una base de Datos.

Antes de entrar en materia de lo que es un Índice para una base de datos partamos de que la gran mayoría de la humanidad se ha enfrentado a la búsqueda de un tema en un libro, lo que primero se nos enseñó para no tener que leer todo el libro hasta encontrar lo que buscamos fue ubicar la sección llamada "índice" la cual relaciona dos elementos: un Texto y una Página en la que encontraríamos el contenido relacionado al Texto que buscamos. Lo anterior es sencillo de hacer si en el librero sólo existe un libro, pero si en el librero existen varios libros con el mismo título, entonces tengo que encontrar al libro por el autor para luego encontrar el índice que me lleve al texto deseado. Con éste ejemplo vemos que trabajar de esa manera me permitirá llegar al texto deseado pero haciendo una revisión "muy pesada" de todos los datos involucrados en un librero (no digamos que pasaría en una biblioteca). En una base de datos computarizada no es la excepción si bien puedo buscar por un campo llave (p. ej. Estado de la República Mexicana, Primer Apellido, Nombre, etc.), estoy buscando dentro del registro entero de mi base de datos, la cual la debo de cargar en memoria de la computadora, y para los letrados en computación se entiende que llevar datos ente disco y memoria hace que las computadoras se "alenten", por ello mejorando este proceso se encontró que contar con un elemento como un fichero bibliográfico que agiliza la búsqueda de un libro en una biblioteca, así los índices de una base de datos nacen para hacer más ágil la búsqueda dentro de una base de datos, ya que es una estructura de datos más ligera que el todo de una base de datos y permite hacer búsquedas más fáciles dentro de éstas.

¿Qué es un índice de una base de datos? Técnicamente lo definimos como una estructura de datos cuyo objeto es mejorar la respuesta ante una búsqueda de un dato, para ello utiliza un identificador para cada registro dentro de cada tabla que

conforma la base de datos, es decir, en vez de buscar directamente en una tabla un dato, se busca el índice dentro de su tabla y este devuelve el registro completo dentro de la posición que guarda el índice, además no sólo sirven para la búsqueda de registros sino también para el ordenamiento de los datos como mejor sea requerido por el usuario del sistema.

Las tablas de índices normalmente al ser más pequeñas se pueden mantener en la memoria de la computadora lo que hace que el desempeño de búsqueda en el sistema que usa índices es más eficiente que aquellos que tienen que operar con uso de disco y memoria para operar los datos.

Centrados en el objeto de este estudio y considerando que la base de datos sugerida para el manejo de la Base Histórica de Certificados emitidos entre los años 1982 y 2000, sugerimos el uso de índices, ya que si bien la estructura es la siguiente:

- a. Clave de Estado (Campo Llave)
- b. Clave del Centro de Trabajo
- c. Primer Apellido, antes Apellido Paterno (Campo Llave)
- d. Segundo Apellido, antes Apellido Materno (Campo Llave)
- e. Nombre(s) (Campo Llave)
- f. RFE
- g. Matrícula
- h. CURP
- i. Nivel
- j. Promedio
- k. Literal
- l. Folio
- m. Libro
- n. Foja
- o. Fecha de Conclusión (dd/mm/aaaa)
- p. Fecha de Elaboración (dd/mm/aaaa)
- q. Observaciones

Entendemos que la posibilidad de que más de un educando se apellide Gutiérrez, Pérez, Rodríguez, López (algunos de los apellidos más populares en el país), nos permite pensar en crear una tabla de índices por Primer Apellido que agilice el ordenamiento y la búsqueda de los registros por cada estado.

