

**UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL**

MODELO DE SERVICIO DE BI CORPORATIVO PARA COLBÚN S.A.

PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN
INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE LA INFORMACIÓN

PABLO WLADIMIR HIDALGO RIVAS

PROFESOR GUÍA:
SEBASTIÁN RÍOS PEREZ

MIEMBROS DE LA COMISION:
LUCIANO VILLARROEL PARRA
FELIPE AGUILERA VALENZUELA

SANTIAGO DE CHILE

2019

RESUMEN EJECUTIVO

El presente documento es realizado en la empresa Colbún S.A. (generadora y distribuidora de energía eléctrica), bajo la Gerencia de TI.

El caso de estudio nace a partir de una serie de eventos del mercado eléctrico tanto interno como externo las que en definitiva terminan por trasladar a las compañías eléctricas desde un escenario ciertamente impasible a uno altamente competitivo y dinámico.

Desde esta mirada y bajo el contexto del valor de la información, nace este trabajo a partir de diversas derivadas tales como: aumento de necesidades de análisis y estudios para detectar oportunidades o riesgos del negocio; aumento de necesidades acceso a datos de distinta naturaleza y fuente de origen; aumento de volumetría de datos; disminución de tiempo para la toma de decisiones; aumento de intencionalidad por anticiparse a los hechos, a la automatización y optimización de procesos; etc.

El problema radica en que la compañía no cuenta con la capacidad para hacer frente a este nuevo escenario de demanda de tratamiento de la información. Las definiciones de accesos, tiempos de respuesta, equipo de trabajo y plataforma tecnológica disponible, entre otros, no están preparados para apoyar este desafío organizacional.

La oportunidad y propuesta de valor es generar un diseño de habilitación de capacidad que permita atender la nueva demanda de explotación de la información. Los elementos que componen esta preparación se concentran bajo un framework conceptual de servicio de Business Intelligence que define para la organización un conjunto prácticas, metodologías, arquitecturas y herramientas que se hacen cargo del nuevo escenario de la organización. En lo principal como aspectos fuerza de este trabajo, se relevan estándares de Arquitectura Tecnológica moderna de BI y Gestión de los Datos, junto con ello un roadmap de madurez y una mirada de procesos de negocios que compromete la conexión con las estructuras orgánicas de la compañía.

Como resultado, se obtiene una visión de camino de cómo la empresa podría transformarse con liderazgo de datos para aportar al desafío de una organización costo-eficiente. Junto con lo anterior, la propuesta ejecuta una prueba de concepto (PoC) con un caso de uso real de la compañía para validar aspectos de la arquitectura propuesta.

Palabras Claves (Arquitectura de BI, Data Management, Data Governance, Analítica de Datos)

DEDICATORIA

A mi esposa Claudia e hija Amalia. Este trabajo no habría sido posible sin su amor, comprensión, apoyo y ayuda. Claudia: gracias por tu paciencia y soportar una o dos micro malhumoradas de estrés, creo que el doctorado lo dejaré para otra vida, te amo infinito. Amalia: perdón por haberme dormido mientras jugábamos, pero entre la tesis, tus pilas inagotables y tu mal dormir, no tenía más opción, te compensaré construyendo esta tesis en versión cuento, te amo infinito.

A mi madre Alicia y padre Juan, quienes, con su sabiduría de padres, supieron entender mi ausencia y apoyar mi meta. Los compensaré con una semana de nieta, para empatizar.

A mis hermanas y hermano, quienes, con la suerte de tenerme, les agradezco la confianza depositada, aunque habría deseado otro tipo de depósito.

AGRADECIMIENTOS

A mis compañeros y amigos del MBE, Claudio Rodríguez y Roberto Rojas, con quienes compartí muchos momentos de humor, aunque pocos de estudio.

A mi profesor guía Ezequiel Muñoz, quién con su experiencia y amigable guía colaboró a la convergencia de este proyecto.

A Laura Sáez y Ana María Valenzuela, quienes, con su preocupación y vocación ayudaron a retener la esperanza.

A todos quienes aportaron a que este proyecto finalizara, gracias totales.

Pablo Wladimir Hidalgo Rivas

TABLA DE CONTENIDO

1	CAPITULO 1: INTRODUCCIÓN	9
1.1	Motivación del Proyecto	9
1.2	Problemática del Proyecto	9
1.3	Objetivos del Proyecto	9
1.3.1	Objetivo general.....	9
1.3.2	Objetivos específicos	10
1.4	Justificación y Relevancia del Proyecto	10
1.4.1	Justificación	10
1.4.2	Relevancia.....	10
1.5	Alcance	11
1.5.1	Dentro de alcance.....	11
1.5.2	Fuera del alcance	11
1.6	Solución Propuesta	11
2	CAPITULO 2: MARCO TEÓRICO	12
2.1	Metodología de Ingeniería de Negocios.....	12
2.1.1	Patrones de Arquitectura y de Procesos de Negocio (PPN).....	12
2.2	Analítica de Datos (Data Analytics)	13
2.2.1	Inteligencia de Negocios (Business Intelligence - BI)	13
2.2.2	Análisis de Negocios (Business Analytics – BA).....	14
2.2.3	Big Data.....	14
2.3	Data Management.....	16
2.3.1	Estándar DAMA	16
2.4	Procesos de extracción de datos	19
2.4.1	Estándar CRISP-DM.....	20
2.5	Arquitecturas de referencia	21
2.5.1	Arquitectura Visión ORACLE	21
2.5.2	Arquitectura Visión IBM	22
2.5.3	Arquitectura por Configuración (Hybrid Cloud)	25
2.5.4	Arquitectura por Pipelines (Lambda)	26
2.6	Modelos de madurez y esquemas de referencia.....	27
2.6.1	Progresión de Capacidades Analíticas	27
2.6.2	Gestión de Datos Maestros	29

2.6.3	Tecnologías emergentes	30
2.6.4	Naturaleza de los datos	31
2.7	Organización y Personas	32
2.7.1	Centro de Competencia de Inteligencia de Negocios (BICC)	32
2.7.2	Roles del equipo BI.....	33
3	CAPITULO 3: ACERCA DE LA INSTITUCIÓN	35
3.1	El Negocio de Colbún S.A.	35
3.1.1	Plantas eléctricas.....	36
3.1.2	Clientes.....	36
3.1.3	Estrategia Corporativa	37
3.1.4	Modelo de Negocio.....	38
3.2	Delta de Hax de Colbún S.A.....	39
3.3	Modelo Porter de Colbún S.A.....	40
3.4	Análisis F.O.D.A. de Colbún S.A.	41
4	CAPITULO 4: PROYECTO.....	42
4.1	Levantamiento de la situación actual	42
4.1.1	Arquitectura de Macro Procesos actual	42
4.1.2	Plataforma y Arquitectura de BI actual	46
4.1.3	Recursos relacionados en situación actual.....	48
4.2	Diagnóstico de la Situación Actual	49
4.2.1	Identificación de los problemas en situación actual	49
4.2.2	Análisis de los problemas en situación actual.....	50
4.3	Análisis de Alternativas	51
4.4	Propuesta de Solución	52
4.4.1	Framework de BI Corporativo	53
4.4.2	PoC de validación.....	74
4.5	Evaluación económica	79
4.5.1	Beneficios	79
4.5.2	Desembolsos	80
4.5.3	Horizonte de Evaluación.....	80
4.5.4	Tasa de Descuento.....	80
4.5.5	Flujo de Caja.....	80

4.6	Plan de Implementación y Acción	82
4.6.1	Plan de acción	82
4.6.2	Gestión del Cambio	82
4.6.3	Criterios de aceptación	84
4.7	Riesgos Asociados al Proyecto	85
5	CAPITULO 5: CONCLUSIONES Y TRABAJOS FUTUROS.....	87
5.1	Conclusiones.....	87
5.2	Trabajos futuros	88
6	CAPITULO 6: BIBLIOGRAFÍA	89
7	CAPITULO 7: ANEXOS	91
7.1	Estructura Organizacional de Colbún.....	91
7.2	Etapas de elaboración del proyecto	92
7.3	Benchmark de plataformas de mercado para el BI	93
7.4	Diagrama Clúster para el Big Data corporativo	98
7.5	Landscape tecnológico.....	99

INDICE DE IMÁGENES

Ilustración 1: Conceptualización de Data Analytics. Elaboración propia	13
Ilustración 2: Dimensiones y Framework del modelo DAMA	17
Ilustración 3: Proceso CRISP-DM	20
Ilustración 4: Arquitectura de Referencia Visión ORACLE	21
Ilustración 5: Arquitectura de Referencia Visión IBM	22
Ilustración 6: Arquitectura de Referencia Hybrid Cloud	25
Ilustración 7: Arquitectura de Referencia Lambda.....	26
Ilustración 8: Modelo de Progresión de Capacidades Analíticas de Gartner (2016)	27
Ilustración 9: Niveles de análisis para la ventaja competitiva	28
Ilustración 10: Modelo de Madurez del Master Data Management. Gartner 2015	29
Ilustración 11: Ciclo de Tecnologías Emergentes. Gartner	30
Ilustración 12: Composición de la Variedad, Volumen y Velocidad de los datos.....	31
Ilustración 13: Centro de Competencia de Inteligencia de Negocios (BICC - Gartner) ..	33
Ilustración 14: Colbún en Cifras (Memoria anual 2018)	35
Ilustración 15: Centrales de Colbún	36
Ilustración 16: Clientes de Colbún.....	36
Ilustración 17: Pilares Estratégicos	37
Ilustración 18: Modelo de Negocio de Colbún	38
Ilustración 19: Modelo Delta de Hax.....	39
Ilustración 20: Arquitectura de Macro Procesos: Nivel 0	42
Ilustración 21, Arquitectura de Macro Procesos: Nivel 1	43
Ilustración 22: Arquitectura de Macro Procesos: Nivel 2	44
Ilustración 23: Proceso de Demanda de Desarrollo de Nuevos Productos / Aplicativos	45
Ilustración 24: Arquitectura de Capacidades actual	46
Ilustración 25: Actual Arquitectura de BI.....	47
Ilustración 26: Framework Modelo de Servicio BI para Colbún	53
Ilustración 27: Canvas de Servicio BI para Colbún	54
Ilustración 28: Arquitectura de Macro Procesos: Nivel 2	56
Ilustración 29: Arquitectura de Macro Procesos: Nivel 3	57
Ilustración 30: Proceso de Demanda de Desarrollo de Nuevos Productos BI.....	58
Ilustración 31: Proceso de Demanda Habilitación de Ambientes y Ejecución de Productos BI.....	59
Ilustración 32: DataFirst Method de IBM, como metodología para la Priorización	60
Ilustración 33: Arquitectura BI propuesta – Perspectiva agnóstica.....	64
Ilustración 34: Arquitectura BI propuesta – Perspectiva General	65
Ilustración 35: Estructura Organizacional propuesta	66
Ilustración 36: Diagrama de Gestión de Calidad de Datos - DAMA	69
Ilustración 37: Capturas del Caso de Uso Informe al Directorio (Real)	74
Ilustración 38: Diagrama de Proceso Caso Uso Informe al Directorio.....	75
Ilustración 39: Arquitectura de PoC de validación – Uso de Capacidades	76
Ilustración 40: Arquitectura de PoC de validación - Herramientas de Implementación ..	77

Ilustración 41: Capturas del Caso de Uso Informe al Directorio (PoC).....	78
Ilustración 42: Flujo de caja de la propuesta	81
Ilustración 43: Principal Estructura Organizacional de Colbún	91
Ilustración 44: Magic Quadrant for Data Integration Tools	94
Ilustración 45: Magic Quadrant for Data Management Solutions for Analytics	94
Ilustración 46: Magic Quadrant for Analytics and Business Intelligence Platforms.....	95
Ilustración 47: Magic Quadrant for Data Quality Tools	95
Ilustración 48: Magic Quadrant for Data Science and Machine Learning	96
Ilustración 49: The Forrester Wave™: Big Data Fabric, Q2 2018	96
Ilustración 50: Clúster para Big Data Colbún.....	98
Ilustración 51: Landscape Big Data 2012	99
Ilustración 52: Landscape Big Data 2018	99

INDICE DE TABLAS

Tabla 1: Análisis de las 5 Fuerzas de Porter	40
Tabla 2: Análisis F.O.D.A.	41
Tabla 3: Análisis de alternativas.....	51
Tabla 4: Indicadores de desempeño propuestos.....	55
Tabla 5: Roadmap de madurez, Gobierno y Cultura	71
Tabla 6: Roadmap de madurez, Arquitectura, Tecnología y Capacidades	72
Tabla 7: Roadmap de madurez, Personas y Habilidades	73
Tabla 8: Estrategia de Gestión de Cambio.....	84
Tabla 9: Tabla riesgos, impacto y mitigación.....	86
Tabla 10: Resumen Informes Data Analytics Gartner & Forrester	97

1 CAPITULO 1: INTRODUCCIÓN

Colbún es una empresa chilena del segmento “utility” dedicada principalmente a la generación eléctrica. La empresa se creó en 1982 y en la actualidad posee 26 centrales de generación en Chile y Perú, a través de las cuales cuenta con una capacidad instalada total de 3.893 MW distribuidos en distintos tipos de tecnologías de generación. En la empresa trabajan en total cerca de 1.000 personas.

El presente proyecto se refiere al tema de dotar de capacidad para atender casos de usos del mundo de la Inteligencia de Negocios, la Analítica y el Big Data, esto bajo estructuras de gobierno, arquitectura y procedimientos de dirijan una dirección sostenible y con visión corporativa.

Vale la precisión que este trabajo hace distinción entre conceptos del BI Tradicional (Business Intelligence) y BI Analítico (Business Analytics), sin embargo y para simplicidad lectora se establece el criterio genérico de “BI” como el concepto que representa a ambos paradigmas.

1.1 Motivación del Proyecto

La principal motivación del proyecto es, a través de la explotación de los datos, aportar al proceso de transformación de empresa competitiva. Contando ya con una incipiente demanda por acceder de manera rápida y oportuna a diversas fuentes de datos para obtener indicadores, trazabilidad, explicación a eventos, etc. y, además, con una latente demanda por generación de valor adicional a través de la predicción de eventos y el procesamiento de grandes volúmenes de información.

1.2 Problemática del Proyecto

El trabajo busca establecer a través de un marco de trabajo (framework) los principales elementos que deben asentarse para garantizar una adecuada explotación de los datos. Dicho framework identifica a: la Tecnología, las Arquitecturas, las Personas y los Procedimientos como pilares estratégicos del modelo, que a su vez “evoluciona” en un roadmap de madurez.

1.3 Objetivos del Proyecto

Los objetivos que abordará el siguiente proyecto, son los expresados a continuación:

1.3.1 *Objetivo general*

Apoyar objetivo corporativo de ser una empresa competitiva, generando propuesta estratégica para instalar un modelo de servicio BI corporativo a través de un framework conceptual que introduce diversas prácticas, estándares, metodologías, arquitecturas y herramientas para la explotación eficiente de los datos.

1.3.2 **Objetivos específicos**

A partir del objetivo general, se plantean los siguientes objetivos específicos:

- Diseño y Definiciones de Procesos que indiquen cómo levantar, validar, aprobar y ejecutar las iniciativas de BI
- Diseño de Arquitectura y Selección Tecnológica que sustenten la capacidad de implementar las iniciativas de BI
- Diseño de Estructura Organizacional que dirija, controle y administre la inversión de BI

1.4 **Justificación y Relevancia del Proyecto**

El presente proyecto resulta relevante para la organización, ya que le permite sumar capacidad inexistente y aprovechar los beneficios de explotar los datos, en efecto:

1.4.1 **Justificación**

La analítica de datos se ha instalado hoy en día como un arma fundamental para lograr los objetivos de las empresas, siendo los datos en sí un activo estratégico dentro de las organizaciones.

En la actualidad, la empresa vive una profunda y transversal transformación por eficientar sus procesos y sobrevivir en un mercado que se ha vuelto altamente competitivo y cambiante. Lo anterior, trajo consigo la necesidad de contar con mejores procesos de toma de decisiones, más rápidos y eficientes, junto con estar más atentos a los riesgos y condiciones del mercado, incluso adelantarse a estos. El problema radica en que la compañía, desde el punto de vista tecnológico, no cuenta con la capacidad instalada para hacer frente a este nuevo escenario, sus sistemas de información son intrínsecamente transaccionales y la cultura promedio desconoce el valor que se puede generar a partir de los datos.

Este proyecto entonces va en la perfecta línea de la estrategia corporativa de contar con procesos eficientes, aportando componentes que administran las necesidades de información para desarrollarlas sobre herramientas y arquitecturas que agilizan la entrega de valor con toma de decisiones más rápidas, con ello liberando el actual esfuerzo humano existente entre la recopilación de los datos, su tratamiento y posterior consolidación.

1.4.2 **Relevancia**

El presente trabajo por una parte aporta con una estrategia clara y amplia de cómo la organización puede sacar ventaja a partir de sus datos. Por otra parte, la alta gerencia ha posicionado dentro de sus objetivos corporativos el instaurar “Procesos eficientes con tecnologías robustas y digitales” como instancia clave para ser costo-eficiente. A partir de esto, es que este trabajo tiene un efecto relevante sobre las necesidades de la empresa.

1.5 Alcance

1.5.1 *Dentro de alcance*

Los siguientes aspectos serán considerados como elementos integrantes del proyecto.

- Generar propuesta para instalar la capacidad que permita brindar soluciones eficientes de inteligencia de negocios
- Establecer ruta de madurez que permita conocer la evolución de las capacidades
- Implementación PoC que permita validar parte de la arquitectura propuesta

1.5.2 *Fuera del alcance*

Los siguientes aspectos serán considerados como elementos integrantes del proyecto.

- Levantamiento formal de necesidades del negocio
- Adquisición de herramientas tecnológicas
- Implementación de soluciones tecnológicas productivas
- Implementación de metodologías, estructuras de organizativas, procedimientos y modelos de gobierno

1.6 Solución Propuesta

El proyecto propone un “Modelo de Servicio de BI Corporativo” que primeramente instaure el concepto y la relevancia de la explotación de los datos, luego propone un framework conceptual para concentrar las dimensiones más relevantes de la Inteligencia de Negocios, desde el punto de vista de los procesos, las tecnologías, las arquitecturas y las personas. Finalmente, validar aspectos técnico-prácticos de la propuesta mediante la implementación de una prueba de concepto (PoC) que involucre recolección de datos de distintas fuentes.

2 CAPITULO 2: MARCO TEÓRICO

La actual situación de la compañía, en términos de demanda de BI no cubierta, obliga a plantear una serie de “componentes” estándares que permitan transitar por un camino probado por la academia y la industria, es por ello que a continuación se relevan una serie de Metodologías y Modelos que son tomados para la conformación de la propuesta de proyecto. Si bien es cierto para el tópico de BI existe gran cantidad de Metodologías y Modelos de diferente índole, para este trabajo solo se han seleccionado un conjunto acotado, pero suficientemente representativo.

2.1 Metodología de Ingeniería de Negocios

Parte de la propuesta de este proyecto corresponde a la metodología descrita en el libro “Ingeniería de Negocios: Diseño Integrado de Servicios, sus Procesos y Apoyo TI” (Óscar Barros, 2016), el cual propone modelar los procesos de negocios alineados con la estrategia de las empresas, definiendo para ello Patrones de Negocio (PN) y Patrones de Arquitectura y de Procesos de Negocio (PPN), que en términos subyacentes están presentes en todas las empresas. Esta metodología interfiere de manera holística la empresa considerando todos los aspectos que la componen: su modelo de negocios, estrategia, capacidades, procesos y recursos relacionados. En específico, la propuesta de este proyecto guarda relación con PPN, por lo que a continuación una descripción general:

2.1.1 *Patrones de Arquitectura y de Procesos de Negocio (PPN)*

Cualquier diseño hecho a partir de los Patrones de Negocio (PN) puede ser convertido en un correspondiente diseño de procesos por medio de la instanciación o especialización de uno de los Patrones de Arquitectura y de Procesos de Negocios. Éstos se basan en extensiva experiencia de diseño de procesos realizada en cientos de casos reales y comparten la idea de que existen cuatro agrupaciones de procesos llamados macroprocesos, que existen en cualquier organización, ellos son:

Macroproceso 1 (Macro 1): Conjunto de procesos que ejecutan la producción de bienes y/o servicios de la empresa, el cual va desde que se interactúa con el cliente para generar requerimientos hasta que éstos han sido satisfechos. Este macroproceso se asemeja a lo que otros autores denominan cadena de valor, sugerido por Porter (1996).

Macroproceso 2 (Macro 2): Conjunto de procesos que desarrollan las nuevas Capacidades que la empresa requiere para ser competitiva: los nuevos productos y servicios, incluyendo Modelos de Negocios, que una empresa requiere para mantenerse vigente en el mercado; la infraestructura necesaria para poder producir y operar los productos, incluyendo la infraestructura TI; y los nuevos Procesos de Negocios que aseguren la efectividad operacional y creación de valor para los clientes, estableciendo

como consecuencia los sistemas basados en TI necesarios. En este macroproceso se lleva a cabo la innovación requerida.

Macroproceso 3 (Macro 3): Planificación del negocio, que comprende el conjunto de procesos necesarios para definir el curso futuro de la organización en la forma de estrategias, que se materializan en planes y programas.

Macroproceso 4 (Macro 4): Conjunto de procesos de apoyo que manejan los recursos necesarios para que los anteriores operen. Hay cuatro versiones que se pueden definir a priori: para recursos financieros, humanos, infraestructura y materiales.

Los Macroprocesos entregan una estructura integrada y coherente para el buen funcionamiento de la organización.

2.2 Analítica de Datos (Data Analytics)

El Data Analytics es término utilizado para englobar las técnicas y procedimientos utilizados para el “tratamiento” de datos, con el objetivo final de obtener información relevante para la planificación de estrategias y la toma de decisiones empresariales. Gartner lo define como “*Analytics* ha surgido como un término para englobar las diferentes variedades de iniciativas de Inteligencia de Negocios y aplicaciones relacionadas” (Gartner, 2018)

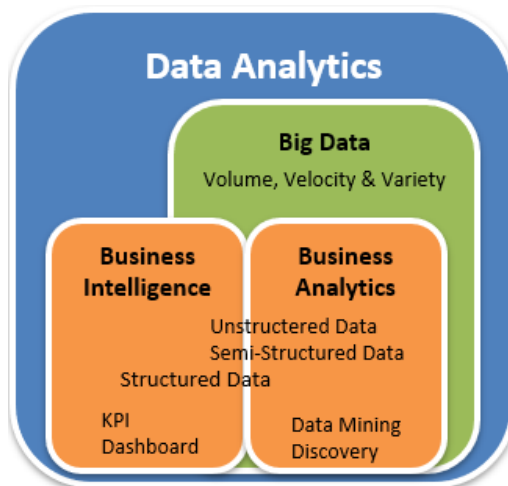


Ilustración 1: Conceptualización de Data Analytics.
Elaboración propia

Las referencias bibliográficas aún no apuntan a un estándar preciso en cuanto a su definición, sin embargo, el término es concluyente que está compuesto por 3 paradigmas: Business Intelligence, Business Analytics y el Big Data, los cuales serán descritos a continuación

2.2.1 Inteligencia de Negocios (Business Intelligence - BI)

El término Inteligencia de Negocios o Business Intelligence como lo conocemos hoy en día se le atribuye a Howard Dressner (Gartner, 1989) y lo define como “*Conceptos y métodos para mejorar la toma de decisiones empresariales mediante el uso de un sistema de soporte basado en hechos*”, y nace como resultado del creciente análisis de

la información recolectada por las empresas a través de las mejoras en la tecnología de almacenamiento. La meta del BI es facilitar la toma de decisiones en una empresa a partir de la información existente, el BI se concentra en la información pasada de la empresa, para obtener resultados a través de datos estructurados.

2.2.2 **Análisis de Negocios (Business Analytics – BA)**

El Business Analytics es un concepto que surge a partir de las necesidades y oportunidades creadas (al igual que el BI) por la gran cantidad de información que las empresas estaban siendo capaces de recopilar, surge como una extensión del BI ya que apunta a la planificación de decisiones futuras, “*refers to the skills, technologies, practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning*” (Beller, & Barnett, 2009), el BA puede abarcar diversas fuentes de datos (datos que no sean necesariamente estructurados) para obtener resultados predictivos que permitan reconocer tendencias que faciliten un análisis desde múltiples perspectivas.

Una de las técnicas más usadas es la de Minería de datos (Data Mining – DM). Bajo este concepto existe el del aprendizaje automático, el cual abarca una serie de algoritmos que permiten aprender de los datos para descubrir patrones de comportamiento en estos. Existe una serie de algoritmos de aprendizaje ampliamente utilizados, agrupados en 2 grandes categorías generales: algoritmos de aprendizaje supervisado y algoritmos de aprendizaje no supervisado.

- **Algoritmos de aprendizaje supervisado.** Este tipo de algoritmos tienen conocimiento a priori de los datos, generalmente a través de una etiqueta (target) que identifica lo que se desea predecir. Estos algoritmos generalmente son utilizados en labores predictivas. Algunos de los algoritmos más utilizados bajo este concepto son: Árboles de decisión, K-nn (K Vecinos más cercanos), Clasificación Naive Bayes, Regresión por mínimos cuadrados, Regresión logística, Support Vector Machine, Métodos ensemble
- **Algoritmos de aprendizaje no supervisado.** Estos algoritmos no tienen conocimiento a priori de los datos, por lo que se debe descubrir relaciones implícitas en los datos. Estos algoritmos generalmente son utilizados en labores descriptivas. Algunos de los algoritmos más utilizados son: Algoritmos de clusterización (Ej. K-means), Análisis de componentes principales, Valor Singular de descomposición, Análisis de componentes independientes

2.2.3 **Big Data**

Dada la naturaleza abstracta del término “Big Data” se plantearán varias definiciones de distintas fuentes para poder definir en su complejidad a que se refiere cuando se habla de “Big Data”.

Según la OED (Oxford English Dictionary) Big data es “Datos de gran tamaño, al punto que su manipulación y manejo presentan grandes desafíos de logística; (también) La rama la informática que involucra dicha data.” (Oxford English Dictionary, 2019)

Michel Cox y David Ellsworth (1997) publicaron *“Application-controlled demand paging for out-of-core visualization”*, donde hacen referencia al término Big Data como *“Las visualizaciones proveen un desafío interesante para los sistemas de computación: los sets de datos son generalmente muy grandes, comprometiendo las capacidades de la memoria principal, disco local, e incluso discos remotos. Llamamos a éste el problema de Big Data. Cuando los sets de datos no caben en la memoria principal (procesador), o cuando no caben en el disco local, la solución más común es obtener más recursos”*. (Cox, & Ellsworth, 1997)

En 2001 Doug Lanely en su estudio *“3D Data Management: Controlling Data Volume, Velocity, and Variety”* describió las tres V's (volumen, velocidad y variedad), que se usan hoy en día en la mayoría de las definiciones elementales de Big data, como desafíos claves del manejo de datos para las empresas (Lanely, 2001)

2.3 Data Management

La “Gestión de Datos”¹ tiene cabida dentro de una definición superior que se llama: “Gestión de la Información Empresarial”², que es una disciplina integradora para estructurar, describir y gobernar los activos de información a través de los límites organizativos y tecnológicos para mejorar la eficiencia, promover la transparencia y permitir la visión empresarial.

Diversos son los estándares que definen prácticas para la gestión del conocimiento de la información, estableciendo para los datos diversos aspectos tales como su significado, ubicación física, características, uso, probidad, etc. Algunos de estos estándares son:

- Data Management Body of Knowledge (DMBOK). DAMA Association
- Data Management Maturity (DMM) Model. CMMI Institute
- The Data Governance Council. IBM

En particular, para el desarrollo de este proyecto se ha seleccionado a DAMA, entre otros aspectos por ser una comunidad internacional fundada hace más de 20 años y cuya documentación (DMBOK) ha sido escrita por centenares de profesionales ligados a la gestión de los datos. Siendo lo anterior garantía de agnosticismo, madurez y constante evolución, permitiendo una referencia válida para este trabajo.

2.3.1 Estándar DAMA

El término “Data Management” se refiere a la necesidad y los beneficios que trae el tener una gestión de datos eficaz, con lo cual surge la comunidad DAMA³ que propone una guía de las mejores prácticas para el manejo de datos en el DMBOK⁴ (Earley, & Henderson, 2017) el cual propone un sistema de gestión de datos que tiene once áreas de conocimiento o disciplinas las cuales se representan en la siguiente ilustración.

¹ Data Management (DM)

² Enterprise Information Management (EIM)

³ Data Management Association

⁴ Data Management Book of Knowledge

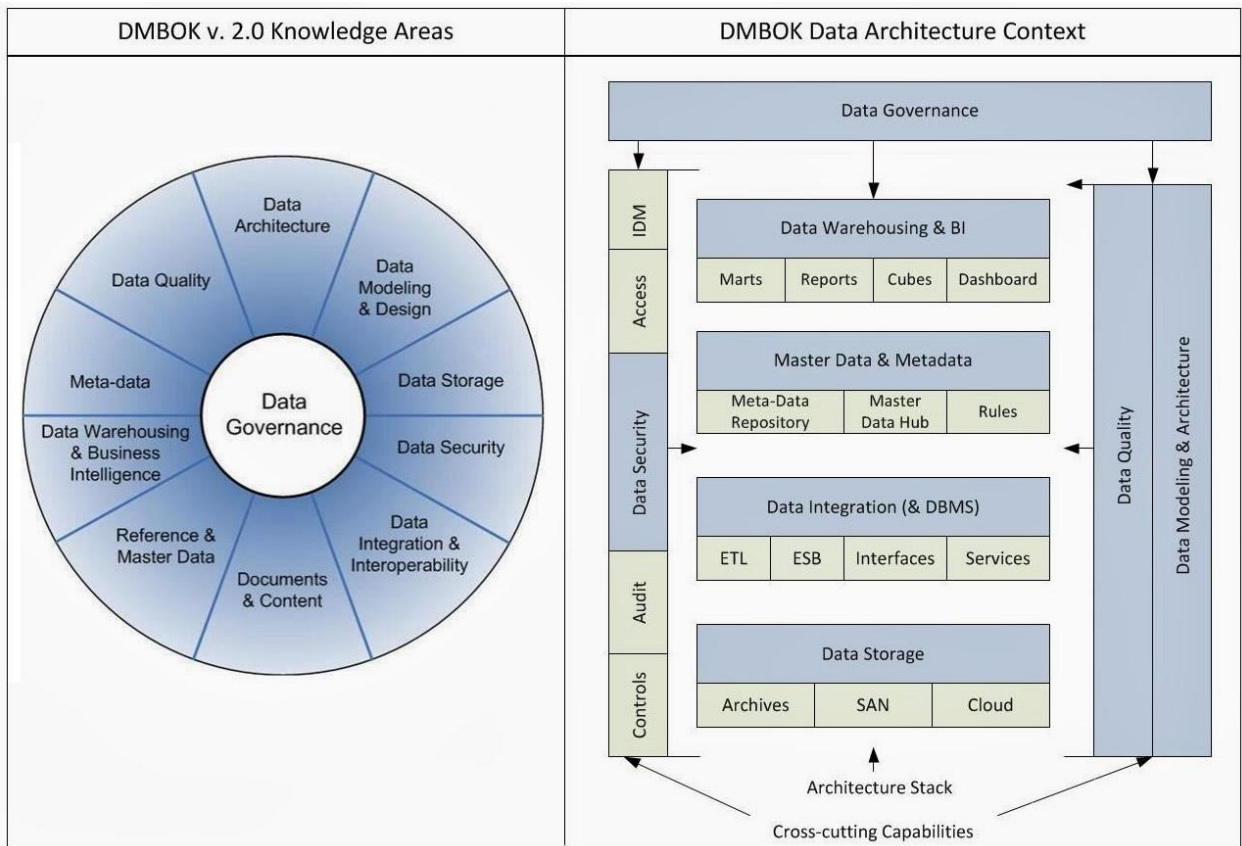


Ilustración 2: Dimensiones y Framework del modelo DAMA

A continuación, un resumen de las once áreas de conocimiento de la Gestión de Datos (DMBoK):

1. **Data Governance:** El Gobierno de los Datos se define como el ejercicio de autoridad y control (planificación, monitoreo y cumplimiento) sobre la gestión de los activos de datos. Interactúa e influye en cada una de las diez dimensiones de gestión de datos.
2. **Data Architecture:** La Arquitectura de Datos se define como un puente entre la estrategia empresarial y la ejecución de la tecnología, actuando como agente de cambio, transformación y agilidad, para el despliegue de los productos y servicios. Las actividades que se cubren en esta dimensión son:
 - Entender las necesidades de información de la organización, para poder desarrollar y mantener el modelo de datos empresarial
 - Definir y mantener la arquitectura de Base de Datos
 - Definir y mantener la arquitectura de Integración de Datos
 - Definir y mantener la arquitectura de Big Data y Analítica
 - Definir y mantener la arquitectura de Metadatos
3. **Data Modeling & Design:** El Modelado y Diseño de Datos es el proceso de descubrir, analizar y determinar los requisitos de datos, y luego representar y comunicar estos requisitos de datos en una forma precisa llamada modelo de

datos. Los modelos de datos representan y permiten a una organización comprender sus activos de datos.

