



**El futuro
es de todos**

**Gobierno
de Colombia**

**EVALUACIÓN DE PILOTOS
Piloto MONITOREO Y ANALÍTICA**

**OFICINA ASESORA DE PLANEACIÓN
Monitoreo y Evaluación**

LINA MARCELA GALLEGO RUIZ
Jefe Oficina Asesora de Planeación

Claudia Paola Bustos Martínez
Martha Liliana Cruz Bermúdez
Jorge Claudio Cotes Martínez
Profesionales de Monitoreo y Evaluación

BOGOTÁ D.C., DICIEMBRE DE 2019



1. INFORMACIÓN GENERAL

Nombre del Piloto: Piloto MONITOREO Y ANALÍTICA
Fecha de Inicio: 01 de octubre de 2019
Fecha terminación: 19 de diciembre de 2019
Fecha de evaluación: 30 de diciembre de 2019
Proponente: GRUPO DINAMIZA S.A.S.
Contacto: Luis Fernando Duque
E-Mail: luisfernando@dinamiza.co
Teléfono de contacto: 3124493712
Otras entidades: NA
Componentes: NA
Costo para COMPUTADORES PARA EDUCAR: Gastos de recurso en Nube Azure y Licenciamiento P.B.I.

2. JUSTIFICACIÓN

Desde el inicio del programa COMPUTADORES PARA EDUCAR ha realizado la entrega de terminales a las sedes educativas, y desde el año 2014 se cuenta con agentes de seguridad instalados en las terminales donadas, los cuales capturan información como Georreferenciación, Geocercas, activación y desactivación de equipos, generando gran cantidad de información. Actualmente es necesario realizar el aprovechamiento de la información capturada, mediante la definición de alternativas de solución para la obtención, procesamiento y análisis de la información, que le permita a la institución la toma de decisiones basadas en datos.

En este sentido, y para fortalecer el proceso de monitoreo y analítica, se hizo necesario adelantar un piloto que permitiera diseñar e implementar, una solución basada en nube utilizando herramientas de Microsoft Azure para conectarse a diferentes fuentes de datos, unificar la información, depurarla y prepararla para ser consumida por un modelo de analítica descriptiva que permita resolver preguntas propias del negocio, basándose en la información histórica con la que cuenta el agente de seguridad Absolute y otras fuentes de información interna.

Lo anterior permitirá establecer posibles acciones frente a la analítica de la información, en pro de la toma de decisiones basadas en datos.

3. OBJETIVOS INICIALES DEL PILOTO

General

Diseñar e implementar a manera de piloto, una solución basada en nube utilizando herramientas de Microsoft Azure para conectarse a diferentes fuentes de datos, unificar la información, depurarla y prepararla para ser consumida por un modelo de analítica descriptiva que permita resolver preguntas propias del negocio.

Específicos



1. Realizar pruebas con el reporte generado por al menos un (1) agente de seguridad instalado en las terminales donadas por COMPUTADORES PARA EDUCAR, realizando el análisis de los reportes generados por este.
2. Analizar las herramientas antirrobo para determinar algunas de las características y/o especificaciones necesarias para los equipos donados por COMPUTADORES PARA EDUCAR.
3. Reconocer mecanismos, metodologías y recursos necesarios para realizar una implementación masiva de herramientas de monitoreo y uso de equipos donados por COMPUTADORES PARA EDUCAR.
4. Establecer las variables mínimas necesarias a monitorear en las terminales donadas por COMPUTADORES PARA EDUCAR en las sedes educativas oficiales del país.
5. Definir la utilidad y la viabilidad técnica de la obtención y análisis de información, en función de:
 - Rendimiento de software y hardware de las máquinas.
 - Grado de uso de software y hardware.
 - Uso por parte de docentes y estudiantes.
 - Necesidad de conectividad y otra infraestructura.

4. CONCEPTUALIZACIÓN

Este piloto se realiza teniendo en cuenta la metodología CRISP-DM (por sus siglas en inglés Cross Industry Standard Process for Data mining), aplicada a un caso de negocio de la entidad, tiene como finalidad, facilitar el entendimiento del uso de técnicas de analítica básica utilizando herramientas en nube, para este caso herramientas que corren sobre Microsoft Azure.