Ejemplo de ello sería ejemplificado de la siguiente manera (fig. 1):

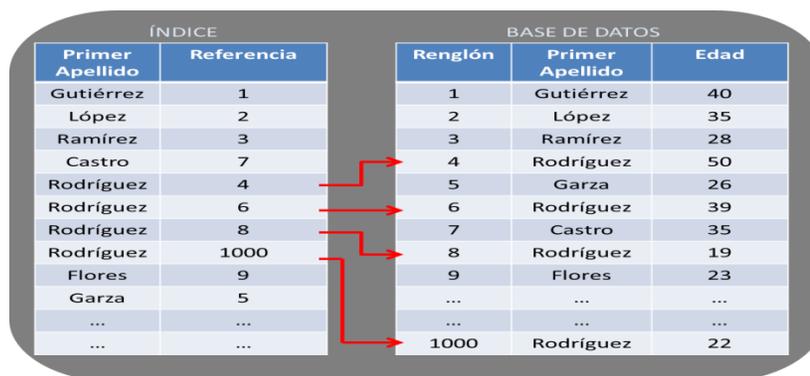


Fig. 1

En donde vemos que la tabla ÍNDICE opera con menos columnas lo que la hace más fácil permanecer en memoria de la computadora vs. la BASE DE DATOS que reside principalmente en disco, y para este ejemplo contempla todas las columnas que maneja el sistema de base de datos. Para acceder a un registro o varios de ellos, con sólo buscar por la columna "Primer Apellido" el apellido en cuestión está dentro de la tabla INDICE el cual me indica por medio de una referencia directamente al número de registro dentro de la base de datos, que de otra forma tendría que ordenarse o usar métodos de búsqueda menos eficientes al uso de índices.

Dentro del entorno tecnológico del INEA tenemos entendido que usan SQL Server de Microsoft como plataforma de bases de datos, el cual cuenta con la funcionalidad de crear índices, por lo que sugerimos que esta práctica se adopte en el caso de que se automatice la administración de la Base Histórica de Certificados emitidos durante el período 1982-2000.

Finalmente, el usar índices en el manejo de una base de datos puede mejorar el tiempo de búsqueda ¿de cuánto ahorro estamos hablando? Eso depende de la complejidad de la base de datos, pero en casos "muy sencillos" de búsqueda usando índices el tiempo se puede reducir en rangos de 70%, pero para tener el dato exacto a este respecto, se debe realizar las pruebas de desempeño contra la base de datos en la que se está trabajando.

2. Reporte con estadísticas de registros históricos

Esta sección debe permitir saber cómo determinar un reporte dinámico de accesos a la información de registros históricos a través de estadísticos en diferentes aspectos de la base de registros.

Ahora bien, antes de desarrollar este tema mismo que complementa las conclusiones ofrecidas en el cuarto entregable del presente estudio, es importante entender que un dato por sí solo no necesariamente es información, es decir, sin un contexto UN dato no es una clara representación de algo lo que se debe entender que no posee información, por ejemplo, si yo presento el número "10", en este momento tengo un dato y la pregunta inminente sería ¿10 qué?, pero si damos el contexto de cupo máximo por curso igual a 10, en ese momento tengo información compuesta por variables llamadas:

- Número Máximo de Asistentes en un curso
- 10 (diez), es el valor a la variable mencionada en el punto anterior

Y ambos datos en conjunto generan la información requerida para quien la necesita, ya sea para exponerla o para quien desea consumirla. De la misma manera si digo que 10 es la calificación máxima de un examen del INEA, los datos que tengo son:

- Valor máximo de la calificación que otorga el INEA
- 10 (diez), es el valor a la variable mencionada en el punto anterior

Así de esta manera para el segundo ejemplo la información que tengo es que un alumno de 10 es aquel que tiene un aprovechamiento total de ese curso que se ha examinado. De esta manera vemos que un mismo dato, el valor 10 para ambos ejemplos, hasta que se combinan con una descripción de una variable a la que se asigna dicho valor, es que obtenemos información.

