

Inspira Crea Transforma

Comité de Gobierno de Datos y Analítica



Direccionamiento Estratégico | Dirección de Informática

Mayo 2020

AGENDA DEL COMITÉ

1

Balance de las socializaciones del proyecto

2

Reestructuración del plan de trabajo

3

Avances en cada uno de los proyectos asociados

→ Diagnóstico inicial del uso de la información

→ Calidad de datos

→ Políticas

→ Arquitectura de BI

Socializaciones del programa con las distintas dependencias

15 — Socializaciones realizadas

- Sistemas Integrados de Gestión y Control
- Dirección de Informática
- Admisiones y Registro
- Oficina de Relaciones Internacionales
- Desarrollo Estudiantil

4 — Socializaciones pendientes

Nuevas socializaciones

- Oficina para el éxito del estudiante
- Mercadeo
- Desarrollo Institucional y Egresados
- EAFIT Llanogrande
- Vic. Sostenibilidad y Proyección Social
- Departamento de Prácticas Profesionales
- DIAF
- Innovación
- Secretaría General
- Idiomas

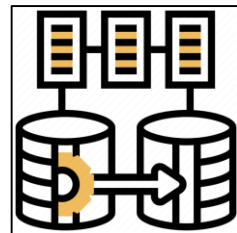
Proceso de aprendizaje con las dependencias



