MISO
Maestría en Ingeniería de Software

Proyecto: Propiedades de los Alpes

2024

Descripción del proyecto del curso

MISO
Maestría en Ingeniería de Software

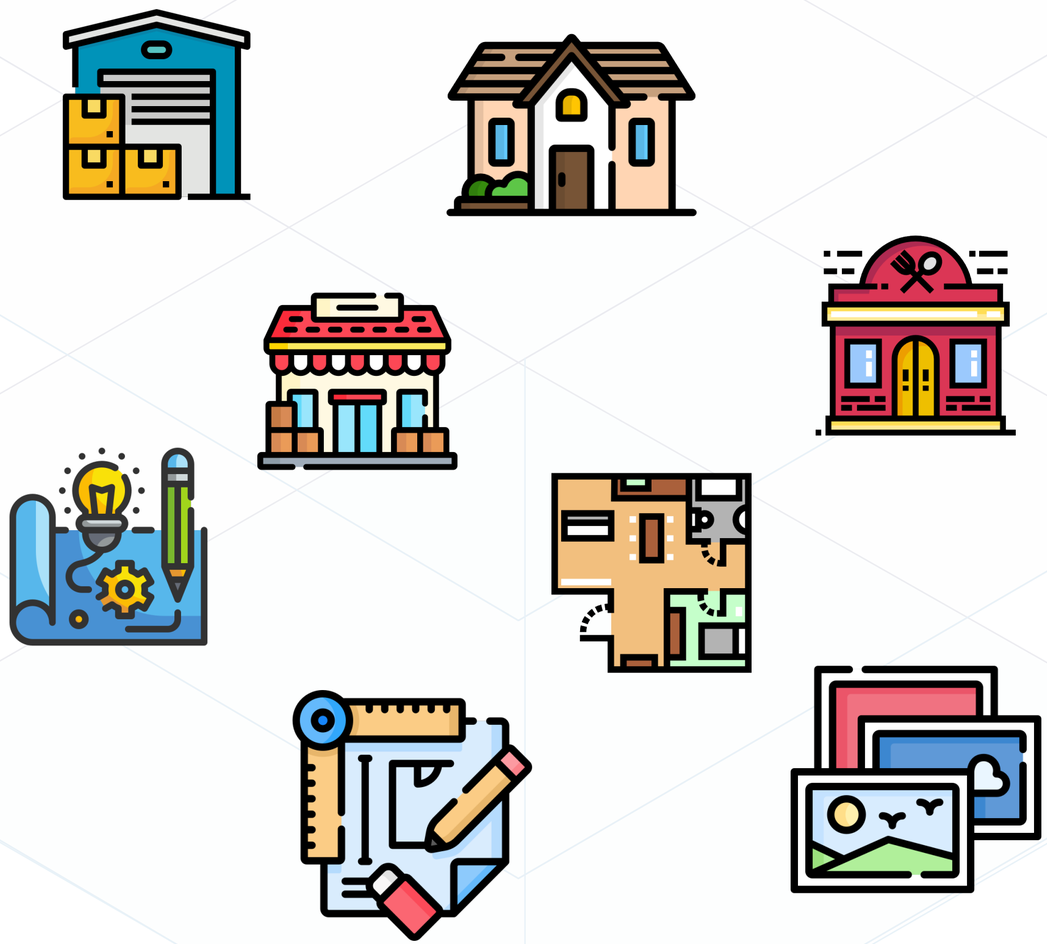
Propiedades de los Alpes

Contexto

- Propiedades de los Alpes (PDA) es una compañía colombiana constituida en 2002. Su enfoque principal es proveer información sobre bienes raíces comerciales tales como detalles de las propiedades (tamaño, tipo de construcción, número de pisos, fotografías, etc), listados de alquiler, venta, subarrendamiento; información precisa sobre arrendatarios; transacciones sobre arrendamientos; transacciones y comparativos de ventas.
- La compañía inicio operaciones en Bogotá, siendo este su primer mercado. En el transcurso de 10 años, se expandieron a las principales ciudades comerciales del territorio colombiano, tales como Medellín, Cali, Bucaramanga, Cartagena, Santa Marta y Barranquilla.
- En el año 2015, PDA recibió financiación para poder comenzar su expansión global. El objetivo de esta expansión es poder proveer los servicios mencionados anteriormente en diferentes mercados en Latino América, Europa, Asia y Oceanía.

Propiedades de los Alpes

Servicios



- **Características de propiedades**
  - El equipo de investigación de campo realiza inspecciones físicas de estructuras de oficinas, industriales y minoristas en cada área metropolitana.
  - Para cada propiedad, se investigan las características físicas clave, que incluyen: tamaño, tipo de construcción, zonificación, tamaño del lote, número de pisos, estacionamiento y más.
  - Se recopilan fotografías exteriores originales para las necesidades de mercado e informes de los usuarios.

Servicios

- **Disponibilidad**
  - El equipo de investigación de listados para cada mercado identifica, ingresa y mantiene los listados (en los mercados donde estén disponibles) de alquiler, subarrendamiento, venta y trabajo conjunto.
  - El sistema procesa actualizaciones en tiempo real de los principales CRM y el equipo de investigación de listados se comunica mensualmente de manera proactiva con agentes y propietarios para recopilar información adicional y garantizar que los listados sean lo más oportunos posible.