4. **Data Storage:** El Almacenamiento de los Datos incluyen el diseño, la implementación y el soporte de datos almacenados, buscando maximizar su valor a lo largo de su ciclo de vida, desde la creación/adquisición hasta la eliminación.
5. **Data Security:** La Seguridad de Datos incluye la planificación, el desarrollo y la ejecución de políticas y procedimientos de seguridad para proporcionar autenticación, autorización, acceso y auditoría adecuados de los activos de datos e información. Los detalles de la seguridad de los datos (por ejemplo: qué datos deben protegerse) difieren entre industrias y países, sin embargo, el objetivo de las prácticas de seguridad de datos es el mismo: proteger los activos de información en alineación con las regulaciones de privacidad y confidencialidad, acuerdos contractuales y requisitos comerciales.
6. **Data Integration & Interoperability:** La Integración e Interoperabilidad de Datos (DII) describe procesos relacionados con el movimiento y la consolidación de datos. La integración consolida los datos en formas consistentes, ya sean físicas o virtuales. La interoperabilidad de datos es la capacidad de comunicación de múltiples sistemas.
7. **Documents & Contents:** La Gestión de Documentos y Contenidos implica controlar la captura, el almacenamiento, el acceso y el uso de datos e información fuera de las bases de datos. Se centra en mantener la calidad, integridad y el acceso a documentos e información no estructurada o semiestructurada, similar la Gestión de Operaciones de Datos para bases de datos relacionales.
8. **Reference & Master Data:** En esta dimensión se hace diferencia entre los Datos de Referencia y los Datos Maestros, en conjunto buscando aportar una visión 360° de la información. La Gestión de Datos Maestros (MDM) implica el control de los valores e identificadores de datos maestros que permiten el uso constante, a través de los sistemas, de los datos sobre las entidades esenciales. La Gestión de Datos de Referencia (RDM) conlleva el control de los valores de dominio definidos y sus definiciones.
9. **Data Warehousing & Business Intelligence:** Un Data Warehouse (DW), Bodega de Datos o Almacén de Datos es una combinación de dos componentes principales. El primero es una base de datos integrada para el soporte a la toma de decisiones. El segundo está relacionado con los programas computacionales utilizados para recolectar, limpiar, transformar y almacenar datos de diversas fuentes operacionales y externas. Ambas partes se combinan para soportar los requerimientos históricos, analíticos y de Inteligencia de Negocios.

Esta dimensión se refiere a la recopilación, integración y presentación de datos para propósitos de análisis de negocio y toma de decisiones. Se compone de actividades que soportan todas las fases del ciclo de vida de soporte a las

decisiones que provee contexto, mueve y transforma datos desde las fuentes a un repositorio destino común y luego proporciona diversos medios de acceso, manipulación y presentación de reportes sobre un conjunto de datos integrado

10. **Meta-Data**: trata de integrar, controlar y proporcionar metadatos.

11. **Data Quality**: a través de la que se define, controla y mejora la calidad de los datos

2.4 Procesos de extracción de datos

En el marco de los procesos que definen “movimiento” de datos desde un origen a un destino para propósitos de integración y/o aplicar modelos matemáticos, se encuentran: el ETL⁵, KDD⁶ y CRISP-DM⁷. El primero, no reviste aspectos analíticos, sino que se enfoca sólo en la Extracción, Transformación y Carga de datos. El segundo, representa un Proceso de Extracción de Conocimiento, el cual define un flujo direccional para encontrar un “modelo” válido, útil y entendible que describa patrones de acuerdo a la información. El tercero en tanto, persigue lo mismo que el segundo, pero adicionando una serie de interrelaciones adicionales que permiten corregir eventuales errores cometidos en etapas tempranas del proceso. Adicionalmente a lo anterior, en un estudio comparativo de metodologías para minería de datos (Moine, Haedo, & Gordillo, 2011) se indica que CRISP–DM es la guía de referencia más utilizada en el desarrollo de proyectos de Minería de Datos.

Por lo anteriormente descrito es que CRISP–DM será el método relevado para ser incorporado dentro de la propuesta metodología de este proyecto como estándar para proyectos de Minería de Datos, por ello a continuación se describen las etapas del proceso:

⁵ Extract, Transform and Load

⁶ Knowledge Discovery in Databases

⁷ Cross Industry Standard Process for Data Mining

2.4.1 Estándar CRISP-DM

Esta metodología considera 6 etapas, las cuales se describen a continuación:

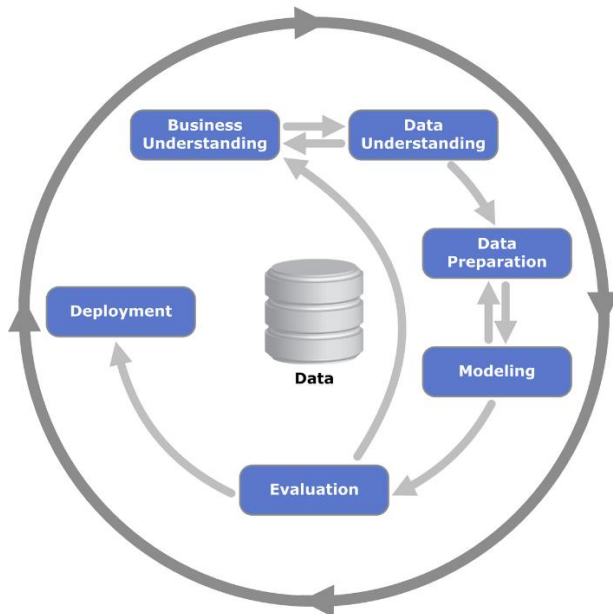


Ilustración 3: Proceso CRISP-DM

- **Comprender el negocio**

En esta fase se debe comprender los objetivos y requisitos del negocio desde una perspectiva de empresa. Se debe tener claridad respecto a lo que persigue conseguir el proyecto en función de lo que la empresa busca.

- **Comprender los datos**

En esta fase se considera un primer acercamiento a los datos, revisar su calidad, consistencia, estructura y detalles que permitan comprender el problema en relación a los datos.

- **Preparación de los datos**

Esta fase se realizan procesos de selección de campos a utilizar, limpieza de datos, agregación de nuevas fuentes, eliminación de campos que no se utilizarán y cambios de formato para adecuar al nuevo modelo.

- **Modelado**

Esta fase es donde se debe seleccionar el modelo que se aplicará sobre los datos y que permitirá resolver la problemática planteada. Por tanto, se debe tener en cuenta el resultado que se persigue, seleccionar el modelo adecuado para la resolución del problema, construir el modelo, evaluar los resultados obtenidos luego de la aplicación del modelo.

- **Evaluación**

En esta fase se evalúa el o los modelos y los resultados obtenidos de estos. En caso de que pueda ser mejorado el proceso que se ha realizado hasta esta fase, es posible volver a una fase previa para corregir alguna omisión o algo que permitiera un mejor resultado del modelo previamente aplicado.

- **Despliegue**

Esta fase es la que permite mostrar los resultados obtenidos. El analista sugiere lo que es necesario realizar para cumplir con los objetivos trazados de acuerdo a lo que los modelos han indicado y las empresas pueden tomar acciones de acuerdo a sus procesos y estrategia de negocios.

2.5 Arquitecturas de referencia

Muchas y variadas son las formas que adoptan las arquitecturas para la explotación de los datos, no existiendo un estándar único, este trabajo ha seleccionado 4 layout de arquitecturas referenciales para analizar y conformar la propia

2.5.1 Arquitectura Visión ORACLE

La siguiente ilustración muestra la visión Oracle. (2016), que a juicio propio da una mirada que representa los usos que debe tener cada capa en el marco de una arquitectura Data Analytics

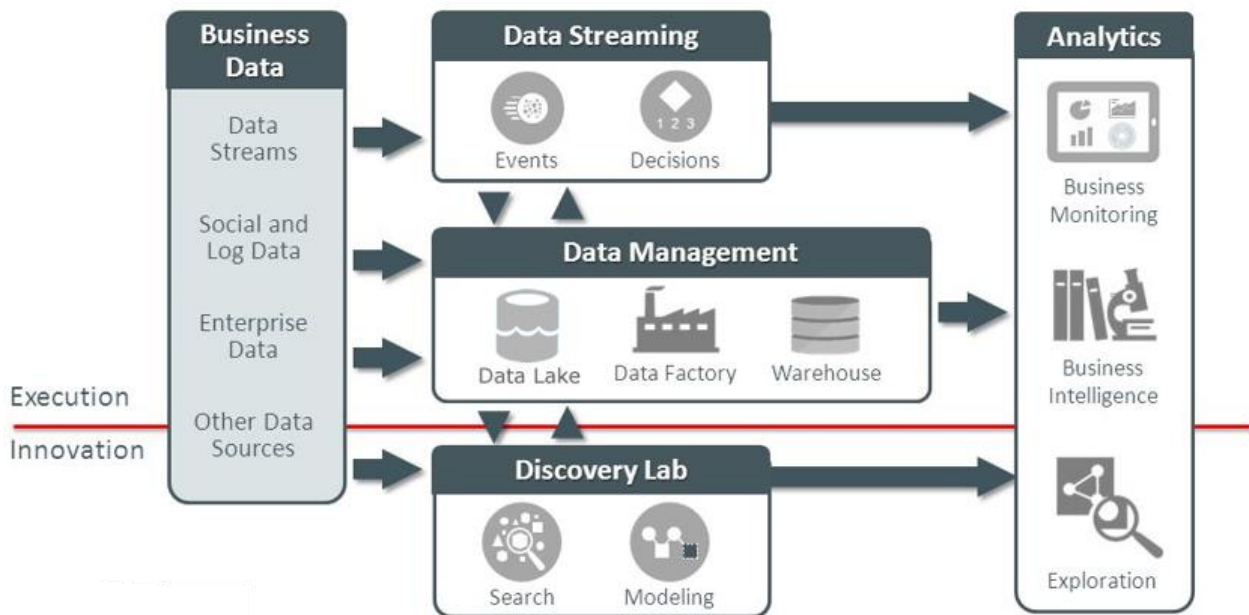


Ilustración 4: Arquitectura de Referencia Visión ORACLE

Este diseño habla de componentes en torno a:

Usos:

1. **Execution:** que somete a los datos a procesos sincronizados para el movimiento y despliegue de datos.
2. **Innovation:** que somete a los datos a procesos matemáticos y científicos para el descubrimiento de valor.

Capas:

1. **Data Streaming:** tratamiento de datos en línea
2. **Data Management:** tratamiento de datos recolectados
3. **Discovery Lab:** tratamiento de datos para la invención
4. **Analytics:** tratamiento de datos para la visualización y análisis

2.5.2 Arquitectura Visión IBM

La siguiente ilustración muestra la visión IBM (Ouyang, & Moura, 2017), que a juicio propio da una mirada que representa las suficiencias o capacidades que debe tener cada capa en el marco de una arquitectura Data Analytics.

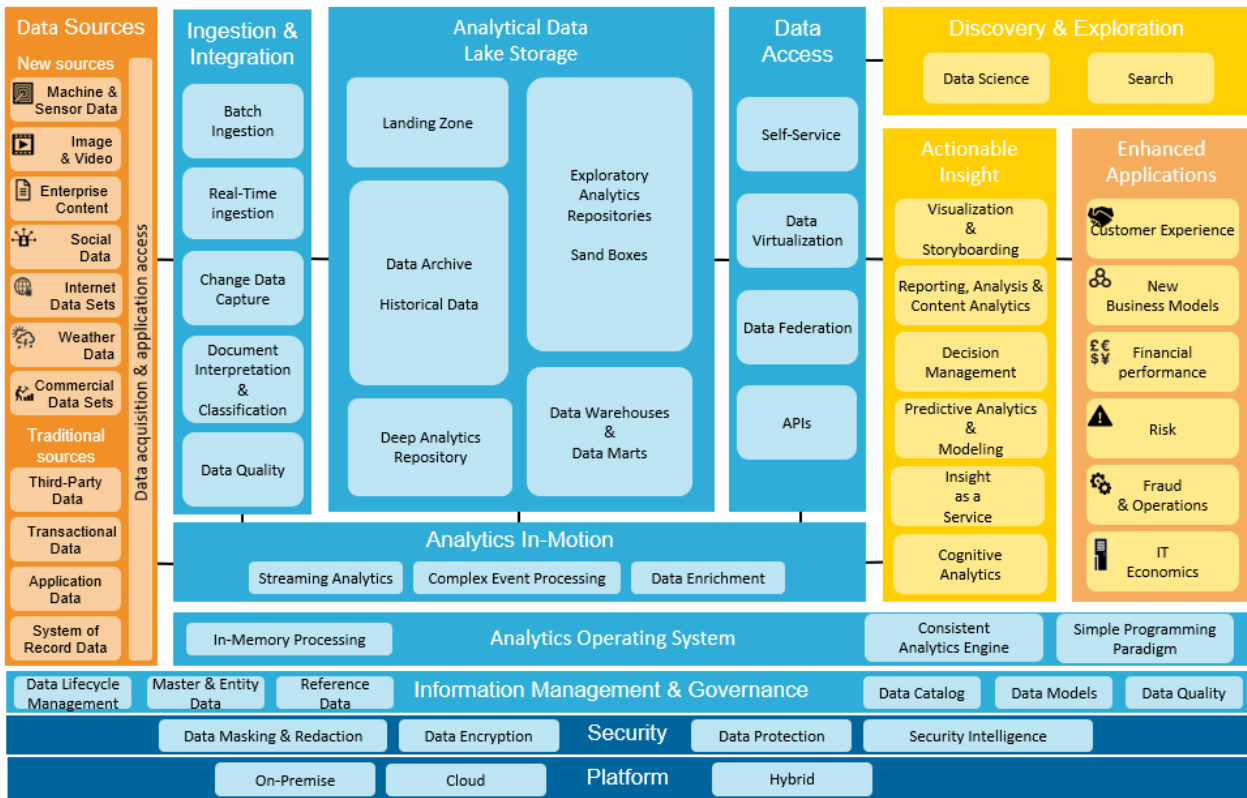


Ilustración 5: Arquitectura de Referencia Visión IBM

A continuación, por cada una de las 11 capas que componen esta arquitectura de referencia, se presenta una descripción de las capacidades más importantes

1. Ingestion & Integration

- **Batch Ingestion:** importación de datos en pequeños pedazos durante intervalos periódicos de tiempo.
- **Real Time Ingestion:** importación de datos es importado a medida que es naturalmente emitido por la fuente.
- **Change Data Capture:** captura de los cambios hechos desde la fuente de datos, para luego replicarlos en otro lugar.
- **Document Interpretation & Classification:** Se le da un sentido a la data y se la clasifica de foma correspondiente.
- **Data Quality:** capacidad de medir qué tan adecuado son los datos dependiendo del uso que se le quiera dar en un determinado contexto

2. Analytical Data Lake Storage

- **Landing Zone:** zona de datos temporales
- **Data Archive / Historical Data:** zona de datos históricos
- **Deep Analytics Repository:** zona de datos donde se guardan los datos limpios y resultados de las analíticas

- Explanatory Analytics Repositories / Sand Boxes: -
 - **Data Warehouses & Data Marts:** Repositorios de información correspondientes a las áreas de la empresa o productos implementados
- 3. Analytics In-Motion**
- **Streaming Analytics:** capacidad de tomar datos de eventos y realizar acciones analíticas en “tiempo real”. Es una forma avanzada de CEP
 - **Complex Event Processing (CEP):** capacidad de tomar datos de eventos y realizar aplicar reglas de negocio o cálculos en “tiempo real” para obtener datos de mayor complejidad
 - **Data Enrichment:** capacidad de realzar o refinar la “información cruda”
- 4. Analytics Operating System**
- **In-Memory Processing (BI):** procesamiento de datos almacenados “en memoria” para aumentar nivel de respuesta a “tiempo real”
 - **Constant Analytics Engine:** se refiere a un “motor de análisis de datos” el cual funciona de forma constante para el procesamiento de datos en “tiempo real”.
- 5. Data Access**
- **Self-Service (BI):** entrega de permisos de diseño para usuarios finales con el fin de que desplieguen sus propios reportes y análisis en el marco de una arquitectura y portafolio determinada
 - **Data Visualization:** representaciones graficas de información para facilitar la comprensión de los datos
 - **Data Federation:** se refiere a la capacidad de integración de datos que permite acceder a diferentes fuentes bajo una visión común
 - **APIs (Application Programming Interface):** procesos, funciones y métodos que brinda una determinada biblioteca de programación a modo de abstracción para que sea empleada por otros sistemas.
- 6. Discovery & Exploration**
- **Data Science:** Es el estudio de dónde viene la información, qué representa y cómo puede ser convertida en un recurso de valor.
 - **Search:** Se refiere al proceso de extraer valor a partir de los datos recolectados
- 7. Actionable Insight**
- **Visualization & Storyboarding:** acción de presentar la información de forma gráfica y relatada con dinámica de negocio
 - **Reporting, Analysis & Content Analytics:** creación de Reportería, Análisis y Analíticas de contenido.
 - **Decision Management:** se refiere al manejo de la toma de decisiones.
 - **Predictive Analytics & Modeling:** se refiere a las analíticas y modelos que permiten prever algún suceso relevante
 - **Insight as a Service:** solución o servicio que ofrece un “atajo” o una propuesta perceptiva sobre algún evento determinado, entregando pasos necesarios para llegar a algún resultado deseado
 - **Cognitive Analytics:** uso de tecnologías inteligentes para reunir, extraer, manejar y controlar tanto datos como fuentes de ellos. Se usa

frecuentemente para acciones que replican a un humano como texto a voz o voz a texto

8. Enhanced Applications

- **Customer Experience:** se refiere a cómo un cliente percibe determinado evento
- **New Business Models:** se refiere a nuevos modelos de negocios
- **Financial Performance:** se refiere al desempeño en el área financiera
- **Risk:** se refiere a los riesgos asociados a alguna acción
- **Fraud & Operations:** se refiere al manejo de fraudes y operaciones
- **IT Economics:** se refiere al área de Tecnologías de Información y su economía

9. Information Management & Governance

- **Data Lifecycle Management:** política del ciclo de vida bajo el cual se maneja el flujo de información de un sistema.
- Master & Entity Data: -
- **Reference Data:** tipo de data que es usada para categorizar otra data
- **Data Catalog:** catálogo de las mediciones compuesta por meta data, definiciones, sinónimos, índices, etc. que permiten una mejor identificación de los datos
- **Data Models:** se refiere a la interrelación lógica y flujo de datos entre diferentes elementos de datos que son requeridos para diferentes procesos de negocios
- **Data Quality:** se refiere a la medición de las propiedades de los datos desde diferentes perspectivas para monitorear la confianza y estado de estos

10. Security

- **Data Masking & Redaction:** el enmascaramiento de datos es el proceso de reemplazar la data autentica con información inauténtica que tenga la misma estructura. En tanto, redacción es el proceso de tachar o remover información sensible a ser identificada.
- **Data Encryption:** es el proceso mediante el cual se “traducen” datos a alguna otra forma o código
- **Data Protection:** se refiere a la protección de datos mediante alguna forma o clave
- **Security Intelligence:** se refiere a la recolección, evaluación y respuesta a datos generados por la red de la organización, para monitorear potenciales amenazas de seguridad en tiempo real

11. Plataforma

- **On-Premise:** hace referencia a una plataforma ubicado físicamente dentro (o a una fibra) de las instalaciones de la organización
- **Cloud:** hace referencia a una plataforma ubicado en algún repositorio virtual
- **Hybrid:** se refiere a una plataforma que utiliza tanto plataformas on-premise como plataformas cloud, dependiendo de la situación, la práctica habitual es que la información clave de la empresa queda on-premise y la información pública en cloud

2.5.3 Arquitectura por Configuración (Hybrid Cloud)

Otro tipo de arquitectura de referencia, de acuerdo con la madurez de las tecnologías actuales, es la “Hybrid Cloud” (IBM, 2017), el Instituto Nacional de Estándares y Tecnológicos (NIST)⁸ lo describe como “dos o más infraestructuras independientes en la nube (privadas, comunitarias o públicas) que se mantienen como entidades únicas, pero están conectadas por estándares o tecnologías exclusivas/privadas que faciliten datos y portabilidad de aplicaciones.”. Se dividen en tres grandes grupos: Public Cloud, Dedicated Cloud, y On-premise.

- **Public Cloud:** Provee escalabilidad, tarificación por uso, flexibilidad e independencia de la localización. Para datos que la organización desea compartir de forma pública - ej. Estadística operacional, indexadores, etc.
- **Dedicated Cloud:** Provee seguridad y la protección de datos. Para datos que la organización desea compartir de forma interna - ej. Modelos estadísticos, catálogo de metadato, indicadores, etc.
- **On-premise:** Provee mayor control sobre los procesos de datos. Para datos que son de gran volumetría o altamente sensibles – ej. propiedad intelectual, datos de ganancias, datos de sensores, etc.

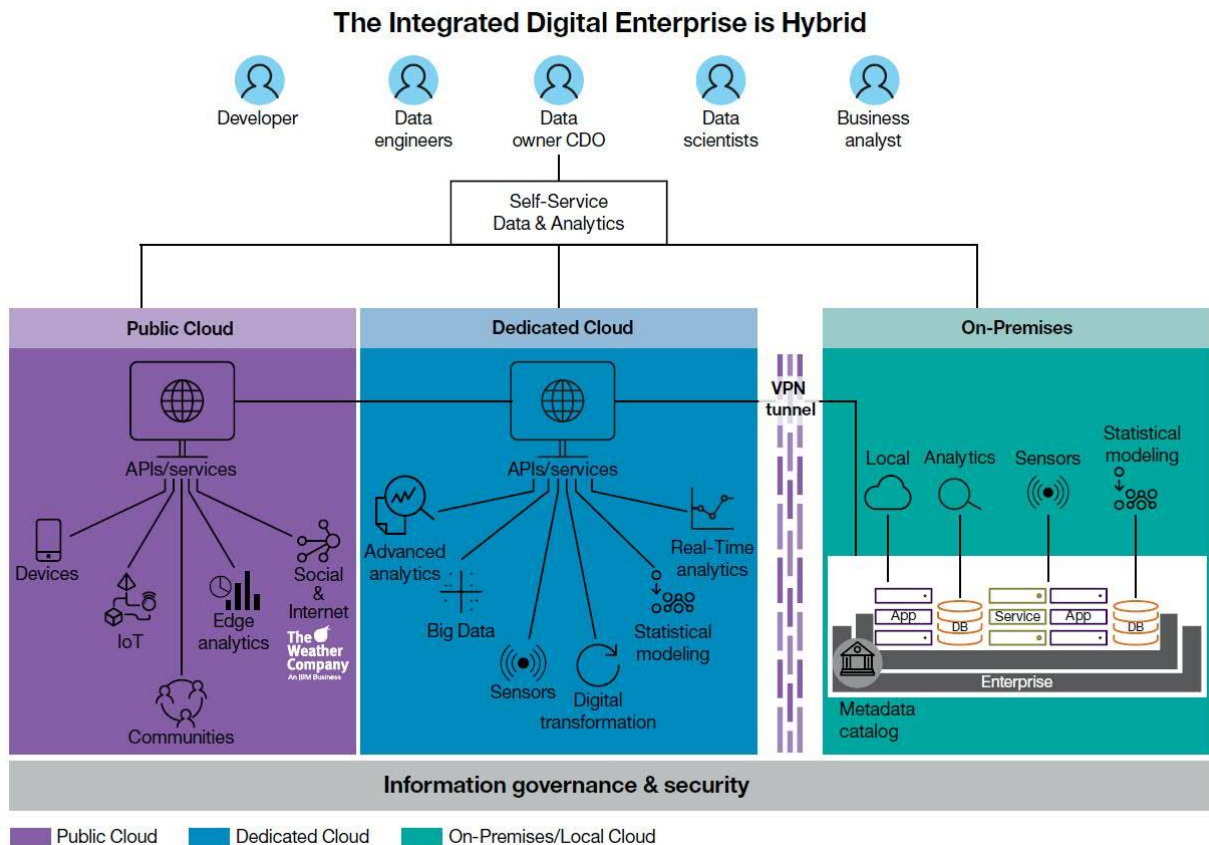


Ilustración 6: Arquitectura de Referencia Hybrid Cloud

⁸ National Institute of Standards and Technology

Diversos artículos tecnológicos describen el valor que los entornos de Nube Híbrida pueden proporcionar a las organizaciones que intentan equilibrar las demandas de tiempo de respuesta, costos y disponibilidad.

2.5.4 **Arquitectura por Pipelines (Lambda)**

Desde el punto de vista de tratamiento de datos es posible mencionar a la “Arquitectura Lambda” (Marz, 2012). En alto nivel de representación, esta arquitectura propone un sistema de procesamiento de datos basado en dos tipos:

1. **Modo batch:** permite procesar volúmenes de datos en tiempos programados (ej. cada “x” minutos, horas o días)
2. **Modo stream o tiempo (semi)-real:** permite procesar datos casi al instante en que estos son producidos (ejemplo: cada “x” milisegundos o segundos)

La siguiente ilustración muestra una representación de alto nivel de esta arquitectura a la cual se le ha adicionado elementos propicios para la analítica de datos (Tejada, Buck, & Wasson, 2018).

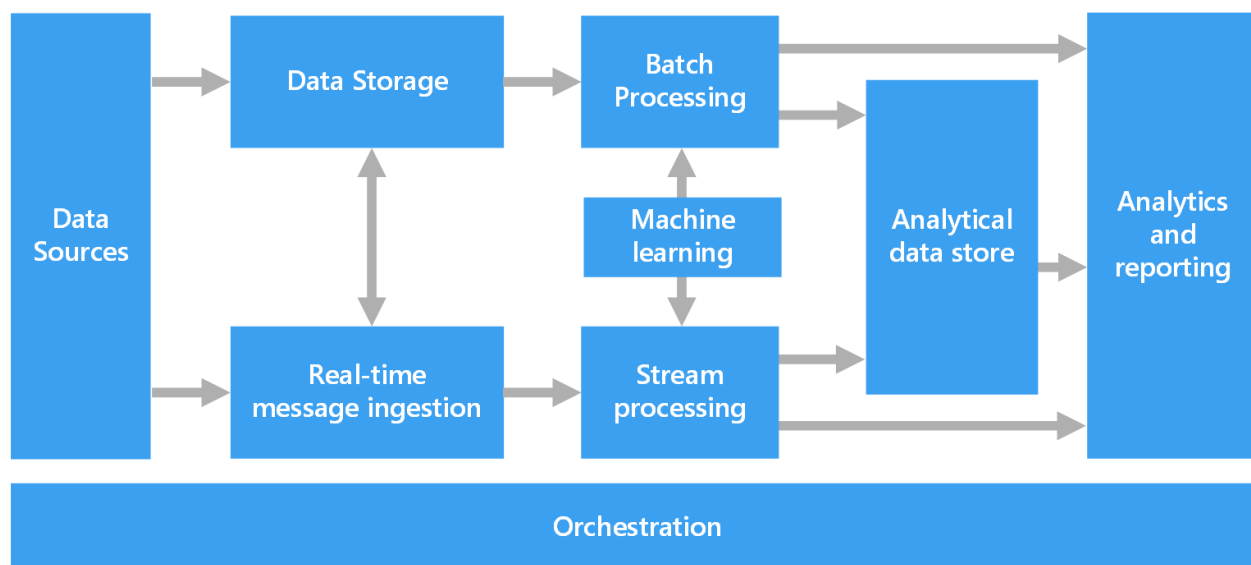


Ilustración 7: Arquitectura de Referencia Lambda

2.6 Modelos de madurez y esquemas de referencia

Para esta sección se han seleccionado una serie de modelos y esquemas que representan tanto la evolución de un proceso a través del tiempo para ir representando los desarrollos en diferentes aspectos, como clasificaciones que aportan a la distinción de tecnologías y datos bajo un cierto contexto.

2.6.1 Progresión de Capacidades Analíticas

La siguiente ilustración evidencia el impacto en la cultura de la toma de decisiones basadas en datos.

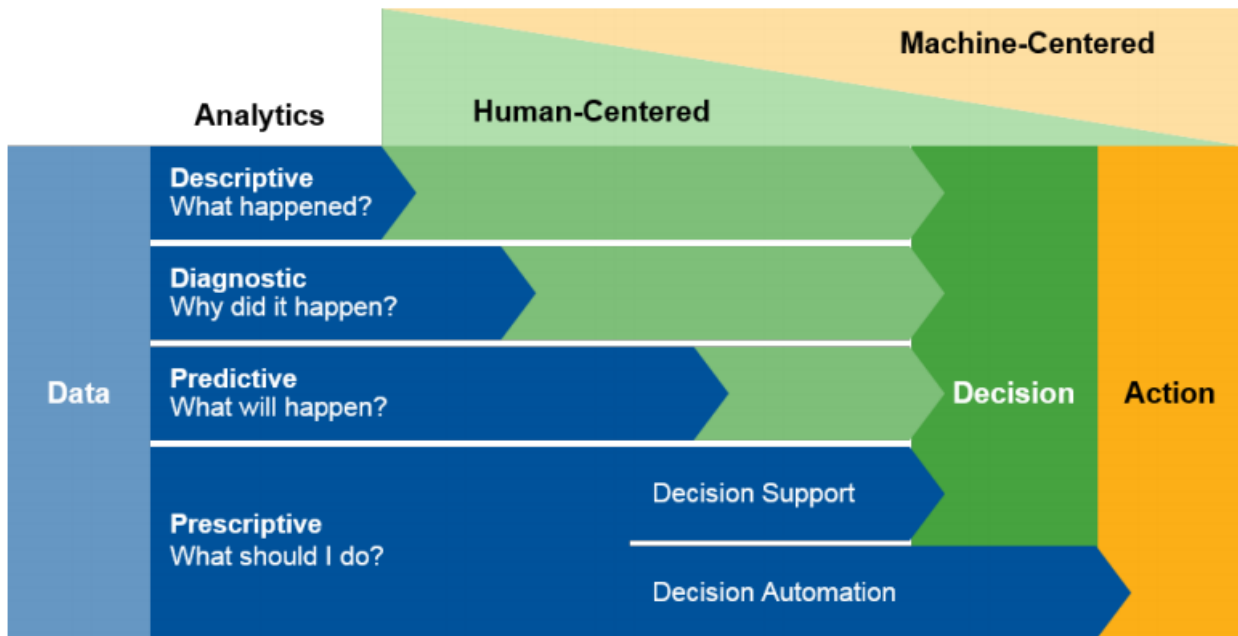


Ilustración 8: Modelo de Progresión de Capacidades Analíticas de Gartner (2016)

A continuación, una descripción general de las dimensiones de análisis.

