

E-BOOK

De Microsoft Fabric bouwstenen
van Maestro

Axians

Hogeweg 41
5301 LJ Zaltbommel
Tel: +31 88 988 90 10
axians.nl/maestro

The best
of ICT with
a human
touch



Microsoft Fabric

“Microsoft Fabric biedt vrijwel alle benodigde tools om data uit verschillende bronnen op te halen, deze te bewerken en vervolgens opnieuw te gebruiken.”

Microsoft Fabric is een SaaS-oplossing van Microsoft die vrijwel alle benodigde tools biedt om data uit verschillende bronnen op te halen, deze te bewerken en vervolgens opnieuw te gebruiken. Op het eerste gezicht lijkt Fabric eenvoudig te gebruiken dankzij de vriendelijke graphical user interface. Desondanks moeten er bij de inrichting van een Data & AI platform met Microsoft Fabric heel veel keuzes worden gemaakt. Kies je bijvoorbeeld voor een Dataflow of een Data Pipeline om data op te halen? Ook is het belangrijk om een logische gelaagdheid in de data op te bouwen, bijvoorbeeld met het Medallion-principe. De vraag is hoe je dit in de praktijk toepast en welke waarde je eruit haalt. Je zult ook moeten beslissen waar je deze

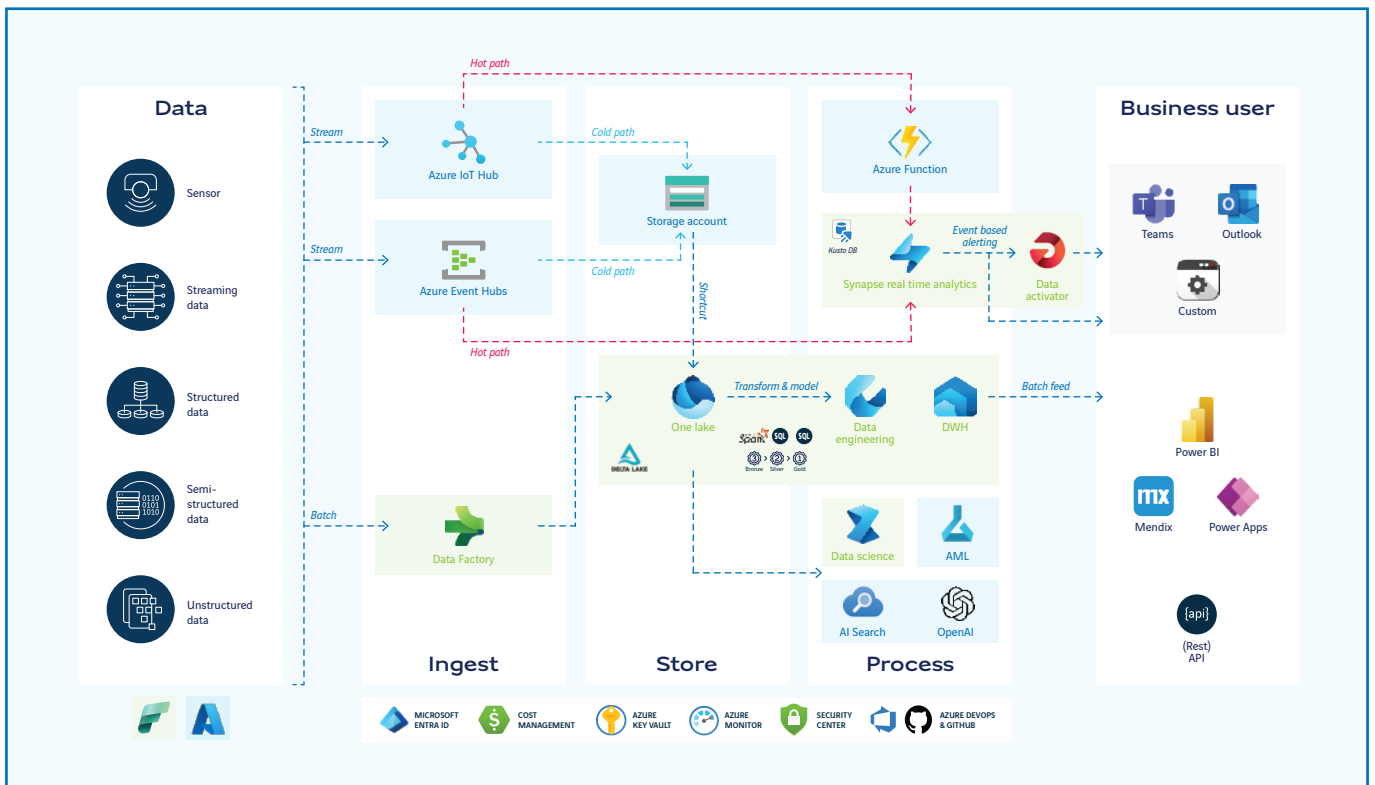
verschillende datalagen opslaat. Kies je voor het Lakehouse of ga je ook gebruik maken van datawarehousing? Met Fabric heb je ook de mogelijkheid om het data mesh-principe te omarmen. Dit biedt krachtige ondersteuning voor data governance, maar is ook complex en tijdrovend om in te richten. Kortom, er zijn tal van componenten beschikbaar binnen Microsoft Fabric om uiteindelijk tot het ideale datafundament voor jouw organisatie te komen. Maar welke componenten heeft jouw organisatie écht nodig? Hoe richt je deze technisch in? En hoe zorg je ervoor dat de omgeving veilig is en blijft? Dit zijn nog steeds relevante vragen en het kost veel tijd om deze randvoorwaarden te onderzoeken.