Entendido lo anterior, podemos entender que existen relaciones de mayor complejidad en donde podemos correlacionar dos o más datos, lo cual dará información en medida de entender el proceso de correlación, por ejemplo, si busco dentro de la estructura sugerida para la administración de la Base Histórica de Certificados del INEA y considerando el ejemplo de la tabla de la fig. 1, información sobre la mayor densidad por estado con personas mayores a 30 años, en este momento tengo que mis datos son: CLAVE ESTADO y EDAD de los alumnos con edad igual o mayor de 30. Para encontrar estos datos tengo que buscar dentro de las tablas de cada estado registros donde el registros EDAD tenga un número mayor o igual a 30 lo cual sumará en uno un contador asociado a "Contador CLAVE ESTADO"

en donde suceda esto, que para lo mostrado en la fig. 1 el "Contador CLAVE ESTADO" será igual a 5 (cinco), que deberá interpretarse que la información que obtuvimos es que 5 alumnos son mayores de 30 años.

Explicado lo anterior, cuando hablamos de un reporte dinámico este no es limitado a tener UN solo reporte, es decir, dependiendo los datos que asocio dentro de una base de datos me puede dar tantos reportes dinámicos como mi necesidad de servicio, negocio o medio para el que solicito la información lo requiera, pero para facilitar este trabajo si se cuenta con el apoyo de un sistema como el que entregamos en nuestra entrega previa podemos entonces contar con soportes de la siguiente índole con sus respectivos beneficios:

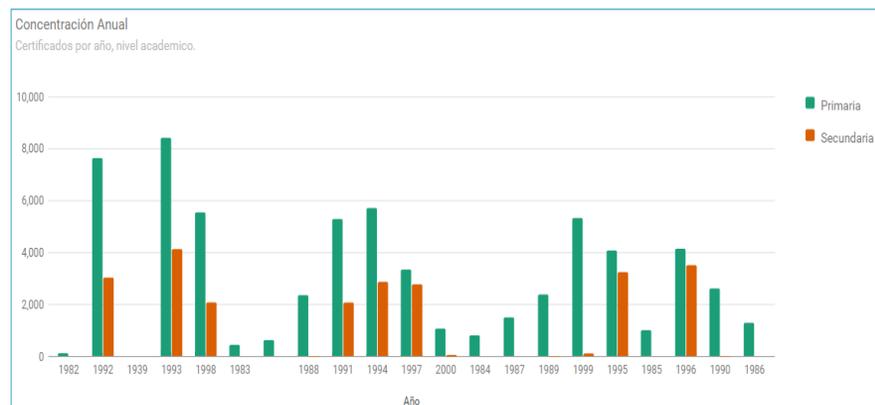
1. Análisis de acciones sobre la base de datos (altas, bajas y cambios), lo que denotará que tanto trabajo se dejó hacer en su momento al solicitar datos a los Educandos, por estado y/o por nivel educativo, y en paralelo se podrá determinar la actividad de actualización que se realiza para mantener lo mejor posible la base histórica de certificados del período comprendido entre 1982 y 2000. Así también se podrá tener:

- 1.1. Un control de registros totales por estado y por nivel educativo
- 1.2. Identificación de registros duplicados
- 1.3. Certeza y confiabilidad de la información

2. Control estadístico por nivel educativo por cada estado

En principio este reporte considerando únicamente los datos de la Base Histórica podríamos verla de la siguiente manera:

Dashboard - Anualizado.



Y ya fuera con datos por estado, por región de la república o todo el país, nos dejaría ver la actividad de certificación que el INEA tuvo en esos años, pero como se ha explicado en otros entregables, si de manera dinámica pudiéramos correlacionar estos datos contra datos de la población objetivo del INEA, podríamos entender que actualmente la emisión de certificados es constante, creciente o decreciente con base a por ejemplo:

- la explosión demográfica de los estados
- el índice de migración que en ellos se manifiesta
- el índice de mortandad

De esta manera el ajuste a los programas del INEA se pueden dar con base en datos duros y con marca de tendencias a partir de un universo mayor de años a los que ahora el INEA cuenta con los sistemas que no operan los datos de la Base Histórica de Certificados emitidos durante los años 1982 al 2000.