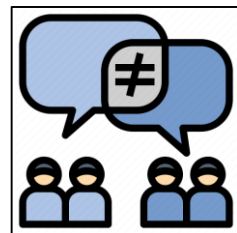
Complejidad



Responsabilidad compartida



Proceso de migración



Lenguaje

Replantear el alcance de los proyectos

Identificación del Gobierno actual

Uso de los datos e información

Calidad de los datos

Políticas

Analítica

Arquitectura de BI

Infraestructura tecnológica

Reportes institucionales

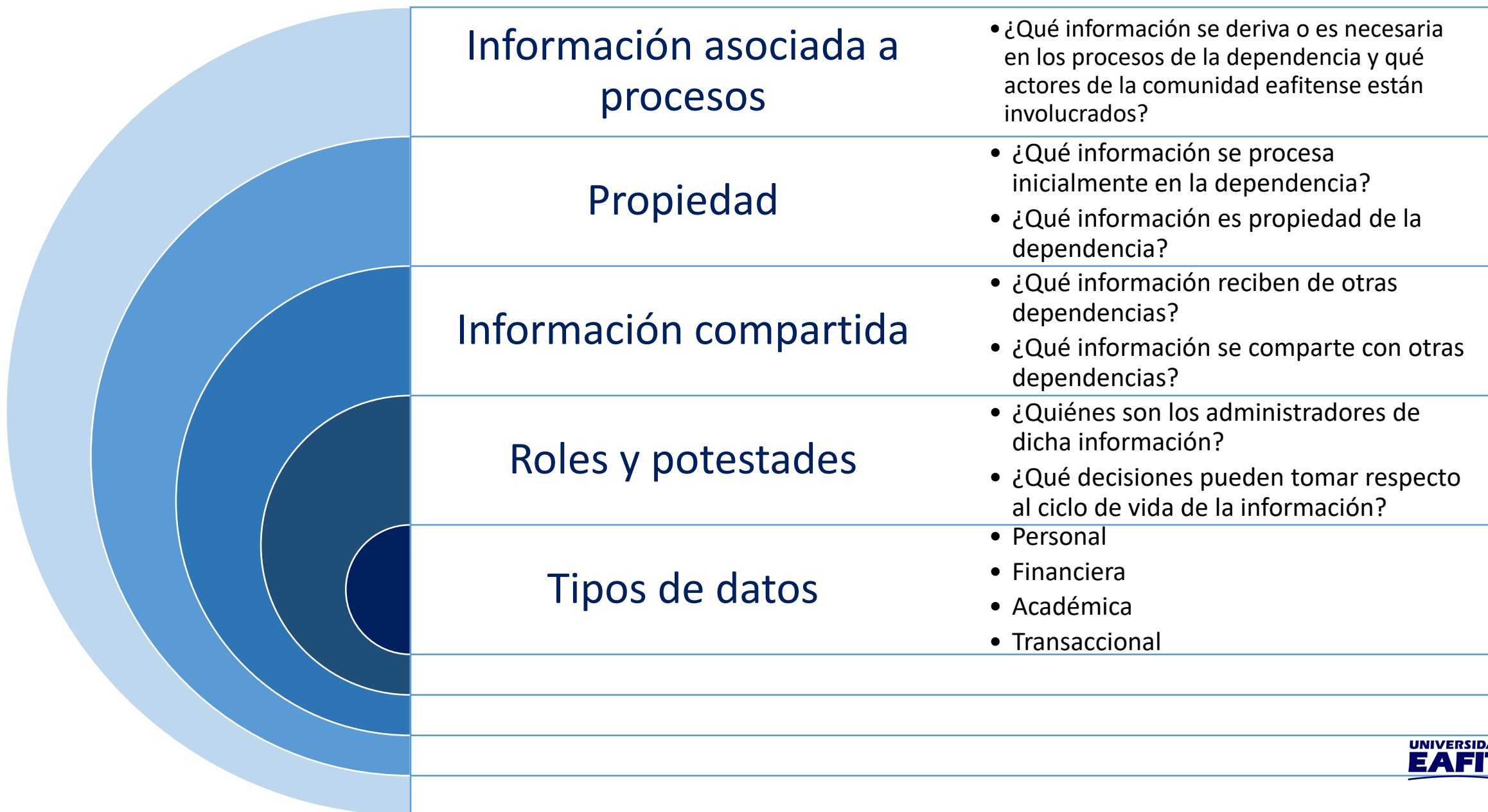
Visualizador institucional

Data Warehouse

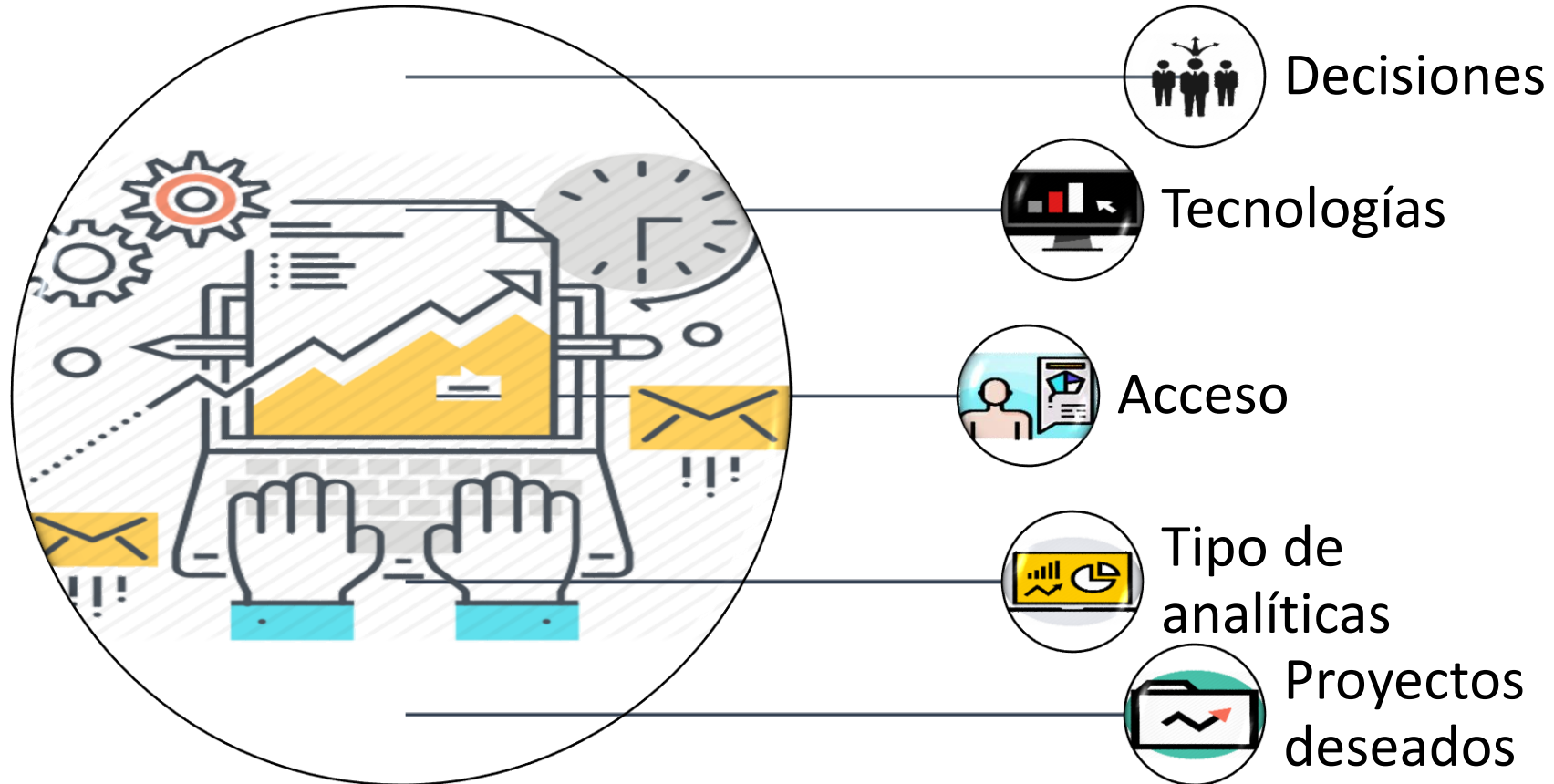
② Reestructuración

Plan de trabajo

Información



Analíticas



Reportes

¿Qué es analítica?



La analítica es el descubrimiento, la interpretación
la comunicación de patrones significativos en los
datos.

“ understood as the connection between data
and effective decision making within an
organization”



2002, Emerging Trends in Business Analytics, Ron Kohavi

Tipos de analíticas

¿Qué?

- Analíticas descriptivas

¿Por qué?

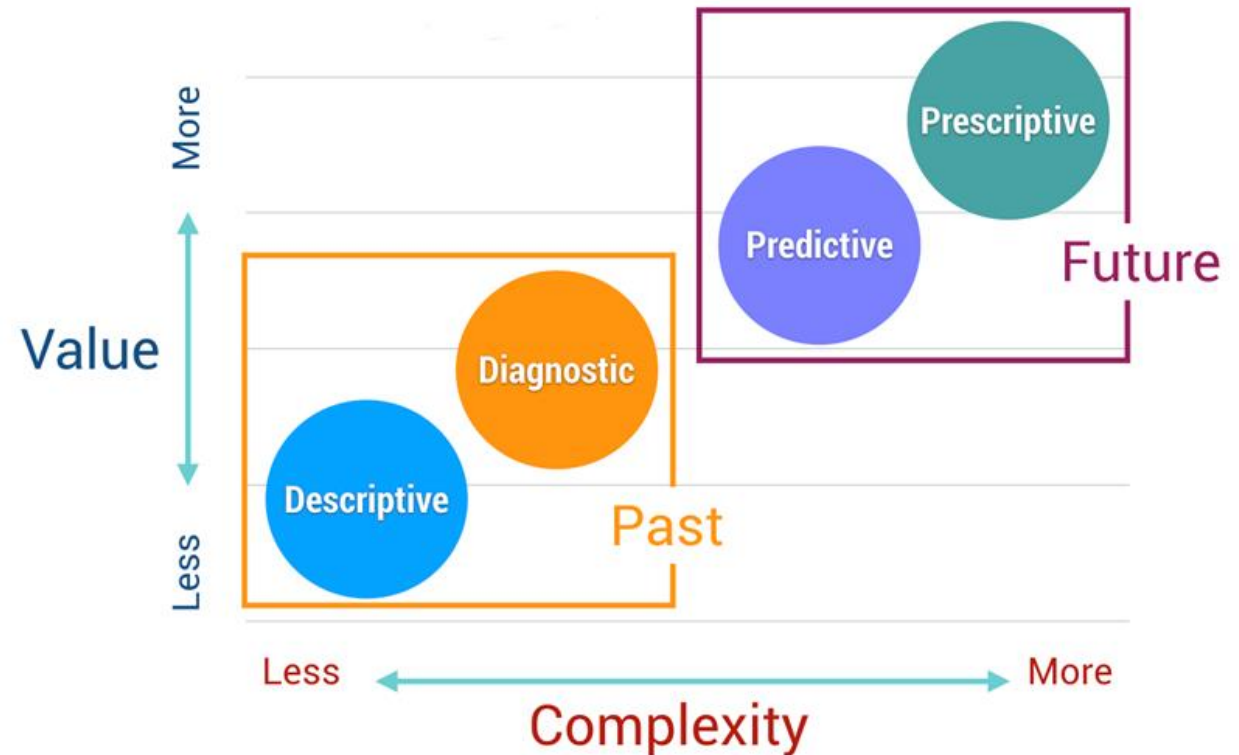
- Analíticas diagnósticas

Si...

