

Mòdul 1: Introducció al BI

Índex de continguts

Conceptes del BI	2
Característiques dels programes de BI	3
Cicle de vida de les dades.....	4
Plataforma de dades.....	5
Beneficis.....	5
Anàlisi de dades.....	6
Categories de l'anàlisi de dades.....	6
Rols en les dades.....	6
Beneficis.....	7
Creació d'un sistema de BI	8
Principals programes de BI.....	10
Introducció a Power BI	12
Què és Power BI?.....	12
Components de Power BI.....	12
Flux de treball a Power BI.....	13

Conceptes del BI

Si intentem resumir al màxim, podríem dir que un programa informàtic de Business Intelligence (BI) és una eina empresarial que permet transformar les dades i la informació en coneixement de valor per a l'organització.

Entrant més en detall, un programari de BI permet recopilar, processar i convertir les dades en informació de valor per poder analitzar-la. Per tant, permet crear un ecosistema de dades d'origen divers que genera un coneixement per ajudar a la presa de decisions en les organitzacions.

Hem vist com en els últims anys han aparegut nombroses eines per extreure i transformar les dades d'una organització en informació valuosa que es desplega en l'anomenada cadena de valor.

El concepte de cadena de valor va ser descrit i popularitzat per Michael E. Porter en l'obra *Competitive Advantage: Creating and sustaining superior performance*. L'autor el defineix com un model teòric que permet descriure les activitats que generen valor en una organització.

A partir d'aquesta descripció, sembla evident que el BI respon a aquesta necessitat i que, necessàriament, ha patit i patirà moltes evolucions al llarg del temps per tal d'anar-se adaptant.

Resulta complicat establir una **definició del concepte de BI**, però per fer-la tan àmplia com sigui possible podria explicar-se com: **el conjunt de mètodes, eines, aplicacions, pràctiques i capacitats que s'enfoquen en la transformació de les dades i la informació en un coneixement que permet prendre més i millors decisions als membres d'una organització.**

Un cop definit, de manera àmplia, el concepte de BI, podem enfocar-nos en altres conceptes que formen part d'aquest ecosistema i amb els quals haurem de familiaritzar-nos al llarg del curs.

Alguns d'ells seran treballats i comentats àmpliament al llarg del curs mentre que d'altres es presenten només com a elements dins del conjunt del BI. Es detallen a continuació, els elements principals:

- **Magatzem de dades (data warehouse):** tal com el seu nom indica, és el repositori on emmagatzemar les dades de l'empresa i serveix per alimentar els sistemes de BI.

- **Sistemes de reporting (informes):** són utilitzats per mostrar els resultats de l'anàlisi de les dades i la informació. Normalment, és el coneixement que arriba a qui ha de prendre decisions de manera que l'ajudi en la presa d'aquestes.
- **Anàlisi de dades:** dins d'aquest concepte s'inclouen les visualitzacions gràfiques, l'aplicació de tècniques de Machine Learning (anàlisi predictiva) i altres operacions sobre les dades.
- **Dashboard o quadre de comandament:** és un panell on es mostra informació clau de l'empresa. Sol ser un resum de tot el coneixement que aporta el sistema de BI de l'organització i està, habitualment, pensat per als directius de les empreses.

Característiques dels programes de BI

Connectivitat de fonts de dades: les capacitats de connectivitat de fonts de dades permeten als usuaris connectar-se i ingerir dades estructurades contingudes en diversos tipus de plataformes d'emmagatzematge, tan locals com al núvol.

Preparació de dades: Inclou la transformació i manipulació de les dades pel anàlisi i presentació d'aquestes. Aquest procés inclou la neteja de dades, la integració de diferents fonts i la creació de models analítics (com mesures, conjunts, grups i jerarquies definits per l'usuari).

Visualització de dades: la visualització de dades implica la compatibilitat amb taulers altament interactius i l'exploració de dades a través de la manipulació d'imatges i gràfics. S'inclou una gran varietat d'opcions de visualització que van més enllà dels gràfics circulars, de barres i de línies, com mapes de calor i d'arbres, mapes geogràfics, diagrames de dispersió i altres elements visuals de propòsit especial.

Narració de dades: la narració de dades és la capacitat de combinar la visualització interactiva de dades amb tècniques narratives per empaquetar i brindar informació en una format convincent i fàcil d'entendre per presentar-la als destinataris finals.