- **Análisis Descriptivo:** Se aborda el problema de lo sucedido en el pasado para obtener información sobre cómo ser más eficiente en el futuro. Por ejemplo: la estadística de producción de una planta.
- **Análisis Diagnóstico (o Inquisitivo):** Los datos pasados se comparan con la respuesta a la pregunta de ¿Por qué sucedió algo? Estos análisis ayudan a profundizar en la causa-efecto para ver los patrones emergentes y tener una comprensión más profunda de un problema en particular. Por ejemplo: analizar la producción de una planta por distintas perspectivas (categorías, perfiles, etc.), realizar comparaciones con otras variables y/o con otras plantas.
- **Análisis Predictivo:** Describe la probabilidad de ocurrencia de un evento. Utiliza los descubrimientos del Análisis Descriptivo y de Diagnóstico para predecir tendencias futuras, siendo esto una aproximación que puede o no ser asertivo. Se basa en modelos estadísticos. Por ejemplo: modelo matemático basado en el comportamiento histórico de fallas que predice su comportamiento futuro.
- **Análisis Prescriptivo:** se consiguen recomendaciones sobre las acciones que se han de seguir para reducir costos y aumentar beneficios. La disciplina toma como

base todo el conocimiento y técnicas de los Análisis Descriptivos, Diagnóstico y Predictivos (clasificación, predicción y segmentación); sumando además técnicas de simulación y optimización. A grandes rasgos, un Sistema Prescriptivo toma información del negocio, predice en base a dicha información qué impacto tendrán las diferentes políticas o acciones que se vayan a tomar, y escoge la política con mayor beneficio.

Otro de los modelos de referencia en torno a las capacidades analíticas, es el que se presenta a continuación, donde se hace la distinción de las capacidades en torno al BI (Business Intelligence o Analítica tradicional) y el BA (Business Analytics o Analítica avanzada)



Ilustración 9: Niveles de análisis para la ventaja competitiva

Los términos de “BI” y “BA” ya fueron descritos en la sección de “Analítica de Datos (Data Analytics)”. En términos simples y generales la diferencia se centra en que la primera se aboca a la revisión del pasado, mientras que la segunda incorpora modelos matemáticos para predecir y optimizar de cara al futuro.

2.6.2 Gestión de Datos Maestros

La Gestión de Datos Maestros (Master Data Management) es la capacidad de crear un vínculo común entre los datos claves de la organización para crear una única fuente de referencia, tratando de eliminar redundancias que se producen con la existencia de múltiples fuentes de información. Este vínculo se establece a través de un conjunto de conductas y tecnologías que permiten crear y mantener datos maestros coherentes, completos, contextuales y precisos en todas las partes donde estén presentes (aplicaciones, bases de datos, procesos, etc.)

La adopción del de la Gestión de Datos Maestros, al igual que la mayoría de los procesos, requiere de un desarrollo progresivo en la empresa, en el cual se va adaptando a nuevas formas de distribuir/ordenar las responsabilidades en torno a la gestión de la información. Partiendo desde un nivel inicial 1 hasta un nivel optimizado 5. La siguiente ilustración representa dicho proceso.

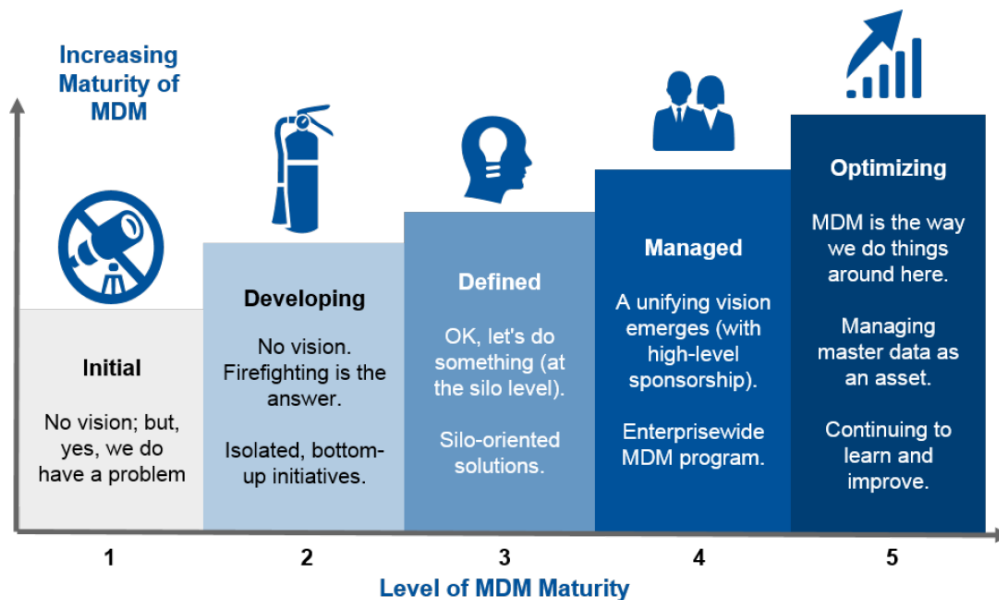


Ilustración 10: Modelo de Madurez del Master Data Management. Gartner 2015

A continuación, una descripción general de la ilustración:

- **Inicial:** No hay visión ni estrategia, tampoco existe el rol administrador de datos, las aplicaciones funcionan como silos
- **En desarrollo:** Iniciativas de visión comienzan a surgir, surge la cultura del administrador de datos, existen soluciones parciales de calidad de datos
- **Definido:** Visión estratégica pero limitada, uso exitoso de métricas a nivel de dominio, solución MDM cubriendo solo un dominio
- **Administrado:** Visión unificadora para MDM, organización con grupos de administradores de datos centralizadas o, mejores prácticas compartidas en toda la organización.
- **Optimizado:** Visión es un facilitador clave del éxito comercial, gobierno y administración empresarial bien establecida.

2.6.3 Tecnologías emergentes

El artículo de Gartner llamado “El Ciclo de sobre expectativas para Tecnologías Emergentes” reúne información de más de 2.000 tecnologías sintetizadas en un conjunto de 35 tendencias tecnológicas emergentes. Este ciclo de expectativas se concentra específicamente en un conjunto de tecnologías que se muestran prometedoras en entregar un alto grado de ventaja competitiva dentro de los próximos 5 o 10 años.” (Gartner, 2018)

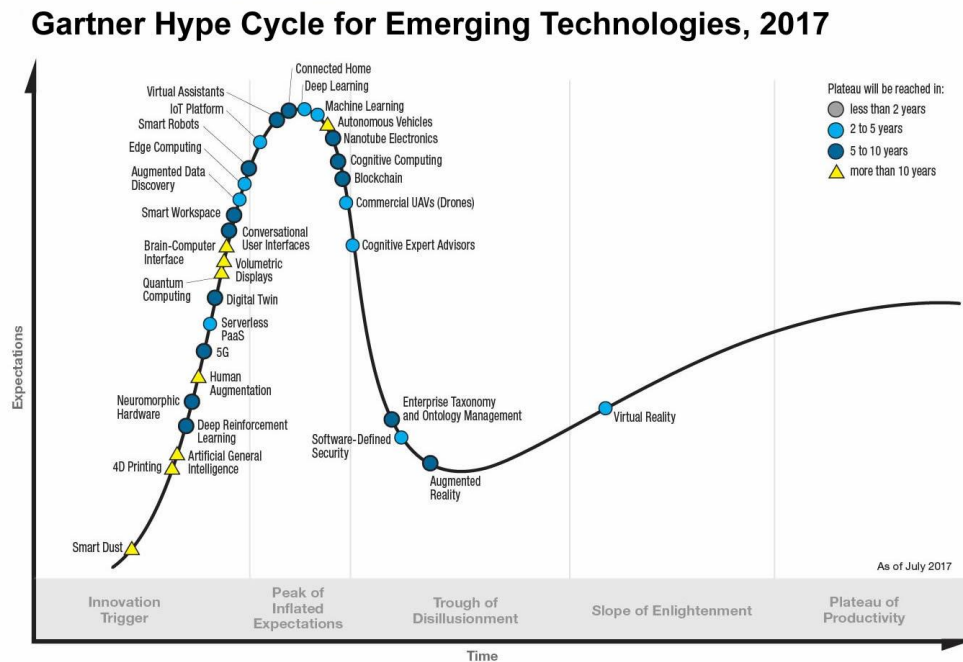


Ilustración 11: Ciclo de Tecnologías Emergentes. Gartner

A continuación, una descripción general del gráfico:

- **Lanzamiento (Innovation Trigger):** Se refiere a avances tecnológicos potenciales, se cautiva el interés de los medios, aun cuando no existen productos que apliquen la tecnología y la viabilidad comercial no está comprobada.
- **Pico de expectativas sobre-dimensionadas (Peak of Inflated Expectations):** La publicidad temprana produce una serie de historias de éxito, pero también de fracasos, otras en tanto esperan a ver su evolución.
- **Abismo de desilusión (Trough of Disillusionment):** El interés disminuye cuando no existen nuevos experimentos e implementaciones que utilicen la tecnología.
- **Rampa de consolidación (Slope of Enlightenment):** Los ejemplos de cómo la tecnología puede beneficiar a las empresas empiezan a cristalizar y ser más ampliamente comprendida. Un mayor número de empresas invierten en la tecnología mientras que las empresas más conservadoras se mantienen cautelosas.
- **Meseta de productividad (Plateau of Productivity):** La adopción generalizada comienza a despegar. La amplia aplicabilidad y relevancia de la tecnología en el mercado están dando resultados.

2.6.4 Naturaleza de los datos

Para el análisis de la naturaleza de datos se presentan tres variables: volumen, variedad y velocidad. La complejidad de su tratamiento se clasifica en: estructurados, semi estructurados y desestructurados. Como lo propone el Big Data, a medida que los datos van aumentando, la variedad también se incrementa y con ello se va haciendo más difícil mantenerlos estructurados.

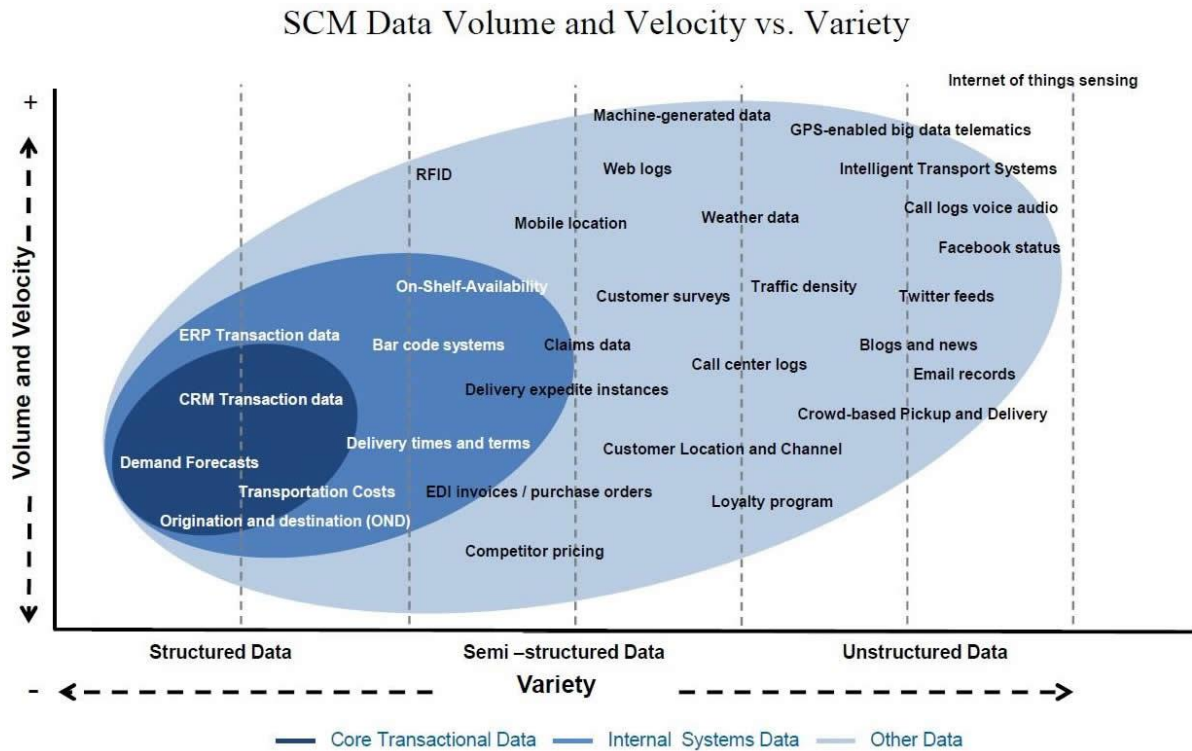


Ilustración 12: Composición de la Variedad, Volumen y Velocidad de los datos

En lo principal, se definen 3 tipos de variedades:

- **Datos estructurados:** son aquellos que tienen definida su longitud, formato y tamaño. Típicamente se almacenan en bases de datos relacionales y planillas de cálculo, los cuales se pueden consultar y procesar en forma de diseño fijo
- **Datos no estructurados:** son aquellos que tienen forma o estructura desconocida. Se almacenan en múltiples formatos como documentos, correos electrónicos, archivos multimedia (imagen, audio o video), etc.
- **Datos semi-estructurados:** son aquellos estructurados, pero no definidos, por ejemplo: archivo XML, HTML, JSON, etc.

2.7 Organización y Personas

Para esta sección se han seleccionado aspectos organizativos y de perfiles que son típicamente considerados al momento de conformar los equipos de trabajo en torno al BI.

2.7.1 Centro de Competencia de Inteligencia de Negocios (BICC)

Para el correcto funcionamiento de las Herramientas de BI en una empresa es necesario establecer un orden y distribución tanto del personal adecuado como de las herramientas disponibles, en la publicación *“Business Intelligence Competency Centers: A Team Approach to Competitive Advantage”* (Miller, Bräutigam, & Gerlach, 2006), se presentan seis desafíos que conlleva la implementación exitosa de una instancia de BI:

- **Desafíos de información (“Data Challenges”)**: El acceso, almacenamiento y uso de forma efectiva de los datos disponibles.
- **Desafíos Tecnológicos (“Technology Challenges”)**: Mantener una cadena de valor y comunicación entre las diferentes plataformas de información existentes.
- **Desafíos de Proceso (“Process Challenges”)**: Establecer los procesos adecuados para el correcto funcionamiento de una estrategia de BI.
- **Desafíos Estratégicos (“Strategy Challenges”)**: Unificar los diferentes desarrollos de BI e intereses de las áreas de una empresa bajo un solo ente supervisor.
- **Desafíos de Usuarios (“Users Challenges”)**: Definir usos y necesidades de las diferentes áreas para distribuir la información de manera eficiente.
- **Desafíos Culturales (“Cultural Challenges”)**: Lograr una cultura de toma de decisiones basada en la información.

Para solucionar estos problemas es que surge la estrategia BICC que propone la implementación de un equipo “multifuncional” que cuente con una estructura organizacional permanente, que se encargue de la correcta implementación de estrategias BI dentro de la empresa, Gartner lo define como *“A **business intelligence competency center** (BICC) develops the overall strategic plan and priorities for BI. It also defines requirements, such as data quality and governance and fulfills the role of promoting the use of BI”* (Gartner, 2012).

Bajo una mirada genérica, el BICC se encarga del desarrollo del Plan Estratégico y la promoción del BI.

La siguiente ilustración identifica las habilidades que deben estar presentes en los equipos de trabajo y las distintas estrategias de despliegue en la organización.

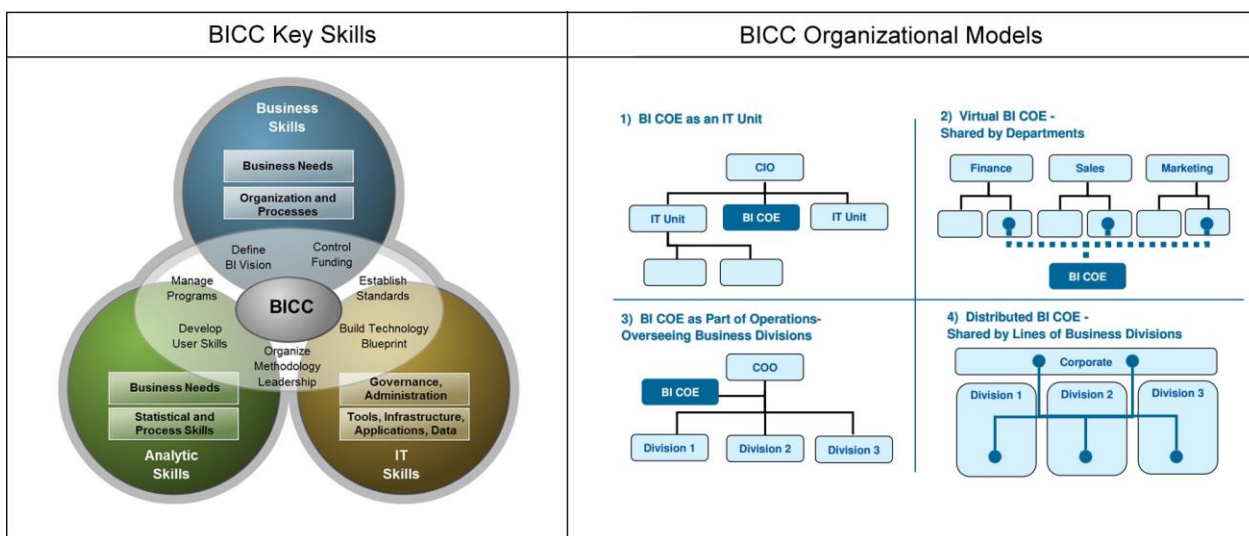


Ilustración 13: Centro de Competencia de Inteligencia de Negocios (BICC - Gartner)

2.7.2 Roles del equipo BI

A continuación, se describen los principales roles que se establecen en el marco de los equipos de trabajo para el BI.

- **Data Architect**

El Arquitecto de datos es el responsable de la diseñar, implementar y mantener la arquitectura de datos de la organización, se encargan de ver como los datos serán almacenados y dependiendo de los usos que se le den como serán integrados con otras aplicaciones. El Data Management Body of Knowledge (DAMA-DMBOK, 2017) lo describe sus funciones como *“provides a standard common business vocabulary, expresses strategic data requirements, outlines high level integrated designs to meet these requirements, and aligns with enterprise strategy and related business architecture.”*

- **Data Engineer**

El Ingeniero de Datos se encarga de proporcionar los datos de una manera accesible y apropiada a los usuarios y Data Scientists. Es un perfil especializado en infraestructura Big Data. Desarrolla y explota técnicas, procesos, herramientas y métodos que deben servir para el desarrollo de aplicaciones Big Data. Tiene un gran conocimiento en gestión de bases de datos, arquitecturas de clúster, lenguajes de programación y sistemas de procesamiento de datos

- **Data Analyst**

El Analista de Datos se encarga de analizar los datos obtenidos y los contrasta con las necesidades de la organización para poder obtener información relevante, es el encargado de recoger y distribuir las necesidades de negocio para los Data Scientist.

- **Data Scientist**

El Científico de Datos es el encargado de extraer información desde los datos aplicando diferentes tipos de estrategias matemáticas y estadísticas entre otras, debe tener

conocimientos tanto informáticos como matemáticos además de comprender el funcionamiento de la organización para aplicar las metodologías necesarias en cada caso.

- **Data Manager**

El Administrador de Datos es quien se encarga de asegurar la operación y el uso eficiente y efectivo de la información, definiendo protocolos y estándares para garantizar la calidad e interoperabilidad de los datos. El perfil está relacionado a diversas disciplinas de la gestión de datos, tales como: Data Governance, Data Modeling y Master Data Management, entre otros.

3 CAPITULO 3: ACERCA DE LA INSTITUCIÓN

Colbún S.A. es una empresa de generación eléctrica que opera en una de las 3 zonas de abastecimiento energético en Chile, el Sistema Eléctrico Nacional (SEN), cuya extensión es de Arica a Chiloé. Posee 23 centrales hidroeléctricas y 9 termoeléctricas distribuidas en las zonas centro y sur de Chile, siendo el tercer actor más importante del mercado (Memoria Anual Colbún S.A. 2018).

3.1 El Negocio de Colbún S.A.

Colbún S.A es una empresa de generación de energía eléctrica, cuyo producto es inyectado al SEN, además de abastecer directamente a diversas empresas productivas y de servicios. La siguiente ilustración presenta a la compañía en cifras



Ilustración 14: Colbún en Cifras (Memoria anual 2018)

La distribución de ingresos de Colbún está dada por los siguientes conceptos:

- **Energía (74%):** venta de energía a clientes o al mercado spot
- **Potencia (11%):** el mercado eléctrico chileno paga a los generadores por poner capacidad a disposición del sistema
- **Peajes (11%):** se refiere a los ingresos por las líneas de transmisión que la compañía tiene
- **Otros (4%):** entradas principalmente por ventas de bonos de carbono.

3.1.1 Plantas eléctricas

En la siguiente ilustración se presentan las centrales por tecnología de Colbún:

Thermal: 2,271 MW

Hydro¹: 1,630 MW



Ilustración 15: Centrales de Colbúnⁱ

3.1.2 Clientes

La energía eléctrica producida es inyectada a Clientes Regulados (distribuidoras de electricidad) y a Clientes No Regulados o Libres (empresas mineras, industriales y estatales).



Ilustración 16: Clientes de Colbúnⁱⁱⁱ

3.1.3 Estrategia Corporativa

La compañía tiene preocupación continua por ser una empresa de altos estándares de gestión y calidad. Cada año se definen las directrices a seguir a partir de la Visión, Misión y Pilares Estratégicos, estas luego bajan a Objetivos Corporativos que en sí mismo contienen a las iniciativas, proyectos y tareas que cada colaborador debe realizar, estableciendo para esto metas e indicadores que permiten un “qué hacer” alineado. Lo anterior da forma a la Estrategia Corporativa.

A continuación, se presenta la primera capa de la Estrategia Corporativa, esto es la Visión, Misión y Pilares Estratégicos:

- **Visión:** Ser una empresa chilena referente en la generación de energía segura, competitiva y sostenible.

Somos la única empresa eléctrica de capitales 100% chilenos y eso nos hace diferentes. Nos lleva a tener un mayor compromiso con lo que pasa en nuestro país y a tener una voz potente en cuanto a la definición de la política energética que Chile necesita. Nuestro sueño es destacarnos por ser “ejemplo” de cómo se hacen las cosas en la industria de la generación.

- **Misión:** Generar valor de largo plazo desarrollando y gestionando activos de infraestructura energética, integrando con excelencia las dimensiones económicas, técnicas, medioambientales y sociales.

Construimos y gestionamos una matriz diversificada que nos permita aportar un suministro de energía confiable al sistema, con niveles de precios que fomenten el desarrollo económico del país. Y todo en un marco de respeto y cuidado del medio ambiente, construyendo relaciones duraderas y de confianza mutua con nuestros grupos de interés.

- **Pilares Estratégicos:** Para el año 2018 la alta gerencia definió 3 pilares estratégicos: “Organización Comprometida”, “Clientes Fidelizados” y “Soluciones que agregan valor”, esta última contiene el Objetivo Corporativo de “**Procesos eficientes con tecnologías robustas y digitales**”, bajo el cual este trabajo se convierte en iniciativa. A continuación, la ilustración que lo representa:



Ilustración 17: Pilares Estratégicos ^{iv}

3.1.4 **Modelo de Negocio**

El modelo de negocio que la compañía declara a través de su sitio web intranet contempla los siguientes puntos:



Ilustración 18: Modelo de Negocio de Colbún

El modelo de negocios de la compañía se centra en ser un referente nacional en la generación eléctrica segura, competitiva y sustentable. Para ello, la empresa define 3 líneas de negocio: Generación (como actividad principal), Desarrollo de infraestructura energética (nuevas plantas) y Comercialización de Energías y Servicios (ventas a clientes y mercado spot); lo anterior con una Gestión de Sostenibilidad a lo largo de toda su cadena de valor.

La Fórmula de Utilidades está dada por ingresos de Energía, Potencia, Transmisión y otros, puntos que ya fueron detallados en el apartado de “El Negocio de Colbún S.A.”. Mientras que los principales costos están dados por Compras al mercado spot, Combustibles, Transmisión y otros.

Los Recursos Claves en tanto, tienen que ver con las principales áreas que apalancan la continuidad operacional del negocio.

3.2 Delta de Hax de Colbún S.A.

La compañía está inmersa dentro de un mercado regulado, con clientes que suscriben contratos de largo plazo y en donde, además, existe un organismo establecido por ley que coordina la operación del sistema (Coordinador Eléctrico Nacional, CEN). Bajo los anteriores antecedentes es posible asegurar que bajo la mirada del Modelo Delta de Hax, la empresa se enmarca en la estrategia de “**Liderazgo en Costo**”, en efecto, ya que la misión, visión y objetivos corporativos están alineados hacia una producción eficiente y económica.



Ilustración 19: Modelo Delta de Hax

3.3 Modelo Porter de Colbún S.A.

Basado en los esquemas conceptuales del “Diamante de Competitividad de Porter”, a continuación, se presenta un resumen del análisis de las cinco fuerzas:

Fuerza Competitiva	Nivel de Atractivo
<p>Barreras de Entrada</p> <p>El sector eléctrico tiene fuertes barreras de entrada. Por un lado, se necesita un fuerte nivel de inversiones para la construcción de centrales y por otro lado el modelo de negocios y normativa vigente resulta ser altamente compleja. A lo anterior se suma el emergente rechazo ciudadano a la construcción de nuevas plantas generadoras.</p>	Alta
<p>Poder Negociador de los Proveedores</p> <p>Para las centrales hidráulicas los costos de proveedores son bajos en comparación a las centrales térmicas. En general, proveedores de combustibles fósiles son pocos en el mundo y los precios están sujetos a variables tanto especulativas como operaciones y no dependen de una negociación de mercado, además, las inversiones en las centrales ya están hechas y no es fácil venderlas (liquidarlas) ni convertirlas a otro combustible.</p>	Alto
<p>Poder de Negociación de los Clientes</p> <p>Los clientes de las empresas generadoras son, por un lado, las distribuidoras, y por otro los grandes clientes industriales (mineros, por ejemplo). En general, no tienen muchas alternativas dónde elegir y son por naturaleza a muy largo plazo (10 a 20 años). Pese a que existen procesos de licitación y/o acuerdos bilaterales, solo ahí pueden manifestar cierto poder de negociación, pero una vez determinado el proveedor, no es fácil cambiarse de él.</p> <p>Sin embargo, gran parte de los contratos de clientes estaría finalizando para el 2020, esto en el mercado en general, por lo que se prevé que para entonces los clientes tendrán un mayor poder de negociación del que tienen hoy en día.</p>	Bajo / Medio
<p>Amenaza de Sustitutos</p> <p>Hoy en día la electricidad prácticamente no posee sustitutos cercanos, más bien la sustitución que podía enfrentar la industria es la autoproducción para el caso de clientes industriales con capacidad de mantener sus propias unidades generadoras, los cuales son muy pocos en Chile.</p>	Baja
<p>Intensidad de la Rivalidad</p> <p>El no poder afectar el nivel de demanda, hace que la competencia sea muy fuerte por diversos aspectos: conseguir nuevos contratos, conseguir bajos precios en combustibles, eficiencia operacional, financiamiento, etc. Esto además se intensifica con el hecho que es el Coordinador Eléctrico Nacional quien determina los precios de transacción de energía entre las generadoras.</p>	Alta

Tabla 1: Análisis de las 5 Fuerzas de Porter

3.4 Análisis F.O.D.A. de Colbún S.A.

A continuación, un análisis F.O.D.A. de la compañía:

		Positivo	Negativo
		Fortalezas	Debilidades
Interno		<ul style="list-style-type: none"> • Rentabilidad creciente. • Segunda empresa nacional con mayor producción de energía. • Conocimiento del sector eléctrico permite la expansión de la compañía. • Obtención de financiamientos gracias a la trayectoria y solidez económica de la empresa. • Solidez de principales accionistas (Matte) y operador eléctrico extranjero (Suez Energy). • Matriz energética diversificada. • Múltiples reconocimientos de Responsabilidad Social y una buena relación con la comunidad apoya la continuidad operacional. 	<ul style="list-style-type: none"> • Plantas de generación se encuentran a grandes distancias, lo que dificulta el abastecimiento de combustible y la operación centralizada. • Escenario es fuertemente competitivo. De hecho, pueden enfrentarse resultados operacionales negativos ante malas decisiones. • Altas probabilidades de rechazo del sector social por tratarse de empresa de un importante grupo económico del país.
		Oportunidades	Amenazas
Externo		<ul style="list-style-type: none"> • La demanda eléctrica ha crecido un 6,1% promedio anual en los últimos cinco años, lo que genera un potencial importante de crecimiento. 	<ul style="list-style-type: none"> • Alta variabilidad en de las condiciones climáticas afectan la estabilidad de la rentabilidad. • Alta regulación y vigilancia del mercado impide maximizar la producción. • Legislación obliga a asumir costo sistémico en caso de que la competencia quiebre. • Modificaciones en la legislación vigente. • Riesgo en estabilidad de precios y suministro de combustibles.

Tabla 2: Análisis F.O.D.A.

4 CAPITULO 4: PROYECTO

El siguiente capítulo presenta al lector el proceso de investigación que se llevó a cabo para la elaboración de este proyecto. Considerando las condiciones generales de la institución y la información con la que se cuenta para construir una propuesta razonable que dé respuesta a los objetivos planteados en este proyecto.

4.1 Levantamiento de la situación actual

Para analizar el estado sin proyecto, se hará primeramente una bajada hacia el proceso de atención de demanda de “Nuevos Productos / Aplicativos”, esto mediante la metodología de Ingeniería de Negocios propuesta por el autor Oscar Barros, luego, el escenario de arquitectura BI bajo el cual se enfrentan los requerimientos del negocio.

4.1.1 *Arquitectura de Macro Procesos actual*

La siguiente “Ilustración 20” representa la arquitectura de macroprocesos de la Gerencia de TI, en su nivel general. La macro 1 se destaca en color ya que se apertura en la “Ilustración 21”.