Introductie Maestro

Met het Maestro dataplatform van Axians zijn de antwoorden op al deze vragen al afgevangen en geïntegreerd in het ontwerp. Hierdoor biedt Maestro een quick start voor de realisatie van een toekomstbestendig Data & AI platform op maat met Microsoft technologie. De inrichting van Maestro vindt grotendeels plaats middels de Infrastructure-as-Code opzet. Dit bespaart veel tijd, kosten en risico's op onnodige (menselijke) fouten.

Bij het platform hoort altijd een Foundation wat ervoor zorgt dat de Analytics-oplossing goed landt op de Azure-omgeving. Hiermee borgen we de toegang en veiligheid. Aanvullend op de Foundation bieden we Maestro in twee Microsoft varianten aan: Azure Synapse en Microsoft Fabric. In dit document lichten we componenten en ontwerpkeuzes toe van Maestro met Microsoft Fabric.



Figuur 1: De Microsoft Fabric bouwstenen van Maestro

Data ontsluiten en verladen met Maestro

“Maestro biedt de mogelijkheid om uiteenlopende data te verwerken binnen het centrale dataplatform, waardoor extra tools overbodig worden.”

Tegenwoordig komen we allerlei verschillende typen data uit diverse bronnen tegen. Om strategisch de juiste keuzes te maken en diepere inzichten te verkrijgen, wil je alle soorten data en bronnen kunnen verwerken. Maestro biedt de mogelijkheid om deze uiteenlopende data te verwerken binnen het centrale dataplatform. Dit maakt extra tools overbodig. Een cruciaal bouwsteen van Maestro is **Onelake**, de centrale opslaglocatie voor alle datatypes. De grote voordelen? Geen datasilo's en tijdrovende integraties meer, verbeterde datakwaliteit en datamanagement én de data kan tevens voor andere analytics toepassingen gebruikt worden.

Om data te verladen bevat Maestro standaardontwerpen voor connectoren. Deze zijn vaak al eerder ontwikkeld of kunnen indien nodig worden gemaakt. Deze connectoren, ook wel pipelines genoemd, worden gemaakt met **Data Factory** binnen Fabric. De Data Factory pipelines halen data uit de bron en brengen deze naar Onelake, waarbij Data Factory de planning en orkestratie van deze activiteiten faciliteert. Een connector zorgt ervoor dat zowel data als metadata uit de bron worden opgehaald en naar Onelake worden gebracht. Metadata biedt inzicht in de

eigenschappen en structuur van de data en kan deze verkrijgen, bijvoorbeeld door time traveling toe te passen. Via Dynamic Management Views verkrijgen we inzicht in de metadata, zoals de tabellen in de database en hun schema's. De metadata wordt opgeslagen in een hub-database, waardoor het bronsysteem minder belast wordt en we met één query volledige metadata kunnen ophalen. Dit stelt ons ook in staat om de geschiedenis van de metadata bij te houden en snel in te grijpen bij fouten.

De connectoren zijn met behulp van Data Factory en notebooks zo ontworpen dat data parallel geladen kan worden, deze functionaliteit is niet standaard mogelijk in Microsoft Fabric. **Parallel laden** voorkomt dat tabelverladingsprocessen op elkaar moeten wachten. Zeker als je steeds meer data moet verladen is deze functionaliteit belangrijk. De pipelines hoeven niet op elkaar te wachten, wat de efficiëntie verhoogt. Als je niet kiest voor parallel laden, kan het verladen soms te lang duren en impact hebben op de werking van een dataplatform. Daarnaast kunnen we met de Maestro connectoren kiezen welke **specifieke objecten** uit de bron geladen moeten worden. Daarmee zorgen we dat alleen de objecten worden opgehaald waar behoefte aan is. Op een later moment toch nog extra objecten ophalen? Dan kan dat heel eenvoudig worden aangegeven. Ook voorzien de Maestro connectoren in het **incrementeel verladen** van data. Daarmee is het niet nodig om continu de volledige dataset op te halen maar halen we alleen de data op waar een wijziging op heeft plaats gevonden, en dat scheelt weer veel compute.