Servicios

- **Transacciones**
  - En los mercados donde están disponibles, se adquiere e ingresa los arrendamientos registrados para oficinas y propiedades industriales. Esta información se monitorea y actualiza semanalmente para dar cuenta de nuevos arrendamientos, extensiones, vencimientos, etc.
  - Para los mercados donde los arrendamientos registrados no están disponibles, por medio procesos de ingeniería de datos se extrae y procesa los datos de diferentes fuentes, incluyendo usuarios que envían puntos de datos de transacciones para acceder a información confirmada de transacciones de arrendamiento.

Servicios

- **Información de inquilinos**
  - Revisa los sitios web de las empresas, los sitios de redes sociales y los registros comerciales para verificar la información de las direcciones, recopilar múltiples contactos clave, identificar grupos industriales, ver inquilinos con próximas fechas de vencimiento, valor de alquileres, entre otros.
  - El equipo de investigación de inquilinos llama directamente para verificar la información de los inquilinos sobre propiedades clave en cada área del mercado.

Servicios

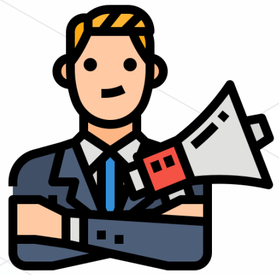
- **Comparativos de ventas**
  - Se cargan las transacciones de ventas históricas para cada mercado para garantizar que los usuarios puedan realizar de inmediato análisis de ventas de propiedades y de mercado.
  - Para nuevas ventas, el equipo de ventas comparables identifica al verdadero comprador y vendedor de la propiedad y recopila información más profunda que incluye materiales de mercado y noticias en los medios.
  - El investigador de campo del mercado local realiza una nueva inspección para garantizar que mostramos la estructura tal como estaba en el momento de la venta.

Servicios

- **Análisis de mercado y propiedades**
  - Se permite a los usuarios desglosar de propiedad principal, la participación de mercado de las agencias, las tendencias históricas de arrendamiento y venta, y más en un área geográfica específica o dentro de un conjunto de propiedades competitivas.
  - Los usuario pueden comparar las tarifas de venta promedio de una propiedad en cuestión, la duración promedio de los acuerdos, las tasas de desocupación y más con edificios dentro de la misma área.

Clientes

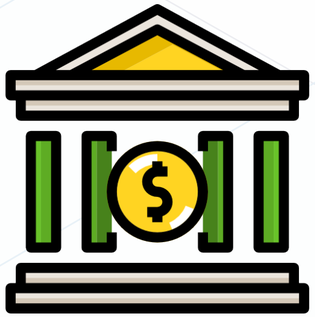
Grupo de inversión en activos inmobiliarios



Agencias de representación de propietarios

Firmas para asesoría de inquilinos

Peritos



Instituciones financieras

Instituciones gubernamentales

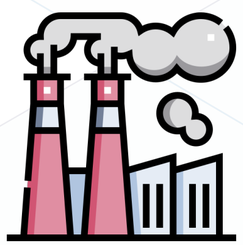
Tipos de propiedades

Minoristas

Las propiedades minoristas son espacios que se utilizan para vender o comercializar bienes y servicios de consumo. Abarcan desde tiendas individuales hasta grandes centros comerciales o centros comerciales.

Oficinas

Son espacios que se utilizan con fines comerciales e incluyen grandes torres de oficinas de varios pisos hasta espacios suburbanos más pequeños. Los bienes raíces de oficinas incluirán instalaciones adicionales para los inquilinos, cocinas, salas de reuniones, baños y espacios para escritorios.



