

Datawarehouse optimaliseert besluitvormingsproces voor investeerders

DOOR JOLANDA DE GROOT

Een datawarehouse biedt vele mogelijkheden om data te analyseren en zo de besluitvorming op diverse niveaus te ondersteunen en te verbeteren. Vastgoedbelegger Bouwinvest maakt gebruik van een datawarehouse en ondervindt hier veel voordelen van. Financial Investigator sprak met Marleen Bosma, Head of Global Research & Strategic Advisory bij Bouwinvest.

Wat is een datawarehouse?

‘Een datawarehouse is in essentie een grote verzameling centraal beheerde data afkomstig van zowel interne als externe bronnen. Een datawarehouse wordt gebruikt om de besluitvorming op diverse niveaus (operationeel, tactisch en strategisch) te ondersteunen en te verbeteren door middel

van de aggregatie en combinatie van data, analysetechnieken en reporting. Het gaat erom de ‘business intelligence’ te vergroten, processen efficiënter te maken en de kans op fouten te verkleinen.

Voor ons specifiek geldt dat we naast een datawarehouse ook een ‘analyse-omgeving’ hebben opgezet, waarmee grote hoeveelheden data vanuit het datawarehouse kunnen worden opgehaald. Binnen deze omgeving kunnen we nieuwe analysetechnieken toepassen, zoals machine learning.’

Op welke wijze maken jullie binnen het beleggingsproces gebruik van het datawarehouse?

‘Op velerlei wijze. We brengen markten in beeld, genereren forecasts en voeren scenario-analyses uit. Dit alles om onze strategische en tactische visie nog beter te kunnen formuleren. Het feit dat de wereld complexer, sneller, non-lineair en meer geïntegreerd is dan voorheen, vraagt om een andere benadering. Door grote hoeveelheden data, een goede infrastructuur en geavanceerde analysetechnieken te combineren, zijn we nu veel beter in staat om nieuwe ontwikkelingen te monitoren en hier pro-actief op in te spelen. Momenteel zijn we de koppeling tussen portefeuille-data en marktdata aan het neerzetten. Het uiteindelijke doel is beleggingsbeslissingen te optimaliseren voor onze investeerders.’

Hoe bepalen jullie de input voor het datawarehouse?

‘Wij hebben gekozen voor een unieke ‘tailor made’-oplossing, die we hebben opgezet in nauwe samenwerking met een

Een goede masterdata-validatie waarborgt dat er in het vervolgproces geen vervuiling plaatsvindt.



MARLEEN BOSMA

Het gaat erom de 'business intelligence' te vergroten, processen efficiënter te maken en de kans op fouten te verkleinen.

externe partij. Wij hebben de 'infrastructuur' aangelegd, waaraan vervolgens nieuwe bronnen gekoppeld kunnen worden. Zo zijn we niet afhankelijk van een externe partij, maar kunnen we veel in huis doen en verder doorontwikkelen. Onze externe partner heeft ons meegenomen in de vakgebieden 'master data management' en 'data science', maar vervolgens hebben we deze expertise zelf in huis gehaald. Dit stelt ons in staat om op alle facetten van het beleggingsproces gebruik te maken van het datawarehouse.'

Hoe gaan jullie om met inconsistenties of tegenstrijdigheden in data?

'In dergelijke gevallen is het essentieel dat het Master Data Management (MDM) in de basis goed is neergezet, dat er een duidelijke data-governance is en dat dit geheel organisatiebreed is afgestemd. Dit betekent niet veel meer dan nagaan welke checks er worden uitgevoerd op de data die het datawarehouse ingaan, zodat er geen 'vervuiling' ontstaat. We hebben bepaald welke ruwe data we als leidend zien bij inconsistenties, maar hetzelfde geldt voor definities, naamgeving en regionale indelingen. In de praktijk bleken er verschillende definities gehanteerd te worden voor Londen of Parijs. Als je naar de 'administrative boundaries' kijkt, vallen de steden onder meerdere regio's. Dan is de vraag: 'Welke neem je aan als basis?' In Nederland hebben we te maken met periodieke gemeentelijke herindelingen. Ook hierop moeten de data worden aangepast. Hiervoor hebben we een model ontwikkeld. Maar er zijn ook meerdere varianten voor yield. Het is essentieel om de juiste definitie te bepalen en toe te passen. Externe data worden immers vaak in verschillende eenheden aangeleverd. Dan is het van belang dat de data worden omgevormd naar volledig uitgeschreven getallen, met consistentie in de notatie.