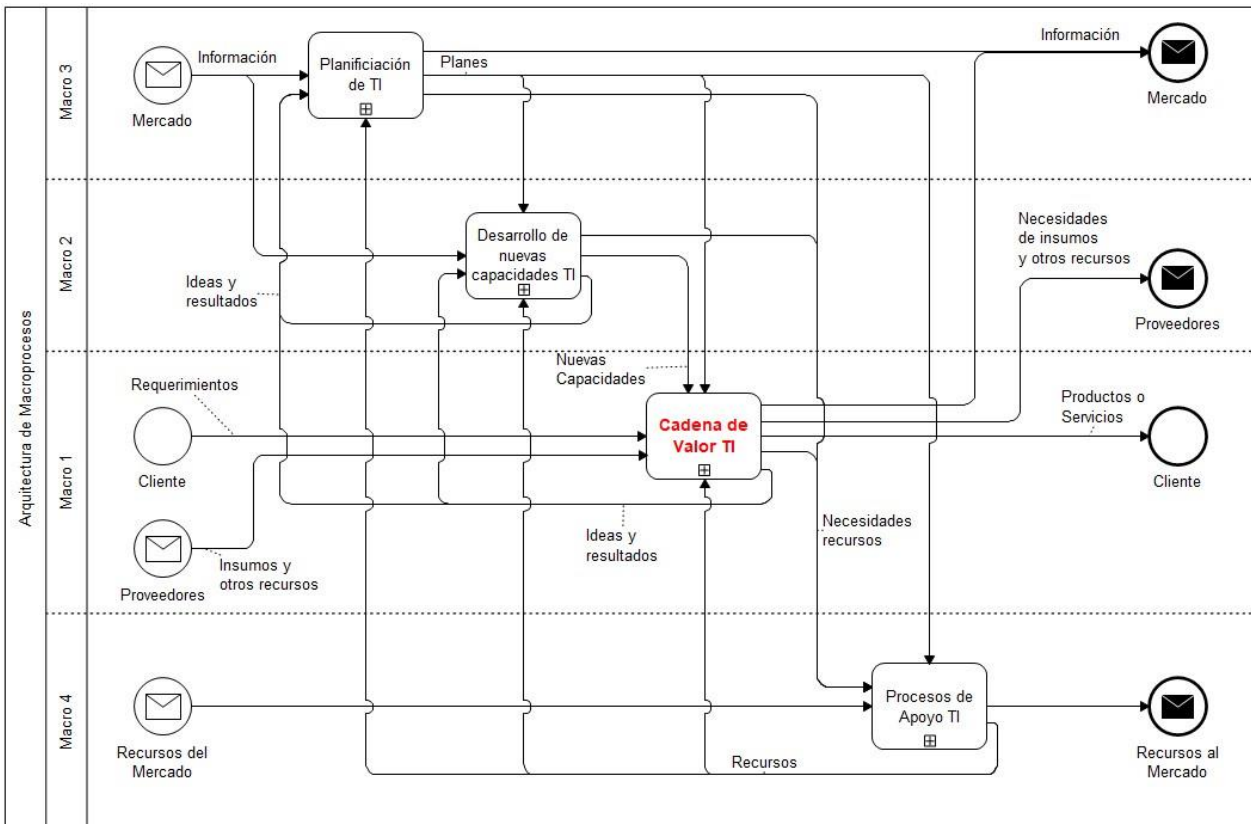


Ilustración 20: Arquitectura de Macro Procesos: Nivel 0

La “Ilustración 21” corresponde a la apertura de la “Cadena de Valor TI”. La cadena de valor “Gestión de Suministro de Productos y Servicios” se destaca en color ya que se apertura en la “Ilustración 22”

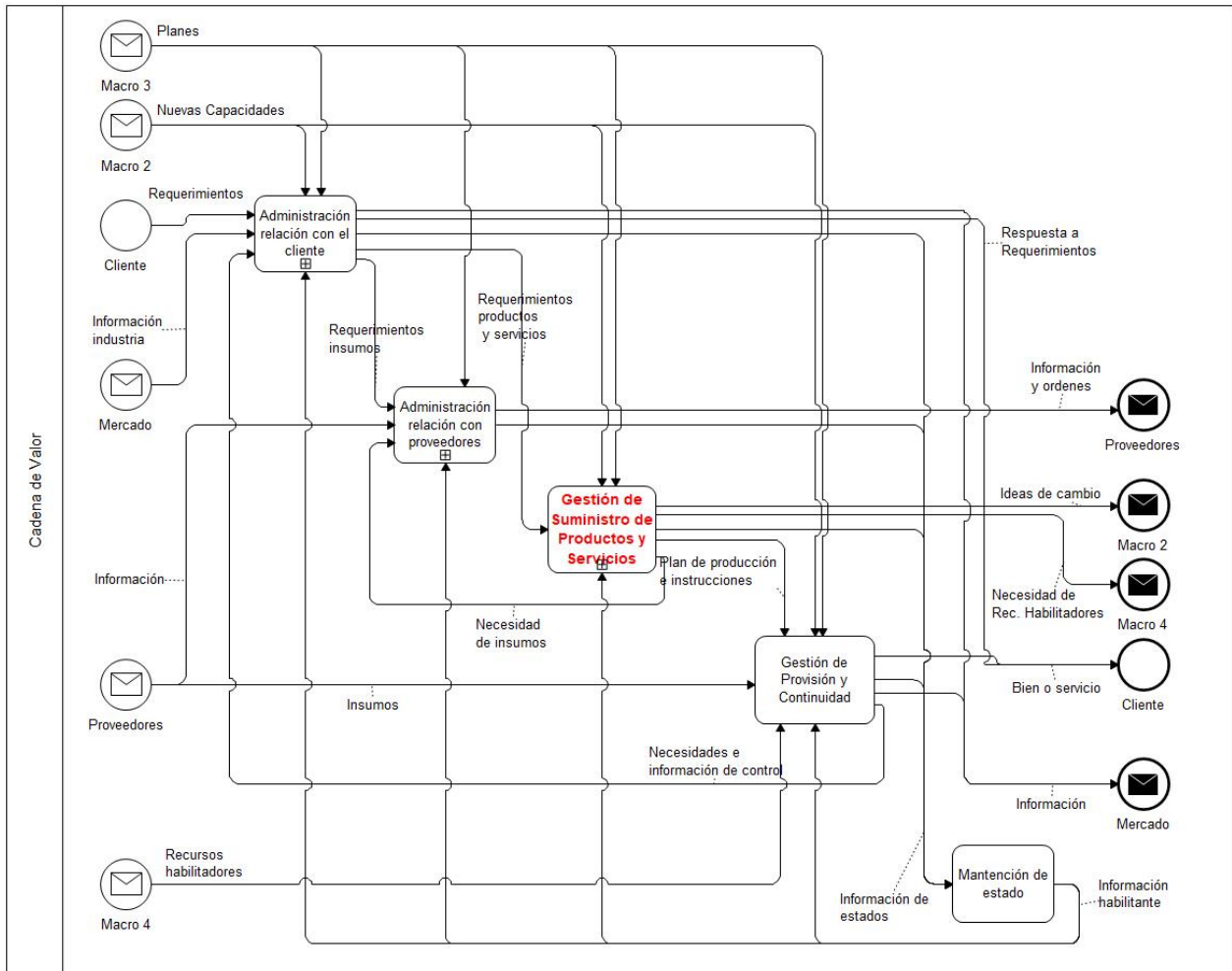


Ilustración 21, Arquitectura de Macro Procesos: Nivel 1

La “Ilustración 22” corresponde a la apertura de la cadena de valor “Gestión de Suministro de Productos y Servicios”. Es posible apreciar dos macro procesos que por su naturaleza sistémica la compañía las mantiene separadas: Plataformas Corporativas y Plataformas Operacionales, la primera hace referencia a los sistemas SAP y sistemas legados, mientras que la segunda hace referencia a los sistemas de control del negocio como lo son los SCADA, Pi System y Prime Read.

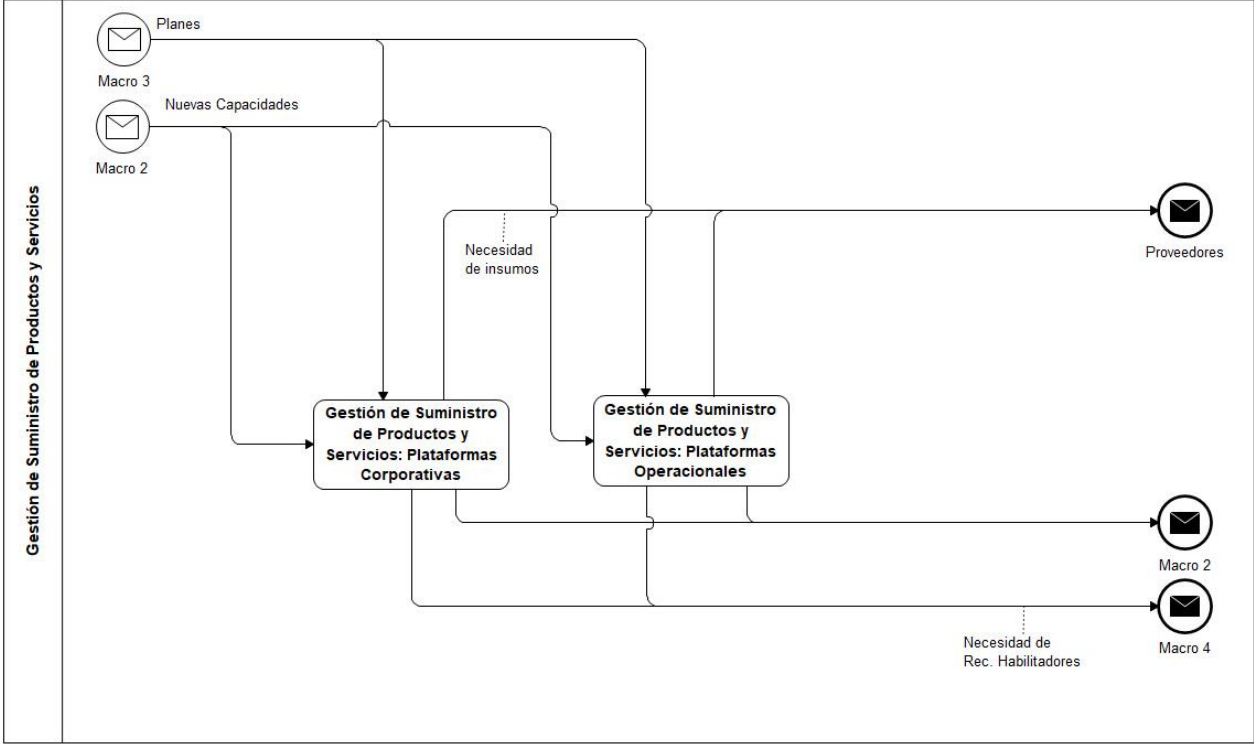


Ilustración 22: Arquitectura de Macro Procesos: Nivel 2

La “Ilustración 23” a la apertura de los macro procesos de “Gestión de Suministro de Productos y Servicios”, esto es, el proceso en sí que sigue las iniciativas de nuevos productos o aplicativos.

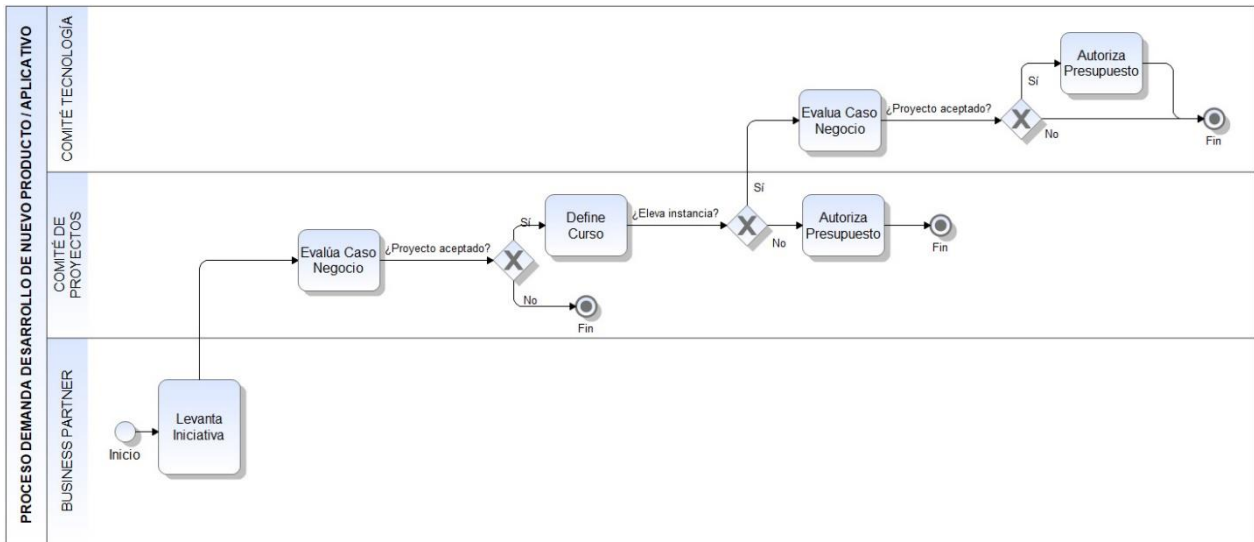


Ilustración 23: Proceso de Demanda de Desarrollo de Nuevos Productos / Aplicativos

De acuerdo con esto, desde ya posible concluir que el proceso de atención de la demanda no hace distinción si la iniciativa corresponde a un desarrollo de software o un producto de BI.

4.1.2 Plataforma y Arquitectura de BI actual

A continuación, y de acuerdo con la arquitectura de referencia expuesta en el marco teórico, se presenta la siguiente ilustración como resultado de un levantamiento de capacidades instaladas de manera completa o parcial, junto con ello el GAP en base a un estado “To Be” que la documentación respectiva sugiere como práctica estándar.

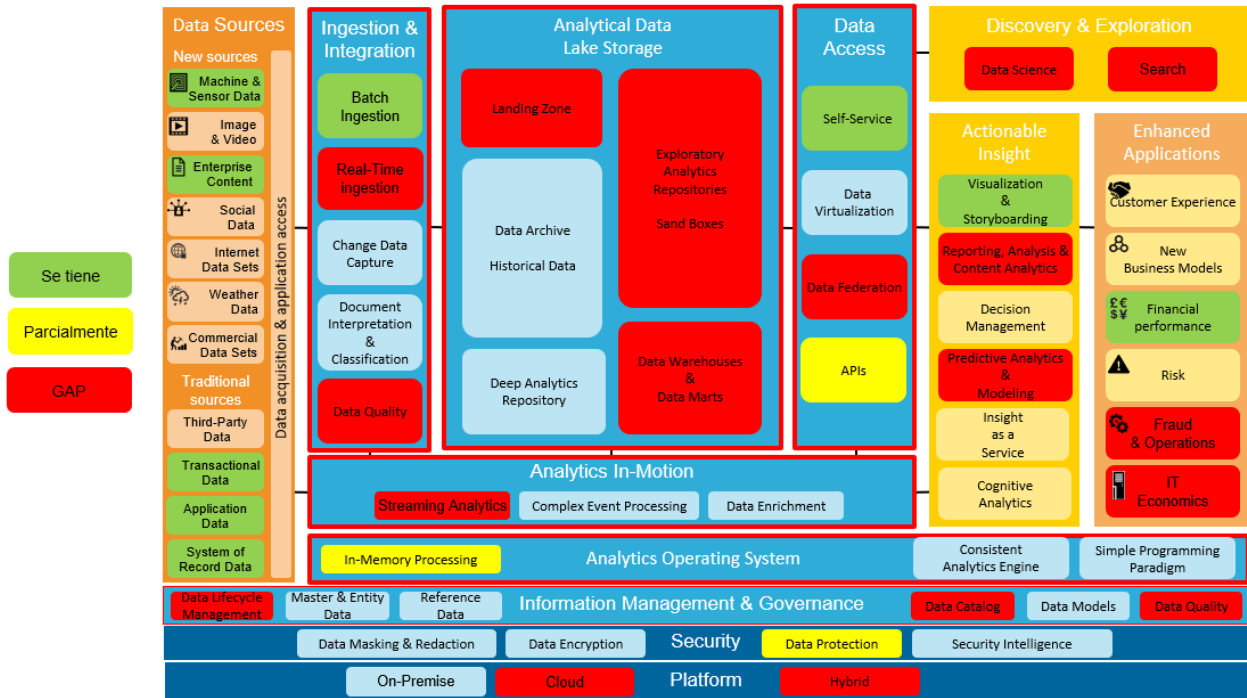


Ilustración 24: Arquitectura de Capacidades actual

Es posible ver claramente el GAP (en rojo) las capacidades no disponibles en la compañía, pudiéndose concluir desde ya, una condición ciertamente subestándar.

Una adecuación simplificada a la arquitectura de referencia presentada anteriormente en términos de capas y acondicionada para fines del proyecto, es como la que se presenta a continuación:

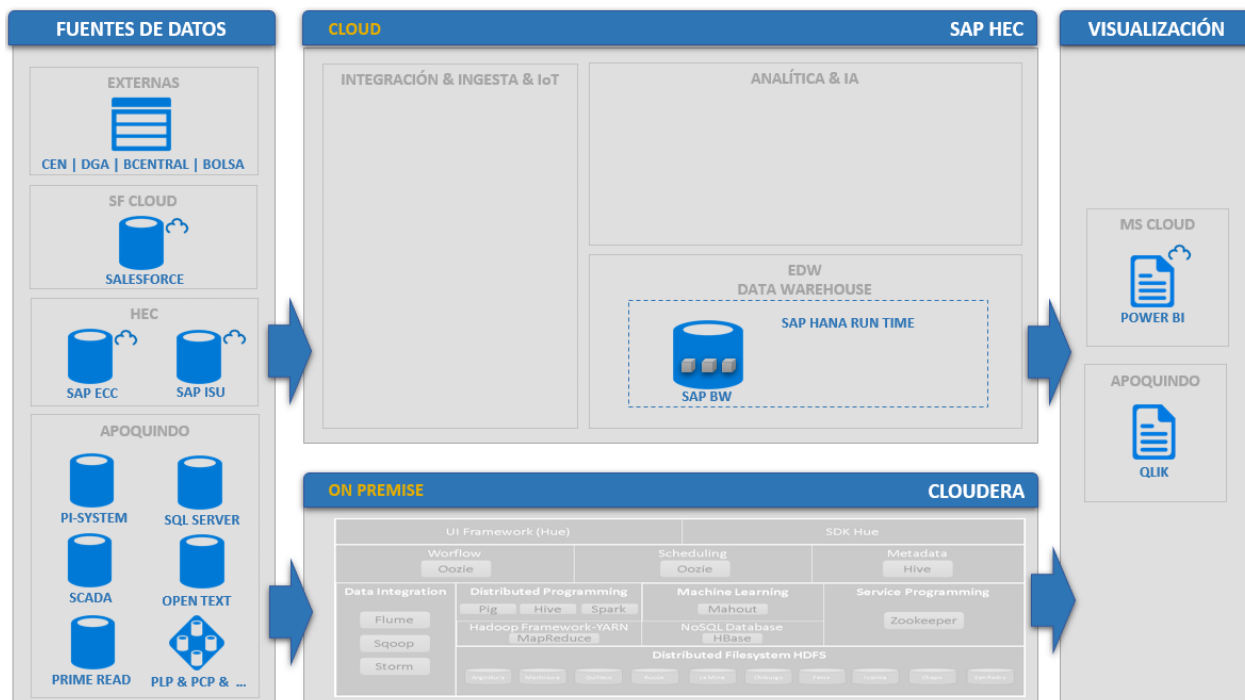


Ilustración 25: Actual Arquitectura de BI

En la ilustración anterior, en lo principal es posible observar una arquitectura tecnológica no preparada para la integración de múltiples fuentes de datos (salvo para datos SAP con modalidad cubos), el tratamiento de grandes volúmenes de datos y la analítica avanzada. Adicionalmente a lo anterior, sin gobierno. Se tomará luego esta ilustración para representar la situación “con proyecto”.

Respecto de las líneas tecnológicas de la compañía, estas van por plataformas SAP y Microsoft, en tanto que las herramientas de visualización son Qlik y PowerBI. También referencia importante para Cludera, empaquetado comercial de Hadoop, que se encuentra en proceso de prueba a nivel departamental.

4.1.3 *Recursos relacionados en situación actual*

Un catastro interno arrojó que la compañía cuenta con cierto personal distribuido en distintas áreas de la compañía con skills relacionadas a los que típicamente se conjugan en proyectos o estructuras de trabajo del BI. El listado es el siguiente:

- En áreas del Negocio
 - 2 con Diplomado en Big Data
 - 1 con experiencia senior en Data Science y manejo de plataforma Cloudera
 - 3 con experiencia senior en desarrollo herramienta Power BI
- En área de TI
 - 1 con Diplomado en Big Data
 - 1 con Diplomado en BI
 - 2 con experiencia senior en desarrollo en herramienta Power BI
 - 2 con experiencia senior en desarrollo en herramienta QlikView

Lo anterior, al menos servirá como estrategia para establecer comités de BI o armar estructuras colaborativas del tipo “champions users” o “usuarios claves” que ayuden a la culturización, gobierno, habilitación y adopción de los cambios técnico-culturales.

4.2 Diagnóstico de la Situación Actual

Esta sección hace una introspectiva para responder las consecuencias de la situación actual, identificando los problemas, los afectados, las limitaciones y oportunidades.

4.2.1 *Identificación de los problemas en situación actual*

Algunos de los problemas detectados son los siguientes:

- Problemáticas respecto del negocio y sus nuevas complejidades:
 - Interconexión Sistema Integrado Norte Grande (SING) – Sistema Interconectado Central (SIC): la unificación de ambos sistemas dio paso al Sistema Eléctrico Nacional (SEN) que contiene el 99,3% de la capacidad instalada en Chile
 - Internacionalización de la compañía: Colbún adquiere la Central Térmica Fénix Power (ubicada en Chilca, Perú)
 - Entrada tecnología ERNC: políticas públicas y comerciales fomentan inversión en proyectos de energía renovable no convencional (ERNC). Lo anterior se basa en que este tipo de fuente es más sustentable y económica que las fuentes tradicionales
 - Aumento de competencia: cambio en la regulación amplió el segmento de clientes con los que las generadoras pueden establecer contratos comerciales.

Lo anterior habla de un nuevo escenario más complejo, amplio y competitivo, con tecnologías más eficientes que hacen disminuir los precios y entrada de nuevos competidores que buscan tomar participación de mercado. Lo anterior obliga per-se a ser una compañía eficiente

- Problemáticas respecto de los datos: como efecto completo al nuevo escenario de mercado:
 - Aumento de volumetría de datos
 - Aumento de necesidades de análisis y estudios
 - Disminución del tiempo para toma de decisiones
 - Aumento de necesidades de procesamiento de datos
- Problemáticas respecto estándares internos
 - Compañía carece de un modelo corporativo de BI o estructura orgánica que haga frente a las nuevas necesidades del negocio.
 - Plataforma tecnológica corporativa no preparada
 - Herramientas de reportabilidad “ad-hoc” de lenta y costosa adaptación
 - Herramientas de reportabilidad “self-service” con desarrollo sólo a nivel departamental, sin gobierno
 - Herramientas para el procesamiento volumétrico de datos y el desarrollo de la analítica con desarrollo experimental
 - Falta de estándares técnicos y metodológicos para la explotación de los datos

Lo anterior habla de una compañía no preparada para el nuevo escenario de mercado que exige prácticas, metodologías y herramientas que faciliten y aceleren la toma de decisiones.

4.2.2 *Análisis de los problemas en situación actual*

A continuación, se describen quiénes son los principales afectados con esta situación, las limitaciones que ha provocado, pero también las oportunidades que ofrece.

4.2.2.1 ¿Quiénes son los más afectados?

Con la actual condición probablemente subestándar para este tipo de compañía, o al menos con una cultura principalmente transaccional, el principal afectado es el negocio en general, ya que diversas áreas han estado restringidas de explorar o requerir soluciones que van más allá de un reporte ad-hoc. Adicionalmente, también hay un tema de imagen compañía que no adecua sus plataformas tecnológicas hacia las nuevas tendencias del mercado.

4.2.2.2 Limitaciones detectadas

A continuación, se destacan algunos puntos relacionados con la limitación que impone la actual “condición transaccional” de la compañía:

- Falta de estructura organizacional responsable del BI inhabilita o merma a las áreas del negocio a la automatización y digitalización de sus procesos
- Inorgánica tecnológica y falta de estándares obligan al aumento de esfuerzo humano para asumir labores de manipulación de datos (integraciones manuales, procesos de limpieza de datos, consolidaciones, etc.), restando espacio al análisis y acciones de mejora.

4.2.2.3 Oportunidades detectadas

A continuación, se destacan algunos puntos relacionados con la oportunidad que abre la línea del BI Corporativo

- Reducción de tiempos de los usuarios en el acceso a la información
- Adelantarse a problemáticas del mercado con modelos de pronóstico
- Posibilidad de trabajar con grandes volúmenes de información

4.3 Análisis de Alternativas

A continuación, se exponen 3 alternativas al proyecto que en su momento analizaron en la interna corporativa, junto con ello se documenta la resolución llegada.

Alternativa	Criterio Evaluación		Resolución
	Costo	Factibilidad	
<p>Mantener situación</p> <p>Al no contar con plataforma y equipo para atender la demanda, podría permitirse que las propias áreas mandantes tomen la determinación que mejor les parezca.</p>	<p>Alto</p> <p>La descentralización y no gobierno abrirían opción a la adquisición de diversas tecnologías de mercado o desarrollos locales que inhibirían opción de economía de escala o reutilización de conocimiento y soluciones</p>	<p>Baja</p> <p>La nueva estructura organizacional del departamento de TI tiene dentro de sus objetivos tomar la dirección de las decisiones tecnológicas de la compañía</p>	<p>Inviabile</p> <p>Diversas organizaciones mundiales declaran que en esta era de transformación digital existen 2 tipos de empresas: las que se adaptan para darle valor a sus datos y las que desaparecerán</p>
<p>Derivación de la demanda</p> <p>Esto es externalizar la demanda, no solamente con recursos hombres, sino que con tecnología y servicios</p>	<p>Alto</p> <p>La externalización “end to end”, es decir, desde el levantamiento hasta la habilitación en producción de una solución, pasando por la implementación y pruebas, representaría un alto costo en toda la cadena de operación</p>	<p>Baja</p> <p>El modelo de desarrollo de soluciones tecnológicas que ha dado resultado a la compañía es contar con equipos multidisciplinarios que combinen sus conocimientos para diseñar la solución, luego se externaliza la implementación. No está dentro la habitual dinámica la externalización de todo el flujo.</p>	<p>Bajo atractivo</p> <p>La visión de la compañía es que mantención y operación que no genere valor pueda ser externalizada, BI no es un servicio con estas características por lo que no representa factibilidad de llevarse a cabo.</p>
<p>Implementar Proyecto</p> <p>Tomar esta propuesta y llevarla a implementación empresarial</p>	<p>Alto</p> <p>La inversión sin duda podría ser considerada alta ya que requiere adquirir tecnología nueva, habilitarla, capacitar, administrar, etc.</p>	<p>Alta</p> <p>La compañía ha manifestado su intención de abrirse camino a una visión digital y automatizada. “Sólo” exige que se hagan los análisis acabados y que las recomendaciones de acción estén plenamente validadas y respaldadas</p>	<p>Alto atractivo</p> <p>No hacer algo al respecto o entregar las llaves del know-house a terceros no es interesante para la organización, pero si lo es el desarrollar la capacidad que permitirá contar con un activo estratégico que genere valor y oportunidades al negocio</p>

Tabla 3: Análisis de alternativas

4.4 Propuesta de Solución

En función de lo planteado en la descripción de la problemática, el levantamiento de la situación actual y el marco teórico, la propuesta de solución es la conformación de estructuras organizacionales y procedimentales, que, en conjunto con la inclusión de nuevas tecnologías, permita a la organización contar con la capacidad de resolver casos de usos relacionados a oportunidades de la analítica de datos. Lo anterior, con visión de madurez paulatina. La especificación de la propuesta se recoge a través de un framework conceptual que responde con los elementos habilitadores requeridos para el desafío.

Los detalles de este framework serán explicados en las siguientes secciones.

4.4.1 Framework de BI Corporativo

Como se comentó anteriormente, el núcleo central del proyecto se centra en un framework conceptual que reúne una serie de componentes habilitadores que apalancan la transformación de la compañía desde una organización “transaccional” hacia una organización “Data Driven”, es decir, que el desarrollo del negocio esté direccionado o influenciado por la explotación de los datos.

La siguiente ilustración presenta el esquema de la solución.

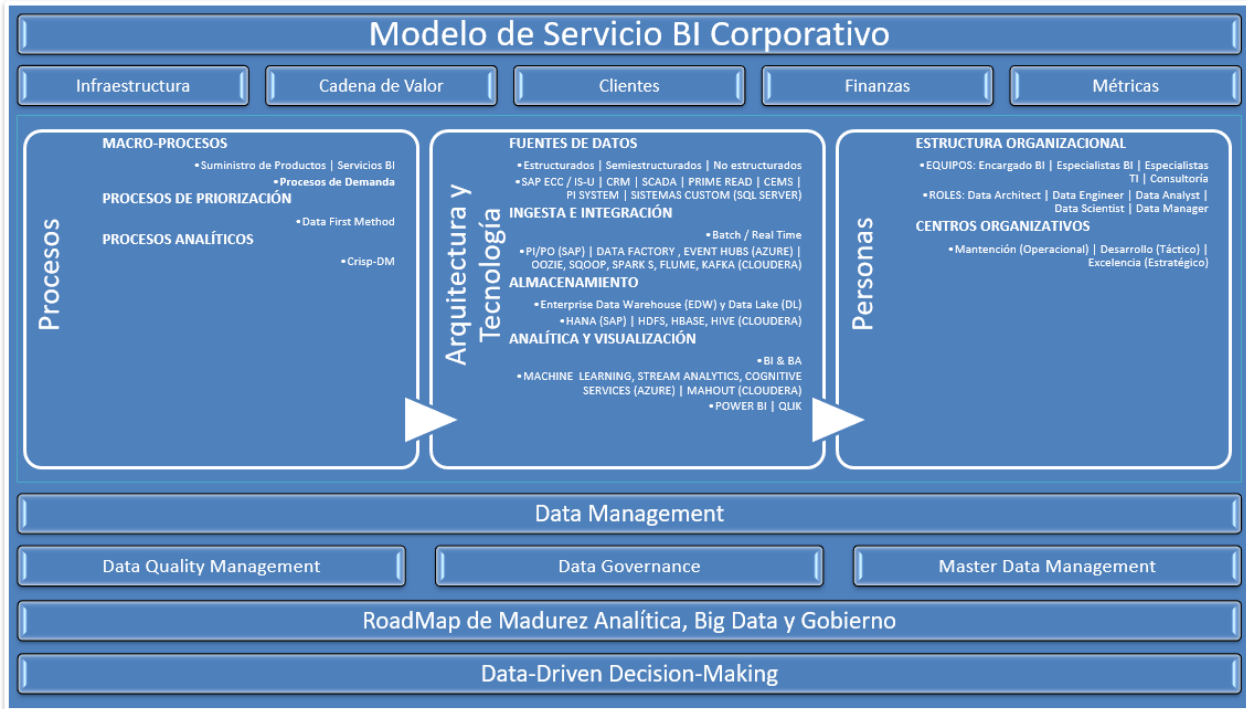


Ilustración 26: Framework Modelo de Servicio BI para Colbún

En las siguientes secciones se explicará cada una de las piezas involucradas en este framework, que en lectura general, este intenta responder diversas preguntas en torno al servicio de BI corporativo:

¿Para qué? - ¿Cómo? - ¿Qué? - ¿Cuánto? - ¿Con quién?

¿Sobre qué? - ¿Con qué?

¿Cuándo?

4.4.1.1 Canvas de Servicio

Para describir / estructurar la lógica del “modelo de negocio”, se utiliza a continuación el estándar Canvas, el cual provee una serie de módulos que permiten dar una visión global de manera rápida y sencilla de la propuesta de servicio.

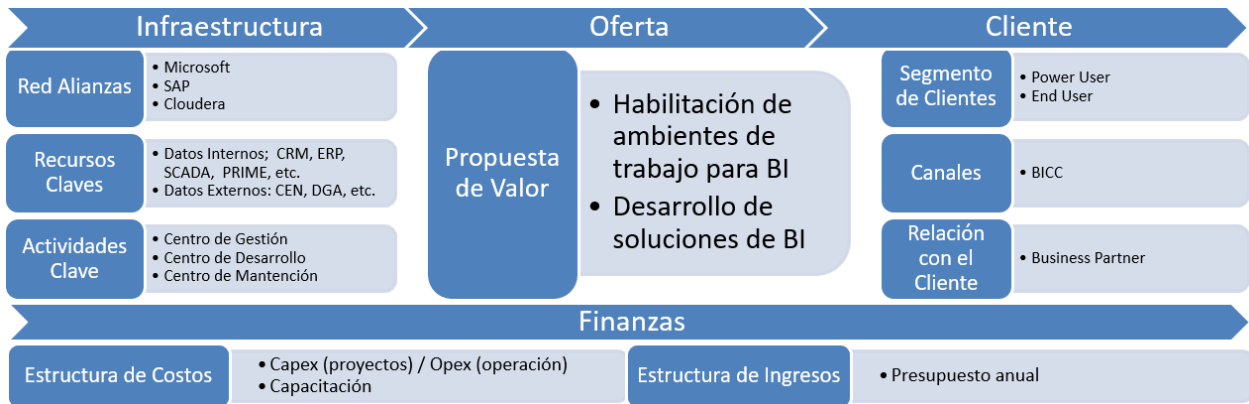


Ilustración 27: Canvas de Servicio BI para Colbún

4.4.1.2 Métricas de Servicio

El framework también se hace cargo de proponer una serie de indicadores para identificar si el servicio está prestando la utilidad que la compañía requiere.