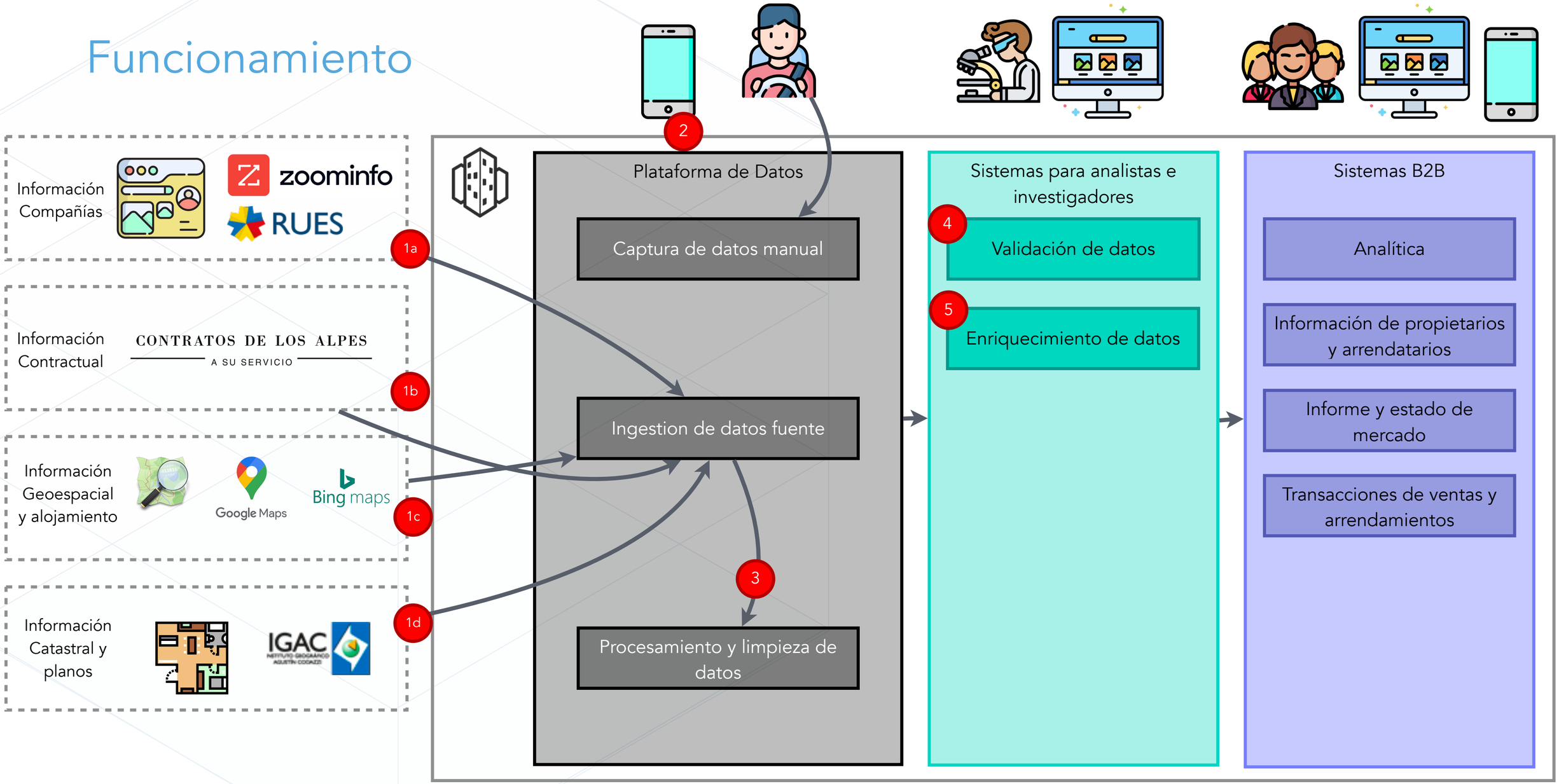
Industrial

Los espacios industriales pueden ser bodegas extremadamente grandes que se utilizan con fines de fabricación, almacenamiento y logística. En el otro extremo del espectro, las propiedades industriales pueden ser más pequeñas y, a veces, tener una combinación de espacio para bodegas y oficinas.

Uso Especializado

Espacios donde se desarrollan actividades especificas tales como inmuebles que contienen instalaciones sanitarias, hospitales, centros médicos, laboratorios, entre otros.

Funcionamiento



Flujo de trabajo

1. Lo primero que PDA necesita es capturar la información de las diferentes entidades/atributos/conceptos necesarios para proveer una vista centralizada de los diferentes servicios previamente nombrados. Esta información puede llegar por dos mecanismos: **pull** o **push**. Cuando se usa un método push, agentes van a las fuentes de información (websites, APIs, llamada, etc) para extraer la información y persistirla en un data lake para subsecuentemente ser procesada. Tenga en cuenta el concepto de **agente**, se usa para describir tanto un sistema de computo como los investigadores humanos que recolectan la información de forma manual o semi-automatizada (tal vez con ayuda de algunos scripts básicos o herramientas low-code). En el caso de un método push, los datos son enviados desde la fuente a algún API de PDA o un sistema de almacenamiento de objetos, como un Bucket o servidor FTP. Algunos de la información recolectada incluye:
    a. **Información sobre compañías:** Uno de los conceptos principales es la recolección de las compañías en un mercado. Tenga en cuenta que conceptos como arrendatario, inquilino, vendedor, comprador, entre otros, hacen referencia a una compañía desenvolviendo dicho papel en una transacción. Dicha información normalmente es de acceso público en muchos de los mercados, en otro casos es necesario llegar a acuerdos con entidades gubernamentales y en otros se pueden usar servicios de tercero que proveen dicha información. En el caso de Colombia el servicio del **Registro Unico Empresarial** (RUES) nos provee con información de registro de la compañía.
    *** Nota: Tenga en cuenta que en otros mercados se deberá acceder a diferentes servicios, por lo que el sistema debe ser lo suficientemente adaptable. Así mismo, se puede usar algunos servicios de tercero como ZoomInfo en mercados donde el gobierno no provea dichos servicios tecnológicos. En algunos casos es necesario complementar dicha información usando el website, redes sociales y demás información de acceso público de la compañía.**

Flujo de trabajo

1.