Cómo parte de la metodología se establecieron las siguientes preguntas de negocio, que buscan analizar diferentes variables identificadas en los datos existentes para el presente piloto:

- ¿Es posible conocer la causa raíz de los mantenimientos en una zona determinada?
- ¿Cuál es el contenido de mayor uso en las terminales? (Docentes, estudiantes y personal administrativo)
- ¿Qué tanto se usan las terminales?
- ¿Los contenidos que se encuentran instalados en las terminales fueron colocados por el usuario final?
- ¿Tipificación adecuada de los tipos de casos que reciben cualquiera de los niveles de soporte técnico realizados en los equipos?

Así mismo se establecieron los criterios de éxito, los cuales dependen directamente de la calidad de los datos y la existencia de datos relacionados en las bases disponibles para la realización del piloto.

- Identificar al menos 10 problemas de mantenimiento con varios niveles de desagregación (zonificación, tipo de terminal, marca, entre otros).
- Identificar al menos 20 contenidos de mayor uso.



- Establecer rangos de uso de las terminales
- Identificar contenidos instalados por los docentes en las sedes.
- Identificación de tipos de casos que reciben cualquiera de los niveles de soporte técnico realizados en los equipos.

Los recursos requeridos para el desarrollo del presente piloto son:

Recursos de TI

Crédito en nube para configuración del modelo que se diseñe, proporcionado por COMPUTADORES PARA EDUCAR.

Recurso humano

Un (1) Ingeniero para configuración de herramientas de nube, proporcionado por COMPUTADORES PARA EDUCAR y Dinamiza.

Un (1) Ingeniero para extracción, transformación, cargue y preparación de información, proporcionado por Dinamiza.

Un (1) ingeniero para modelado de datos y visualización de resultados, proporcionado por Dinamiza.

Un (1) recurso de negocio para apoyo al entendimiento y validación de resultados, proporcionado por COMPUTADORES PARA EDUCAR y Dinamiza.

Fuentes de datos

Se utilizarán las siguientes fuentes de datos:

- Datos de mantenimiento provenientes de la BD “Casos soporte técnico 2014-2018”.
- Datos de dispositivos provenientes de la nube de Absolute.
- Datos de despacho de COMPUTADORES PARA EDUCAR.

Recopilación de datos

A continuación, se presentan los datos disponibles para la ejecución del piloto.

Ítem	Nombre fuente	Cantidad de registros	Peso	Formato
1	Reportes	148.865	42.7M	XLS
2	Reporte Monitoreo	+65.000.000	4.29G	XLS

Diccionario de datos

Los campos descripción y observaciones, aunque vacíos, se dejan para resaltar la importancia de contar con ellos para ejercicios futuros.



Tabla 1. Reportes

ITEM	Campo	Descripción	Observaciones	Requerido
1	ESN			SI
2	UserName			SI
3	UserMail			SI
4	PCName			SI
5	HostName			SI
6	Department			SI
7	Serial			SI
8	Asset			SI
9	Make			SI
10	Model			SI
11	AssignedUserName			SI
12	lastAssetCall			SI
13	Group1			NO
14	Group2			NO
15	Group3			NO
16	Group4			NO
17	Group5			NO
18	Group6			NO
19	Group7			NO
20	Group8			NO
21	Group9			NO
22	Group10			NO
23	Group11			NO
24	Group12			NO
25	Group13			NO
26	Group14			NO
27	Group15			NO
28	Group16			NO
29	Group17			NO
30	Group18			NO
31	Group19			NO
32	Group20			NO
33	LocalIP			SI
34	ProxyIP			SI

De la “Tabla1. Reportes” se seleccionan los siguientes datos para la ejecución del piloto, tomando la decisión de utilizar la totalidad de los registros, sin embargo, solo se utilizarán los campos seleccionados para construir el modelo.