Ámbito	Medida de desempeño	Unidad de Medida
BI Tradicional: Automatización de reportabilidad	<ul style="list-style-type: none"> Capacidad de reducir tiempo y esfuerzo a través de: <ul style="list-style-type: none"> Recolección y disponibilización de datos Construcción de reportería para la gestión 	HH, \$
BI Analítico: Construcción de modelos de Predicción, Minería, Optimización, etc.	<ul style="list-style-type: none"> Capacidad de generar mejoras y oportunidades a través de <ul style="list-style-type: none"> Modelos matemáticos que permitan anticiparse a eventos, descubrimiento de patrones Capacidad de computo que acorte los tiempos de procesamiento de información 	HH, \$
Gestión	<ul style="list-style-type: none"> Capacidad de ejecutar proyectos en Plazo y Costos establecidos Capacidad de abordar los casos de negocio propuestos con las herramientas soportadas 	%

Desarrollo de conocimiento y habilidades	<ul style="list-style-type: none"> Asistencia a instancias de aprendizaje tales como seminarios, workshop, meetup, cursos, talleres, etc. 	N°
Soporte	<ul style="list-style-type: none"> Tiempo de resolución de requerimientos e incidentes bajo un escenario de servicio de mantención de estado 	T

Tabla 4: Indicadores de desempeño propuestos

Estos indicadores, solo requieren de una planilla de cálculo para su sistematización, salvo Proyectos y Soporte, que son seguidos bajo herramientas corporativas (Project Server y Service-Now respectivamente). De acuerdo a las actuales instancias de seguimiento de la Gerencia de TI, estos indicadores pueden ser presentados en reunión mensual de “Gestión Operacional” que convoca a todas las jefaturas de área para rendir cuenta de las actividades realizadas y por realizar.

4.4.1.3 Pilar de Procesos

Este primer pilar presentado en el framework, se enfoca principalmente en describir los siguientes procesos metodológicos:

- **Procesos de Demanda:** indica cómo es el flujo de atención para la cadena de valor de este servicio.
- **Proceso de Priorización:** para el contexto de BI, indica cómo hacer un levantamiento de casos de usos y llevarlos a priorización.
- **Procesos analíticos:** indica cómo orientar los trabajos de implementación de minería de datos
- **Esquemas de Explotación:** indica cómo y con qué se implementan los tipos de proyectos que el servicio atenderá (reportabilidad, Analítica, IoT, Big Data, etc.)

4.4.1.3.1 Arquitectura de Macro Procesos

Para llegar a los **Procesos de Demanda**, se utilizará la Metodología de Ingeniería de Negocios descrita en el marco teórico.

El rediseño del proceso comienza a visualizarse a partir del nivel 2, es decir, que lo descrito en la “Ilustración 20: Arquitectura de Macro Procesos: Nivel 0” e “Ilustración 21, Arquitectura de Macro Procesos: Nivel 1” de la situación actual, permanecen sin alteración.

El primer cambio propuesto aparece en la siguiente ilustración, correspondiendo a la creación de la línea “Gestión de Suministro de Productos y Servicios: Plataformas BI”, este macroproceso aloja el seguimiento a los procesos relacionados con la entrega del servicio a la compañía.

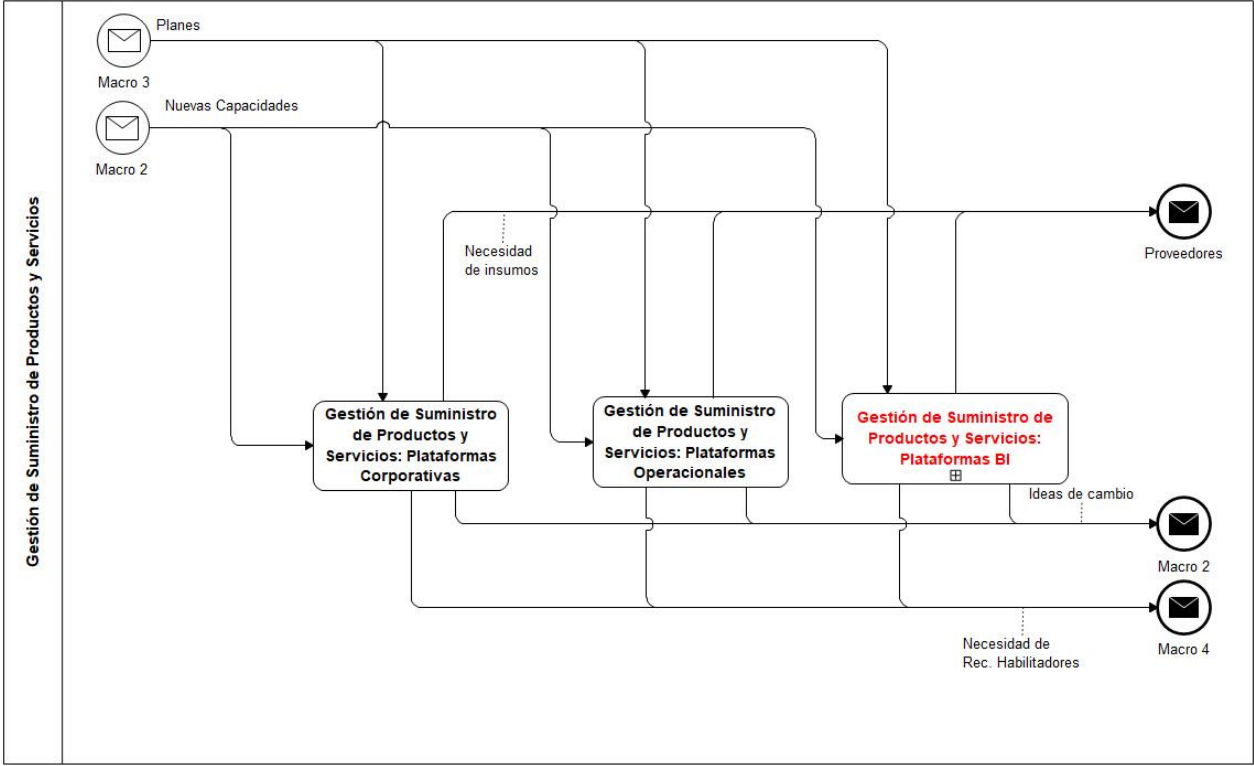


Ilustración 28: Arquitectura de Macro Procesos: Nivel 2

(Véase situación actual en página: 44)

En apertura al nuevo macroprocesos de “*Ilustración 28*”, se presenta en la siguiente figura, donde se propone la creación de dos procesos de negocios: “Desarrollo de Nuevos Productos BI” y “Habilitación de Ambientes y Ejecución de Productos BI”, los cuales corresponden a la cadena de valor del servicio, declarados en la precedente sección de “*Canvas de Servicio*”

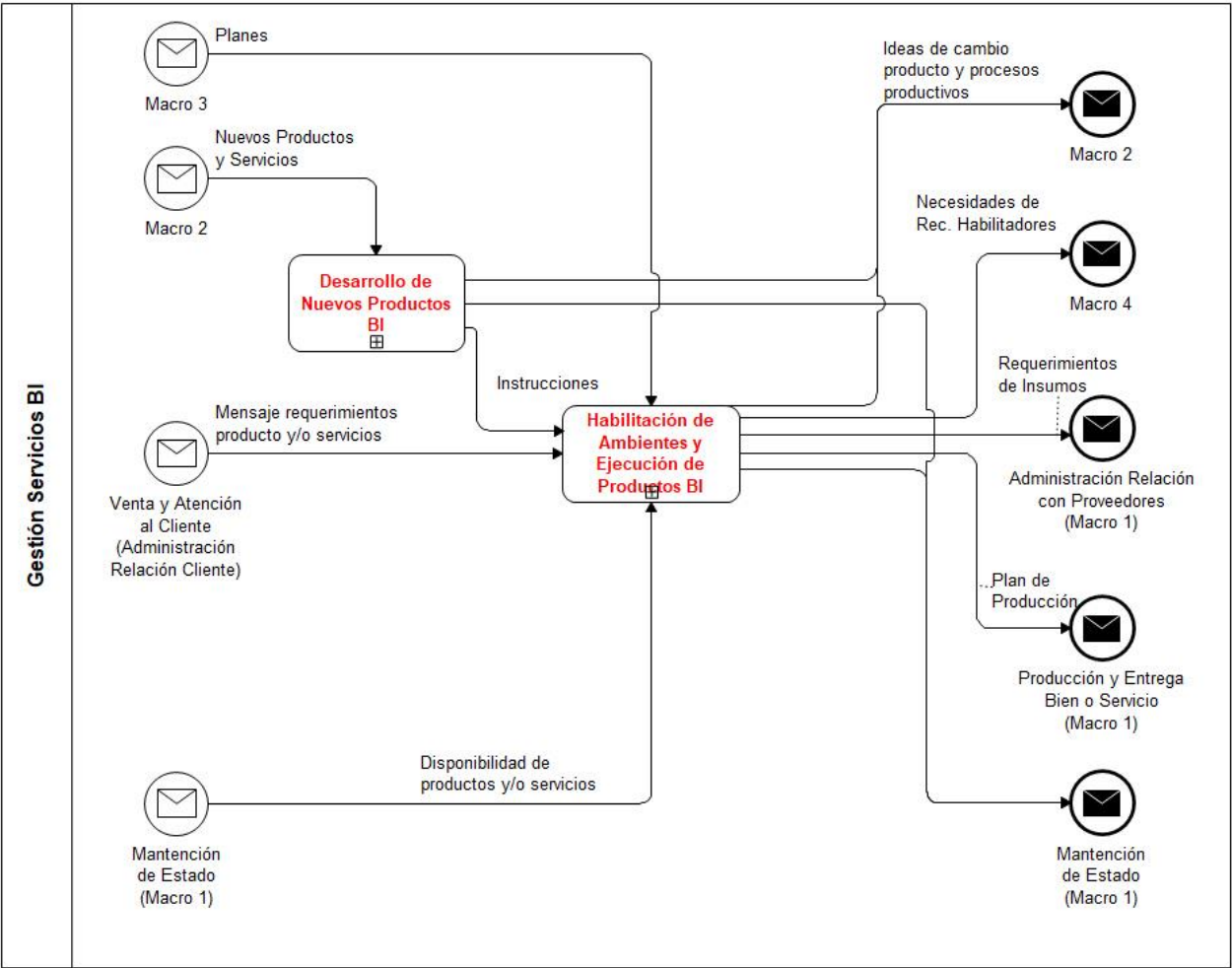


Ilustración 29: Arquitectura de Macro Procesos: Nivel 3

Las siguientes ilustraciones proponen los Procesos de Flujos de Aprobación para iniciativas del BI, tanto para “Nuevos Productos BI” como para la “Habilitación de Ambientes y Ejecución de Productos BI”.

Se entenderá por “Nuevos Productos”, típicamente a las solicitudes de nuevos proyectos de BI. Por ejemplo: construcción de procesos de recolección y disponibilización de datos (data mart), sistemas de reportería con integración de fuentes de datos, creación de modelos analíticos, etc.

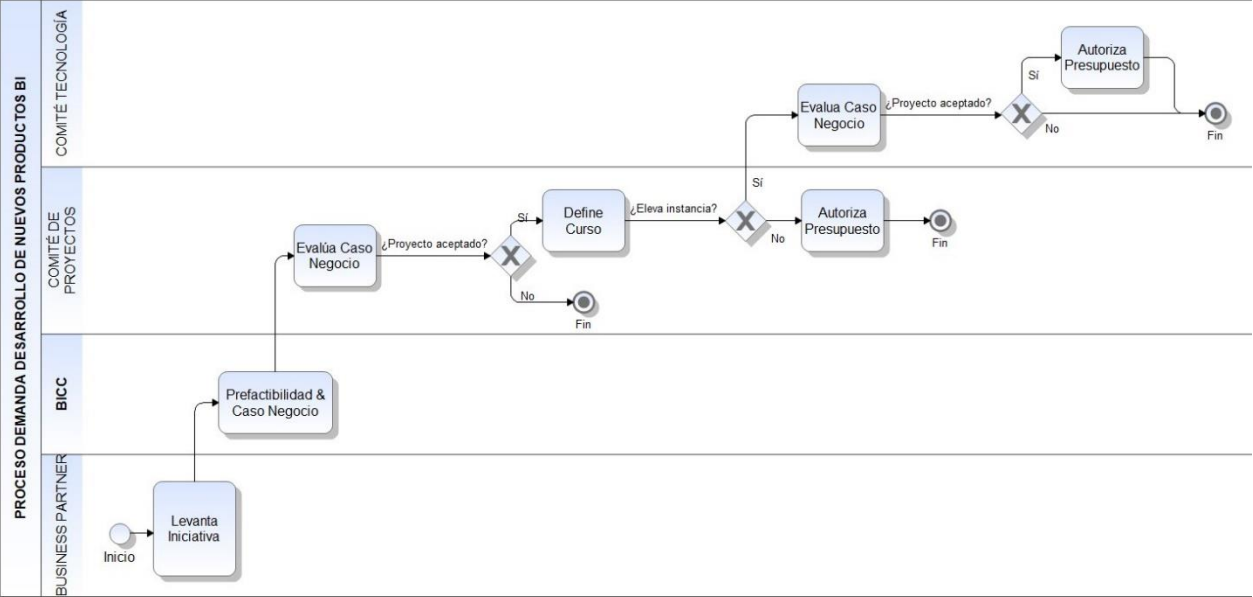


Ilustración 30: Proceso de Demanda de Desarrollo de Nuevos Productos BI

Se entenderá por “Habilitación de Ambientes y Ejecución de Productos”, típicamente a las solicitudes de uso de productos ya existentes o explotación que requiera intervenciones significativamente menores a los de un nuevo proyecto, en la organización se les conoce como “Evolutivos”. Por ejemplo: acceso a fuentes de datos, acceso o suscripción sistemas de reportería, acceso a modelos analíticos, también todo lo anterior con modificaciones que no requieran llamar a instancias superiores de financiamiento, de acceso a especialistas como PMO, Jefatura de Proyectos, etc.

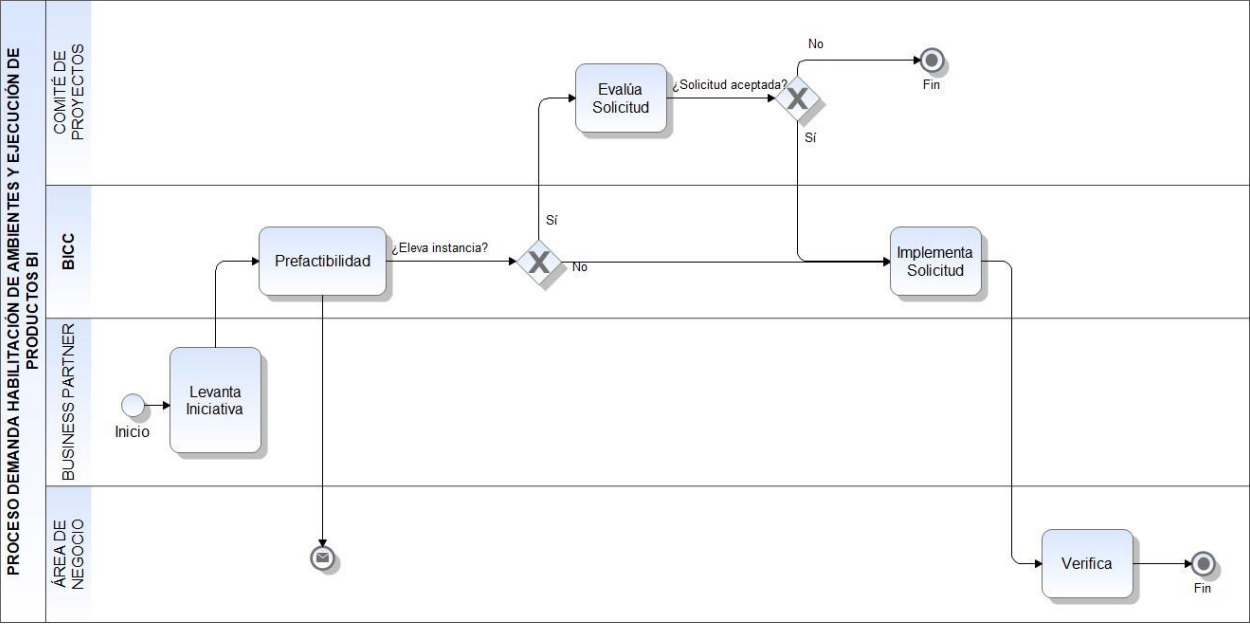


Ilustración 31: Proceso de Demanda Habilitación de Ambientes y Ejecución de Productos BI

4.4.1.3.2 *Proceso de Descubrimiento y Priorización de Demanda*

Siguiendo bajo el contexto de “Procesos”, esta sección propone la incorporación de una metodología para el Descubrimiento (o levantamiento) y Priorización de oportunidades de negocios ligadas a la analítica de datos y alineadas al plan estratégico de habilitar a la compañía con características “Data Driven”.

El DataFirst Method es una metodología desarrollada por IBM, agnóstica a la tecnología, que provee una planificación para lograr un negocio dirigido por datos o “Data Driven”. Permite entender de mejor manera la situación actual de la empresa y cuáles son los siguientes pasos para dirigir, los cuales se dividen en un roadmap de iniciativas analíticas ejecutables, cuantificables y que presenten una oportunidad de negocio a corto, mediano y largo plazo.

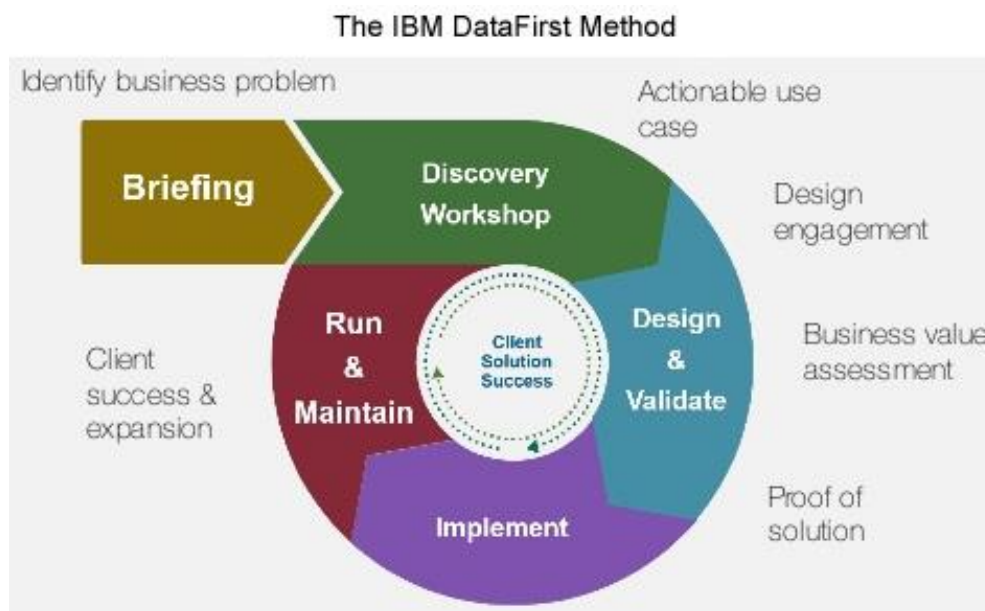


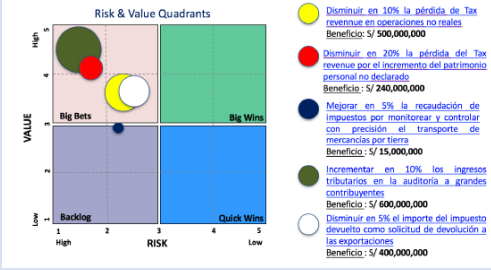
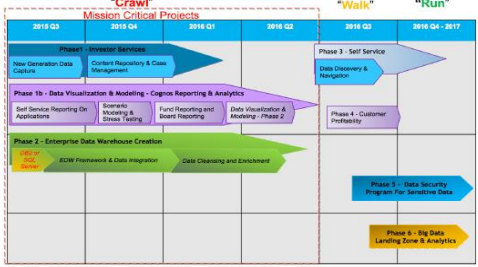
Ilustración 32: DataFirst Method de IBM, como metodología para la Priorización

A grandes rasgos esta metodología funciona de la siguiente manera:

- Bajo la actividad de “**Briefing**”, se explica la visión y la metodología
- A través de una actividad “**Discovery Workshop**”, se realiza una evaluación para conocer el estado de madurez para determinar qué tan cerca/lejos se está de ser “data driven”, junto con ello se realiza un proceso de “descubrimiento” de casos de usos accionables. Algunas de las actividades que forman parte del Discovery Workshop se encuentran se presentan a continuación:
 - Definir Metas y Problemas: para poder definir las cualidades de la solución
 - Mapeo de Infraestructura: para conocer capacidades tecnológicas instaladas
 - Identificar grupos y usuarios: para poder identificar quienes son los usuarios claves
 - Crear Casos de Uso: para contextualizar los problemas identificados
 - Priorizar Necesidades: para separar los problemas importantes

- Identificar las Brechas y Problemas: para encontrar nuevos problemas y buscar soluciones
- Definir MVP: Como prueba de validación

Los entregables de esta fase son:

Entregable	Captura de Ejemplo
<p>Casos de uso detectados, detallados y su Priorización</p>	 <p>Risk & Value Quadrants</p> <ul style="list-style-type: none"> ● Disminuir en 10% la pérdida de Tax revenue en operaciones no reales Beneficio: \$/ 500,000,000 ● Disminuir en 20% la pérdida del Tax revenue por el incremento del patrimonio personal no declarado Beneficio: \$/ 240,000,000 ● Mejorar en 5% la recaudación de impuestos por monitorear y controlar con precisión el transporte de mercancías por tierra Beneficio: \$/ 15,000,000 ● Incrementar en 10% los ingresos tributarios en la auditoría a grandes contribuyentes Beneficio: \$/ 600,000,000 ● Disminuir en 5% el importe del impuesto devuelto como solicitud de devolución a las exportaciones Beneficio: \$/ 400,000,000
<p>Estado de madurez de la organización</p>	 <p>Maturity</p> <p>Operations BI & Data Warehousing Self-Service Analytics New Business Models</p> <p>Cost Reduction Modernization Insight-Driven Transformation</p> <p>current state desired state</p> <p>Most Are Here</p>
<p>Roadmap u Hoja de Ruta de posibles proyectos</p>	 <p>"Crawl" Mission Critical Projects "Walk" "Run"</p> <p>2018 Q3 2018 Q4 2019 Q1 2019 Q2 2019 Q3 2019 Q4 - 2017</p> <p>Phase 1 - Internal Analytics</p> <p>Phase 1b - Data Visualization & Modeling - Cognos Reporting & Analytics</p> <p>Phase 2 - Enterprise Data Warehouse Creation</p> <p>Phase 3 - Self Service Data Discovery & Hierarchy</p> <p>Phase 4 - Customer Profitability</p> <p>Phase 5 - Data Security Program For Sensitive Data</p> <p>Phase 6 - Big Data Lending Tools & Analytics</p>

- Luego, en el marco la actividad **“Design & Validate”** se realiza un plan de priorización para que a partir de su estado inicial la empresa pueda lograr su objetivo de transformación.
- Las actividades de **“Implement”** y **“Run & Maintain”** corresponden a la eventualidad que las iniciativas entren a proceso de desarrollo.

Esta metodología bien podría repetirse anualmente con motivo de cubrir las oportunidades de mejora actualizadas a los objetivos estratégicos corporativos.

4.4.1.3.3 **Procesos Analíticos**

Para esta sección, no es otra cosa que seguir el **“Estándar CRISP-DM”** de tratamiento de datos para la conformación de modelos analíticos.

4.4.1.4 Pilar de Arquitectura y Tecnología

Este segundo pilar presentado en el framework, se enfoca principalmente en describir las plataformas tecnológicas y la arquitectura bajo la cual se atenderá la demanda del BI, a continuación, se describen ambos aspectos

4.4.1.4.1 *Plataforma tecnológica*

Luego de la revisión del análisis realizado en la sección de “Benchmark de plataformas de mercado para el BI”, a lo descrito en sección de “Etapas de elaboración del proyecto” y a los conceptos expresados en “Arquitectura por Configuración (Hybrid Cloud)”, sumado a docenas de horas de discusiones técnicas, el diseño define como propuesta tecnológica el uso de:

- **Solución Public Cloud:**
 - **Microsoft Azure** como solución de explotación de datos en Cloud. Proporciona una gama de servicios bajo el tipo: Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS) y Infraestructure-as-a-Service (IaaS). Se incluyen productos para cómputo (máquinas virtuales, contenedores, etc.), almacenamiento de datos, analítica, redes, IoT, seguridad, desarrollo web y móvil, etc.
 - *Necesario para habilitar al negocio con soluciones empaquetadas y simplificadas en complejidad*
 - *Necesario para acceder a un amplio y “económico” mercado de consultoría en desarrollo*
 - *Para cubrir herramientas y servicios no habilitados en solución local*
 - *Elasticidad y abstracción a complejidades de mantenimiento que permitirán ir paulatinamente adoptando su uso (véase dimensión de <<Alojamiento>> en sección de “RoadMap de Arquitectura y Tecnología”*
 - *Uso de servicios de almacenamiento para “data fría” o de respaldo*
- **Solución Dedicated Cloud:**
 - **SAP BW (en HEC)** como solución “Enterprise Data Warehouse” (EDW) para la convergencia de datos para la gestión, KPIs corporativos, reportabilidad integrada, etc. Tiene la obligación particular de ser del tipo “Full Use”, que corresponde a una condición comercial que, entre otros, permite que los “datos SAP” puedan ser usados fuera del ecosistema de esta marca (por ejemplo, llevarlos a otra nube) o bien ser consumidos por herramientas externas (por ejemplo, soluciones de visualización o de analítica)
 - *Necesario para salvaguardar el linaje, robustez y gobierno de los datos financiero contable de la compañía*
- **Solución On-premises:**
 - **Cloudera** como solución de explotación de datos on-premises y conformación del “Data Lake” (DL) corporativo. Esta solución corresponde a un empaquetado comercial del ecosistema de herramientas Apache

Hadoop, el cual se especializa para trabajos con grandes volúmenes de datos (Big Data)

- *Necesario para las plataformas operaciones Pi System, SCADA y PrimeRead de la compañía*
 - *Solución ya se encontraba en etapa de prueba y validación satisfactoria*
 - *Rentabilización de infraestructura existente y con capacidad ociosa. La compañía posee 2 Data Center de alto estándar.*
 - *Adquisición de aprendizaje técnico y monetizado de las implementaciones.*
 - *La mayor volumetría de los datos corporativos se encuentra en ambientes locales*
- **Soluciones de reportabilidad:**
 - **Power BI y QlikView** como soluciones de reportabilidad self-service. Proporcionan un servicio de análisis de datos presentarlos a través de paneles
 - *Necesario para habilitar a las áreas del negocio con herramientas de autoatención para la presentación de datos. BI Tradicional*
 - *Corresponden a productos de la línea corporativa*
 - *Herramientas ya han proliferado por distintos departamentos de la compañía.*

Estas plataformas tecnológicas en conjunto forman la base para atender la demanda no cubierta por la situación actual.

4.4.1.4.2 *Arquitectura Tecnológica*

Bajo esta sección se presenta una propuesta de arquitectura destinada al almacenamiento, movimiento y recuperación de datos basada en las arquitecturas de referencia del marco teórico. Dos perspectivas para explicar el mismo fin:

- **Perspectiva agnóstica**

Con la siguiente perspectiva de arquitectura propuesta, se indican principalmente las capas y flujos de datos concernientes a usos habituales en los BI modernos, esto con independencia a las herramientas.

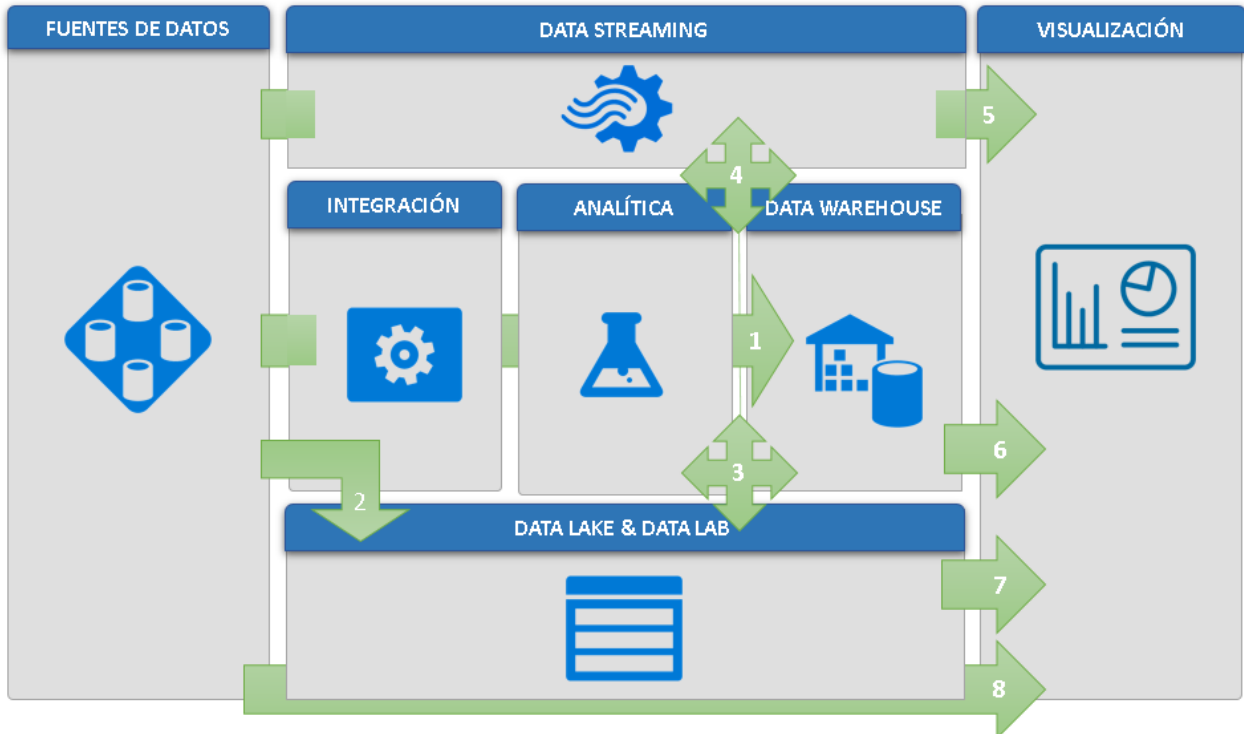


Ilustración 33: Arquitectura BI propuesta – Perspectiva agnóstica

Este diseño de arquitectura recoge y satisface las arquitecturas de referencia planteadas en el Marco Teórico, contando con:

- Diseño híbrido que permite equilibrar las demandas de tiempo de respuesta, costos y disponibilidad
- Capacidades suficientes para tratar la analítica de datos y el manejo de grandes volúmenes de información
- Tratamiento de datos en línea y recolección programada

A continuación, una descripción por flujo:

[1] **ETL OLAP**: caso de datos de recolección OLAP para centralizarse en el Warehouse. Ejemplo: cubo de MM (Gestión de Materiales)

[2] **ETL Convencional**: caso de datos de recolección para centralizarse en el Data Lake.