Consulta en llenguatge natural: la capacitat de consulta en llenguatge natural (NLQ) permet als usuaris fer preguntes sobre les dades utilitzant termes que s'escriuen en un quadre de recerca o es parlen.

Informació automatitzada: un atribut central de l'analítica augmentada és la capacitat d'aplicar tècniques d'aprenentatge automàtic (ML) per generar informació automàticament per als usuaris finals (per exemple, mitjançant la identificació dels atributs més importants en un conjunt de dades).

Catàleg d'anàlisi: es refereix a la capacitat del producte per mostrar contingut analític perquè sigui fàcil de trobar i consumir. El catàleg es pot buscar i fa recomanacions als usuaris.

Generació d'informes: la funció de generació d'informes proporciona informes paginats que es poden programar i ampliar a una gran comunitat d'usuaris.

Integració de la ciència de dades: capacitats que permeten el desenvolupament augmentat i la creació de prototips de models de dades i aprenentatge automàtic (DSML) per part de científics de dades ciutadans i científiques de dades amb integració en l'ecosistema més ampli de ciència de dades i aprenentatge automàtic.

Magatzem de mètriques: la capacitat de proporcionar una capa virtualitzada que permet als usuaris crear i definir mètriques com a codi, governar aquestes mètriques des dels magatzems de dades i donar servei a totes les aplicacions d'anàlisi, ciència de dades i negocis posteriors. Això també inclou capacitats com la gestió d'objectius.

Col·laboració: la col·laboració d'anàlisi és l'aplicació de capacitats de col·laboració a fluxos de treball d'anàlisi per a organitzacions que volen proporcionar un entorn en el qual un ampli espectre d'usuaris pugui coproduir simultàniament un projecte d'anàlisi.

Governança: les capacitats de governança rastregen l'ús i administren com es comparteix i promou la informació.

Cicle de vida de les dades

Des del moment de creació o generació de les dades segueixen un cicle de vida fins que es publiquen per a la seva distribució.

Les fases no són completament lineals, i en molts casos poden superposar-se o interactuar entre si i no totes les fases són necessàries en tots els casos.

Les principals etapes del cicle de vida són les següents:

- **Captura:** les dades generades s'emmagatzemen en arxius que emmagatzemen dades en format estructurat, bases de dades relacionals, etc. En altres casos cal buscar-los i accedir-hi a través de procediments com una API o mitjançant scraping.
- **Processament:** l'objectiu d'aquesta etapa és construir el model de dades. Abans de procedir a l'anàlisi cal processar les dades per identificar problemes de qualitat, seleccionar les que formaran part de l'anàlisi i també per calcular noves variables a partir dels originals.
- **Anàlisi:** l'objectiu d'aquesta etapa és extreure coneixement de les dades, en forma de models que permetin explicar relacions, detectar patrons i tendències, detectar anomalies, fer prediccions, etc.
- **Visualització:** aquesta etapa permet complementar l'anàlisi mitjançant visualitzacions que proporcionin coneixement sobre les dades i alhora que mostrin les seves característiques més rellevants i respondre a preguntes concretes de l'anàlisi.

- **Publicació:** les dades processades, els models analítics o els informes construïts s'han de publicar perquè puguin ser explotats per a tota l'organització.