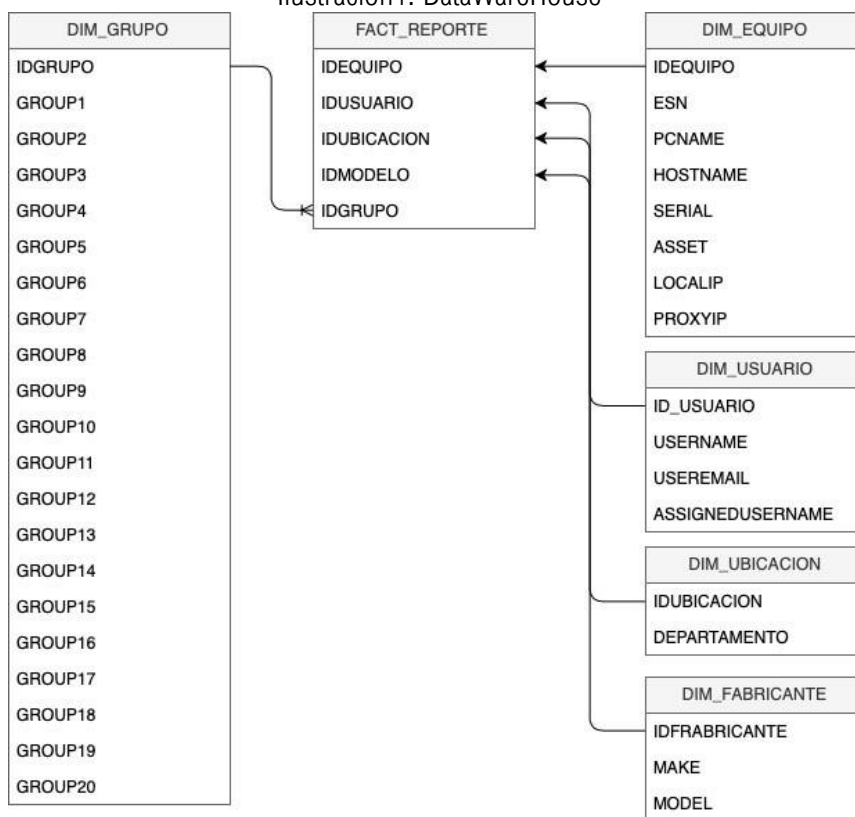
ITEM	Campo	Validación
1	ESN	SI
2	UserName	SI
3	UserMail	SI
4	PCName	SI
5	HostName	SI
6	Department	SI
7	Serial	SI



8	Asset	SI
9	Make	SI
10	Model	SI
11	AssignedUserName	SI
12	lastAssetCall	SI
13	LocalIP	SI
14	ProxyIP	SI

Para la construcción del modelo de datos se utilizará una estructura de datos del tipo DataWareHouse, dando formato a los datos de tipo numérico y tipo fecha, a fin de poder realizar los filtros asociados a rangos de fecha y otros, como se muestra en la siguiente ilustración:

Ilustración1. DataWareHouse



5. RESULTADOS Y ANÁLISIS

Como resultado del ejercicio, se tiene que la data proporcionada no ayuda a responder las preguntas de negocio. No obstante, se tomó la decisión de seguir adelante para ver como fluyen los datos desde una fuente hasta un tablero de control y se presentan los resultados como se indica a continuación:

Grupo de recursos

Cuenta de almacenamiento



El futuro
es de todos

Gobierno
de Colombia

Microsoft Azure portal showing the 'datamontoreo' container. The interface includes a search bar, navigation pane, and a table of blobs.