[3] y [4] **Analítica y Centralización**: caso de datos para aplicarse algún tipo de analítica y/o concentrarse en fuente centralizada

[5] **Visualización desde el STREAM**: caso de datos en línea. Para presentaciones de monitoreo de señales

[6] **Visualización desde el EDW**: caso de datos para gestión. Para presentaciones de tableros de indicadores corporativos

[7] **Visualización desde el DATA LAKE**: caso de datos crudos y/o volumétricos. Para presentaciones exploratorias. Ejemplo: Datos de redes sociales (Facebook y Twitter), resultados de escenarios de proyección de largo plazo (PLP)

[8] **Visualización desde el ORIGEN**: caso de datos directo del origen. Para presentaciones donde la integración es nula o baja. Requiere precaución de no afectar la transaccionalidad. Ejemplo: Datos de una fuente operacional (SQL)

- **Perspectiva tecnológica**

Con la siguiente perspectiva de arquitectura de BI propuesta, se indican las “herramientas convocadas” para apoyar los procesos de tratamiento de datos en la compañía. En similitud con la “*Ilustración 25: Actual Arquitectura de BI*”, ahora se refiere a una situación con proyecto que contempla las tecnologías habilitadoras

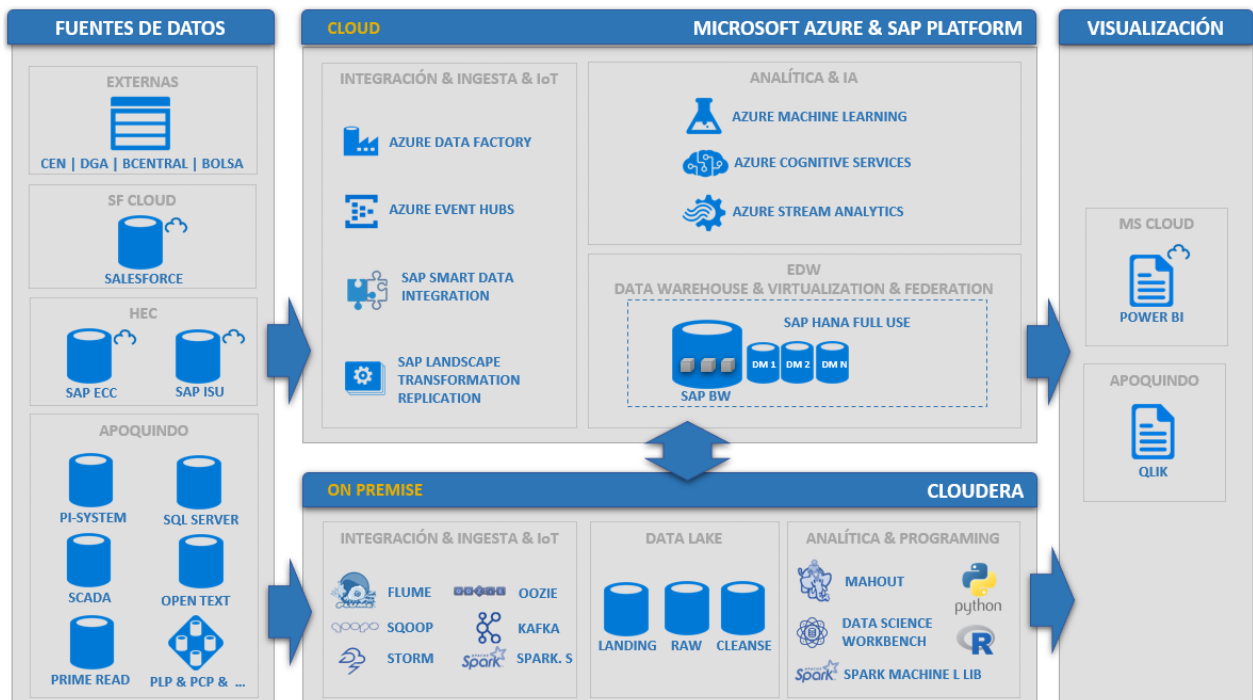


Ilustración 34: Arquitectura BI propuesta – Perspectiva General

4.4.1.5 Pilar de Personas

Corresponde al tercer pilar presentado en el framework. A continuación, se presenta una Estructura Organizacional para el Gobierno del BI (capturando la ideología del BICC) y las funciones de los Centros Organizativos.

4.4.1.5.1 Estructura Organizacional

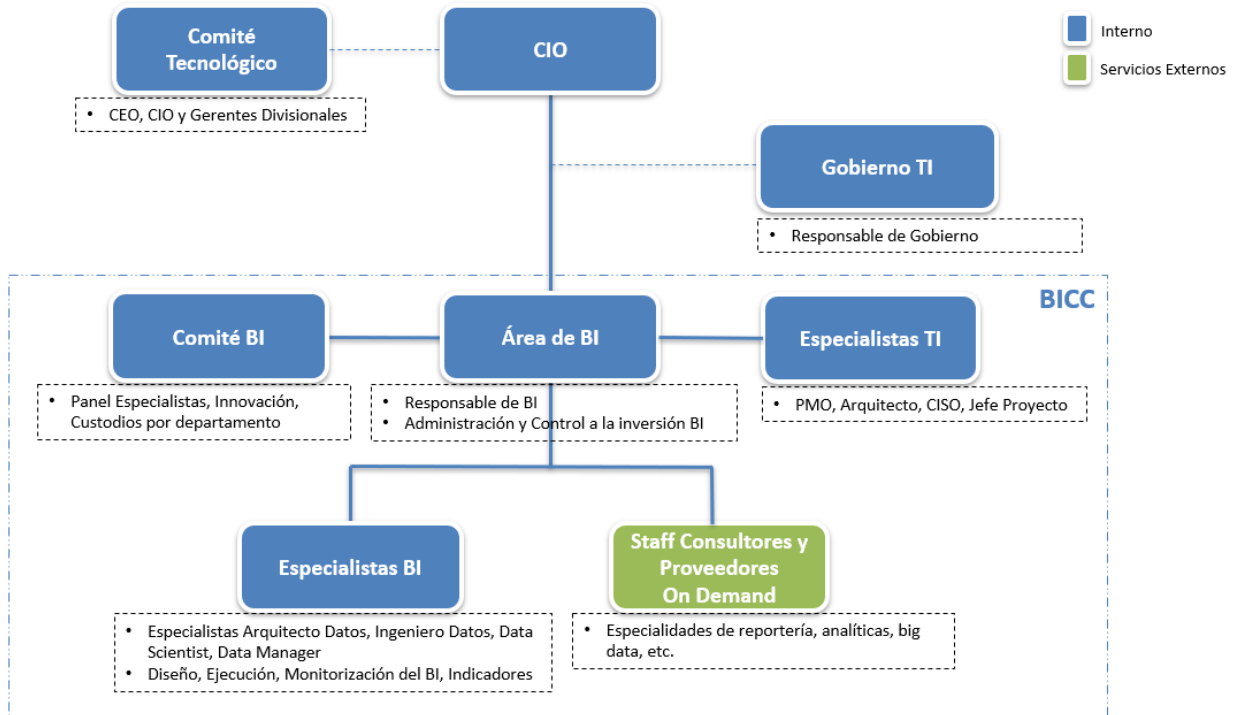


Ilustración 35: Estructura Organizacional propuesta

De acuerdo con lo presentado en la sección de “Organización y Personas”, esta estructura organizacional propuesta corresponde al modelo de “BI como una unidad de TI”, pero también tiene connotación del modelo “BI Virtual Compartido por Departamentos” ya que se reconocen elementos importantes dentro de la organización que cuentan con skills de interés a la causa y/o han llevado adelante iniciativas que, si bien han sido departamentales, estas han aportado valor a la organización

4.4.1.5.1 Centros Organizativos

Bajo este equipo organizativo que promueve el valor y el potencial de BI en la organización, se propone cubrir diversas funciones bajo 3 centros organizativos:

- **Centro Mantenición (Operacional):** encargado de la mantención de estado
 - Atención solicitudes e incidentes
 - Base de conocimiento y procedimientos
 - Monitoreo e indicadores de desempeño

- **Centro Desarrollo (Táctico):** encargado del desarrollo técnico
 - Diseño y Prototipados
 - Patrones de diseño ETL y Visualización
 - Modelado de información
 - Seguimiento al Desarrollo
- **Centro de Excelencia (Estratégico):** encargado del desarrollo estratégico
 - Gestión de la Demanda
 - Prepara Caso de Negocio para presentar a Comité de BI
 - Prepara Propuesta de Proyecto para presentar a Comité de Proyectos
 - Gestión del Desarrollo
 - Gestión del Data Management y Gobierno
 - Gestión de la capacitación y entrenamiento

4.4.1.6 **Data Management**

El proceso de implementación de un programa de “Data Management” es un proceso integral y profundo en la organización, las 11 dimensiones que plantea el marco teórico cubren diversos aspectos técnicos, procedimentales y estratégicos que permiten crear la capacidad para gestionar el conocimiento que hay sobre la información.

4.4.1.6.1 **Data Quality Management**

Para los alcances de este proyecto investigativo y bajo el entendido que la apertura a todos los aspectos que releva esta disciplina requeriría por si solo una tesis equivalente a esta, sólo se presentará un mayor detalle para la dimensión de “**Gestión de la Calidad de Datos (Data Quality Management)**”. El objetivo de la dimensión anteriormente comentada es ser “un proceso continuo para definir los parámetros para especificar los niveles aceptables de calidad de datos para satisfacer las necesidades del negocio y de garantizar que la calidad de datos cumple con estos niveles” (DMBok, 2017). Implica diversos procesos tales como: analizar la calidad de los datos, identificación de anomalías, definición de reglas de calidad, procesos de revisión, etc.

La siguiente ilustración representa un diagrama formal de contexto de Gestión de la Calidad de Datos:

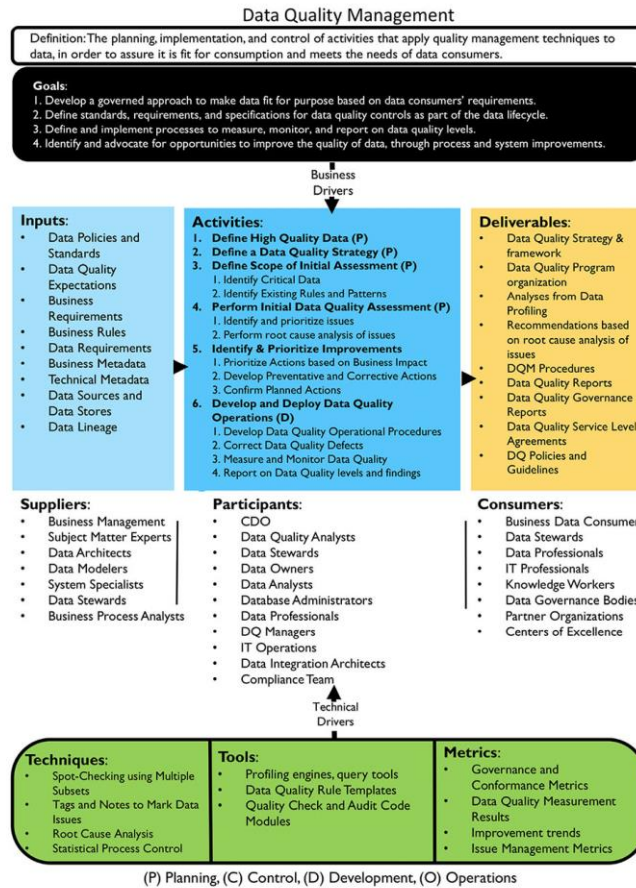


Figure 91 Context Diagram: Data Quality

Ilustración 36: Diagrama de Gestión de Calidad de Datos - DAMA

En particular y para efectos de establecer un nexo con la realidad corporativa a través del framework presentado, se entrega a continuación de lo que sería una definición para la gestión de calidad de una variable crítica dentro de la organización: **Generación Bruta (MWh)**, variable que es utilizada en diversas instancias de reportabilidad e indicadores, lo anterior bajo la metodología DAMA

→	#	ACTIVIDADES	ACCIÓN COLBÚN	→
ENTRADAS				ENTREGAS PRIMARIAS
<ul style="list-style-type: none"> • Fuentes de datos • Expectativas de calidad de datos 	1	Desarrollar y Promover Conciencia de Calidad de Datos (O)	<p><i>Presentación Modelo Gestión Datos</i></p> <ul style="list-style-type: none"> • Identificar Roles y responsabilidades. • Relacionar la calidad de los datos con el desempeño personal. 	<ul style="list-style-type: none"> • Mejora en la calidad de los datos • Gestión del dato • Acuerdo de niveles de servicio de calidad del dato
	2	Definir los requerimientos de calidad de datos (D)	<i>Documento de requerimiento de calidad de datos</i>	
	3	Perfilar, Analizar y Evaluar la calidad de datos (D)	<p><i>Someter conjunto de datos a un análisis empírico.</i></p> <p><i>Evaluar con experto anomalías encontradas y el impacto en el negocio.</i></p>	

	4	Definir las métricas de calidad de datos (P)	<i>A partir de los impactos identificados, se debe enumerar los requerimientos de calidad de datos, con el fin de asociar las reglas del negocio que le serán aplicados y sus procesos para medir la conformidad.</i>	
	5	Definir Reglas del Negocio para calidad del dato (P)	<i>Definir reglas por (MÓDULO VALIDACIÓN DE DATOS SICOP): Rango de cumplimiento: Máximas tolerables. Correlación: Generación con horas de operación.</i>	
	6	Probar y validar los requerimientos de calidad de datos (D)	<i>Aplicar y revisar las reglas definidas.</i>	
	7	Establecer y monitorear los niveles de servicio para la calidad de datos (P)	<i>Definir ANS considerando:</i> <ul style="list-style-type: none"> • Elementos cubiertos • Umbrales de aceptabilidad • Expectativas de calidad. • Métodos de medición • Usuarios involucrados 	
	8	Medir y monitorear Continuatamente la calidad de datos (C)	<i>Utilizar el módulo de validación de datos para medir la calidad del dato, considerando la granularidad "conjunto de datos".</i>	
	9	Administrar problemas de Calidad de Datos (C)	<i>Notificación y seguimiento de los incidentes asociados a la calidad del dato (SERVICE NOW)</i>	
	10	Corregir y limpiar los defectos de calidad de datos (O)	<ul style="list-style-type: none"> • Eliminación de causa raíz. • Corrección de datos de forma automatizada (PROCEDIMIENTOS ALMACENADOS). 	
	11	Diseñar e implementar Procedimientos Operacionales DQM (D)	<i>El equipo de Operaciones de datos debe diseñar procedimientos de dejar operando las actividades de:</i> <ul style="list-style-type: none"> • Inspección y vigilancia • Diagnóstico y evaluación de alternativas de remediación • Resolución del problema • Presentación de Informes (POWER BI) 	
	12	Monitorear Procedimientos y Rendimientos Operacionales DQM	<i>Monitorizar el seguimiento de los incidentes, velando por el cumplimiento de los ANS</i>	
PROVEEDORES	PARTICIPANTES		HERRAMIENTAS	CONSUMIDORES
	<ul style="list-style-type: none"> • Analista de calidad del dato • Expertos áreas temáticas • Administradores de datos 		<ul style="list-style-type: none"> • Módulo de Validación de Datos SICOP • Service Now • Power BI 	
				MÉTRICOS

4.4.1.7 RoadMap de Madurez

Como componente transversal, el framework contempla un roadmap de madurez el cual proyecta los estados por los que la organización debería transitar en los próximos tres años. Este proceso de transformación ha sido categorizado en los siguientes estados:

- **Data Collector:** Recolectora de datos, con foco en la integración y combinación de las fuentes de datos, junto con el comienzo de conocer el concepto del BI
- **Data Aware:** Consciente de los datos, es decir, la organización conocerá las potencialidades y limitaciones de la explotación del dato. La culturización de la analítica ya debería estar instalada
- **Data Driven:** Impulsada por datos, significa que la organización toma decisiones estratégicas basadas en análisis de datos

4.4.1.7.1 RoadMap de Gobierno y Cultura

A continuación, el roadMap evolutivo en términos de Gobierno y Cultura

Dimensión	Fragmentado [2018]	Data Collector [2019]	Data Adware [2020]	Data Driven [2021] →
Gobierno de datos	<ul style="list-style-type: none"> • Discusión en torno al dato (propietario, público-privado, validez, existencia) • Departamental y no formalizado 	<ul style="list-style-type: none"> • Comienzan a surgir las políticas y estándares comunes, KPI administrado centralmente y administración de seguridad • Comienza a surgir la Democratización y Alfabetización de la información 	<ul style="list-style-type: none"> • Visión unificadora (con patrocinio de alto nivel) del programa de Gestión de Datos en toda la empresa • Agenda y prioridades acordadas, normalización de datos e inicio del cambio del sistema de origen 	<ul style="list-style-type: none"> • La Gestión de Datos como un activo. Continuando para aprender y mejorar. Es la forma en que se hacen las cosas • Estructura organizacional de gobierno
Cultura y Alfabetización	<ul style="list-style-type: none"> • Valor de los datos subestimados y decisiones "suficientemente buenas" • No hay visión. Se lucha contra "incendios". Iniciativas aisladas, normalmente construidas de abajo hacia arriba. 	<ul style="list-style-type: none"> • Los tomadores de decisión inciden cada vez más en la decisión basada en datos 	<ul style="list-style-type: none"> • Capacidad de madurez que identifica oportunidades, capacidades y limitaciones de BI 	<ul style="list-style-type: none"> • Modelo de Negocio basado en datos • Propietarios & administradores de datos

Tabla 5: Roadmap de madurez, **Gobierno y Cultura**

4.4.1.7.2 RoadMap de Arquitectura y Tecnología

A continuación, el roadMap evolutivo bajo aspectos tecnológicos y de arquitectura

Dimensión	Fragmentado [2018]	Data Collector [2019]	Data Adware [2020]	Data Driven [2021] →
Arquitectura BI	<ul style="list-style-type: none"> Sin definición 	<ul style="list-style-type: none"> Montaje de Arquitectura empresarial Definición de Arquitecturas de Explotación (Reportaría, Analítica, IA, IoT, HPC) 	<ul style="list-style-type: none"> Arquitectura Core, self-service y gobernada 	<ul style="list-style-type: none"> Arquitectura e Infraestructura optimizada (Bus de Integración, EDW, Data Lake, Capacidad de computo elástico, etc.)
Fuentes de Datos (*1)	<ul style="list-style-type: none"> Procesos ETL establecidos para unas pocas fuentes de datos primarios (estructurados) Silos de Información Departamentales Datos de aplicaciones transaccionales 	<ul style="list-style-type: none"> Procesos ETL establecidos para todas las fuentes de datos primarios (estructurados) Vocabulario común y esquema de inicio 	<ul style="list-style-type: none"> Procesos ETL establecidos para fuentes de datos primarios y secundarias (estructurados y no estructurados) Múltiples modelos de datos corporativos Catálogo de información 	<ul style="list-style-type: none"> Procesos ETL optimizado y establecido para todo tipo de fuentes de datos (estructurados, semiestructurados, no estructurados, redes sociales, sensores) Federado y Virtualizado
Capacidades tecnológicas	<ul style="list-style-type: none"> BI en Diseño Plataformas de alta gama sub-utilizadas y departamentales (sin gobierno corporativo) 	<ul style="list-style-type: none"> BI Descriptivo full (Reportaría) BI Analítico en exploración (modelos matemáticos) Herramientas consolidadas y estandarizadas de gestión de datos 	<ul style="list-style-type: none"> BI Descriptivo self service BI Analítico full Capacidades analíticas ampliadas 	<ul style="list-style-type: none"> Capacidades de focalización especializadas como detección de fraude, análisis cognitivo, etc.
Habilitación tecnológica (*2)	<ul style="list-style-type: none"> Estadística descriptiva 	<ul style="list-style-type: none"> Machine Learning, Deep Learning, IoT Platform 	<ul style="list-style-type: none"> Blockchain, Cognitive Computing, Nanotube Electronics 	<ul style="list-style-type: none"> Virtual Assistants, Connected Home, IA
Casos de Uso	<ul style="list-style-type: none"> Conexión entre fuentes de datos y herramientas de reportabilidad 	<ul style="list-style-type: none"> Reportabilidad Minería de Datos y Minería de Procesos Integración de Sistemas Mantenimiento Predictivo 	<ul style="list-style-type: none"> Mantenimiento Prescriptivo Automatización 	<ul style="list-style-type: none"> Eficiencia Operacional
Análisis	<ul style="list-style-type: none"> Informes e indicadores departamentales 	<ul style="list-style-type: none"> KPIs de la empresa y generación de informes automatizados 	<ul style="list-style-type: none"> Análisis predictivo y prescriptivo, combinando múltiples fuentes de datos 	<ul style="list-style-type: none"> Análisis como servicio
Alojamiento	<ul style="list-style-type: none"> Sólo soluciones transaccionales on-premises 	<ul style="list-style-type: none"> BI Híbrido (Cloud y on-premises) 	<ul style="list-style-type: none"> BI Híbrido en transición (migrando a Cloud de on-premises) 	<ul style="list-style-type: none"> BI productivo Cloud, on-premises como ambiente de QA

Tabla 6: Roadmap de madurez, **Arquitectura, Tecnología y Capacidades**

(*1) De acuerdo con el modelo de madurez presentado en la sección “Naturaleza de los datos”

(*2) De acuerdo con el modelo de madurez presentado en la sección “Tecnologías emergentes”

4.4.1.7.3 RoadMap de Personas y Habilidades

A continuación, el roadMap evolutivo bajo aspectos de las personas y grados de habilidades adquiridas

Dimensión	Fragmentado [2018]	Data Collector [2019]	Data Adware [2020]	Data Driven [2021] →
Equipo de trabajo (*1)	<ul style="list-style-type: none"> Jefes de Proyecto Analistas Informáticos 	<ul style="list-style-type: none"> Jefatura de BI Arquitecto de Datos Ingeniero de Datos 	<ul style="list-style-type: none"> Jefatura de BI & BA Arquitecto de Datos Ingeniero de Datos Data Scientist 	<ul style="list-style-type: none"> Jefatura de BI & BA Arquitecto de Datos Ingeniero de Datos Data Scientist Data Manager
Habilidades	<ul style="list-style-type: none"> Conocimiento SQL, Excel, despliegue de datos 	<ul style="list-style-type: none"> Conocimiento en arquitectura moderna de BI, Análisis de datos estadístico, machine learning 	<ul style="list-style-type: none"> Conocimiento especializado en estadística y análisis de operaciones de negocio 	<ul style="list-style-type: none"> Estructura organizacional de alta competencia en Big Data, Analítica Avanzada y Gobierno de Datos
Grado de Inteligencia (*2)	<ul style="list-style-type: none"> Tipos de análisis: ¿Qué ha pasado?, ¿Cuánto, ¿Cuándo, ¿Dónde?, ¿Dónde está el problema? 	<ul style="list-style-type: none"> Adicional a lo anterior: ¿Qué acciones se requieren?, ¿Por qué ha pasado? 	<ul style="list-style-type: none"> Adicional a lo anterior: ¿Continuará pasando?, ¿Continuará esta tendencia? 	<ul style="list-style-type: none"> Adicional a lo anterior: ¿Qué será lo siguiente que pase? ¿Qué es lo mejor que puede pasar?

Tabla 7: Roadmap de madurez, **Personas y Habilidades**

(*1) De acuerdo con la descripción de responsabilidades presentado en sección de “Roles del equipo BI”

(*2) De acuerdo con el modelo de madurez presentado en la sección de “Progresión de Capacidades Analítica”

4.4.2 PoC de validación

La PoC de validación consiste en automatizar una pequeña parte del “Informe del Directorio” de la compañía, que corresponde a un producto real el cual ha sido declarado como el reporte más relevante de Colbún toda vez que:

- Se usa para explicar mensualmente al directorio el resultado global de la operación del negocio
- Mueve grandes recursos humanos para su elaboración
- Tiene un alto grado de tratamiento manual para obtener los datos de distintas fuentes de información (BD SQLServer, SAP ECC, SAP BW, planillas, etc.), luego preparaciones de datos, limpiezas, consolidaciones y finalmente informe de presentación.

En la siguiente figura, se presentan algunas capturas parciales del informe:

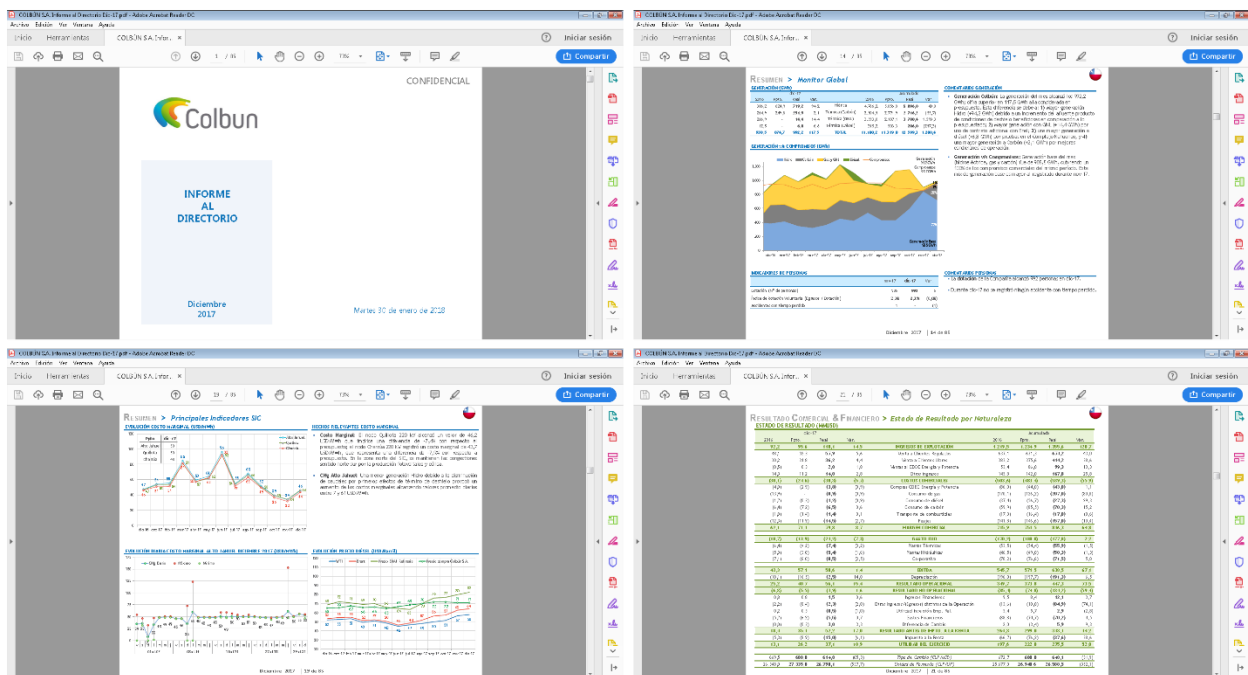


Ilustración 37: Capturas del Caso de Uso Informe al Directorio (Real)

En las siguientes secciones se describirá una revisión general del caso de uso y su implementación en calidad de PoC.

4.4.2.1 Proceso del Caso de Uso y Arquitectura de implementación

Tras un proceso de levantamiento de alto nivel, diagramado en la próxima ilustración, se determinaron las siguientes características de este proceso de reportabilidad:

- Se consolidan análisis de Gasto Fijo, Impuestos, Márgenes, EBITDA, Inversiones y Deuda, Inventario, Operaciones, Medio Ambiente y Cambio Climático, Personas, Seguridad, Comunicaciones, etc.
- Se convocan alrededor de 30 áreas del negocio para su elaboración
- Participan alrededor de 50 especialistas
- Se usan alrededor de 400 HH/mes para su elaboración

- Se usan alrededor de 7 fuentes de datos
- Se tarda 1 semana en completar su elaboración
- El tiempo destinado al análisis no supera las 4 horas

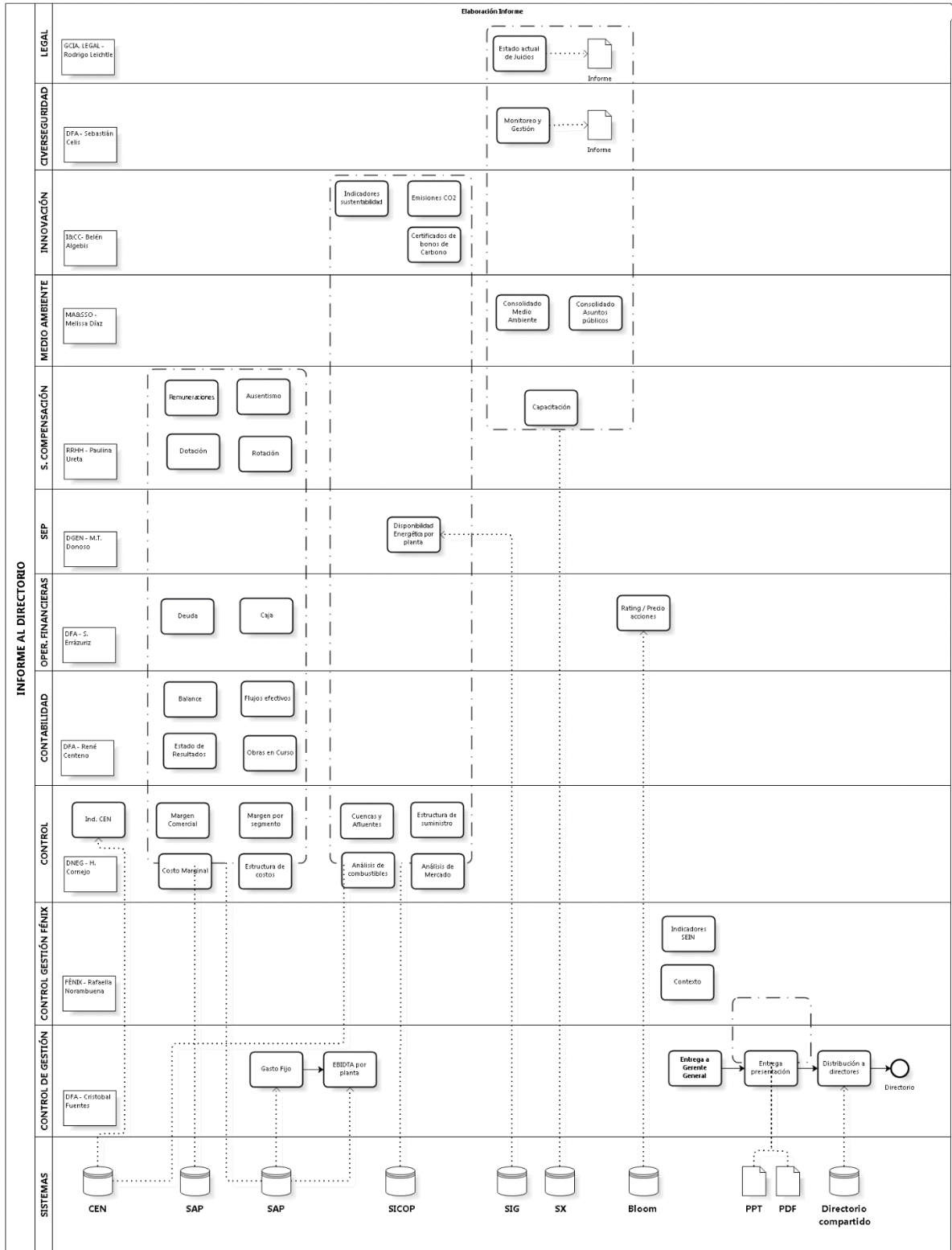


Ilustración 38: Diagrama de Proceso Caso Uso Informe al Directorio

Las primeras 10 pistas (de arriba hacia abajo), corresponden a las áreas encargadas de recopilar la información de otras 20 áreas que no aparecen explícitamente en el diagrama. En tanto, la penúltima pista corresponde área responsable por la consolidación, Control de Gestión, quién a su vez también prepara información para este informe.

Con la anterior información, ya es posible dimensionar que se trata de caso de uso que requiere una intervención para poder acortar los esfuerzos de su elaboración con herramientas que procesen los flujos de datos, desde la integración hasta la presentación, pasando por procesos de limpieza y consolidación de datos.

En base a una de las arquitecturas de referencia presentada en el marco teórico, a continuación, en la siguiente figura se destacan las capacidades que toca este laboratorio experimental.