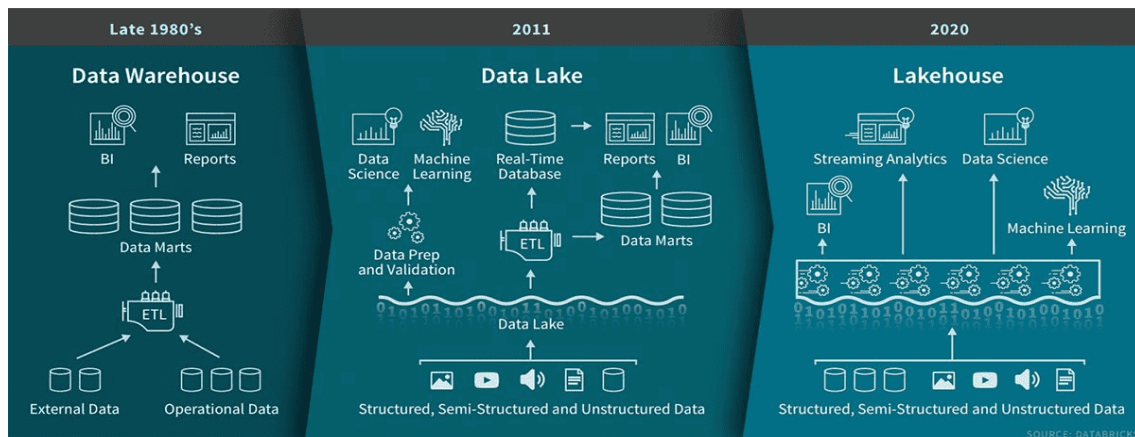
Plataforma de dades

És una solució de programari integrada al núvol, des de la qual es governen, accedeixen i lliuren les dades als usuaris/es, aplicacions o altres tecnologies. En consolidar informació de diferents fonts, una plataforma de dades permet a l'organització un procés de presa de decisions més col·laboratiu.

Beneficis

- **Govern:** Amb una plataforma de dades, una organització pot gestionar millor la seva estratègia de govern de la dada, incloent-hi les dades que s'hi recopilen i qui hi accedeix.
- **Democratització de la informació:** La informació deixa de quedar restringida en compartiments separats i circula per tota l'empresa
- **Avantatge competitiu:** En estandarditzar tant les dades estructurades com les no estructurades, permet a les empreses convertir la dada en un actiu.

En la següent imatge pots veure l'evolució de l'arquitectura de les plataformes de dades al llarg dels anys.



Anàlisi de dades

Les empreses controlades per dades prenen decisions basades en la narrativa que transmeten les seves dades i, en el món actual, les dades no es fan servir en tot el seu potencial, un desafiament al qual la majoria de les empreses s'ha d'enfrontar. L'anàlisi de dades és i ha de ser un aspecte fonamental de totes les organitzacions per ajudar a determinar l'impacte per al seu negoci, i ha d'incloure l'avaluació de l'opinió dels clients, la realització d'investigacions de mercat i productes, i la identificació de tendències o una altra conclusió sobre les dades.

Categories de l'anàlisi de dades

Segons el que responen podem categoritzar l'anàlisi de dades:

- **Què ha succeït?** Crear indicadors clau de rendiment (KPI) a partir de dades històriques. Un exemple d'anàlisi descriptiva és la generació d'informes per proporcionar una visió de les dades financeres i de vendes d'una organització.
- **Per què s'ha produït un esdeveniment?** Identificar anomalies i explicar-les detectant relacions i tendències.
- **Predictiu:** Què ocurrerà en el futur? Utilitza tècniques d'estadística i aprenentatge automàtic (ML): xarxes neuronals, arbres de decisió, regressió.
- **Què fer per aconseguir un objectiu?** Utilitzen tècniques de ML per buscar patrons en les dades. Permeten a les organitzacions prendre decisions basades en dades.
- **Què podria ocórrer si canvien les circumstàncies?** Utilitzen tècniques de ML i de processament de llenguatge natural (NLP) i filtratges específics.

Rols en les dades

Les aplicacions i els projectes actuals poden ser grans i complexos, i sovint impliquen l'ús dels coneixements i les aptituds de diversos individus. Cada individu aporta un talent i una experiència únics, es comparteix l'esforç de treballar de manera conjunta i es coordinen tasques i responsabilitats per veure com un projecte passa del concepte a producció.

Fa poc temps, rols com els d'analistes de negocis i desenvolupadors d'intel·ligència empresarial eren els habituals per al processament i la comprensió de les dades. Però l'augment excessiu de les dades i la varietat en la tipologia d'aquestes, ha provocat que aquests rols evolucionin cap a conjunts d'aptituds més especialitzades que modernitzen i simplifiquen els processos d'enginyeria i anàlisi de dades.

Els rols que podem trobar són:

- **Analista de negocis:** Expert en el negoci. Interpreta les dades presentades a les visualitzacions.
- **Analista de dades:** Recull els requisits. Dissenya el model de dades. Crear els informes. Sovint, les tasques de l'analista de dades i l'analista de negocis poden ser responsabilitat d'una mateixa persona.
- **Enginyer de dades:** Encarregats d'obtenir i netejar les dades. Aprovisionen i configuren la plataforma de dades.
- **Científic de dades:** Elaboren una anàlisi avançada per extreure més valor de les dades.
- **Administrador de base de dades:** Implementa i supervisa les bases de dades.