b. **Información contractual:** La información contractual y de transacciones es aquella que nos puede brindar los movimientos de los vendedores, compradores, arrendatarios e inquilinos. En el caso de Colombia, PDA utiliza los servicios de **Contratos de los Alpes**, empresa que se encarga de proveer servicios de firma digital de contratos. En otros mercados (incluyendo el colombiano) es también común aliarse con empresas que proveen datos listados de propiedades, tales como Metrocuadrado, FincaRaiz, Remax, entre otros. Tenga en cuenta que la mayoría de productos de listado de propiedades se enfocan principalmente en viviendas y propiedades no comerciales (aunque se puede encontrar). En caso de usar un producto de listados es necesario mantener un historial y cadencia correcta, para poder detectar todos los movimientos de las compañías. En este caso la información vital es la dirección o direcciones comerciales de la compañía, número (s) de contacto, fecha de inicio y terminación de contrato, partes y roles (quién es arrendatario e inquilino) y valor de las transacciones.

** Nota: **Cuando se obtiene información de las compañías usando servicios como RUES, se obtiene información básica de localización, normalmente la dirección de registro. Tenga en cuenta que esta dirección, aunque tiene cierto nivel de utilidad, no es lo que PDA desea proveer a sus clientes pues no implica las direcciones de sus propiedades comerciales.**

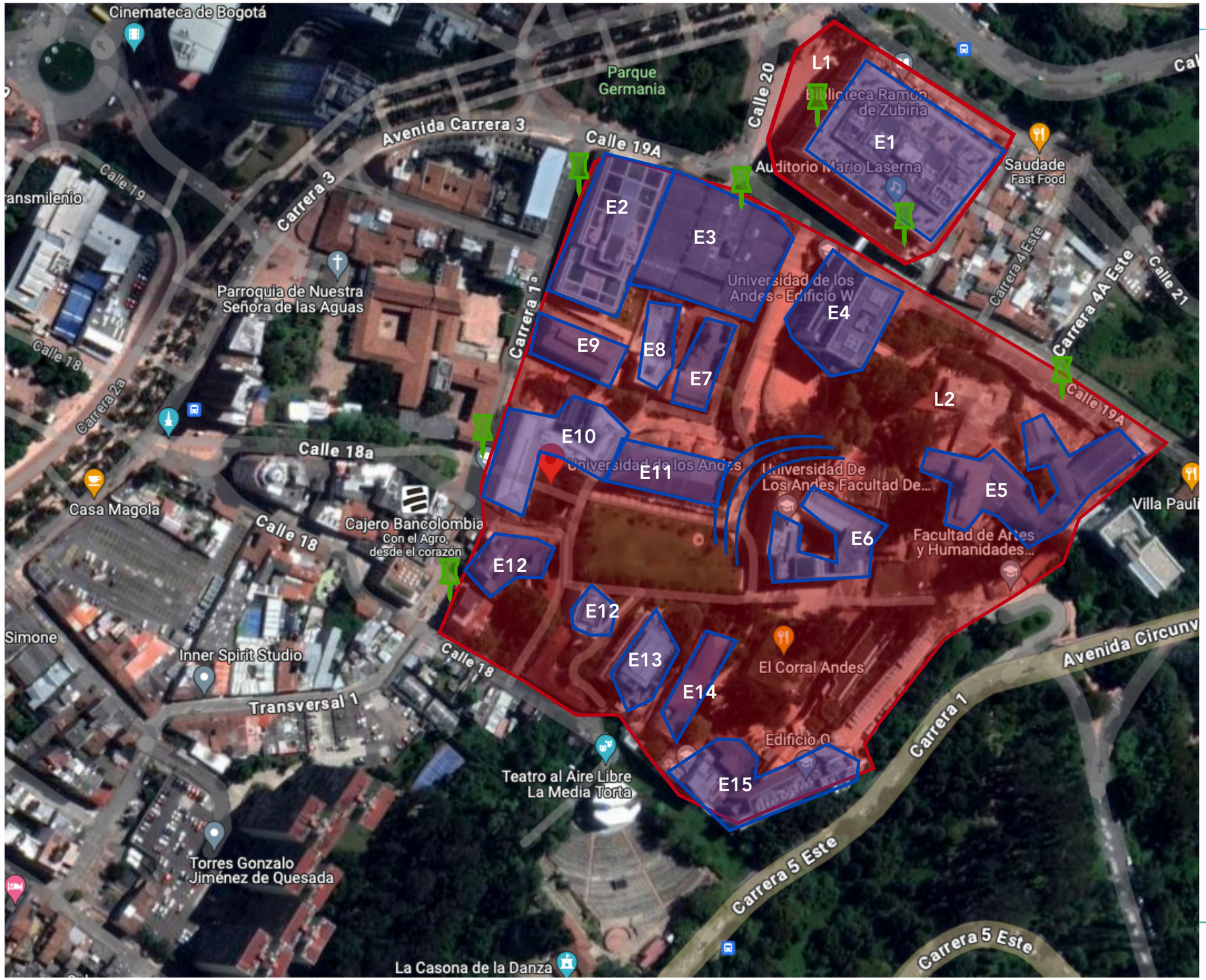
Flujo de trabajo

1.

c. **Geoespacial y alojamiento:** Una parte fundamental de la solución es todo lo relacionado a los sistemas de información geográfica (GIS). En el caso de PDA, es de suma importancia no solamente saber que una compañía X tiene oficinas en las direcciones Q y P, pero también el area del lote, coordenadas, polígonos que describen el lote y edificios, entre otros. Tenga en cuenta que una dirección física puede involucrar una misma propiedad y/o lote. Tome como ejemplo la Universidad de los Andes, en donde el campus principal es un lote con varias direcciones asociadas (a las diferentes entradas) pero representan la misma propiedad. Esta propiedad la cual tiene un único identificador, contiene múltiples edificios con un único identificador.

