**Projectmateriaal**

**Uitgangssituatie: De grote fusie en uitbreiding**

De retailwereld van outdoorartikelen is recent behoorlijk opgeschud. Een voorheen kleine speler, Great Outdoors, heeft met de bouw van een Data Warehouse (DWH) en verschillende dashboards een zeer sterke marktpositie weten te veroveren. Zij dreigt nu niet alleen een van de grootste spelers te worden op de kampeer-, maar als gevolg van een organisatie-uitbreiding ook op de sportmarkt.

Het bedrijf AenC, een kledingdetailhandel in juist deze sportmarkt met meerdere filialen, besluit op advies van hun financieel adviseurs en aandeelhouders te gaan fuseren met 2 organisaties:

* AdventureWorks: een grote fietsenhandelaar. Ook dit bedrijf heeft namelijk als missie om (potentiële) klanten te stimuleren om meer tijd "outdoor" door te brengen.
* Northwind: een grote voedselhandelaar. AenC hoopt op deze manier een alles-in-éénorganisatie te worden voor mensen die (een aantal dagen lang) willen sporten.

Het fusiebedrijf zal *United Outdoors* gaan heten. Men is zich er zeer van bewust dat om de grootste concurrent, Great Outdoors, de baas te zijn, er snelle, maar tegelijkertijd goede beslissingen genomen moeten worden. Het uitvoeren van hedendaagse én voorspellende data-analyses wordt erkend als hét middel zowel de snel- als de juistheid van deze te nemen besluiten te waarborgen. Om dit mogelijk te maken willen de drie bedrijven zo snel mogelijk hun bestaande data-infrastructuur bundelen in een DWH. “Schaalbaarheid” is hier een belangrijke eigenschap: het moet zo makkelijk mogelijk zijn om data uit nieuwe bronnen aan het DWH toe te voegen.

Daarnaast bestaat binnen *United Outdoors* een driestrijd over de verwachtingen omtrent de effectiviteit en uitvoerbaarheid van de bovengenoemde data-analyses.

* Een groep mensen (de “data-pessimisten”) is stellig van mening dat beslissingen vooral genomen moeten worden op basis van gezond verstand, in plaats van “keiharde cijfers” in beschouwing te nemen. De link tussen data-analyses en de bovengenoemde te nemen besluiten zien zij (nog) niet.
* Een groep (de “data-twijfelaars”) ziet wél het nut in van het uitvoeren van data-analyses, maar twijfelt of deze tot correcte beslissingen kunnen leiden, gegeven de hoeveelheden en soorten verzamelde data.
* De derde groep (de “data-optimisten”) wil liever vandaag dan morgen aan de slag met het wegnemen van de zorgen van de bovengenoemde twee groepen. Volgens hen verzamelt United Outdoors voldoende data om in staat te zijn precies de juiste beslissingen te nemen. Bovendien zijn zij van mening dat er genoeg externe databronnen op internet te vinden zijn die in analyses geïntegreerd kunnen worden om tot nóg accuratere voorspellingen te komen.

**Projectopdracht**

Aan jouw projectteam wordt door United Outdoors het volgende gevraagd:

1. Een Plan van Aanpak / andersoortig projectplan met daarin minimaal een taakverdeling, tijdsplanning en samenwerkingsafspraken. De richtlijnen uit dit plan moeten in 3 weken uit te voeren zijn.
2. Een schaal- en overdraagbaar Source Data Model (SDM) in SQL Server Management Studio (SSMS). Hierin zijn alle databasetabellen van alle brondatabases opgenomen, inclusief hun attributen en de correcte primaire-vreemde-sleutelrelaties. Elke te fuseren organisatie heeft één brondatabase.
3. Een schaal- en overdraagbaar Data Warehouse (DWH) in SQL Server Management Studio (SSMS). Hierin is in elk geval een groot deel van de feit- én dimensietabellen uit de brondatabases opgenomen. Alle data die heel erg sterk aan elkaar is gelinkt, is gedenormaliseerd tot één tabel. Hiervoor vinden dus de nodige samenvoegingen plaats, d.m.v. dimensioneel modelleren en implementeren. Denk bijvoorbeeld aan één tabel die álle klantinformatie bevat, één tabel die álle productinformatie bevat, enzovoort.
4. Een dashboard met visualisaties van kwantitatieve sturingsinformatie. Hierin moeten de in-house data-analisten de resultaten van data-analyses inzichtelijk kunnen maken. Een hoog interactief karakter van dit dashboard is zeer gewenst: denk aan zinvolle filters, drill-downstructuren, enzovoort. In eerste instantie kan dit dashboard gemaakt en gepresenteerd worden in Power BI, maar op de langere termijn willen de data-analisten van United Outdoors dat het dashboard ook via een makkelijk toegankelijke webomgeving te raadplegen is.