Beneficis

Atès que les eines d'intel·ligència empresarial acceleren l'anàlisi de la informació i l'avaluació del rendiment, resulten molt valuoses per a les empreses, ja que els ajuden a reduir les ineficiències, a detectar possibles problemes, a buscar noves fonts d'ingressos i a identificar àrees de creixement futur.

Entre els beneficis concrets que experimenten les empreses en utilitzar BI s'inclouen:

- Augment de l'eficiència dels processos operatius.
- Coneixements sobre el comportament i els patrons de compra del client.
- Seguiment precís de les vendes, el màrqueting i el rendiment financer.
- Proves comparatives clares basades en dades històriques i actuals.
- Alertes instantànies sobre anomalies de dades i problemes dels clients.
- Anàlisi que es poden compartir en temps real entre departaments.
- Millorar l'accés a la informació i al coneixement: en transformar les dades en informació i coneixement, aquestes passen a ser molt més accessibles i fàcils d'entendre per part dels diferents membres de l'organització generant així millors resultats que generen noves dades creant un cercle virtuós d'informació i coneixement basat en les dades.
- Es genera un relat únic, consistent, persistent, històric i de qualitat de tot el sistema d'informació d'una organització.
- Es facilita la determinació, la creació i la manipulació de mètriques i indicadors clau de rendiment (KPI) així com indicadors de resultat (KGI) que resulten clau per al funcionament de l'empresa i la consecució dels seus objectius.
- Permet accedir a informació actualitzada tant a nivell agregat com en detall.
- Redueix les diferències entre departaments, sobretot entre el departament de TIC i la resta o la direcció i la resta de l'organització.
- Permet adaptar-se millor al canvi anticipant possibles escenaris futurs.

- Facilita la comprensió de les necessitats dels clients tant potencials com actuals fent més eficient la seva gestió.
- Millora la competitivitat de l'organització de manera transversal fent que la informació important sobresurti, sigui fàcilment accessible i ajudi a la presa de decisions.
- En el passat, les eines d'intel·ligència empresarial eren utilitzades principalment per analistes de dades i usuaris de TI. Ara, les plataformes de BI amb característiques d'autoservei fan que la intel·ligència empresarial estigui disponible per a tothom, des dels executius als equips d'operacions.

Són molts els beneficis d'implantar un sistema de BI, però **¿com sabem si realment cal implantar-lo?**

- Hi ha alguns senyals que permeten identificar la necessitat d'implantar un sistema d'intel·ligència de negoci a l'organització. Algunes d'elles podrien ser:
- Es prenen decisions sense estar prou fonamentades o basades en la intuïció de la direcció.
- Existeixen errors i manques de qualitat en els processos de l'empresa que deriven en problemes d'informació.
- No hi ha comunicació entre departaments o aquesta és dolenta.
- S'usen programes que no són adequats per al correcte manteniment de les dades o la informació de l'organització.
- Hi ha nombroses queixes dels clients que són difícils d'anticipar.
- La publicitat i el màrqueting no aconsegueixen els resultats desitjats.
- Hi ha massa informació al voltant de l'organització que no s'utilitza o no es té capacitat per utilitzar-la.

Creació d'un sistema de BI

Una de les parts més crítiques perquè un sistema de BI tingui èxit és la seva creació i la seva implementació. Aquests passos tenen una gran complexitat i cal treballar-los amb molta profunditat per garantir l'èxit.

Malgrat que la creació d'un sistema de BI depèn molt de les característiques de cada organització, un punt de partida comú acostuma a ser un anàlisi de l'estat de l'organització respecte al BI. Per a això s'acostumen a usar uns models estàndard com el BIMM (Business Intelligence Maturity Model) el qual classifica l'organització a partir de la definició de diferents fases. També serveix, un cop classificada l'organització, per saber quins objectius s'han d'assolir per passar a la següent fase.