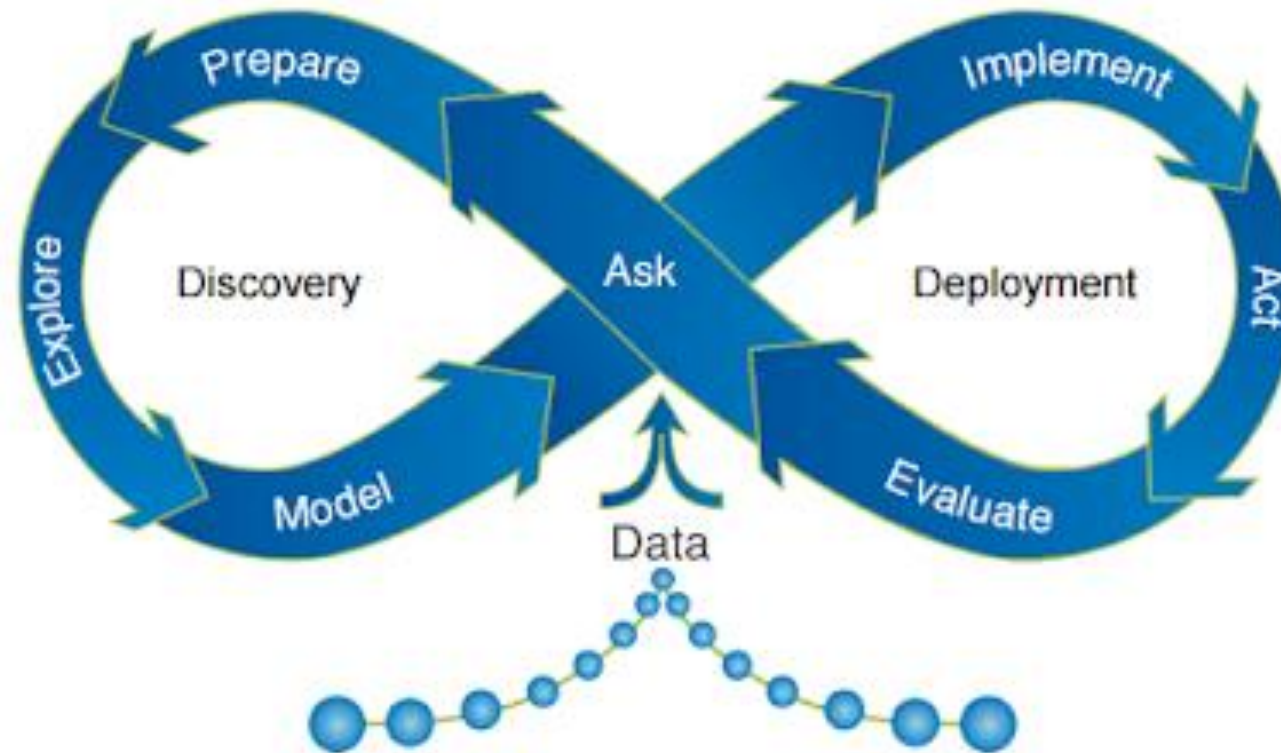
- Analíticas predictivas

¿Cómo?

- Analíticas prescriptivas



Proceso de Analítica

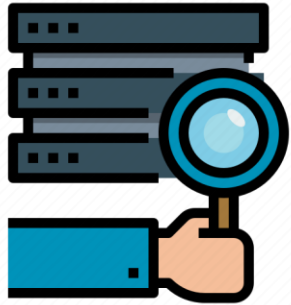


Fuente: SAS - Gestionando el ciclo de vida analítico para la toma de decisiones a gran escala