En otros casos, el concepto de lote, propiedad, dirección y edificio puede cambiar. Piense en el caso del edificio Santo Domingo (SD) o el Mario Laserna (ML). En dicho caso, el lote representa una única propiedad con un único edificio.

Esta información es normalmente adquirida o validada por servicios GIS públicos como el caso de OpenStreetMap o algunos servicios pagos como Google Maps y Bing Maps.



MISO
Maestría en Ingeniería de Software

Lote

Edificio

Dirección

En este caso particular el lote y la propiedad es 1:1. Sin embargo, pueden existir casos en donde un lote puede contar con múltiples propiedades.

Por ejemplo, piense en un conjunto cerrado, en donde la parcela es dividida en lotes más pequeños con propiedades independientes por cada propietario. Cada propiedad puede tener multiples edificaciones, tales como la casa y parqueadero independiente.

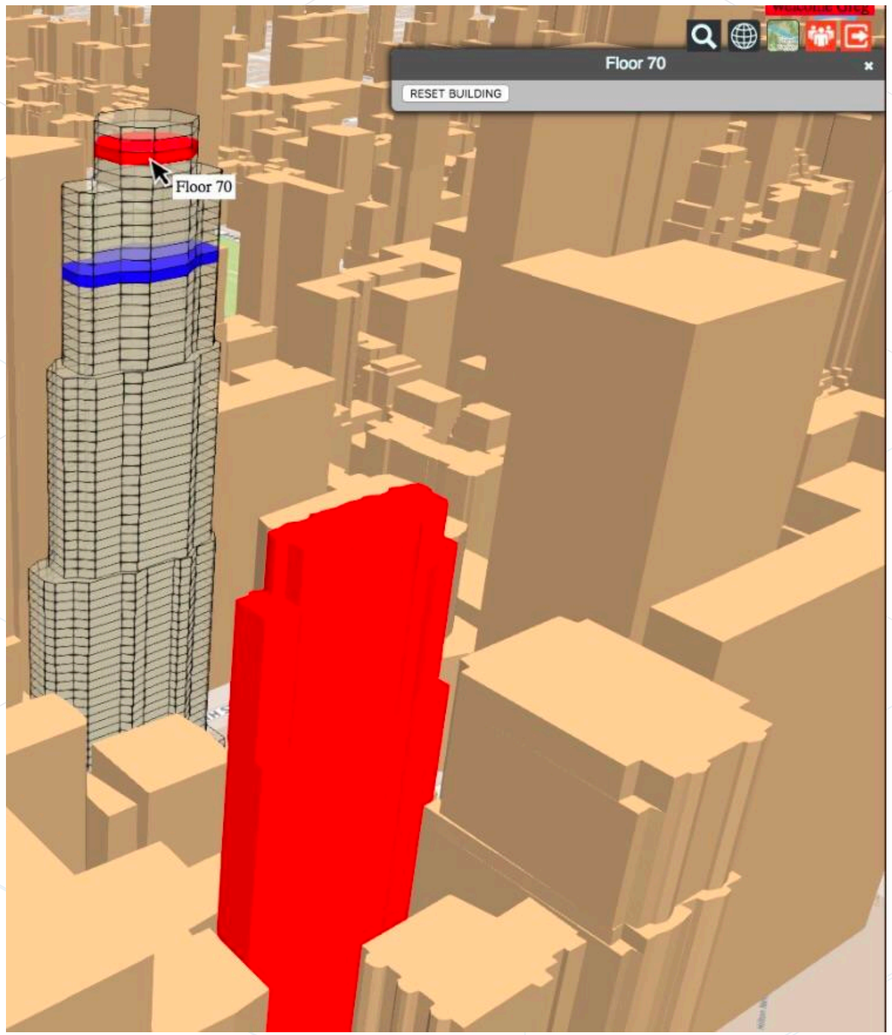
Flujo de trabajo

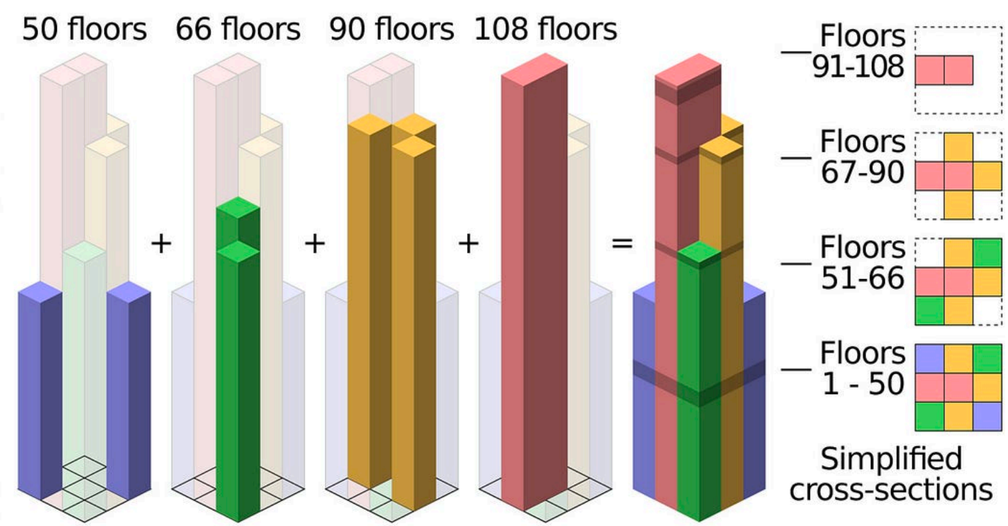
1.

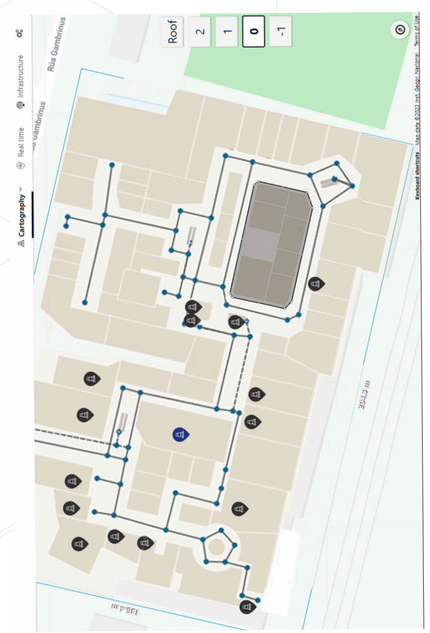
d. **Información catastral y planos:** Aunque los productos GIS nos brinden una gran cantidad de datos, estos nunca contaran con los planos e información detallada de lotes, edificios y pisos. Con la ayuda de esta información, PDA no es solamente capaz de identificar parcelas, lotes, edificios y propiedades, sino además detalles como cantidad de pisos, organización de cubículos/oficinas, áreas de las oficinas y organización de las mismas. Piense en el caso del Centro de Comercio Internacional (Torre Davivienda) en Bogotá. Por medio de juntar la información catastral con las empresas y contractual, podríamos saber que la compañía Contratos de los Alpes se encuentra en el piso 48, oficina 12, la cual cuenta con 10 metros cuadrados por la cual paga un arriendo de $9 500 000 COP mensuales. Al conocer esta información los clientes de PDA pueden conocer y avaluar zonas comerciales casi a granularidad de oficina.