Així, les diferents **fases que componen el BIMM** són:

- **Fase 1:** No hi ha cap sistema d'intel·ligència de negoci implantat a l'empresa. Les dades no es recullen o es recullen de manera errònia. Es perd informació rellevant i no hi ha traspàs de coneixement entre departaments. Les decisions se suporten en la intuïció de la direcció sense usar cap dada ni informació.
- **Fase 2:** Encara no hi ha cap sistema de BI, però les dades són accessibles mentre hi ha un sistema que les recull i alguns usuaris poden accedir a informació rellevant. Vulgarment, es parla de la "fase de l'Excel" on aquest conegut programa realitza més funcions de les que pròpiament hauria de tenir. En aquesta fase alguns usuaris ja detecten la necessitat de millorar aquests processos de gestió de les dades i la informació.
- **Fase 3:** Algunes decisions es comencen a fonamentar en dades. Hi ha alguns usuaris que controlen les dades i elaboren informes que permeten fonamentar decisions a partir d'elles. Aquestes dades són recollides i guardades pels mateixos usuaris sense existir una política concreta de recol·lecció i tractament de dades a l'organització.
- **Fase 4:** La creació del magatzem de dades. Molts usuaris veuen la necessitat de tenir les dades correctament classificades i guardades per tal de treure'n veritable profit i per tal de minimitzar els errors que es poden generar d'una gestió "personal" de les dades. En aquesta fase encara es fan reports de manera individual si bé ja es parla obertament de la creació d'un magatzem de dades (data warehouse).
- **Fase 5:** El magatzem de dades ja és una realitat i es comencen a fer reports de coneixement de manera habitual. Aquesta eina és àmpliament utilitzada i se'n beneficien molts usuaris. Els més avançats ja comencen a detectar les mancances del sistema.
- **Fase 6:** Apareix l'anàlisi OLAP. Les mancances del sistema fan que l'anàlisi OLAP aparegui amb força per poder respondre a preguntes més complexes i elaborades. Així, l'anàlisi OLAP apareix en alguns sectors de l'organització impactant en la presa de decisions i en la manera com es dissenyen i s'apliquen els processos de negoci.
- **Fase 7:** S'aplica un sistema complet d'intel·ligència de negoci. En aquesta fase s'implanten altres processos com la creació de dashboards, processos de qualitat de les dades, etc. Aquests processos impacten en totes les esferes de l'organització i en tots els usuaris/es i s'esdevé una **empresa data-driven**.

Com s'ha comentat, crear un sistema de BI resulta complex i és molt específic de cada organització. També les eines de BI s'han d'adaptar a les necessitats de cada organització i és per això que n'hi ha diverses. La selecció d'una o altra anirà en funció de les necessitats de cada empresa.

Principals programes de BI

La següent imatge correspon al *quadrant màgic de Gartner per a les plataformes d'intel·ligència de negocis i anàlisi d'abril de 2023*. Aquest diagrama és part d'un informe que elabora cada any la consultora *Gartner* a partir d'una investigació de mercat i proporciona una visió panoràmica de les posicions relatives dels principals proveïdors de tecnologia.

Figure 1: Magic Quadrant for Analytics and Business Intelligence Platforms



Source: Gartner (March 2022)

Entre les eines més ben posicionades es troben:

- **Microsoft Power BI:** És la plataforma de BI creada per Microsoft per ajudar analistes i directius a crear informes i quadres de comandament i és la que s'utilitzarà en aquest curs.
- **Qlik:** és un programa de BI al núvol per facilitar la integració i l'anàlisi de dades per a qualsevol mena d'empresa. Gràcies a la seva tecnologia amb intel·ligència artificial, elabora una anàlisi eficient que facilita a les empreses millorar la seva presa de decisions de negoci. Qlik automatitza el processament de dades i les converteix en informació útil en temps real, i les converteix en un quadre de comandament i disposa d'un sistema d'alertes.
- **Salesforce Tableau:** programa de BI intuïtiu i fàcil de manejar que permet a les empreses identificar patrons i tendències a partir de les dades generades en els seus processos de negoci, convertint-los en informació visual ràpidament. Tableau ofereix una arquitectura

de dades flexible, optimitzada, segura i escalable, adaptant-se a les necessitats de cada organització i les àrees funcionals.

En la presentació del mòdul tens tres enllaços. El primer descriu què és el diagrama màgic de Gartner i com funciona. El segon és l'informe publicat per Gartner amb els resultats d'abril de 2023 on s'inclou el diagrama anterior. El tercer enllaç correspon a una pàgina de Gartner on s'analitzen cadascuna de les Plataformes d'Analítica i Intel·ligència de Negocis i permet comparar-les entre si.

Introducció a Power BI

Què és Power BI?