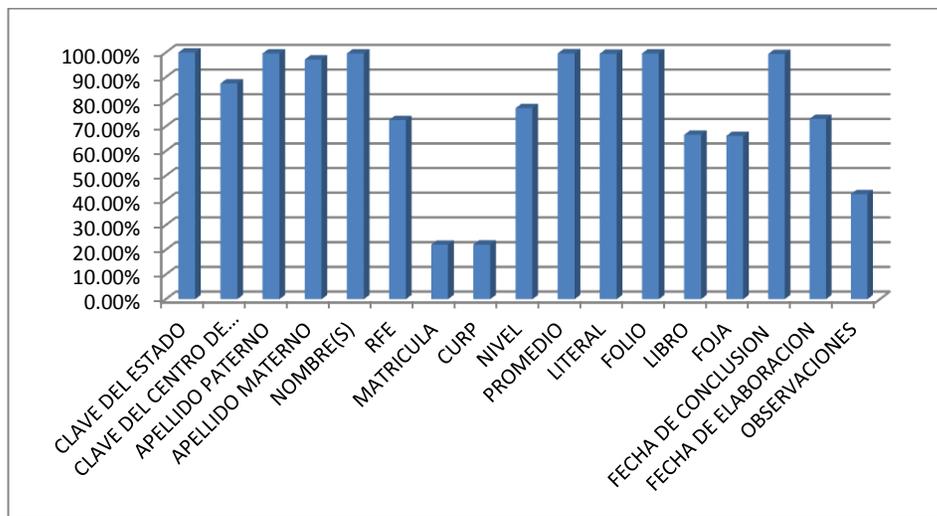
Por ejemplo, considerando datos del Banco Mundial en México en el año 1982 éramos 69.36 millones de mexicanos, mientras que en el año 2000 ya éramos 101.7 millones, es decir, crecimos poblacionalmente 46.63% y estos datos están en esta base histórica, pero al 2016, el mismo Banco Mundial tiene registro de que somos 127.5 millones de mexicanos, es decir, sólo hemos crecido en México 25.37% del año 2000 al 2016. Si esto lo lleváramos a un análisis de cruce entre los datos de esta base histórica de certificados vs. el mismo concepto de registro dentro del sistema SASA, aquí podríamos determinar el alcance que han tenido los programas del INEA y su justificación a partir de otros datos que el mismo instituto pudiera correlacionar a este simple análisis.

3. Control de registros completos vs. incompletos

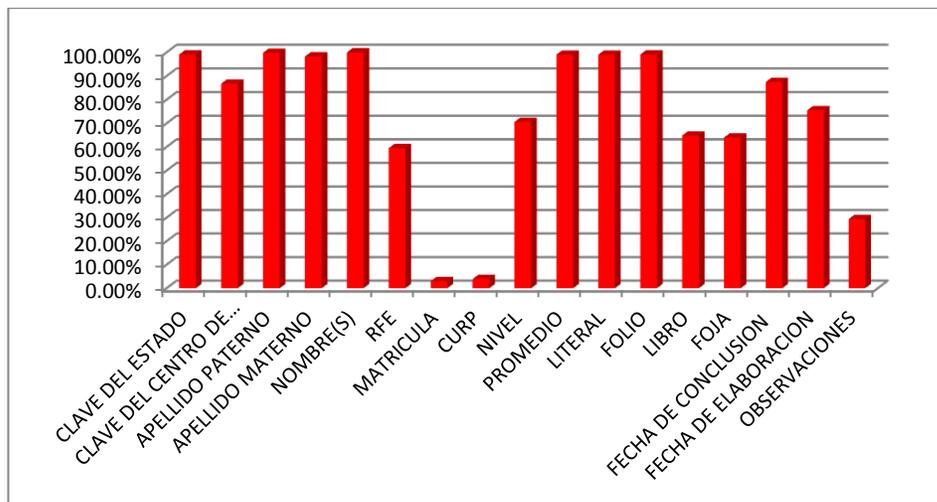
Si hay algo que ha mermado el desarrollo de las áreas de sistemas como áreas de inversión es que en algunos casos el cálculo de almacenamiento de datos tanto en temas tecnológicos como económicos no es preciso, lo que desemboca en recurrentes solicitudes de presupuesto para comprar lo que no se consideró o pedir grandes cantidades de dinero para comprar algo que luego va a quedar demasiado holgado, desperdiciándose la oportunidad de usar esos fondos para otros proyectos.