***** Nota: Dependiendo del mercado, algunos de los datos presentados a los clientes deben ser ofuscados o limitados acorde el tipo de cliente. Por ejemplo, es probable que entidades gubernamentales desean y puedan acceder a cierto nivel de granularidad que clientes como las agencias no puedan (también por cuestión de competencia desleal).**

Flujo de trabajo







Flujo de trabajo

2. Aunque el objetivo principal de PDA es utilizar fuentes ya existentes para la recolección de datos, en la mayoría de mercados siempre será necesaria la captura de datos manual para **poder llegar lo más cercano una cobertura del 100% de todas las propiedades comerciales** (para los clientes de PDA una cobertura menor al 80% de propiedades ya no es confiable). Esta recolección manual de datos pretende rellenar vacíos de algunas zonas, propiedades, datos de compañía, etc, que no fueron posibles de obtener por medio del proceso automatizado, o por el contrario fueron capturados pero no hay seguridad de que los datos sean correctos. Para PDA no es solo cuestión de capturar los datos pero también de garantizar su exactitud. En general, la compañía debe garantizar a sus clientes que los datos son correctos, completos, consistentes, únicos, y oportunos.

   Para lograr lo anteriormente nombrado, PDA cuenta con un equipo de investigadores de campo, los cuales van a las diferentes propiedades comerciales y recopilan los datos faltantes en una aplicación móvil. Algunas de las actividades que los investigadores deben realizar son: toma de fotos del exterior de la propiedad y si es posible del interior; validación de la existencia de la propiedad o cambio de tipo de propiedad; agregar una propiedad o edificio; agregar datos geoespaciales (coordenadas y polígonos) de los edificios; entre otros.

   Así mismo, existen equipos de investigadores para compañías y listados que usando el sistema interno de investigación, para verificar la completitud y validez de los datos ingresados por los investigadores de campo. Estos mismos investigadores pueden hacer actividades de enriquecimiento de datos que no necesiten trabajo de campo tales como llamar a las compañías a verificar datos, consultar sitios web y social media, entre otros.