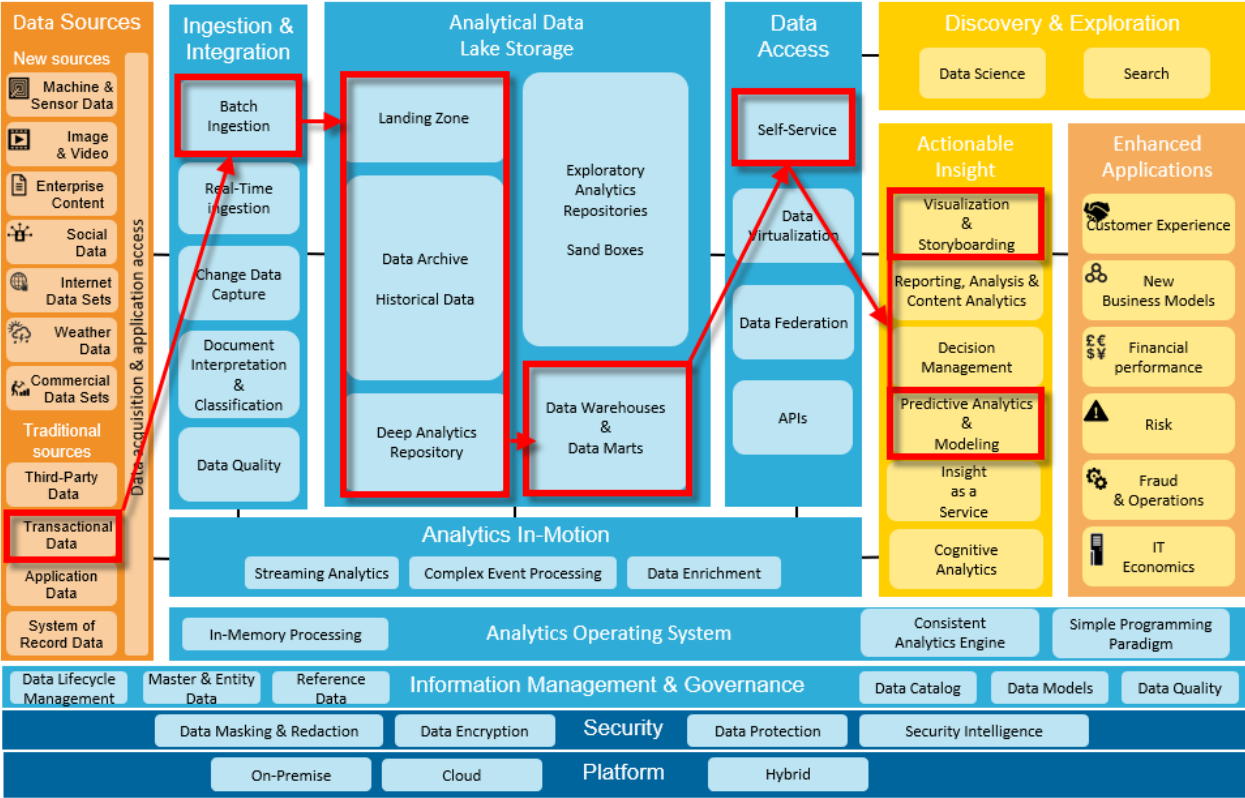


Ilustración 39: Arquitectura de PoC de validación – Uso de Capacidades

En cuanto al uso de herramientas tecnológicas, según plataforma “Hybrid Cloud” definida en la sección de “Plataforma”, es decir, Microsoft Azure como solución Cloud y Cloudera como solución On-premises, a continuación, se presenta una ilustración que describe el tratamiento que tendrán los datos, según prácticas estandarizadas, para ser presentados a los usuarios finales.

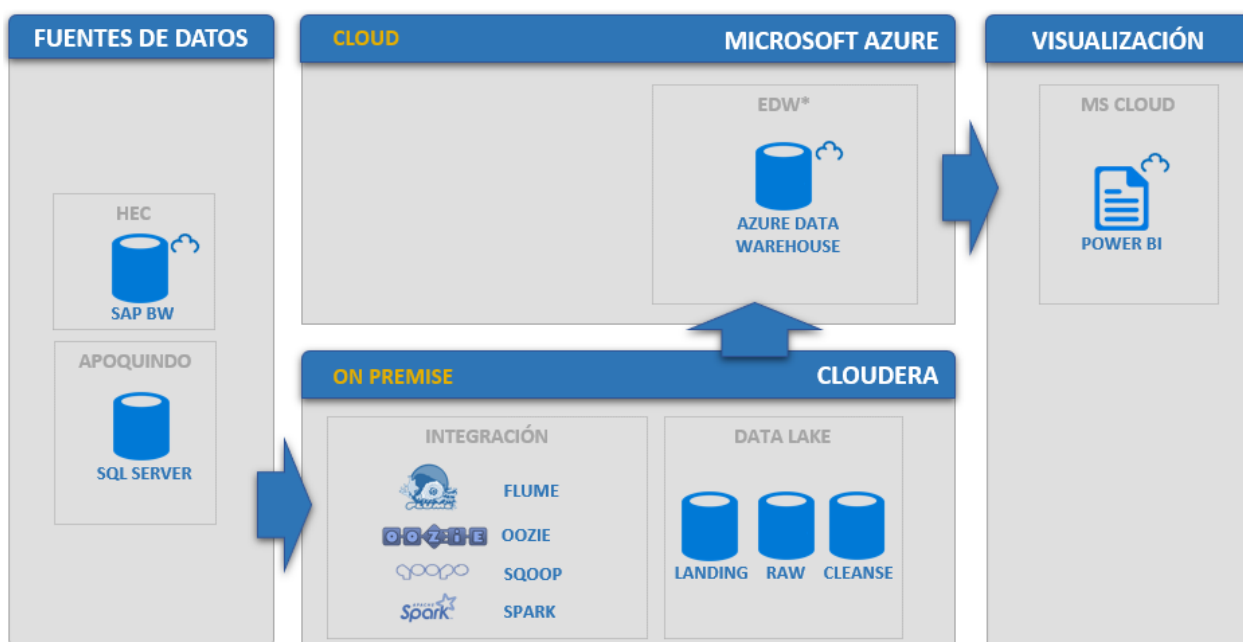


Ilustración 40: Arquitectura de PoC de validación - Herramientas de Implementación

Las herramientas utilizadas se describen en el siguiente relato en función de los procesos de flujo de datos:

Integración

- La integración del tipo “Batch” es resuelta con las herramientas Flume y Sqoop
- El pipeline de extracción es resuelto utilizando Apache Oozie.
- En la “Landing Zone” del Cloudera Data Lake, se almacenan periódica y temporalmente los datos consumidos

Analytical Data Lake Storage

- Se definen 3 zonas de almacenamiento de datos con estructuras cercanas al origen {Landing Zone; Historical Data (Raw); Deep Analytics Repository (Cleanse)}
- Los datos periódicamente extraídos son almacenados en la “Landing Zone”
- Los datos luego son copiados utilizando Spark desde la “Landing Zone” a la “Historical Data”
- Posteriormente, un proceso especializado construido con Spark se encarga de curar los datos y aplicar reglas de negocio para limpieza y validación de datos. Este proceso escribirá los datos ya pre-procesados en la zona de “Deep Analytics Repository” creando tablas en Hive que representan información semántica

Data Warehouse / Data Mart

- Por facilidad en la implementación, la capa “Warehousing” fue dispuesta bajo “Azure Data Warehouse” (EDW*)
- Un proceso construido con Spark toma los datos de Hive y los consolida en un esquema dentro del Data Warehouse. Este Data Mart corresponde a un sub conjunto de datos del Data Warehouse que posee una copia fiel de los datos almacenados en la zona App (Hive) existente en el Cloudera Data Lake.

Visualización

- Finalmente, los datos existentes en el Data Mart son presentados bajo un reporte construido utilizando Power BI.

4.4.2.1 Resultados de la implementación

A continuación, capturas del resultado de la implementación:

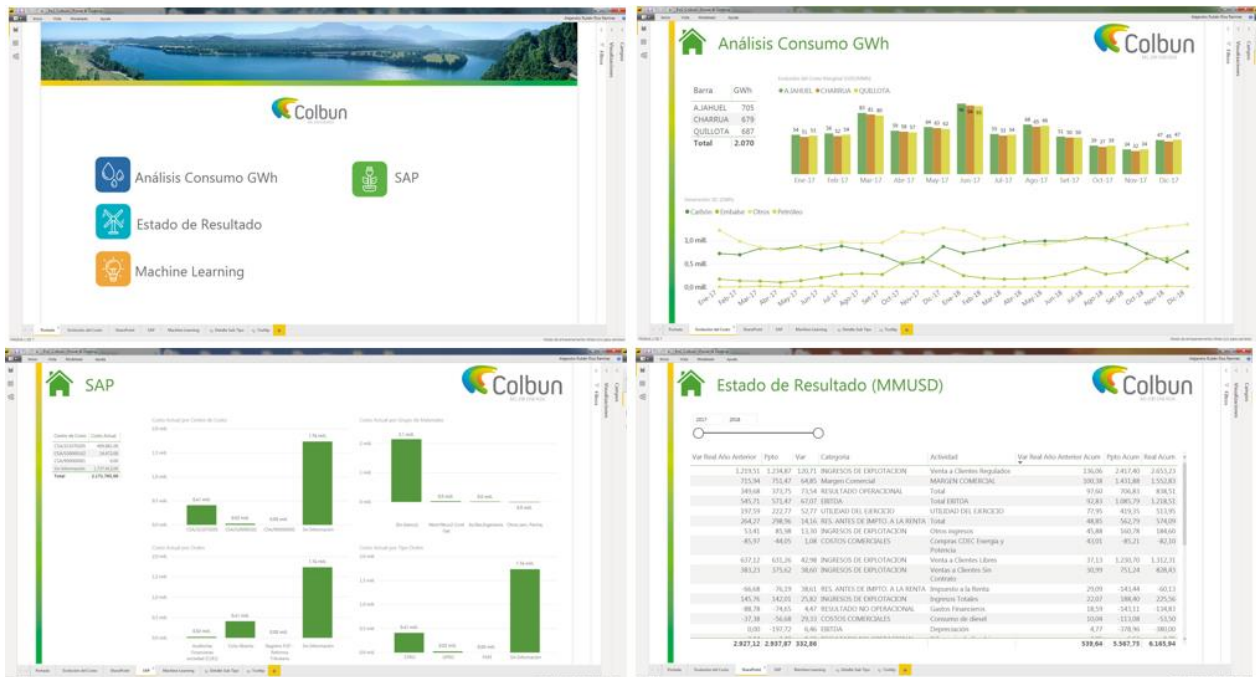


Ilustración 41: Capturas del Caso de Uso Informe al Directorio (PoC)

Como resultado, estas salidas quedan automatizadas para su recolección bajo circuito estándar para recolección de diferentes fuentes, procesamiento y presentación de datos, no requiriendo instancias de entrega de planillas de datos entre áreas. Si bien es cierto las herramientas utilizadas y el proceso definido están sobredimensionados para la PoC, una implementación a escala real de esta y otros casos de uso de explotación de datos sí requerirán realizarse sobre piezas robustas y estándares definidos.

4.5 Evaluación económica

El proyecto tiene como horizonte el montar capacidad técnica y metodológica para atender la demanda de soluciones de inteligencia de negocios, por lo tanto, la iniciativa para ser validada debe ofrecer retorno a la compañía, por lo que a continuación se presentan los aspectos que validan el beneficio de implementar el proyecto.

4.5.1 Beneficios

Los beneficios de la implementación de este proyecto se pueden clasificar en: beneficios cualitativos y cuantitativos:

- **Cualitativos**

Los beneficios cualitativos se encuentran observados del punto de vista de posibles efectos de valor al habilitar la nueva capacidad

- Aumento de la satisfacción de las áreas del negocio por cuanto se genera una estructura que atiende sus necesidades, teniendo a su vez impacto en una mejor percepción de imagen de TI por cuanto toma un rol habilitador. Lo anterior sería posible cuantificarlo a través del uso de actuales instrumentos de medición de servicios que la compañía establece formalmente.
- Facilita una visión integrada y enriquecida para la toma de decisiones por la capacidad de contar con fuentes de datos integradas
- Acelera y sistematiza el proceso de oportunidades de mejora ya que se propone una instancia de “descubrimiento” de necesidades.
- Rentabiliza activos corporativos tecnológicos sub-utilizados haciendo un uso eficiente de recursos
- Fomenta y habilita otros procesos como la Innovación Tecnológica, la Transformación Digital, I+D, etc.

- **Cuantitativos**

Los beneficios cuantitativos se encuentran en directa relación a lo declarado en la sección de “Métricas de Servicio”, esto es capacidad de producir ahorros de costos o ingresos por:

- **BI tradicional o descriptivo**, en donde la automatización integral a los procesos de reportabilidad permiten reducir tiempo de especialistas que hoy dedican altos esfuerzos para la elaboración de reportería, procesos manuales que van desde la descarga de datos desde las fuentes de información, tratamiento de los datos (limpieza, cruce de variables, etc.), consolidaciones y construcción de cuadros de salida, típicamente en presentaciones.
 - En particular, los números que se expresan en flujo de caja, corresponden a una estimación al uso de recursos humanos: ingeniero training (25%), ingeniero especialista (10%) e ingeniero senior (5%), para la construcción de reportabilidad manual, lo anterior para 25 reportes por cada año de evaluación, considerando también sueldos de mercado.
- **BI avanzado o analítico**, en donde existe una alta oportunidad para mejora de procesos, tales como Gestión Comercial, Gestión del Riesgo, Gestión de Operaciones, Gestión Financiera, Gestión Estratégica, etc., a través de la

construcción de modelos matemáticos que permitan un actuar eficiente en las diversas áreas de la organización.

- En particular, los números que se expresan en flujo de caja, corresponden a una estimación de valor esperado por contar con soluciones de modelos predictivos que eviten el 20% (supuesto) de los “costos directos por indisponibilidad de fallas”, que el 2018 marcó 1.58 MUSD (dato área de operaciones).

En la sección de Flujo de Caja se presentará una estimación económica del retorno por la implementación de proyectos de BI, esto según estimaciones de iniciativas por ejecutar.

4.5.2 Desembolsos

Los gastos relacionados a la inversión de implementar este proyecto están principalmente asignados a la renovación y adquisición de equipos computacionales para armar el clúster que permitirá el desarrollo de iniciativas de analítica avanzada y big data, otras líneas de gastos van por los servicios especializados tanto internos como externos con roles como Jefe de Proyecto, Data Architect y Data Engineer, otras especialidades como Data Analyst y Data Scientist no son considerados en esta etapa ya que estos constituyen más valor para la “explotación” que para el “montaje”.

Los gastos relacionados a los costos de operación de continuidad de la capacidad, los cuales están principalmente asignados al licenciamiento de las herramientas y servicios para la mejora continua y mantenimiento de la plataforma.

4.5.3 Horizonte de Evaluación

El horizonte de evaluación es a 3 años, lo cual corresponde a la política de evaluación de proyectos de ámbito tecnológico, mismo periodo para la renovación de hardware por considerarse en obsolescencia.

En cuanto a contratos de prestación de servicios de proveedores, estos tienen revisión cada año.

4.5.4 Tasa de Descuento

La tasa de descuento es una medida financiera que se aplica para determinar el valor actual de un pago futuro y sobre esta base determinar el VAN (valor actual neto). La compañía por política fija este valor al 10% anual para proyectos tecnológicos, por lo tanto, este será el valor a utilizar para este trabajo.

4.5.5 Flujo de Caja

Supuestos y consideraciones

- La inversión en infraestructura podría verse baja en comparación a otras posibles implementaciones en casos similares, pero está explicado porque se hará utilización de hardware ya existente.
- Se declara un concepto de “Tasa ocupación del proyecto”, el cual no hace más que reflejar una línea porcentual de decrecimiento de necesidad de HH de servicios profesionales y necesidades de capacitación. Lo anterior por un concepto de estabilización y madurez competitiva.

Por lo anteriormente mencionado, a continuación, se presenta el flujo de caja del proyecto, estimado en función de los beneficios, desembolsos y supuestos expresados anteriormente

Valores en MUSD	0	1	2	3
Mejora por Operación Extra				
BI Tradicional: Automatización reportabilidad		254	254	254
BI Analítico: Predicción, Minería, Optimización		316	316	316
Costos				
Habilitación infraestructura TI	-150			
HH Servicios externos	-95	-57	-29	-19
Capacitación	-30	-18	-9	-6
Licenciamiento Software Cloudera		-80	-80	-80
Licenciamiento Software Power BI		-50	-50	-50
Mantenimiento infraestructura TI		-8	-8	-8
Resultado operacional (A)	-275	357	394	407
Pérdidas del Ejercicio Anterior		-275	0	0
Resultado no operacional (B)	0	-275	0	0
Utilidad antes de Impuesto (A+B)	-275	81	394	407
Impuesto a la Renta	-58	18	95	102
Utilidad después de Impuesto	-218	63	300	305
Depreciación Activo Fijo Tangible				
Infraestructura servidores		50	50	50
Flujo de Caja Operacional (FCO)	-218	113	350	355
Inversión en Activo Fijo Tangible e Intangible				
Licenciamiento Software Cloudera	-80			
Licenciamiento Software Power BI	-50			
Flujo de Caja Capitales (FCC)	-130	0	0	0
Flujo de Caja Privado Neto (FCO + FCC)	-348	113	350	355
TASA	10%			
VAN	311			

OBSERVACIONES
Ahorro de HH por automatizar reportabilidad
Valor esperado por modelos de analítica avanzada para prevenir fallas
Compra de hardware
Servicios de especialistas
BI, Analítica Avanzada y Big Data
Software para Big Data & Analítica
Software para Reportabilidad
Redes, telecomunicaciones, etc.
La empresa activa este ítem a 3 años

Ilustración 42: Flujo de caja de la propuesta

El VAN resulta ser positivo (311 MUSD), por lo tanto y de acuerdo a los supuestos y consideraciones, el proyecto es rentable.

A considerar, que por alcance no se ha cuantificado la rentabilidad de una buena decisión, ni los errores que pueden contener los procesos manuales, esto por si solo pudiera hacer justificar la inversión.

4.6 Plan de Implementación y Acción

Establecer un plan de acción y definir un proceso de gestión del cambio son fundamentales para introducir debidamente una modificación en la forma de hacer las cosas en cualquier organización. A continuación, un detalle de estos puntos.

4.6.1 Plan de acción

La implementación de esta propuesta de proyecto sugiere los siguientes puntos:

- **Fase 1: Laboratorio (“Test Drive”):** el uso real de las herramientas ofrece la posibilidad de comprobar empíricamente la facilidad de uso, extrapolar costos de implementación a escala, disponibilidad de especialistas, costos de operación, etc. Es por ello que se sugiere una fase donde se tomen algunos casos de uso del negocio y estos sean implementados en calidad de PoC o MVP. Otro beneficio es el poder comprobar el diseño tecnológico y equivocarse a bajo costo.
- **Fase 2: Levantamiento de la demanda:** previo a una fase ya de ejecutar la inversión se sugiere iniciar un proceso de levantamiento de la demanda, como en el sugerido en la sección de “Proceso de Descubrimiento y Priorización” o al menos con los especialistas internos que disponga la organización, esto ayudará a detectar existencia de desacople entre la propuesta y la realidad de los casos de uso, así como también en evitar una infraestructura incorrectamente dimensionada. Otro beneficio es poder contar con una mejor estimación de la evaluación económica ya que se contaría con mayor precisión de las necesidades de implementación.
- **Fase 3: Inversión, montaje y explotación:** habiéndose ejecutado las fases anteriores, se podría estar en condiciones de realizar la respectiva gestión de inversión, con ello planificar el montaje de la infraestructura y posterior explotación.

4.6.2 Gestión del Cambio

Para complementar y reforzar los esfuerzos tecnológicos y de implementación, es de alta importancia asignar esfuerzos a la gestión del cambio, de lo contrario las capacidades que brinda el BI podrían verse limitados o disipados si existe resistencia o si es que no se le considera de valor estratégico.

A continuación, se presentan 10 pasos sugeridos para gestionar correctamente el cambio y la cultura durante la implementación de un proyecto de BI, estos están basados en las sugerencias de Andrew McAfee (co-director del centro de Negocios Digitales de MIT) y John Kotter (experto de la gestión del cambio) bajo el libro de Leading Change.

N°	Paso recomendado	Ejemplo de acción en Colbún
1	Generar conciencia sobre la importancia del Business Intelligence <ul style="list-style-type: none">• Establecer BI como una prioridad estratégica	<ul style="list-style-type: none">• Relevar experiencia de competencia que lleva años ya aplicando estas tecnologías

	<ul style="list-style-type: none"> • Crear un sentido de urgencia con ejemplos de empresas que están generando ventaja competitiva por medio de BI 	
2	Conocer el punto de partida <ul style="list-style-type: none"> • Entender el nivel de madurez de Business Intelligence • Conocer los activos estratégicos 	<ul style="list-style-type: none"> • Levantamiento de fuentes de datos y plataforma tecnológica
3	Definir una visión y alinear al equipo de gerentes de alto nivel <ul style="list-style-type: none"> • La visión debe ser construida en base a los activos estratégicos, tener ambiciones transformativas y tener un objetivo claro • Crea una poderosa alianza estratégica 	<ul style="list-style-type: none"> • Patrocinador ejecutivo: Gerente de TI, Gerente Comercial y Gerente de Operaciones
4	De visión a ejecución <ul style="list-style-type: none"> • Definir metas para medir el éxito • Desarrollar un plan de gobierno corporativo 	<ul style="list-style-type: none"> • Framework contempla un roadmap a 3 años y la implementación de iniciativas de Gobierno
5	Comunicar la importancia y beneficios de las iniciativas <ul style="list-style-type: none"> • Explicar la importancia y los beneficios relevantes para cada individuo • Utilizar todos los canales posibles 	<ul style="list-style-type: none"> • Transmitir los puntos declarados en capítulo de introducción: “Justificación, Relevancia y Beneficios”
6	Ganar el respeto y el derecho a exigir resultados <ul style="list-style-type: none"> • "Walk the talk“, es decir, ser consecuente y demostrar lo que se dice • Éxitos a corto plazo • Identificar y reconocer individuos que estén adoptando exitosamente los cambios 	<ul style="list-style-type: none"> • Implementación de prototipos que validan y demuestran la utilidad del BI
7	Definir comportamientos y cultura <ul style="list-style-type: none"> • Eliminar toda barrera a la adopción • Asegurar información oportuna y confiable por medio de estándares tecnológicos 	<ul style="list-style-type: none"> • Patrocinador ejecutivo busca otros patrocinadores
8	Desarrollar las habilidades esenciales	<ul style="list-style-type: none"> • Indicador de “Desarrollo de conocimiento y

	<ul style="list-style-type: none"> • Desarrollar las habilidades requeridas por medio de centros de aprendizaje o contratación 	habilidades". Sección de " <u>Métricas de Servicio</u> "
9	<p>Alinear incentivos y recompensas</p> <ul style="list-style-type: none"> • Alinear incentivos desde arriba para que efectos se amplifiquen • Establecer recompensas financieras y no financieras 	<ul style="list-style-type: none"> • Implementación de comité que captura grupos de interés en torno al BI
10	<p>Medir todo e iterar</p> <ul style="list-style-type: none"> • Asegurar medir KPIs en diferentes áreas incluyendo financieras, experiencia del cliente, operacionales, capacidades organizacionales, entre otras. • Aprender de los fracasos e iterar 	<ul style="list-style-type: none"> • Definición de "<u>Métricas de Servicio</u>"

Tabla 8: Estrategia de Gestión de Cambio

4.6.3 Criterios de aceptación

Como criterios de aceptación sobre la implementación de este proyecto, se espera a lo menos lo siguiente:

- Eliminación de silos de información, esto es, que existan mecanismos tecnológicos suficientes para extraer o integrar las diversas fuentes de datos corporativas
- Plataforma tecnológica y arquitectura permitan atender la demanda de iniciativas de explotación de los datos, esto es: integraciones, procesamiento de información, analítica y reportabilidad
- Estructura organizacional propuesta o similar con skills ad-hoc de acuerdo con roadmap de madurez

4.7 Riesgos Asociados al Proyecto

Analizando los eventuales sucesos que podrían ocurrir en la implementación de este proyecto, se elaboró una tabla que describe los principales riesgos a los que se podría estar expuesto, el grado de impacto que ello implicaría y la forma sugerida de abordarlos.

Riesgo	Impacto	Mitigación
<p>Desacople entre propuesta y realidad de casos de uso</p> <p><i>El diseño podría quedar subdimensionado o sobredimensionado por no contar con un levantamiento completo de la demanda interna</i></p>	Alto	<p>Levantar formalmente al menos 2 casos de uso candidatos a implementación bajo PoC. Otros casos en tanto serán tema de revisión bajo levantamiento preliminar, es decir, como entendimiento y no necesariamente conducente a resolverse</p> <p>Un mecanismo para disminuir este gap sería ejecutando la actividad descrita en la sección de "<u>Proceso de Descubrimiento y Priorización</u>"</p>
<p>Infraestructura incorrectamente dimensionada</p> <p><i>Las inversiones en este ámbito son altas, y una incorrecta estimación podría afectar la atención de la demanda (caso subdimensionada) o bien, al buen uso de los recursos de la organización (caso sobredimensionada)</i></p>	Alto	<p><i>Misma recomendación que riesgo previo: ejecutar la actividad descrita en la sección de "<u>Proceso de Descubrimiento y Priorización</u>"</i></p>
<p>Silos organizacionales</p> <p><i>Complejidades el montaje de infraestructura tecnológica o incluso el desarrollo de modelos matemáticos, pasan a segundo plano cuando se presentan silos culturales que solo ofrecen barreras de entrada</i></p>	Alto	<p>Contar con un diagnóstico del tipo de silo permitirá una adopción de una estrategia adecuada. Realizar una revisión previa, por si solo es un trabajo que esta fuera del alcance de este proyecto, sin embargo, es un riesgo inherente que no se recomienda dejar de lado.</p> <p><i>"Al eliminar los silos organizacionales, las compañías podrían conseguir la velocidad y la</i></p>

		<i>innovación tecnológica que llegarían en el siglo XXI</i> ⁹
<p>Inadecuada calidad de la información</p> <p><i>El “reflote” o visibilidad de los datos podría traer consigo que su calidad no era lo esperado</i></p>	Alto	La sección que se refiere al “Data Management” consigna una dimensión exclusiva para la calidad de los datos, el cual contempla una serie de actividades para el control y mejoramiento de este ámbito
<p>Incumplimiento a la Seguridad, Privacidad o Normativa</p> <p><i>Los datos transaccionales típicamente están (deberían estar) debidamente resguardados por las plataformas que los gobiernan, pero en el mundo del BI los datos salen de sus ambientes de confort, por lo que se está expuesto a caer en riesgos de este tipo</i></p>	Medio	La sección que se refiere al “ <u>Data Management</u> ” consigna una dimensión exclusiva para la seguridad de los datos, el cual contempla una serie de actividades para el control y resguardo de este ámbito
<p>Inmadurez analítica</p> <p><i>Incorrecta interpretación de patrones o resultados de modelos, causada por la inexperiencia analítica, podría causar sesgos o imprecisiones que afectan el valor del producto</i></p>	Alto	Relevar contar con personal calificado (científicos de datos). Dar énfasis y valor a los procesos de “comprensión” descritos en la sección de “ <u>Estándar CRISP-DM</u> ”
<p>Falta de casos de negocio con retorno significativo.</p> <p><i>El mercado eléctrico se caracteriza por ser altamente regulado y complejo, lo que podría mermar encontrar casos de negocio rentablemente atractivos. También, la poca madurez de la organización en temas analíticos podría mermar la detección de oportunidades de negocio</i></p>	Medio	Inspección de campo con el coordinador (CEN) y otros coordinados, que permitan identificar oportunidades que otros ya han resuelto. Búsqueda de casos de éxito documentados.

Tabla 9: Tabla riesgos, impacto y mitigación

⁹ Jack Welch

5 CAPITULO 5: CONCLUSIONES Y TRABAJOS FUTUROS

5.1 Conclusiones

Un framework como puerta de enlace. Llegar a la conceptualización de esta propuesta fue un largo camino. Por un lado, la organización se encontraba en un estado de madurez que limitaba visualizar tanto las piezas que componen el dominio de la explotación de los datos como su valor resultante, por otro lado, la oferta de mercado especializada en estas temáticas se dedica principalmente a implementaciones de proyectos analíticos. Entonces, este trabajo tuvo un gran aporte para abrir procesos conversacionales tanto al interior de la organización como con actores externos.

Principales problemas. El tiempo para producir entregables al cliente Empresa y al cliente MBE resultó ser un problema en términos de cumplimiento con satisfacción, y es que los entregables para uno no necesariamente son de utilidad para el otro.

Principales facilitadores. Que el proyecto contara con sponsor ejecutivo fue sin duda el facilitador número uno, esto permitió habilitar conversaciones no solo al interior de la organización, sino que con el mercado. Se transformó entonces en un proyecto validado por la organización y al que sin duda se le debía prestar atención por los beneficios que este aportaría.

Implicancias de la implementación. Si bien es cierto se logró consolidar una propuesta integral, en la práctica no todo tiene asegurado su implementación, por ejemplo, temáticas de gobierno de datos aún se encuentra en proceso de “recepción y valorización” de cara al patrocinador ejecutivo, y es que se ve como requisito que previamente se muestren resultados concretos y retornos explícitos antes de apostar a llevarse a cabo nuevas intervenciones estructurales. En definitiva, este diseño y probablemente otro, tiene distintos tiempos de habilitación total o parcial, dependerá entonces del criterio y visión de los tomadores de decisión.

Importancia de los datos. Contar con datos de calidad y procesos que obliguen a la sistematización de datos, en rigor no son parte de la inteligencia de negocios, pero son fundamentales a la hora de abrir líneas de acción respecto de la explotación de los datos ya sea con reportería, construcción de indicadores, analítica avanzada, optimización, etc.

Oportunidades de replicación. El trabajo realizado entregó un marco de referencia para que Colbún iniciara su camino de “Data Driven”. Lo anterior, permite oportunidades de replicación en otras organizaciones que se encuentren en proceso de constituir formalmente una estructura organizacional que se dedique a resolver necesidades de la analítica de datos, sin embargo, se debe precisar que la solución sin duda puede ser

distinta para cada organización, ya que esto depende de su nivel de madurez, su plataforma tecnológica, procesos internos, cultura organizacional, prioridades, visión de valor, etc.

5.2 Trabajos futuros

Piloto que valide el Framework. Si bien es cierto que dentro de la propuesta de solución existe una PoC de validación, por razones de alcance este fue sometido a corroboración sólo desde el punto de vista tecnológico, mas no de los otros aspectos que el framework releva, tales como Métricas, Data Management, Flujo de Aprobación, etc. Entonces, un buen ejercicio futuro sería someter un prototipo a toda la cadena del modelo.

Iniciar proceso de levantamiento de demanda. El presente trabajo representa un diseño teórico-práctico para fijar un marco de referencia para Colbún, pero la implementación a escala corporativa requiere de un proceso serio de levantamiento de demanda, el cual permitirá dimensionar de mejor manera los componentes sindicados en el framework, desde la estructura organizacional requerida, hasta el requerimiento tecnológico, pasando por la evaluación económica que deberá ser sometida a juicio por los tomadores de decisiones de inversión.

Conocer experiencia en la industria. Una forma de enriquecer el diseño a escala real sería consignando entrevistas con otros actores que ya cuentan con la capacidad de plataformas analíticas instalada, esto permite conocer experiencias que podrían acortar procesos de adopción y/o evitar errores propios del proceso de aprendizaje.

Consultoría especializada. Este trabajo no ha pretendido ser el diseño definitivo ni el único camino para lograr transformar la organización en una compañía empoderada de sus datos. Diversos aspectos técnicos, estratégicos y de aceleración pueden ser cubiertos bajo consultoría experimentada que corresponde a la alta gerencia determinar el nivel de especialización e inversión que estén dispuestos a llevar a cabo para una intervención guiada.