Power BI és:

- Una col·lecció de serveis de programari, aplicacions i connectors que funcionen conjuntament.
- Permeten connectar-se amb facilitat als orígens de dades i convertir orígens de dades sense relació entre si en informació coherent, interactiva i atractiva visualment. Les seves dades podrien ser un full de càlcul d'Excel o una col·lecció de magatzems de dades híbrides locals i basades en el núvol.
- Permet visualitzar i descobrir detalls analítics i compartir-los amb qualsevol o amb tots els usuaris que desitja.

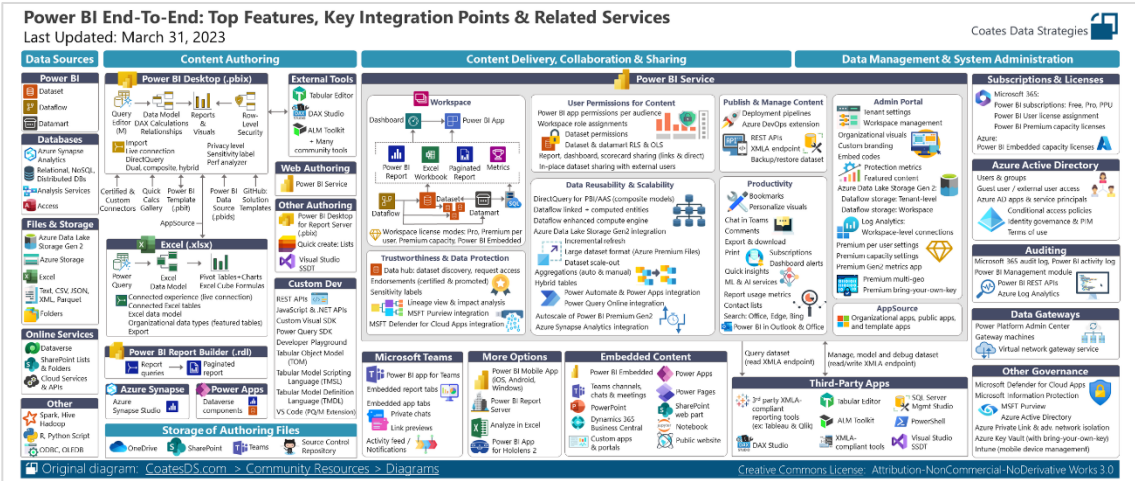
Components de Power BI

Power BI consta de diversos elements que funcionen de manera conjunta. Entre aquests elements es troben:

- **Power BI Desktop**, aplicació d'escriptori de Windows on podem connectar-nos a les dades, transformar-les, modelar-les i visualitzar-les. És una eina gratuïta que s'actualitza un cop al mes i hi ha un fòrum on es recullen les demandes dels usuaris i, les més votades, són implementades per l'equip de disseny del programa.
- **Servei Power BI**, servei de programari com a servei (SaaS) en línia on publiquem els informes i conjunts de dades creats a Power BI Desktop, creem panells i informes. Permet col·laborar i compartir amb la resta de l'organització amb els nostres clients.
- **Power BI Mobile**, aplicacions per a dispositius Windows, iOS i Android. Permeten accedir i treballar amb les dades mentre estem en moviment.
- **Power BI Report Builder**, per crear informes paginats i compartir-los al servei Power BI.
- **Power BI Report Server**, un servidor d'informes local en el qual pot publicar els informes de Power BI, després de crear-los a Power BI Desktop.

En el següent diagrama pots veure tots els components de Power BI així com la seva integració amb altres plataformes com Microsoft 365 i Teams.

En la presentació del mòdul tens un enllaç per descarregar un PDF amb la gràfica.



Flux de treball a Power BI

Existeixen diversos fluxos de treball a Power BI. El flux de treball més comú comença amb Power BI Desktop, on ens connectarem a diferents orígens de dades, els netejarem, transformarem i combinarem utilitzant l'Editor de Power Query. Com a resultat obtindrem el model de dades. A continuació, configurarem els elements de l'informe i un cop estigui llest l'informe el publiquem al servei Power BI.

En el servei de Power BI podrem crear nous informes i panells que podem compartir amb altres usuaris.

Per finalitzar interactuem amb els informes i panells des dels dispositius mòbils utilitzant Power BI Mobile. La següent imatge mostra aquest flux de treball.