Flujo de trabajo

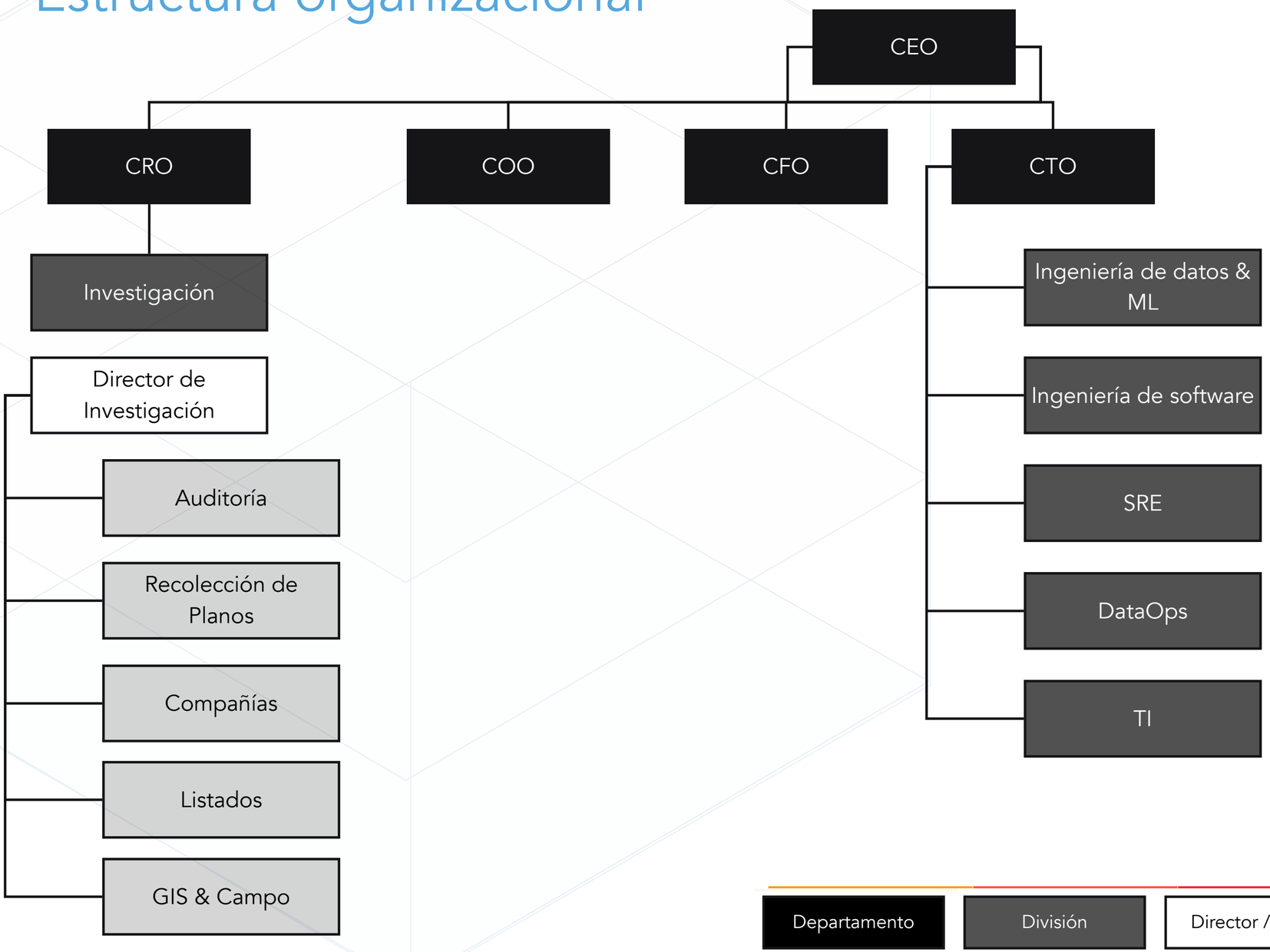
3. Cuando los datos son extraídos de forma automatizada, estos deben pasar por un conjunto de pipelines que contienen reglas de negocio y aplican heurísticas para garantizar una alta calidad en los datos a presentar a los clientes. **Es en esta etapa donde se pueden utilizar técnicas de ingeniería de datos y machine learning para obtener la mayor cantidad de puntos y con la más alta calidad.** Por ejemplo, existen pipelines encargados de hacer el "match" entre compañías y su actual localización usando los contratos de arrendamiento brindados por Contratos de los Alpes. En el caso de los contratos, es muchas veces necesario pasar por procesos de reconocimiento de caracteres (OCR) y eventualmente usar modelos de lenguaje para identificar conceptos como: Quién es el arrendatario, quién es el inquilino, fechas de inicio y terminación del arriendo, dirección de la propiedad bajo arriendo, costo, entre otros.

   Otro ejemplo, es la automatización de llamadas a las compañías para validar datos. El equipo de Machine Learning ha implementado un bot que es capaz de realizar llamadas a las diferentes compañías con base a los datos recolectados y validar que sean correctos. **Este proceso solo aplica en ciertos casos, donde la heurística de calidad por registro es baja y se deben usar métodos "manuales"**. Tenga en cuenta que el resultado de cada llamada también debe ser procesado para entender la semántica de la conversación y extraer los valores necesarios.