6 CAPITULO 6: BIBLIOGRAFÍA

- MICHAEL E. PORTER. 1996. What is Strategy?. Harvard Business Review.
- ROBERT KAPLAN y DAVID NORTON. 2004. Strategy Maps. First Edition. Harvard Business School Press. 324 p.
- MICHAEL E. PORTER. 2008. Las cinco fuerzas competitivas que le dan forma a la estrategia. Harvard Business Review.
- ARNOLD C. HAX. 2003. The Delta Model – A new framework of strategy, Journal of Strategic Management Education1(1). Senate Hall Academic Publishing.
- Barros, O. (2016). Ingeniería de Negocios: Diseño Integrado de Servicios, sus Procesos y Apoyo TI. Santiago, Chile: Business Express Press.
- Gartner. (2018). Big Data Analytics - Gartner Tech Definitions [Glosario "Analytics"]. Recuperado 19 marzo, 2019, de <https://www.gartner.com/it-glossary/analytics>
- Beller, M. J., & Barnett, A. (2009). Next Generation Business Analytics Technology Trends.
- Oxford English Dictionary. (2019); Big Data; Oxford University Press; <http://www.oed.com/view/Entry/18833#eid301162177>
- Gartner. (2019); Big Data; Gartner; <https://www.gartner.com/it-glossary/big-data>
- Cox, M., Ellsworth D. (1997); "Application-controlled demand paging for out-of-core visualization"; 8th IEEE Visualization '97 Conference; https://www.evl.uic.edu/cavern/rg/20040525_renambot/Viz/parallel_volviz/paging_outofcore_viz97.pdf
- Laney D.(2001); 3D Data Management: Controlling Data Volume, Velocity, and Variety; Meta Group; <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
- Earley, S., Henderson, D., & Data Management Association. (2017). DAMA-DMBOK: Data management body of knowledge.
- Marz, N. (2012). Big data lambda architecture. <http://www.databasetube.com/database/big-data-lambda-architecture/>).
- Arquitectura Lambda. Tejada Z.; Buck A.; Wasson M. (2018). Big Data Architectures. <https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/>
- Ouyang, C., & Moura, M. (2017, 4 mayo). Hybrid Cloud Strategy for Big Data and Analytics. Recuperado 19 marzo, 2019, de <https://www.slideshare.net/HadoopSummit/hybrid-cloud-strategy-for-big-data-and-analytics>

Oracle. (2016, marzo). An Enterprise Architect's Guide to Big Data. Recuperado 19 marzo, 2019, de <https://www.oracle.com/technetwork/topics/entarch/articles/oea-big-data-guide-1522052.pdf>

IBM. (2017, marzo). Hybrid cloud for big data and analytics solutions. Recuperado 19 marzo, 2019, de <https://www.ibmbigdatahub.com/whitepaper/hybrid-cloud-big-data-and-analytics-solutions>

Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011). Estudio comparativo de metodologías para minería de datos. In XIII Workshop de Investigadores en Ciencias de la Computación.

Miller, G., Bräutigam, B., & Gerlach, S. (2006). Business Intelligence Competency Centers: A Team Approach to Competitive Advantage. Hoboken: Wiley

Gartner. (2012, 3 diciembre). Business Intelligence Competency Center (BICC) - Gartner IT Glossary [Glosario "BICC"]. Recuperado 19 marzo, 2019, de <https://www.gartner.com/it-glossary/bicc-business-intelligence-competency-center/>

O'Kane, B., & Judah, S. (2015, 5 mayo). [Use the Gartner MDM Maturity Model to Create Your MDM Roadmap]. Recuperado 19 marzo, 2019, de <https://www.gartner.com/doc/3044517/use-gartner-mdm-maturity-model>

O'Kane, B., Judah, S., & Gartner. (2015, 5 mayo). Use the Gartner MDM Maturity Model to Create Your MDM Roadmap. Recuperado 19 marzo, 2019, de <https://www.gartner.com/doc/3044517/use-gartner-mdm-maturity-model>

Gartner; (2018). Hype Cycle for Emerging Technologies; <https://www.gartner.com/en/newsroom/press-releases/2018-08-20-gartner-identifies-five-emerging-technology-trends-that-will-blur-the-lines-between-human-and-machine>

McAfee, A., & Kotter, J. (2012) Leading Change.

Bibliografía complementaria:

Big Data. Zikopoulos P. (2011). Understanding Big Data, Analytics for Enterprise Class Hadoop and Streaming Data. McGraw-Hil.

Data Management. Enfoe M. (2017). Data Management in an Operational Context A study at Volvo Group Trucks Operations (Master Thesis). Lund University School of Economics and Management, Lund, Suecia.

Arquitecturas Referencia Modernas BI. Kosambia, S. (2008). Business Intelligence the Self-Service Way.

Hybrid Cloud. Abramson S., Horka W., Wsniewski L. (2014). A Hybrid Cloud Architecture for a Social Science Research Computing Data Center.

7 CAPITULO 7: ANEXOS

7.1 Estructura Organizacional de Colbún

La estructura organizacional principal de la compañía se presenta en la siguiente ilustración:

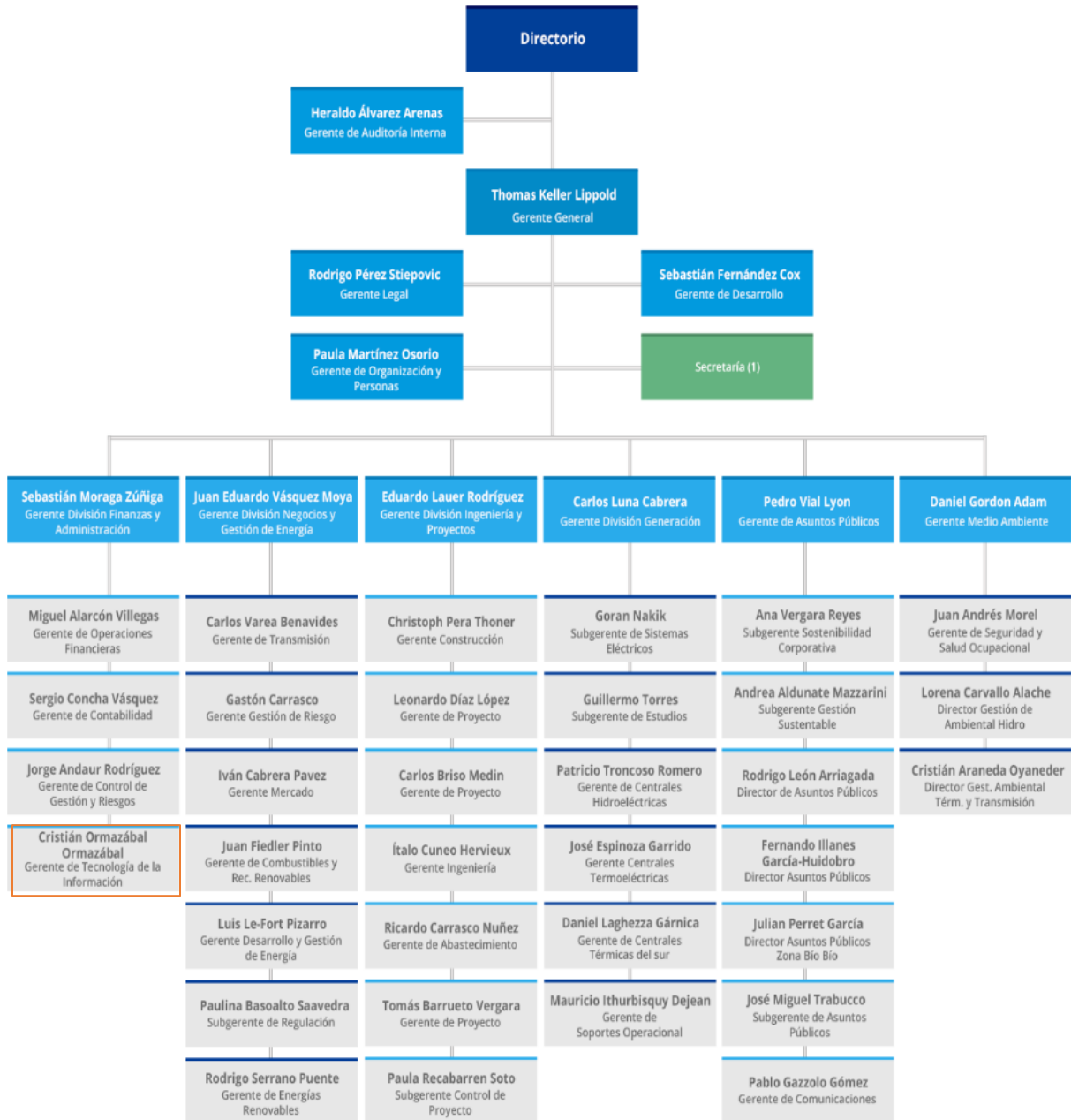


Ilustración 43: Principal Estructura Organizacional de Colbún^v

El presente proyecto de grado es iniciativa directa de la “Gerencia de Tecnología de la Información”, perteneciente a la “División de Finanzas y Administración”. Departamento donde el autor de este proyecto está a cargo de la recientemente creada Área de BI.

7.2 Etapas de elaboración del proyecto

La construcción de esta propuesta de proyecto ha recorrido un largo camino, el cual se declara resumidamente bajo las siguientes etapas:

- **Etapas 1: Interno**
 - Entendimiento de la situación actual
 - Es fundamental iniciar con el conocimiento de la realidad en todas sus dimensiones, en el radar debe estar identificado la cultura organizacional, las plataformas tecnológicas de la compañía, los “dolores” del negocio, lo que buscan los diferentes actores de la cadena de decisión, quienes pueden aportar o incluso dificultar, etc.
 - Entendimiento tecnológico y estado del arte
 - Conocer el estado del arte respecto de con qué y cómo se resuelven los problemas de analítica de datos, otorga visión de campo para identificar las tendencias de quienes ya se encuentran resolviendo problemáticas avanzadas. Para esto fue necesario asistencia a diversos seminarios, workshops, talleres, etc.
 - Identificar los líderes tecnológicos a través de un benchmark permite acotar la búsqueda de herramientas de tratamiento de datos.
 - Conformar una visión interna
 - Antes de sentarse a conversar con los “vendors” es valioso contar con una visión de alto nivel que permita dirigir los esfuerzos, establecer compromisos de responsabilidad y participación, etc.
 - Definir Casos de Uso Piloto priorizados
 - Contar con un levantamiento de problemáticas del negocio es un buen insumo para coronar que se ha completado la fase 1 con un entendimiento acabado de lo que la compañía necesita
- **Etapas 2: Interno y Externo**
 - Inspección tecnológica
 - Convocar a las empresas/especialistas para que presenten su batería de soluciones y propuesta de valor
 - Inspección de campo
 - Saber en qué están las otras industrias, quiénes son los líderes y cuál fue el camino que trazaron
 - Convergencia
 - La iteración de los puntos anteriores puede ser infinita, finalmente se debe llegar al punto en el que la madurez alcanzada dicta que hay que converger, ejercer y analizar resultados

Con lo anterior, no se indica que este sea el camino ideal o recomendado, básicamente se trata del camino que se recorrió para llegar a esta propuesta.

7.3 Benchmark de plataformas de mercado para el BI

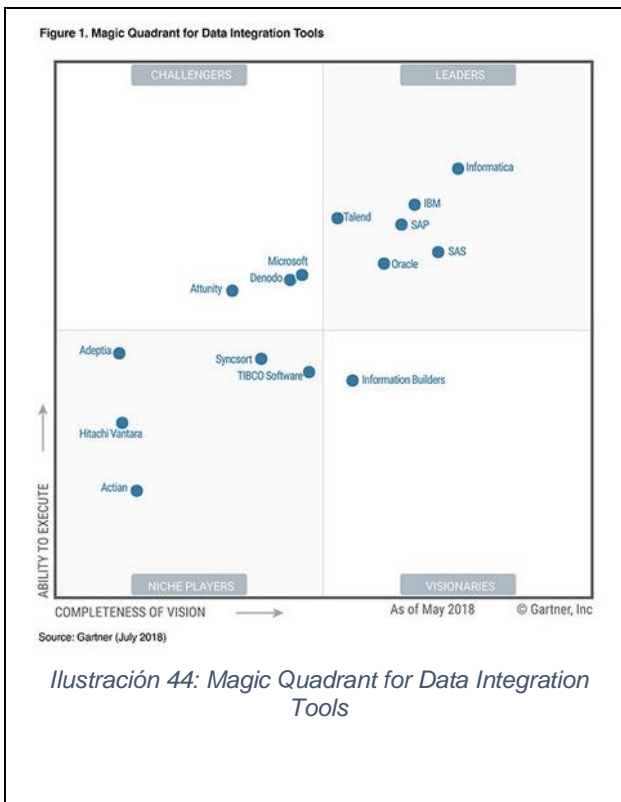
Para comprender este punto es necesario explicar que *Benchmark* en este caso se refiere a la comparación de capacidades de herramientas para distintos tipos de plataformas / compañías.

En el caso de Gartner, este utiliza un gráfico que está formado por dos ejes, el eje X y el eje Y:

- En el **eje X**, define la categoría “integridad de visión” y representa el conocimiento de los proveedores sobre cómo se puede aprovechar el momento actual del mercado para generar valor, tanto para sus clientes como para ellos mismos.
- En el **eje Y** se encuentra la “capacidad de ejecutar”, donde se mide la habilidad de los proveedores para ejecutar con éxito su visión del mercado.

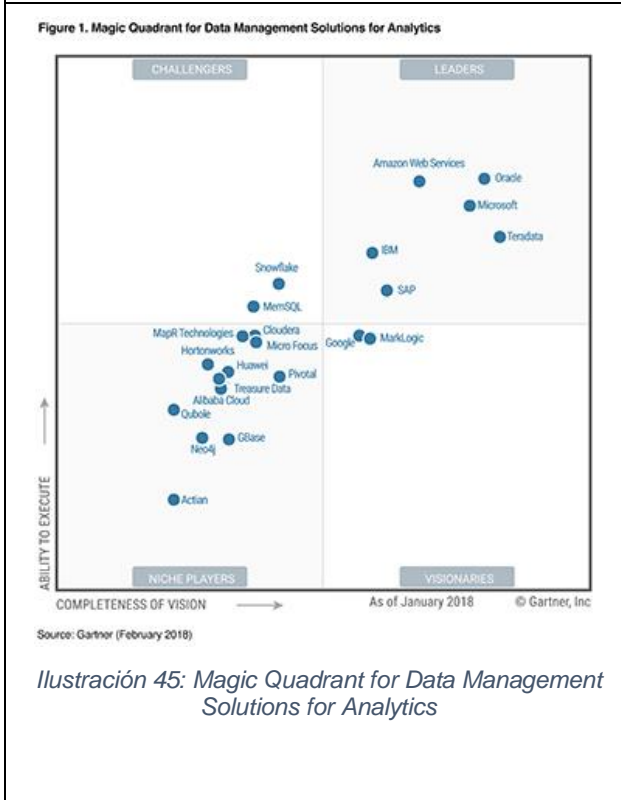
Ambas divisiones fragmentan el cuadrante en cuatro sectores. Ahí es donde se plasman las principales compañías de cada competencia en función de su tipología y la de sus productos: líderes (*leaders*), visionarios (*visionaries*), retadores o aspirantes (*challengers*), y jugadores de nicho (*niche players*).

A continuación, se presentarán gráficos que muestran seis categorías representativas para este documento, acompañando al gráfico se definen los puntos que Gartner evalúa en cada una de estas categorías.



Data Integration Tools

Según Gartner: “El Mercado para las Herramientas de Integración de datos incluye distribuidores que ofrecen productos de software que permiten la construcción en implementación de acceso a datos e infraestructura de entrega para una variedad de escenarios de integración de datos. Para los distribuidores, la demanda de capacidades para la integración de datos tradicional en conjunto con la demanda de soluciones innovativas requiere la distribución robusta y consistente de soluciones altamente desarrolladas. Similarmente, las herramientas de integración de datos opera internamente y se integra con herramientas de datos maestros, herramientas de gobierno de datos, y herramientas de calidad de datos.”¹⁰



Data Management Solutions for Analytics

Según Gartner: “...Definimos *Data Management Solutions for Analytics* (DMSA) con un sistema de software completo que soporta y maneja datos en uno o más sistemas de gestión (generalmente base de datos). DMSAs incluyen optimizaciones específicas para soportar procesos analíticos. Esto incluye, pero no está limitado a, el soporte para procesos relacionales, procesos no relacionales (como procesamiento de gráficos), y *Machine Learning* y lenguajes de programación como Python y R. Los datos no son necesariamente almacenados en una estructura relacional, y múltiples modelos pueden ser usados, por ejemplo, Relacional, XML, JSON, Key-Value, Text, Graph y geoespacial”¹¹

¹⁰ Gartner; (2018) Data Integration Tools; Gartner;

<https://www.gartner.com/reviews/market/data-integration-tools>

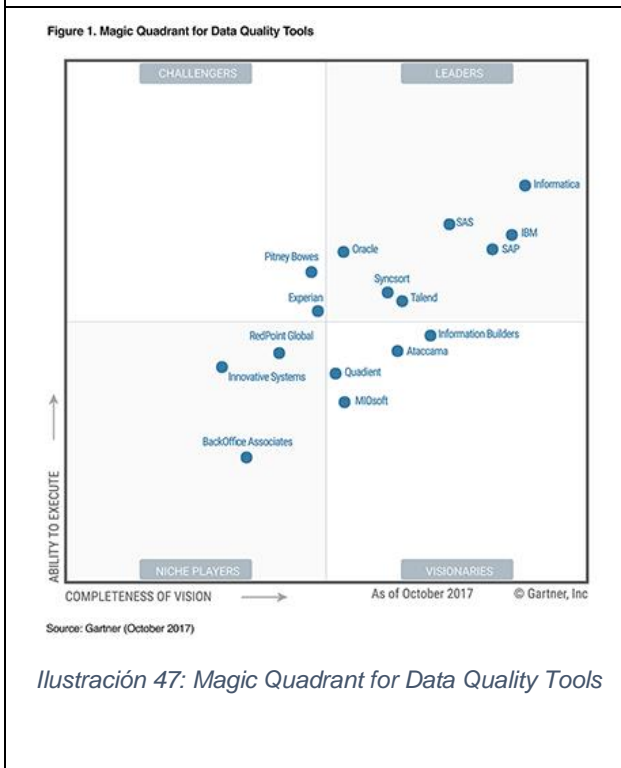
¹¹ Gartner; (2018). Data Management Solutions for Analytics; Gartner

<https://www.gartner.com/reviews/customers-choice/data-warehouse-solutions>



Analytics and Business Intelligence Platforms

Según Gartner: “Una Analítica de Inteligencia de Negocios moderna, soporta el desarrollo de contenido analítico facilitado por tecnologías de Información. Esto es definido por *arquitecturas independientes*¹² que permiten que usuarios sin conocimientos técnicos puedan ejecutar de forma autónoma *flujos de trabajo analíticos*¹³ completos desde accesos de datos, ingesta y preparación para análisis interactivos y la *distribución compartida de ideas*¹⁴.”¹⁵



Data Quality Tools

Se refiere a las herramientas que ayudan a garantizar la calidad del dato lo cual Gartner describe como “La disciplina de garantizar la calidad del dato asegura que la data es *adecuada para el fin*¹⁶ en el contexto de operaciones de negocios, analítica y escenarios de negocios digitales emergentes. Cubre mucha más que solo tecnología, incluye manejo de programas, roles, estructuras organizacionales, usa casos y procesos (como los de monitoreo, reporte y solución de problemas de calidad de datos). También se conecta con iniciativas más amplias en el ámbito del manejo de *gestión de datos empresariales* (EIM), incluyendo información de *gestión de datos maestros* (MDM).”¹⁷

¹² *self-contained architecture*

¹³ full-spectrum analytic workflows

¹⁴ collaborative sharing of insights

¹⁵ Gartner;(2018). Analytics and Business Intelligence Platforms; Gartner;

<https://www.gartner.com/reviews/market/analytics-business-intelligence-platforms>

¹⁶ *Fit to purpose.*

¹⁷ Gartner; (2018). Data Quality Tools; Gartner; <https://www.gartner.com/reviews/market/data-quality-tools>

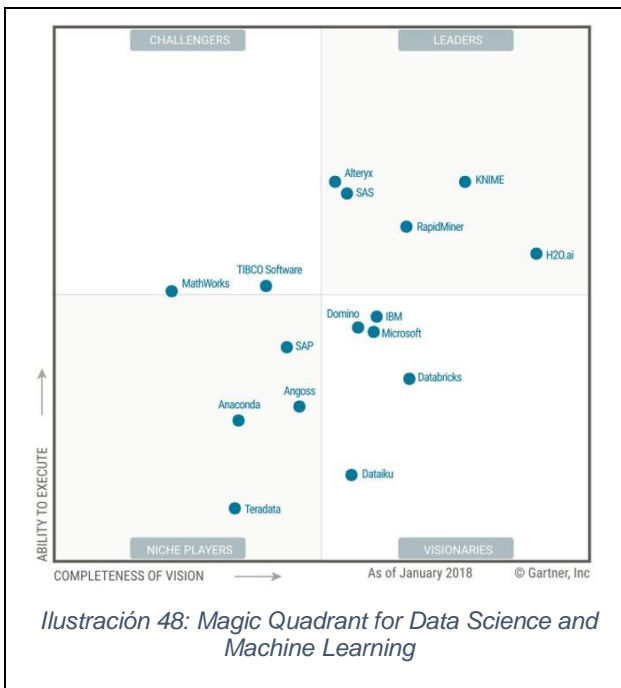


Ilustración 48: Magic Quadrant for Data Science and Machine Learning

Data Science and Machine Learning

Gartner define “Data Science and Machine Learning Platform” como “Aplicaciones de software cohesivas que ofrecen una mezcla de *bloques de construcción básicos*¹⁸ para ambos, crear varios tipos de solución de *data science* e incorporar dicha solución a un proceso de negocio, acompañando infraestructura y productos. Machine Learning es una parte del Data Science que merece especial atención en el proceso de evaluación de estas plataformas”¹⁹

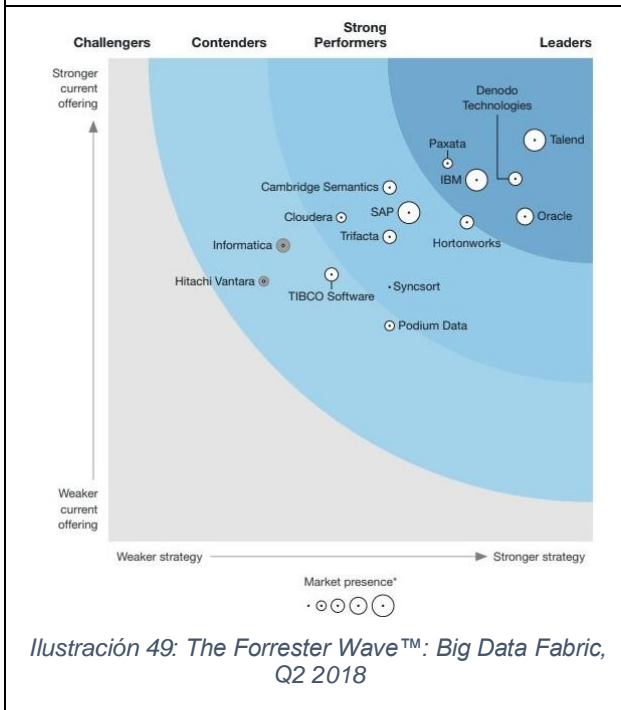


Ilustración 49: The Forrester Wave™: Big Data Fabric, Q2 2018

Big Data Fabric

“Estas plataformas de software aceleran y simplifican la ingestión, integración, descubrimiento y preparación de datos procedentes de múltiples silos creando una capa semántica de datos y automatizando todo el proceso de adecuación de estos.

Este tipo de arquitecturas son ideales para simplificar y agilizar la puesta en marcha de proyectos de análisis en tiempo real, streaming analytics, machine learning o analítica avanzada.”²⁰

¹⁸ Basic Building Blocks

¹⁹ Gartner;(2018).Data Science and Machine Learning; Gartner;

<https://www.gartner.com/reviews/market/data-science-machine-learning-platforms>

²⁰ BI-Spain.com; (2018). The Forrester Wave™: Big Data Fabric, Q2 2018; BI-Spain.com; <https://www.bi-spain.com/articulo/74988/big-data/todos/the-forrester-wavetm-big-data-fabric-q2-2018>

En la siguiente tabla, se presenta un resumen del análisis de las gráficas anteriores, donde se indican cuáles son las compañías que se destacan en la categoría de “Líderes”, tanto para Gartner como Forrester.

INFORME LÍDERES EN BI & BA						
PLATAFORMAS	Gartner: Data Integration Tools	Gartner: Data Quality Tools	Gartner: Data Management Solutions for Analytics	Gartner: Data Science and Machine Learning Platforms	Gartner: Analytics and Business Intelligence Platforms	Forrester wave: Big Data Fabric
IBM	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>
ORACLE	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			<input checked="" type="checkbox"/>
SAP	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>			
TALEND	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				<input checked="" type="checkbox"/>
SAS	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>		
MICROSOFT			<input checked="" type="checkbox"/>		<input checked="" type="checkbox"/>	
INFORMATICA	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>				
TERADATA			<input checked="" type="checkbox"/>			
TABLEAU					<input checked="" type="checkbox"/>	
QLIK					<input checked="" type="checkbox"/>	
KNIME				<input checked="" type="checkbox"/>		
RAPIDMINER				<input checked="" type="checkbox"/>		
ALTERYX				<input checked="" type="checkbox"/>		

Tabla 10: Resumen Informes Data Analytics Gartner & Forrester

En efecto es posible construir diferentes cuadros comparativos, incorporando por ejemplo el desplazamiento que la marca/tecnología haya tenido en los últimos x años, y/o bien incorporando otras categorías; también, adicionar otros informes de Forrester u otras instituciones que emiten similares informes; sin embargo, con esta simple inspección resumen es suficiente para una formar una visión de campo para los propósitos de este proyecto.

7.4 Diagrama Clúster para el Big Data corporativo

El siguiente diagrama corresponde al clúster que se encuentra en proceso de análisis final para el montaje de infraestructura para trabajar problemáticas de Big Data en la organización

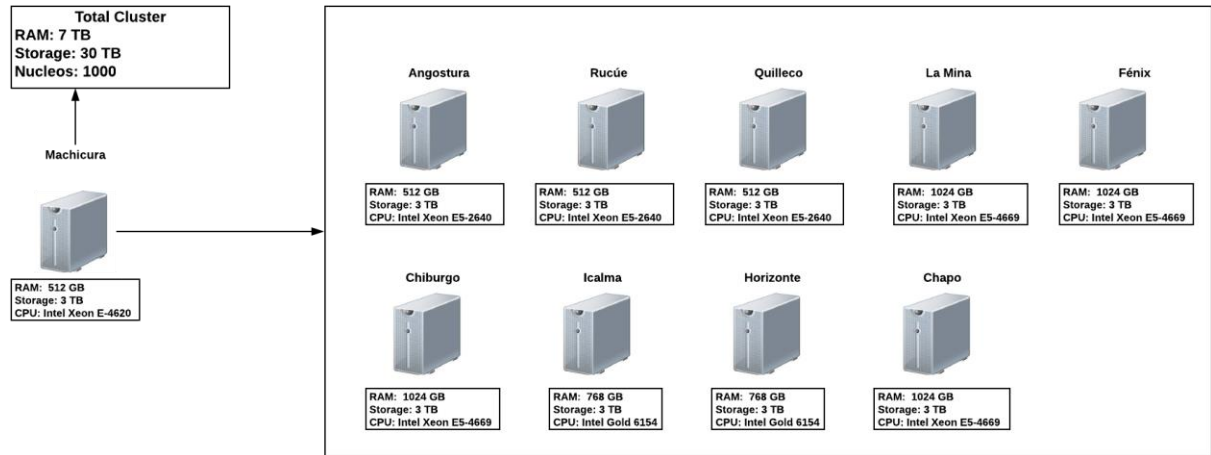


Ilustración 50: Clúster para Big Data Colbún

7.5 Landscape tecnológico

Las siguientes ilustraciones presentan los mosaicos de tecnologías, aplicaciones y empresas para Big Data del 2012 y 2018 respectivamente²¹. Más allá del trabajo de recopilación realizado, la comparación deja en evidencia que con el paso de los años ha existido una considerable proliferación de alternativas, lo que sin duda trae beneficios a las organizaciones que desean oferta de soluciones, pero también abre diversos desafíos como la interoperabilidad, la “elección adecuada”, el dominio de la herramienta, especialistas en el mercado, continuidad de las herramientas, etc. Lo anterior, fue también un reto en este proyecto ya que precisamente uno de sus alcances fue el de definir la tecnología con la que la compañía iba a enfrentar sus oportunidades de explotación de datos.

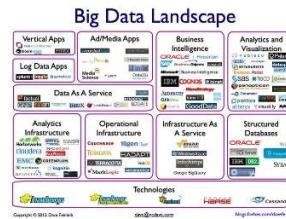


Ilustración 51: Landscape Big Data 2012

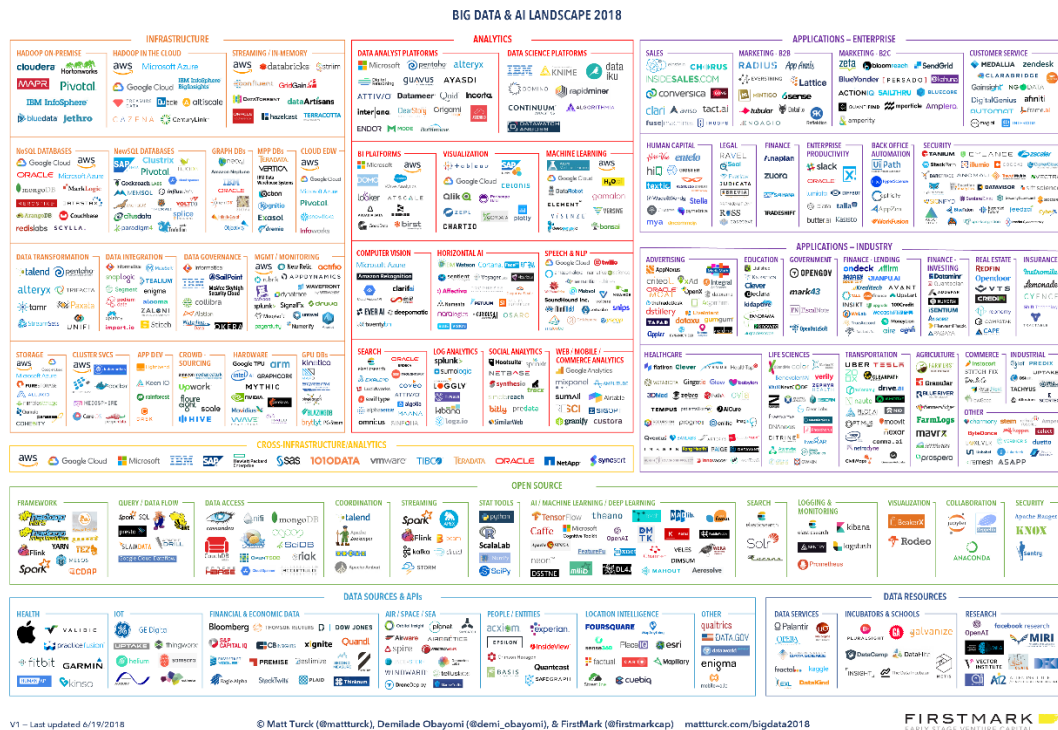


Ilustración 52: Landscape Big Data 2018

²¹ <http://mattturck.com/bigdata2018/>

-
- i Presentación corporativa de la compañía.
 - ii Presentación corporativa de la compañía.
 - iii Presentación corporativa de la compañía.
 - iv Presentación corporativa de la compañía
 - v Sitio Web de Colbún S.A.