Roles y funciones

Extraer Información

- ✓ Ingeniero de datos



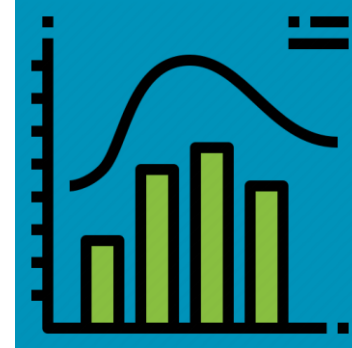
Procesar información

- ✓ Analista de datos



Modelar

- ✓ Científico de datos



Difusión

- ✓ Analista de negocio



3.1

Diagnóstico del uso de la información

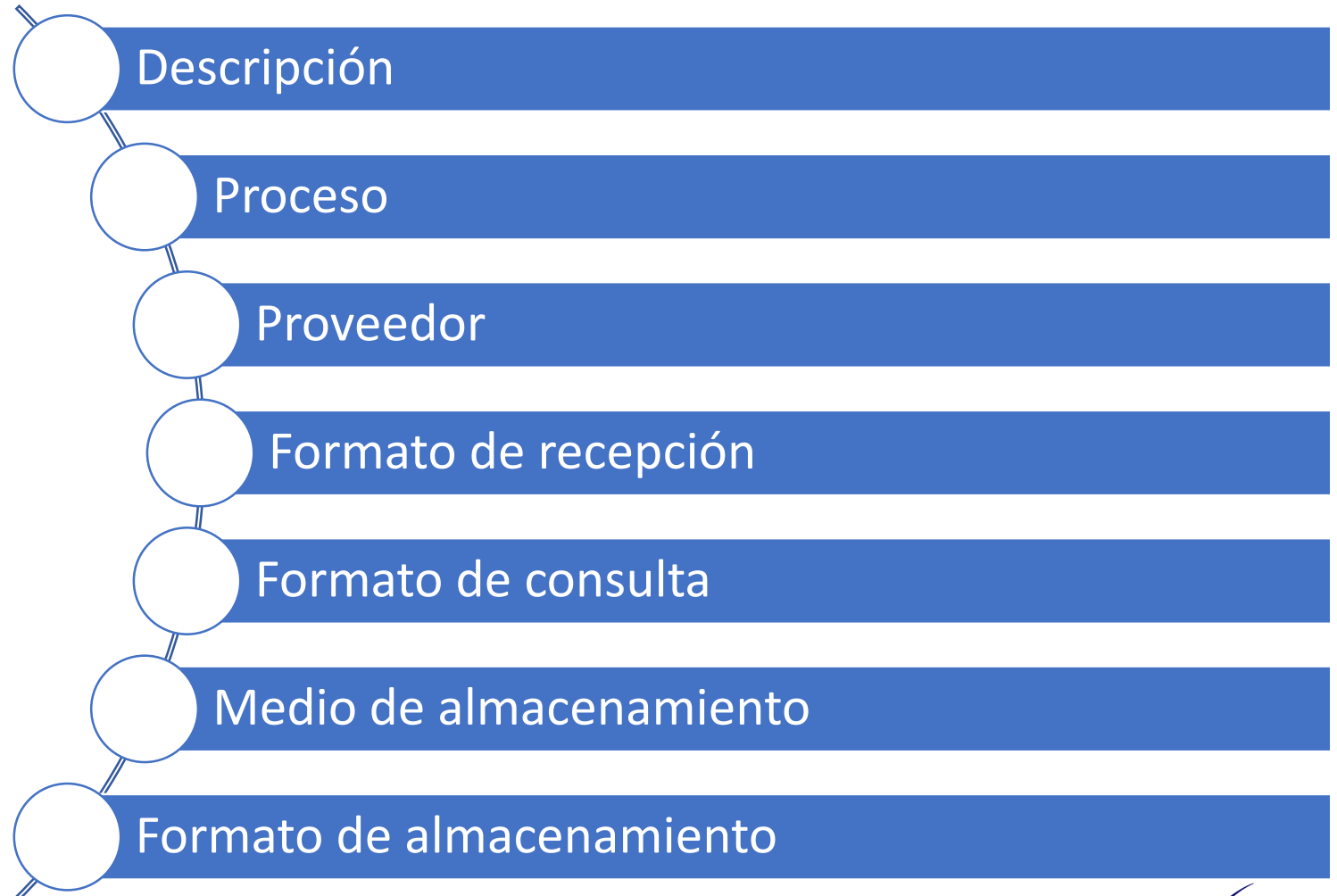




Proceso con las dependencias - Becas



10 Conjuntos de datos
identificados



Inventario fuentes de información - Becas

Descripción del Inventario de						
Nombre de la información	Descripción	Área Responsable de la Información	Proceso en el que esta involucrado el dato	Sedes	Proveedor de la información	Formato de recepción (Medio de recepción o consulta)
Beneficiarios SPP - GE por Programa	Brinda la información de los estudiantes que ingresan a los Programas de SPP y GE por pregrado y quienes se matriculan con estas formas de pago semestralmente	Desarrollo Estudiantil	Estadísticas Generales Acreditación de Programas	Medellín	Sábana de control de la población SPP -GE Construcción propia	No se recibe, es propiedad
Informe para convocatoria de becas	Retorna datos académicos, sociodemográficos, de contacto y formas de pago de estudiantes de la universidad para convocatoria de becas.	Desarrollo Estudiantil	Convocatoria de Becas	Medellín	DINF	Correo
Descuento condonable epm proyección -ppto	Brinda datos de los estudiantes con Fondo EPM como valor matrícula, semestre que le cubre, que le faltan, liquidación, entre otros.	Desarrollo Estudiantil	Presupuesto	Medellín	Apoyo Finaciero	Correo
Comportamiento intersemestrales	Contiene el comportamiento de los estudiantes que matricularon intersemestrales, es decir si retiraron la materia, la cancelaron, la ganaron, la perdieron, los créditos de la materia, entre otros	Desarrollo Estudiantil	Intersemestrales Conciliación con entidades externas	Medellín		

Agenda de reuniones diagnóstico (próxima semana)

- Desarrollo Institucional y Egresados (2da reunión)
Junio 1 – 9:00 am
- Oficina de Relaciones Internacionales (2da reunión)
Junio 2 – 2:00 pm
- Desarrollo Estudiantil (4ra reunión)
Junio 3 – 9:00 am
- Idiomas (2da reunión)
Junio 3 – 11:00 am
- Departamento de Prácticas Profesionales (2da reunión)
Junio 5 – 11:00 pm



3.2

Políticas

Normas y estándares

Identificación de políticas existentes a nivel institucional

Política General de Seguridad Informática

Lineamientos seguridad de la información documental

Política de Protección de Datos personales

Política de Gestión de Datos de la Universidad

Matriz de roles y permisos

¿Existe un marco guía de políticas de Gobierno de Datos?

¿Por dónde comenzar?



Referentes



Elección de un modelo guía de políticas de Gobierno de Datos



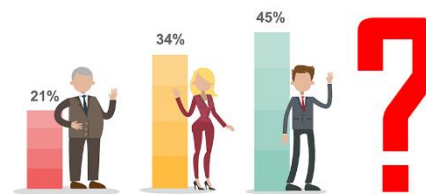
Gobierno de Datos



- Consolidación del programa
- Calidad de datos
- Gestión de acceso
- Analíticas
- Repositorio Institucional



Clasificación de datos



Datos de encuestas



Publicación externa de información



Tecnologías de información



Protección y confidencialidad

¿En qué estamos?

Política de consolidación del programa de Gobierno de Datos

- Misión
- Visión
- Valores
- Objetivos
- Declaración de valor
- Organización - Roles y responsabilidades



Política de seguridad de la información

Comenzando por:

BIA “Business Impact Analysis”

- Inventario inicial de fuentes de información
- Mapeo de procesos
- Líderes
- Información crítica

Como insumos de:

Clasificación de datos

Política de gestión de acceso

Documento de Política de Constitución del Programa

Política de constitución del programa de Gobierno de Datos y Analítica Universidad EAFIT

Misión

La misión del programa de Gobierno de Datos y Analítica es garantizar a la Universidad EAFIT la administración formal y eficiente de sus datos, mediante la definición de normas y responsabilidades sobre el procesamiento, utilización y custodia de la información generada en los distintos procesos institucionales.