Por otra parte y pesando que los cálculos de dimensionamiento del sistema son los correctos, no hay mayor daño a dicho análisis que encontrar que los datos que deben aparecer no están dentro de la base de datos, es decir, el proceso de captura de datos por la razón que sea omite esta captura lo que en automático pudiera hacer entender que dicho dato no es necesario, que no se sabe que hay que pedirlo durante

el proceso de captura o la razón que sea, lo cual afecta el tema de inversiones descrito anteriormente y a su vez impactará la calidad de los reportes que involucren estos datos. Por ejemplo, en los primeros análisis de datos que hicimos para este estudio, percibimos que hay una fuerte carencia de datos de acuerdo a lo descrito en las siguientes gráficas, en donde obtener una correlación a través del RFE, CURP, LIBRO, FOJA o Fecha de Emisión, pudiera dejarme sin posibilidades reales de obtener un dato duro, sólo por el hecho de que al no estar presentes estos datos dentro de la base histórica, eso sólo me permitiría hacer inferencias y proyecciones sobre algún análisis donde estos datos se consideren relevantes.



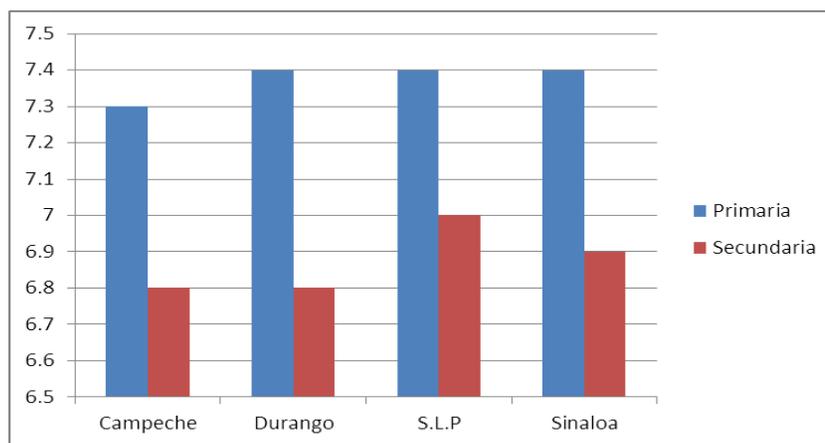
Primaria



Secundaria

4. Mayor visibilidad de los promedios de EDUCANDOS por estado y por nivel (Visibilidad para crear mayores acciones para homologar el desarrollo educacional a cargo del INEA)

La siguiente gráfica toma los datos de las Bases Históricas de los Estados allí mencionados y este ejemplo nos permite ver que hay similitudes entre los valores del Promedio Académico entre niveles y por estado.