Als het MDM eenmaal staat, kosten geautomatiseerde updates in dezelfde dataformats nagenoeg geen tijd meer. We hebben de data vanuit externe providers 'gechallenged' en richting hen ontbrekende of zelfs foutieve data kunnen aangeven, omdat we hier zo nauwkeurig doorheen zijn gegaan.'

Wat is de effectiviteit van de voorspellingen die jullie met het datawarehouse doen?

'Het is nog te vroeg om de voorspellingen te toetsen, maar we zijn veel beter in staat om bijvoorbeeld scenario-analyses uit te voeren of stress-testen uit te voeren. Daarnaast hebben we reeds een efficiencyslag gemaakt in de 'managerial' zin. Waar we voorheen lange tijd bezig

waren met handmatige date-updates, aanpassingen van formules of links in applicaties als Excel, is dit nu een vrijwel geheel geautomatiseerd proces met een kleine foutmarge. De basis van de analyses staat in programmeercode en de data worden periodiek automatisch geüpdatet. De extra tijd die hiermee vrijkomt, kunnen we weer gebruiken voor de doorontwikkeling van analyses, het aanscherpen van resultaten en het verkrijgen van nieuwe inzichten. Daarnaast geeft het ons meer tijd voor een kritische beoordeling van de uitkomsten en om de kwalitatieve discussie te voeren teneinde de beleggingsstrategie te optimaliseren en te monitoren.'

Zijn er (internationale) standaarden voor het verzamelen, organiseren en beheren van data?

'Op onderdelen zijn deze er wel, maar de wijze waarop dit in de praktijk wordt ingericht, kent veel varianten. Uiteindelijk is het doel van belang; welke strategische vragen wil je beantwoorden? Afhankelijk daarvan heb je bepaalde analysemethoden en relevante data nodig. Deze data kunnen op zeer uiteenlopende wijzen gestructureerd zijn. Zo is er bij ons sprake van tijdsreeksen, maar ook van geografische data, hiërarchisch ingedeelde data of statische data. Waar tijdsreeksen (bijvoorbeeld historische rendementen) relatief eenvoudig in het datawarehouse zijn op te nemen, blijken de geografische data uitdagender. Hiërarchisch heb je hierbij te maken met een gecompliceerde boomstructuur, die ook nog eens periodiek verandert. Denk maar aan gemeentelijke herindelingen of Brexit. De datastructuur zal hierop moeten worden afgestemd. Qua beheer is het afhankelijk van het type data, de hoeveelheid data en de update-frequentie of het relevant is dat de data automatisch geüpdatet worden middels een 'push'-functie dan wel via een upload.'

Maken jullie ook gebruik van Artificial Intelligence bij het interpreteren van data?

'Jazeker. In de analyse-omgeving hebben we de mogelijkheid om analyses te programmeren en een breed scala aan machine learning-technieken toe te passen. Deze analyse-technieken worden bijvoorbeeld ingezet bij het beoordelen van de data, het uitzoeken welke parameters van belang zijn, het testen en toetsen van relaties tussen data en het trainen van modellen.

Data science is een complex expertisegebied waarvoor specialistische kennis nodig is. In de vastgoedindustrie is deze van oudsher niet direct aanwezig. Dit is dus een expertise die we hebben moeten opbouwen. Naast 'on the job'-training en kennisontwikkeling via onze externe partner, hebben we ook nieuwe collega's met deze achtergrond in huis gehaald. Een voordeel daarvan is dat we niet verkokeren qua visie. Zo houden we een brede, kritische blik op wat we doen, om zo het beste resultaat te behalen voor de organisatie en onze klanten. En dit zal alleen maar toenemen in de toekomst, naarmate we verder doorontwikkelen.' «