Visión

Desde 2024, el programa de gobierno de datos y analítica será pilar fundamental en el logro de la visión institucional 2030: “Seremos la universidad para todas las generaciones y un ecosistema inteligente en permanente renovación, que conecta propósitos con conocimiento”

Valores

1. Integridad
2. Accesibilidad
3. Privacidad y seguridad
4. Cumplimiento y responsabilidad
5. Integración y eficiencia

Objetivos

1. Crear inventarios y diccionarios de datos que permitan clasificar adecuadamente la información para sus diversos usos
2. Crear un glosario de terminología institucional con el fin de alinear definiciones entre las diferentes dependencias de la Universidad
3. Establecer apropiadamente los roles y responsabilidades respecto al acceso y uso de la información
4. Garantizar la seguridad de los datos incluyendo políticas de confidencialidad y protección de la información contra riesgos y pérdidas
5. Implementar las acciones necesarias para mejorar la calidad de los datos que permitan una mayor precisión y oportunidad para la toma de decisiones informadas
6. Promover una integración eficiente de los sistemas de información institucionales con el fin de coordinar y alinear esfuerzos de las distintas dependencias
7. Alinear los esfuerzos existentes en torno a las analíticas

3.3 Calidad de los datos

An illustration on a light green background with various geometric shapes (triangles, circles, lines) scattered around. In the center, a hand holds a magnifying glass over a document that features a bar chart and a checkmark. To the left, another hand holds a calculator. The overall theme is data analysis and quality control.

Calidad de Datos

Estudiantes

Compleitud

- País de nacimiento
- Departamento nacimiento
- Ciudad nacimiento
- Fecha de nacimiento
- Género
- Estrato



Veracidad

- Coherencia de Género
- Coherencia País de Nacimiento vs Tipo Documento

Total Estudiantes: 102.383

RESULTADOS COMPLETITUD DATOS ESTUDIANTES



12%

12628

Sin país de nacimiento

12%

12530

Sin departamento de nacimiento

12%

12619

Sin ciudad de nacimiento

52%

53050

Sin estrato socioeconómico

RESULTADOS VERACIDAD DE DATOS ESTUDIANTES

54

Participantes con documento de extranjería registran país de nacimiento Colombia



7.3%

7.527

La Fecha de Nacimiento no es valida ejemplo: 2/01/1900

1.228 **1.2%**

El Género no corresponde

Calidad de datos: SNIES-ICFES



19.428

Sin Código AC

Errores MEN

RPC_019	Estudiante de primer curso de programa tecnológico o universitario sin código de prueba SABER 11	10.773
RPC_021	Estudiante de primer curso registra una prueba SABER 11 con longitud inválida	23.892

Calidad de datos: SNIES-ICFES

Optical Character Recognition

Docuware

REPORTE DE RESULTADOS ESTUDIANTE - SABER 11.

PUNTAJE GLOBAL
De 500 puntos posibles, su puntaje global es **360**

¿EN QUÉ PERCENTJE ME ENCUENTRO?
A. Con respecto a los estudiantes del país, usted está

Prueba	De 100 puntos posibles, su puntaje es	¿EN QUÉ PERCENTJE ME ENCUENTRO?
Lectura crítica	73	A. Con respecto a los estudiantes del país, usted está aquí
Matemáticas	72	
Sociales y ciudadanas	73	
Ciencias naturales	68	
Inglés	78	

SU NIVEL DE DESEMPEÑO ES

Prueba	Nivel de desempeño	Descripción
Lectura crítica	4	Para clasificar en este nivel, el estudiante debe: - Resolver problemas de comprensión de interpretación que requieren el uso de inferencias implícitas y explícitas. - Reconocer la estructura de los textos y utilizarla para comprender el contenido. - Reconocer la estructura de los textos y utilizarla para comprender el contenido. - Reconocer la estructura de los textos y utilizarla para comprender el contenido.
Matemáticas	4	Para clasificar en este nivel, el estudiante debe: - Resolver problemas que requieren el uso de inferencias implícitas y explícitas. - Reconocer la estructura de los textos y utilizarla para comprender el contenido. - Reconocer la estructura de los textos y utilizarla para comprender el contenido.

```
pdf_convert("saber 11 x.pdf",format = "png",  
            filenames = "icfesx.png", dpi=200)  
text <- tesseract::ocr("icfesx2.png", engine = eng)  
cat(text)  
pos = regexpr('AC', text)  
substr(text, pos, pos+13)
```

```
> pos = regexpr('AC', text)  
> substr(text, pos, pos+13)  
[1] "AC201620944405"
```

```
****Volver todo el nombre mayuscula y generar codigo de sonido  
foreach y in nom1 nom2 apel ape2{  
  replace `y` = regexpr(`y`, " ", " ")  
  replace `y` = regexpr(`y`, " ", " ")  
  replace `y` = regexpr(`y`, " ", " ")  
  replace `y` = regexpr(`y`, " ", " ")  
}  
  
foreach k in nom1 nom2 apel ape2{  
  replace `k` = ltrim(upper(`k`))  
  gen s_`k` = soundex(`k`)  
  replace s_`k` = substr(s_`k`, 2, 3)  
}
```

Actividades a realizar en Junio

- Finalizar socializaciones del proyecto
- **Uso de metadatos:**
 - Identificación de fuentes reportadas en inventario de información becas
 - Comenzar el diccionario de datos
 - Estructuración de mapeo y gobierno de analíticas
- **Calidad de datos:**
 - Comenzar la documentación de las reglas iniciales de calidad de datos
 - Continuar con el envío de solicitudes de validaciones de información (EPIK-GD)
 - Documentar los resultados de las solicitudes de validación de información
- **Políticas de Gobierno:**
 - Aprobar la política de constitución del programa
 - Establecer mapa de ruta de construcción de políticas siguiendo el marco elegido

Sincronización de actividades con el Gobierno de Datos

- **Proyecto Experience**
- **Comité de resiliencia**

Identificación temprana de dificultades en la virtualidad de las clases

- **Comité de reacción**

Análisis del entorno en educación superior a nivel internacional

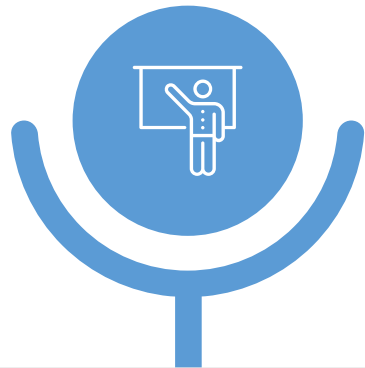
- **Instrumentos de financiación**
- Análisis de información de la herramienta de Microsoft **Edu-Analytics**
- **Migración de EAFIT en Cifras** de visualizador institucional

3.4

Arquitectura de BI



Infraestructura



Proveer la infraestructura tecnológica que soporte las necesidades de BI y Analítica de los procesos y público dentro del alcance (Ciclo de vida del estudiante de Pregrado y Posgrado)

Reportes institucionales



Proveer a la Universidad los reportes complementarios al proyecto EPIK que son necesarios para la operación

Capacidad DataWareHousing



Poner a disposición los recursos tecnológicos requeridos por la institución que permitan construir modelos de BI y analíticos para la toma de decisiones.