Nombre	Modificado	Nivel de acces...	Tipo de blob	Tamaño	Estado de con...
Casos soporte tecnico 2014 2.csv	4/12/2019 19:07:17	Esporádico (inf...	Blob en bloques	3.87 MiB	Disponible ***
Casos soporte tecnico 2014 3.csv	4/12/2019 20:34:21	Esporádico (inf...	Blob en bloques	982.05 KiB	Disponible ***
Casos soporte tecnico 2014 4.csv	5/12/2019 7:59:27	Esporádico (inf...	Blob en bloques	1.82 MiB	Disponible ***
Casos soporte tecnico 2014-2018.xlsx	5/11/2019 17:15:58	Esporádico	Blob en bloques	14 MiB	Disponible ***
Casos soporte tecnico 2014.csv	27/11/2019 17:42:07	Esporádico (inf...	Blob en bloques	6.16 MiB	Disponible ***
Casos soporte tecnico 2015.csv	27/11/2019 17:42:07	Esporádico (inf...	Blob en bloques	3.81 MiB	Disponible ***
Casos soporte tecnico 2016.csv	27/11/2019 17:42:13	Esporádico (inf...	Blob en bloques	6.32 MiB	Disponible ***
Casos soporte tecnico 2017.csv	27/11/2019 17:42:11	Esporádico (inf...	Blob en bloques	5.32 MiB	Disponible ***
Casos soporte tecnico 2018.csv	27/11/2019 17:42:07	Esporádico (inf...	Blob en bloques	10.36 MiB	Disponible ***
Reporte_Absolute.csv	5/11/2019 17:32:23	Esporádico	Blob en bloques	40.86 MiB	Disponible ***

Base de datos

Microsoft Azure portal showing the 'datamontoreo' database. The interface includes a search bar, navigation pane, and a list of tables.

Consultas

- Inicio de sesión
- Nueva consulta
- Abrir consulta
- Comentarios

Consultas

- Consultar
- Cancelar consulta
- Guardar consulta
- Exportar datos como json
- Exportar datos como .csv

Consultas

- Consultar
- Cancelar consulta
- Guardar consulta
- Exportar datos como json
- Exportar datos como .csv

Consultas

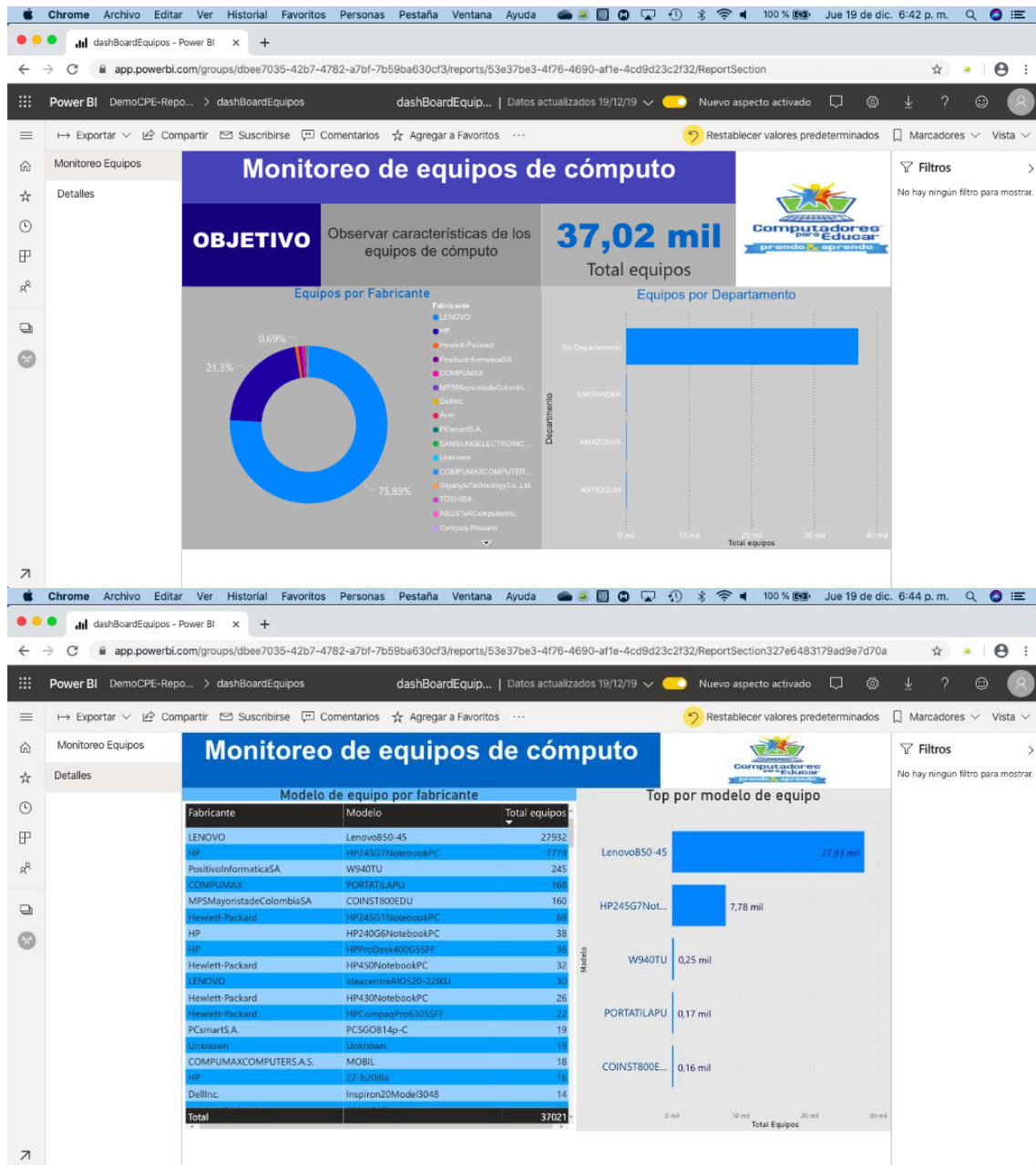
- Consultar
- Cancelar consulta
- Guardar consulta
- Exportar datos como json
- Exportar datos como .csv