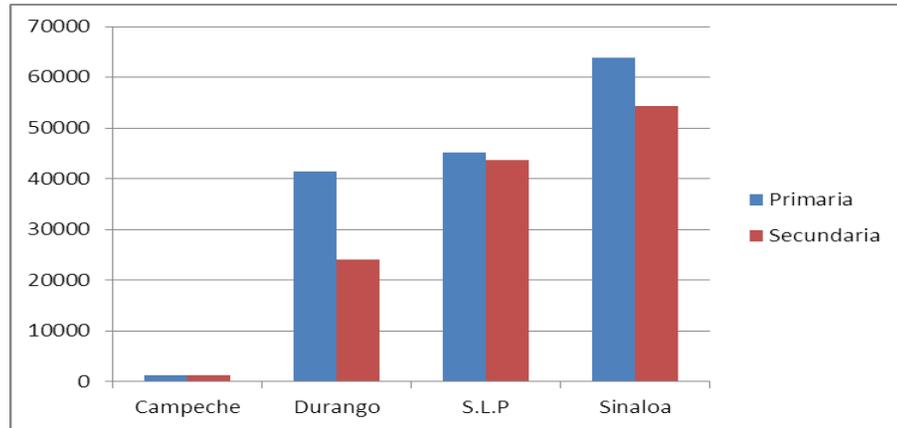


Con lo anterior se puede decir que los programas del INEA en estos estados ha sido homogéneo y que el nivel de aprovechamiento que los Educandos tienen es similar, más sin embargo esto puede generar iniciativas de cómo hacer que el nivel de calificación promedio suba y no decaiga en más de medio punto entre primaria y secundaria.

Otra vista que se puede dar a estos datos, es compararla contra el promedio académico de la población fuera del INEA, con el objeto de dar una explicación pedagógica ante las posibles diferencias que se pudieran analizar.

5. Conocimiento de Certificados emitidos durante el período

La siguiente gráfica toma los datos de las Bases Históricas de los Estados allí mencionados y este ejemplo nos permite ver que hay similitudes entre la cantidad de certificados registrados.



Esta gráfica deja ver que la diferencia en la certificación de educandos cae en promedio por estos estados tomados como ejemplo un 14%, siendo Campeche el más estable mientras que Durango presenta el mayor desfase con un 42% de diferencia entre certificados emitidos en primaria vs. emitidos en secundaria.

Este tipo de análisis al ser extendido de manera dinámica contra otras bases de datos del INEA, por ejemplo la base histórica de certificados almacenada en el sistema SASA a partir del año 2000 permitiría encontrar si el INEA tiene la misma penetración en los estados, permitiendo ubicar por ejemplo, si el nivel de educandos ha crecido o no, tanto en número como proporcionalmente y de la misma manera determinar si el aprovechamiento de los Educandos ha crecido, se ha mantenido o ha decrecido.

Conclusiones

Desprendido del presente estudio concluimos lo siguiente:

- El INEA tienen una oportunidad histórica-operativa de tener en control homogéneo todos los datos asociados a los certificados emitidos desde 1982 y hasta el año 2000 en una plataforma como la que Sid Solutions desarrolló de manera básica y entregada en la 4ª entrega de este estudio, y que en conjunto con desarrollos que permitan explotar esta base histórica con la que el INEA administra a través del sistema SASA, puede generar la suficiente información que de base para consolidar, mejorar y crear estrategias de trabajo y mantener al INEA como pieza clave del desarrollo educativo de México.
- El desarrollar una sistema que maneje la base histórica de certificados dentro de un manejador de bases de datos como lo es SQL Server de Microsoft (estándar en INEA), y el desarrollo de la aplicación para la explotación de datos apoyado de las mejores prácticas técnicas como lo es el uso de índices, harán que los tiempos de búsqueda se reduzcan lo que se traduce en eficiencia operativa de aquello que depende de la identificación de datos y creación de información, como por ejemplo la identificación de un registro para poder emitir un certificado bajo políticas del INEA.

Finalmente agradecemos al INEA habernos solicitado este complemento al trabajo hecho para el **“ESTUDIO PARA LA MEJOR EXPLOTACION DE LA BASE HISTÓRICA DE CERTIFICADOS PARA UNA PRONTA Y EXPEDITA IDENTIFICACIÓN Y MODIFICACIÓN DE DATOS, ASÍ COMO MEJORES NIVELES DE CONTROL Y SEGURIDAD”**, ya que nos dio la posibilidad de explicar de la manera más concreta los beneficios que la tecnología puede traer al INEA, todo ello apoyado en la teoría de sistemas como en la experiencia que en otros proyectos hemos desarrollado en bien de nuestros clientes y México.