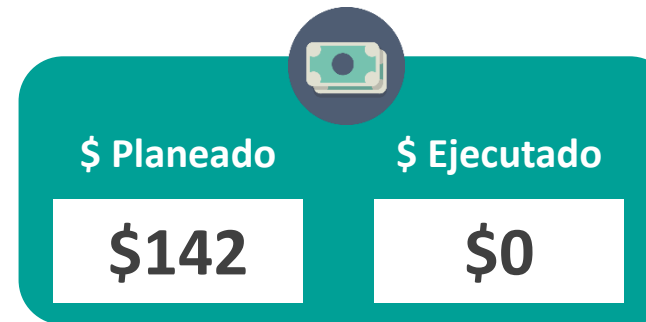
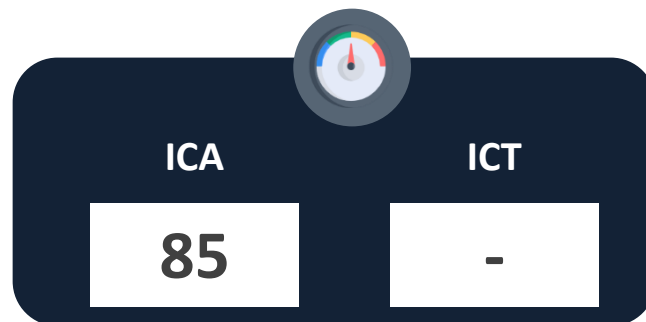
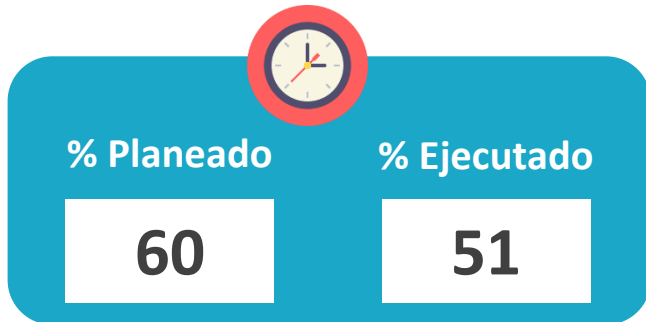
Visualizador



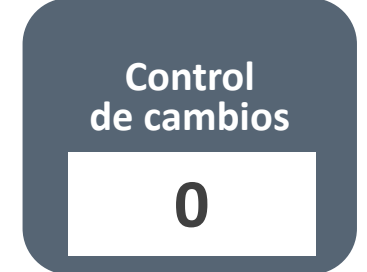
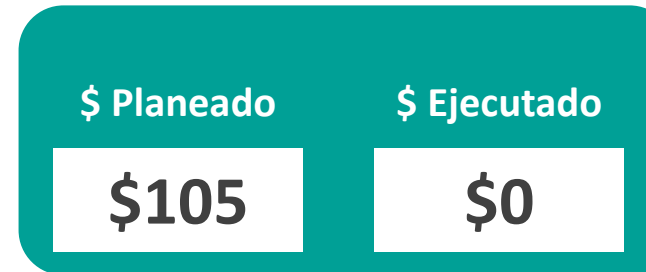
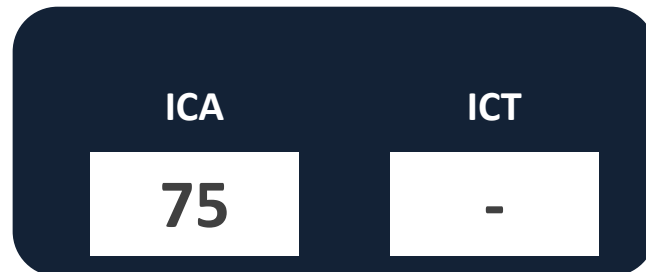
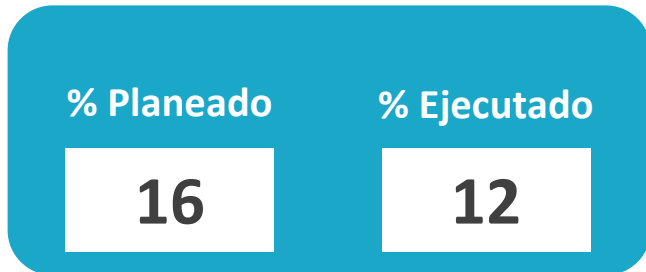
Implementar un visualizador de reportes institucional que soporte las necesidades de los procesos y público dentro del alcance

Indicadores

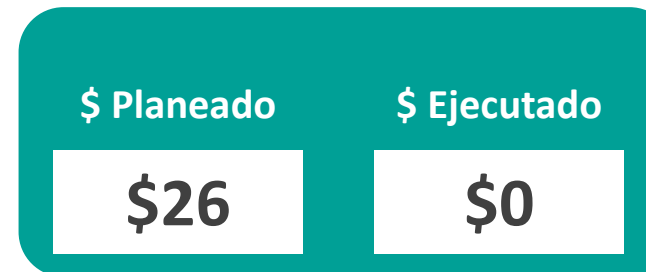
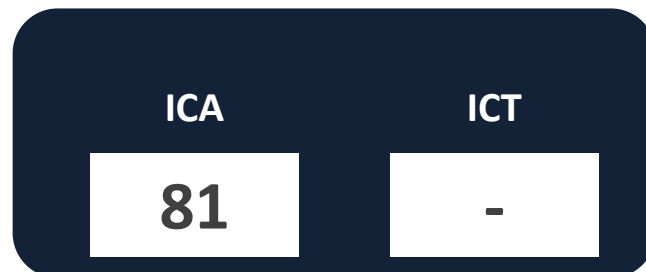
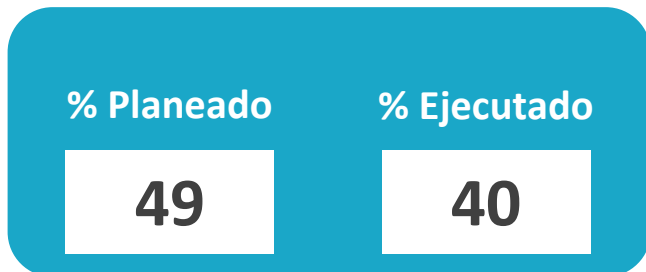
INFRAESTRUC