Visualización mediante tableros de control



El futuro
es de todos

Gobierno
de Colombia





El futuro
es de todos

Gobierno
de Colombia



Adicionalmente y teniendo en cuenta las preguntas de negocio y los criterios de éxito, se tienen los siguientes resultados:

- ¿Es posible conocer la causa raíz de los mantenimientos en una zona determinada?

Criterio de éxito: Identificar al menos 10 problemas de mantenimiento con varios niveles de desagregación (zonificación, tipo de terminal, marca, entre otros).

Resultado: El ejercicio permitió identificar problemas de mantenimiento por zona, terminal y marca para las vigencias 2014 y 2015, sin embargo, para el resto de las vigencias no fue posible debido a la calidad de la data y la variación en la estructura de las BD "Casos soporte técnico 2014-2018".



- ¿Cuál es el contenido de mayor uso en las terminales? (Docentes, estudiantes y personal administrativo)

Criterio de éxito: Identificar al menos 20 contenidos de mayor uso.

Resultado: No fue posible determinar esta información debido a que las fuentes de datos no contaban con información relacionada con el uso de contenidos.

- ¿Qué tanto se usan las terminales?

Criterio de éxito: Establecer rangos de uso de las terminales

Resultado: No fue posible determinar esta información debido a que las fuentes de datos no contaban con información relacionada con el uso de terminales.

- ¿Los contenidos que se encuentran instalados en las terminales fueron colocados por el usuario final?

Criterio de éxito: Identificar contenidos instalados por los docentes en las sedes.

Resultado: No fue posible determinar esta información debido a que las fuentes de datos no contaban con información relacionada con contenidos instalados en las terminales ni tampoco quien los instaló.

- ¿Tipificación adecuada de los tipos de casos que reciben cualquiera de los niveles de soporte técnico realizados en los equipos?

Criterio de éxito: Identificación de tipos de casos que reciben cualquiera de los niveles de soporte técnico realizados en los equipos.

Resultado: El ejercicio permitió identificar la tipificación adecuada de los tipos de casos que reciben cualquiera de los niveles de soporte técnico realizados en los equipos para las vigencias 2014 y 2015, sin embargo, para el resto de las vigencias no fue posible debido a la calidad de la data y la variación en la estructura de las BD “Casos soporte técnico 2014-2018”.