Flujo de trabajo

4. El equipo de auditoría de datos se encarga de revisar información almacenada de forma automatizada. Dado que el volumen de los datos puede ser gigantesco, y el objetivo de automatizar la ingesta de datos es reducir los recursos humanos, los investigadores de auditoría verifican registros que tengan un **índice de confiabilidad** menor a 0.8 (valores van de 0 a 1). La idea es que en muchos casos tenemos una alta confiabilidad de que los datos automatizados son correctos, esto debido a una heurística donde damos pesos a atributos como:
   i. **Fuente:** Si los datos fueron enviados por la entidad oficial del gobierno, sabemos que debe ser correcto. Si por el contrario, los datos son provenientes de un scraper, es probable que la confiabilidad sea menor.
   ii. **Consistencia:** Si vemos que los mismos datos se ven en 3 fuentes distintas podemos asumir que los valores tienen una alta confiabilidad de ser correctos. Si por el contrario, de 3 fuentes todos tienen datos distintos nuestro nivel de confianza disminuye.
   iii. **Completitud:** Si todos los atributos del registro están completos, el nivel de confianza incrementa. Si por el contrario, de 20 campos solo tenemos 5 campos, es altamente probable que este registro pase por una auditoría manual.
5. Finalmente, el equipo de auditoría enriquece los datos faltantes y habilita la visibilidad en el producto de los clientes.

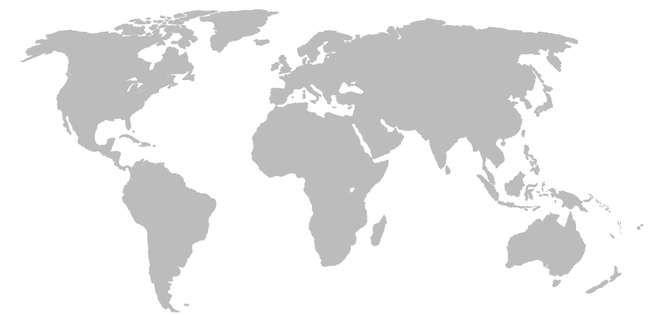
Estructura organizacional



- El equipo de ingeniería y producto (debajo de CTO es de aproximadamente unos **80 empleados**.
- La compañía como un todo tiene un tamaño de más de **380 empleados**.

Estadísticas

Compañías: 1 854 086
Propiedades: ~354 000
Edificios por propiedad: 1 a 3
Contratos anuales: ~400 000



Compañías: 1M a 6M por país
Propiedades: 23K a 500K por país
Edificios por propiedad: 1 a 3
Contratos anuales: 200K a 600K

Clientes estimados: 1% a 2% población total de empresas
Mercado esperado: 48 a 62 países

Estrategia y visión

Expansión global

- Como se mencionó anteriormente, la compañía desea expandir operaciones a nivel mundial. En un principio desea conquistar el mercado latino americano entrando a países como Brasil, Perú, Ecuador, Chile, Argentina y México. Una vez con dichos países, se desea entrar al mercado europeo, MENA, ENEA y Oceanía.
- Actualmente se cuenta con un sistema monolítico el cual comparte base de datos con el sistema de investigadores (para adquisición y validación de datos). Desafortunadamente la arquitectura de datos está altamente acoplada a las reglas de negocio y terminología usada en Colombia, al igual que sus leyes y procesos. Los sistemas también se encuentran fuertemente acoplados, en donde incluso los despliegues deben hacerse en conjunto para evitar efectos secundarios.
- Con este proyecto, esperamos tener un nuevo sistema fresco y listo para todas los nuevos retos. Se quiere tener una arquitectura altamente escalable, modificable y extensible en la parte tecnológica pero también de recursos humanos. El equipo de ingeniería es cada vez más lento y producir código de alta calidad respetando las fechas del Roadmap se ha vuelto una misión imposible.
- Se espera que el SLO para adecuar el sistema a cualquier nuevo mercado sea de 3 a 6 semanas. Actualmente los arquitectos estiman que, dada la arquitectura actual, entrar a un mercado nuevo es equivalente a realizar un proyecto completo con una duración no menor a 6 meses.
- Cabe aclarar que entrar a diferentes mercados involucra retos en el manejo de zonas horarias, latencia de servicios, trabajo remoto y diferencias en las leyes acerca del manejo de los datos. Por tal motivo, el sistema debe estar diseñado de una manera en que permita a equipos de tecnología e investigación trabajar de forma independiente evitando cuellos de botella en desarrollo y despliegue de servicios.

Objetivo

- Su equipo fue contratado para diseñar e implementar la arquitectura para soportar la expansión global a multiples mercados y regiones.
- Para ello, se espera que en los próximos 2 meses usted y su equipo presenten un **documento de arquitectura de solución**.
- El equipo de ingeniería desea también validar sus decisiones de diseño. Por ende, se debe construir **POC/experimentos** para soportar sus decisiones y poderlas presentar al equipo de ingeniería.
- Su documento debe considerar los diferentes puntos de vista y la estructura de equipo que necesita para lograr el cometido.