REPOR INST



VISUALIZADOR



Fecha corte: 26 de mayo de 2020

Infraestructura tecnológica



Conciliación de cronograma con propuesta
Implementación de infraestructura



Referenciación de proveedores, herramientas, ecosistemas y arquitecturas factibles.
Selección del proveedor – En CC



Reportes institucionales



Priorización de reportes a construir en fase 1:
Complementarios EPIK, reemplazo reporteador y reemplazo Discoverer



Inventario de reportes en uso por las dependencias y susceptibles de abordar en fase 1 - Discoverer y reporteador



Capacidad DataWareHousing



Definición de roles y perfiles necesarios para implementación técnica
Planeación con proponentes



Acotamiento del alcance, foco de acción y depuración de redundancias



Visualizador institucional



Migración de reportes
Plan de apropiación con las áreas



Inventario para la transición
Selección herramienta
Referenciación proveedores



Proceso de contratación: Alcance del servicio



Infraestructura

Diseño y despliegue de la arquitectura Cloud para los servicios de BI y Analítica



MPV

Implementación del Mínimo producto Viable en infraestructura desplegada.



Reportes Institucionales

Consultoría, soporte y transferencia de conocimiento: Proyecto de construcción de reportes institucionales para la operación.

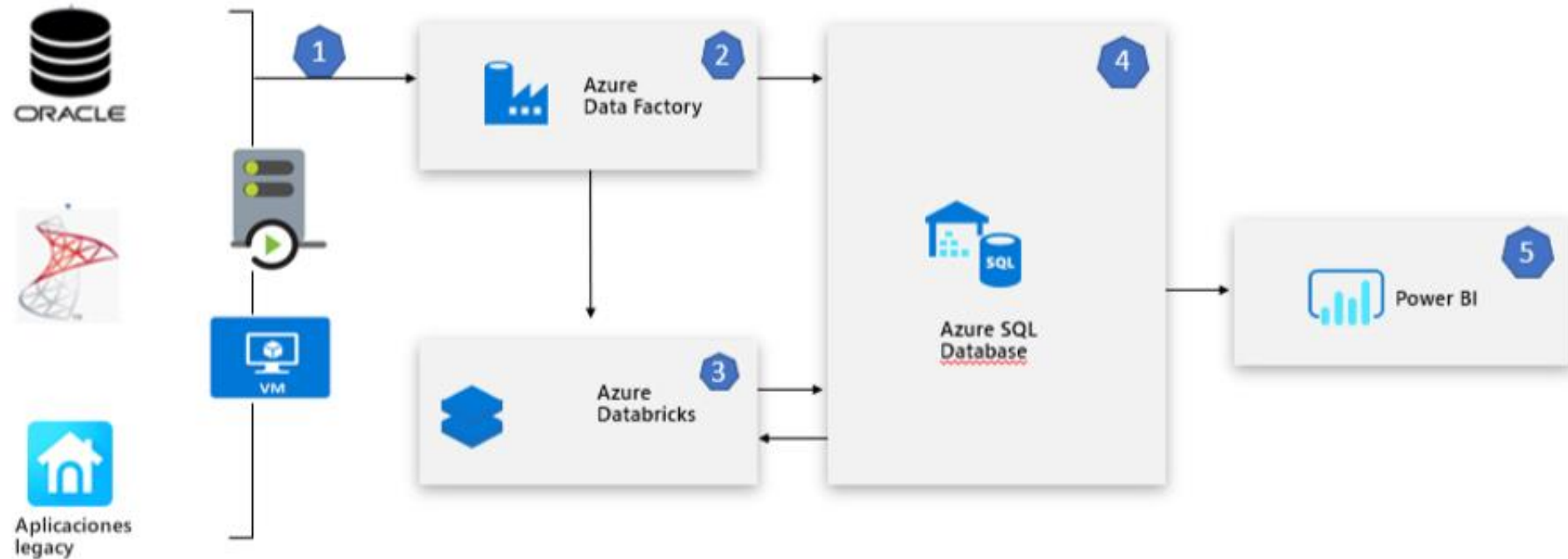


Visualizador Institucional

Consultoría y acompañamiento Transición Tableau - PowerBI



Infraestructura: Arquitectura proyectada



Capa de orígenes

Capa de ingesta

Capa de procesamiento

Capa de almacenamiento

Capa de visualización

Reconfiguración: Mod Analíticos hacia DataWareHousing

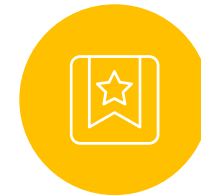
Id necesidades

Definición de perfiles necesarios para el equipo a quien se le hará transferencia de conocimiento en MVP



MVP con proveedor

Acompañamiento en el proceso de implementación



Equipo

Conformación del equipo:

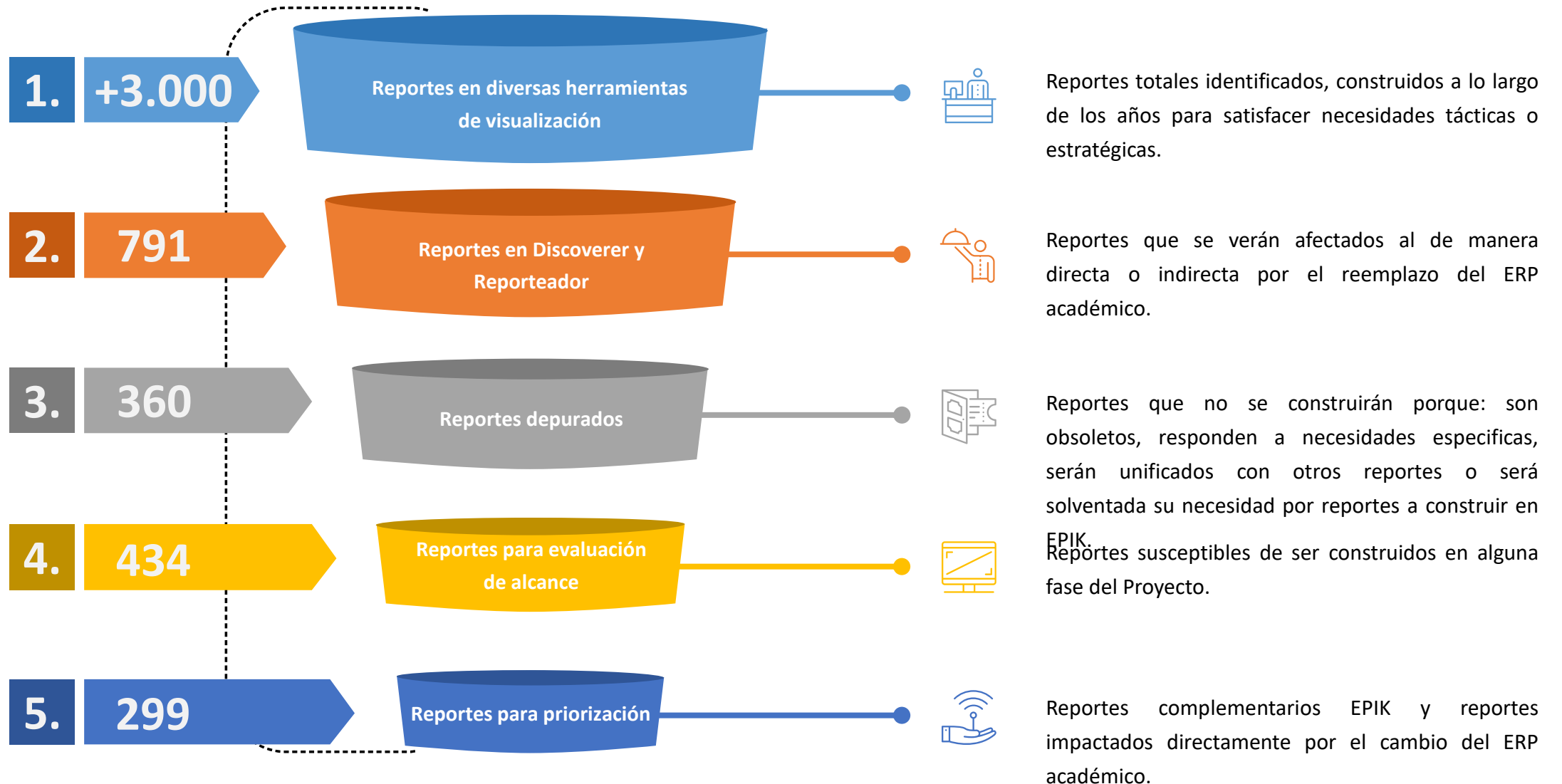
- Identificación de GAPS en cuanto a perfiles en la DINF
- Selección de personas del equipo
- Definición de roles



DWH EAFIT

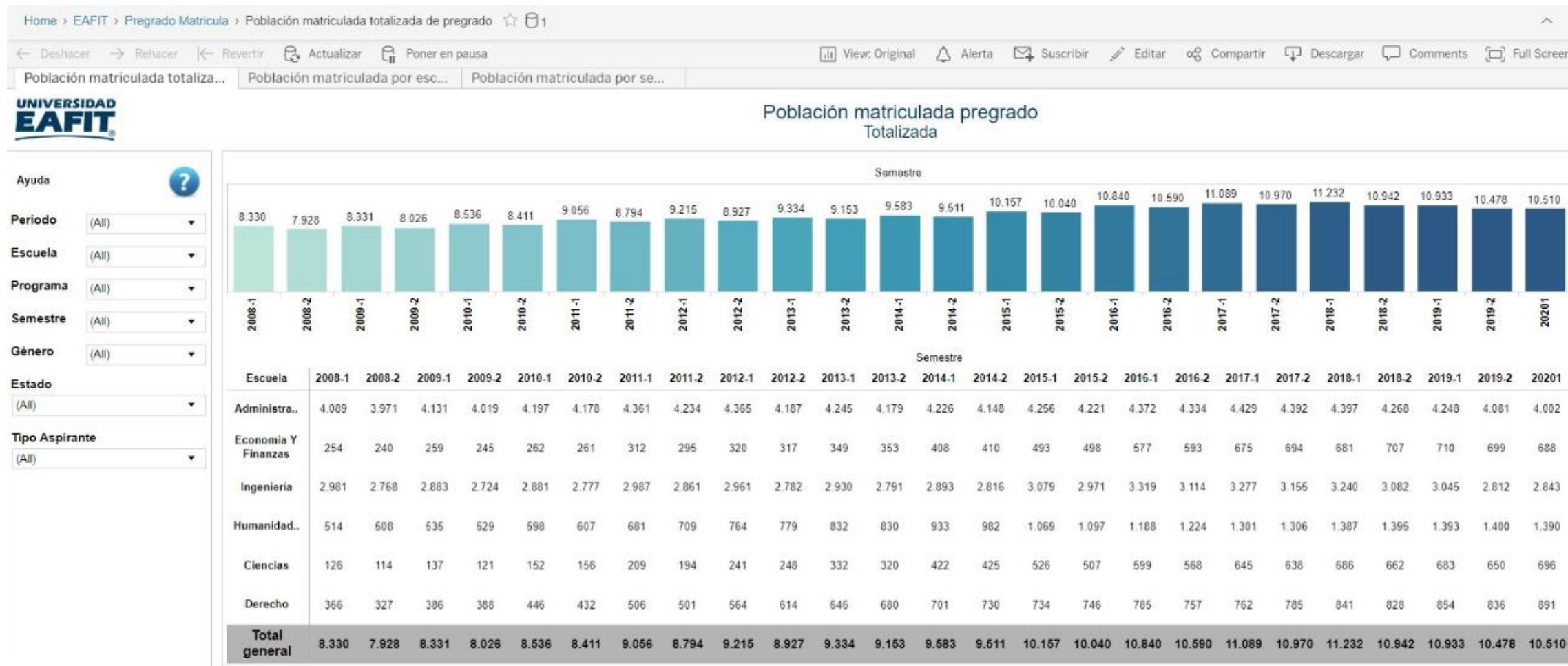
Definición de DataWareHousing EAFIT

Reportes institucionales



Visualizador institucional: La nueva cara

Tableau



Visualizador institucional: La nueva cara

PowerBI



Visualizador institucional: La nueva cara

1. Consolidar un sistema analítico.
2. Reconocer tipologías de programas e IES.
- 3. Evaluar capacidades y logros de aprendizaje.**
4. Consolidar una herramienta de mejoramiento continuo.
5. Generar un sistema que reduzca la subjetividad.
6. Contribuir al fortalecimiento de la institucionalidad.

¡Gracias!

UNIVERSIDAD
EAFIT[®]

60
años