OTROS RESULTADOS

Al visualizar los datos encontramos falencias en la data e inconsistencias en las bases de datos, tales como:

- Falta de estandarización de las variables utilizadas como nombres de departamentos, municipios, estructura de fechas, identificación de columnas vacías, entre otros.
- Estructuración de las bases de datos es variable en cuanto al número de columnas y datos de captura, información incompleta.
- La sistematización de la información no es estandarizada y por tanto presenta inconvenientes para su consolidación. Por ejemplo:
 - Antioquia, ANTIOQUIA, antioquia.
 - Compumax, Compumax S.A.S, Compumax Computer S.A.S.
 - Medellín, MEDELLÍN, Medellin, MEDELLIN.
 - Departamento, Dpo, Dpto, DPTO.



- No fue posible analizar la totalidad de los datos aportados para el piloto, por deficiencias en la calidad de estos.
- No hay unidad en la identificación de las sedes, en algunas no se tenía el número DANE, o el número de radicado por tal razón no fue posible identificar el número de terminales en las sedes ni realizar el cruce de bases de datos para un análisis más completo.
- A nivel institucional no hay una estructura de datos que permita realizar un análisis general del comportamiento de los programas misionales.
- Se desconoce el diccionario de datos existentes en la entidad, y esto dificulta el planteamiento de las preguntas de negocio y la proyección de las actividades de monitoreo entorno a la captura de información para procesos de analítica.

6. CONCLUSIONES

- Al revisar la información y reportes generados por los agentes de seguridad instalados en las terminales donadas por COMPUTADORES PARA EDUCAR, se encontraron hallazgos como:
 - Los agentes de seguridad no capturan información que genere valor agregado a la entidad, en materia de monitoreo
 - El licenciamiento es limitado, por lo cual no permiten acceder a información adicional que podría ser útil para el monitoreo
 - No hay unificación de criterios en la información capturada por los diferentes agentes de seguridad
 - Se identifica la necesidad de contar con agentes de seguridad que capturen la misma información, con la calidad y arquitectura establecida para satisfacer las necesidades institucionales.
- Identificación de herramientas internas que permiten optimizar los procesos de analítica de la entidad tal es el caso de Microsoft Azure, PowerBI.
- Se evidencian las necesidades de fortalecer la calidad de la información generada por los procesos misionales y almacenada en sistemas de información de la entidad.
- Se identifica la necesidad de articular y estandarizar la información generada por las diferentes fuentes de COMPUTADORES PARA EDUCAR.
- Potenciar el uso de la herramienta Azure la cual estaba limitada solo a TI
- Las herramientas utilizadas requieren del conocimiento avanzado de Azure y otros conceptos de herramientas ETL y procesamiento en la nube.
- Hay un desconocimiento con respecto al manejo y potencial de las herramientas de analítica y visualización de Microsoft
- Potenciar la capacidad para procesamiento de datos del equipo de monitoreo.
- Al ejecutar el piloto se observa que la velocidad de respuesta se mejora debido al trabajo con los datos y procesamiento en nube.

7. RECOMENDACIONES.

- Es necesario que la fuente de captura de los datos se realice de manera automática, dando uniformidad a los datos y una calidad aceptable para su procesamiento.
- Se debe ampliar o adquirir un licenciamiento de agentes instalados o a instalar en las terminales donadas por COMPUTADORES PARA EDUCAR, para evitar:



- La ausencia de captura de información que genere valor agregado a la entidad
 - No acceso a la información que capturan los agentes
 - Falta de unidad de criterios en la información capturada.
- Se recomienda aplicar acciones orientadas a:
 1. Realizar un inventario de las fuentes de datos existentes
 2. Para cada fuente se debe identificar la data que resulta de interés para la entidad
 3. Construir los diccionarios de datos respectivos
 4. Definir unas reglas de validación de datos clave
 5. Identificar las fuentes de contraste de datos (Ej. tabla con departamentos y municipios en Colombia - DIVIPOLA)
 6. Construir una estructura dimensional que sirva de base para ejercicios de analítica básica y avanzada.
 7. Actualizar los procesos y procedimientos de la entidad