Extra aandachtspunten

* Tijdens Bouldermomenten wordt kritisch gekeken in hoeverre de 2,5 projectweken **ten volle** zijn benut om het resultaat af te krijgen. Met andere woorden: men staat uitvoerig stil bij de vraag of het projectresultaat in lijn is met wat we verwachten qua hoeveelheid verricht werk in 2,5 week.
* Men is bereid om degenen die een effectieve webomgeving demonstreren een nóg hogere beloning uit te keren, dan degenen die zich beperken tot Power BI Desktop.

**Context**

De context van United Outdoors is als volgt:

* Het bedrijf is nu primair bezig met het regelen van de fusie en geeft jullie volledige autonomie over het opzetten van de data-analyse binnen de hierboven geschetste opdracht
* Omdat de medewerkers van United Outdoors niet beschikbaar zijn zal jouw team de benodigde kennis voor de uitvoering van de opdracht zelfstandig elders moeten inwinnen.
* De (toekomstige) data-analisten van United Outdoors zijn sterk in het bekijken van data, maar niet technisch begaafd. Het gebruik van jargon uit de outdoorbranche wordt door de analisten zeer op prijs gesteld.
* De brondata wordt aangeleverd as-is. United Outdoors kan niet garanderen dat alle data correct is en dat de verschillende bronnen (optimaal) op elkaar aansluiten. Het kan zijn dat op een later moment, eventueel na afloop van het project, geupdate versies van de bronbestanden worden aangeleverd. Uiteraard moeten de dashboards dan automatisch updaten.
* United Outdoors vraagt zowel kwaliteit (zinnige en goed functionerende overzichten in het dashboard) als kwantiteit (zoveel mogelijk relevante overzichten).

**Databestanden**

De volgende databestanden zijn nodig tijdens het uitvoeren van het project:

* De database van **[AenC](https://brightspace.hhs.nl/d2l/common/viewFile.d2lfile/Content/L2NvbnRlbnQvZW5mb3JjZWQvMTAxMDYzLUgtU0UtREVEUy0yNF8yMDI0X1NQUklOR18yL2FlbmMuc3FsaXRl/aenc.sqlite?ou=101063" \t "_self)**, dit is een downloadbaar .sqlite-bestand.
* De database van **[AdventureWorks](https://brightspace.hhs.nl/d2l/common/viewFile.d2lfile/Content/L2NvbnRlbnQvZW5mb3JjZWQvMTAxMDYzLUgtU0UtREVEUy0yNF8yMDI0X1NQUklOR18yL0FkdmVudHVyZVdvcmtzLm1kYg/AdventureWorks.mdb?ou=101063" \t "_self)**, dit is een downloadbaar Access-bestand. Tip voor niet-Windows-gebruikers: helaas wordt MS Access, net als Power BI, alleen ondersteund door Windows. Je kan dit soort bestanden raadplegen met Python Pandas en vervolgens ook exporteren (bijvoorbeeld naar een reeks CSV- of Excelbestanden, of iets dergelijks).
* De database van [**Northwind**](https://brightspace.hhs.nl/d2l/common/viewFile.d2lfile/Content/L2NvbnRlbnQvZW5mb3JjZWQvMTAxMDYzLUgtU0UtREVEUy0yNF8yMDI0X1NQUklOR18yL05vcnRoV2luZC50eHQ/NorthWind.txt?ou=101063), dit is een downloadbaar .txt-bestand. Maak in SQL Server Management Studio een database genaamd "Northwind" aan, open de SQL-editor en kopieer & plak ten slotte de inhoud van dit bestand daarin.