De verwerking van data met Maestro

“Data Factory pipelines binnen Fabric zijn essentieel voor het beheren en automatiseren van datastromen.”

Data factory

Binnen Maestro is Data Factory één van de belangrijkste bouwstenen voor het verwerken van data. Data Factory pipelines binnen Fabric zijn essentieel voor het beheren en automatiseren van datastromen. Ze zorgen voor de efficiënte overdracht en transformatie van data van diverse bronnen naar het centrale opslagpunt: OneLake. Deze pipelines integreren data uit verschillende bronnen, zoals on-premises databases, cloud-opslag, API's en SaaS-toepassingen, waardoor data uit uiteenlopende systemen samenkomt voor analyse in OneLake.

OneLake

Een andere cruciaal bouwsteen binnen het Maestro dataplatform is OneLake, de kern van onze dataverwerkingsoplossing. Door gebruik te maken van Data Factory, zoals eerder beschreven, stroomt data naar OneLake, dat zowel door Microsoft als Axians wordt omarmd als de centrale opslaglocatie binnen het platform. OneLake is een essentieel onderdeel van Microsoft Fabric en helpt organisaties datasilo's te vermijden door gegevens centraal op te slaan. De kenmerken van OneLake zijn:

- **Eén data lake voor de hele organisatie:** OneLake is een enkele, geïntegreerde, logische data lake beschikbaar voor de gehele organisatie. Het ondersteunt elk type bestand - gestructureerd of ongestructureerd - en slaat alle tabulaire gegevens op in Delta Parquet-formaat.

- **Eén kopie van gegevens voor gebruik met meerdere analytische engines:** OneLake richt zich op het verbeteren van samenwerking door één uniforme data lake te bieden voor alle analytische behoeften van de organisatie.

- **Governance en samenwerking:** Binnen een tenant kunnen meerdere domains, sub-domains en workspaces worden gemaakt, waardoor verschillende delen van de organisatie het juiste eigenaarschap en toegang tot hun specifieke data krijgen. Daarmee is het mogelijk om een Data mesh architectuur te omarmen. Het ontwerp van Maestro hanteert in de basis dan ook een data mesh architectuur binnen de medallion architectuur.

Om alles uit OneLake te halen, is het essentieel om de datagelaagdheid, een folderstructuur en andere organisatorische aspecten correct in te richten. Voor het historiseren van data passen we Synapse Data Engineering toe en vervolgens Synapse Data Warehousing om het warehouse op te bouwen en de data klaar te stomen voor Power BI.

Synapse Data engineering

Binnen Synapse Data Engineering op Fabric gebruiken we notebooks om brondata inclusief alle (meta)data wijzigingen historisch te kunnen opslaan. Notebooks bieden een interactieve omgeving waar je code kunt schrijven, uitvoeren en visualiseren om dataprocessen te beheren. Dit stelt je in staat om tijdreizen toe te passen en historische data-analyses uit te voeren, wat waardevolle inzichten oplevert voor strategische besluitvorming. Het gebruik van notebooks bevordert de prestaties en schaalbaarheid van de oplossing. Als je met notebooks aan de slag gaat op Fabric moet er code worden geschreven en moeten deze worden geconfigureerd om de data periodiek op te halen en op te slaan in een historische opslag, bijvoorbeeld een data lake in Delta Lake-formaat. Met Maestro is dit al voor je gebouwd. De Maestro notebooks zijn namelijk “out of the box” beschikbaar voor Maestro gebruikers.



“De Medallion architectuur zorgt voor efficiënte data-integratie en -transformatie.”

Synapse Data warehousing

Met Synapse Data Warehousing binnen Microsoft Fabric bouw je een robuust en schaalbaar datawarehouse op. Ook met Maestro gebruiken we Synapse Data Warehousing om binnen Fabric het datawarehouse op te zetten. Bij het opzetten van je datawarehouse is een ontwerp van het datamodel enorm van belang. Je identificeert hiermee de belangrijkste feiten en dimensies. Het doel hiervan is om de opbouw van het datawarehouse aan te laten sluiten op de behoeften van de eindgebruikers van de data. Je borgt hiermee dat transformaties binnen het datawarehouse worden gedaan en niet binnen de Power BI rapporten. Op deze manier creëer je één versie van de waarheid waar de hele organisatie op kan vertrouwen. Met Synapse Data warehousing maak je via SQL scripts views en tabellen die specifiek gebruikt kunnen worden voor business analytics. Het is belangrijk om de data te verrijken en te modelleren volgens de behoeften van de business, zodat deze klaar is voor analyse. De business views bieden een gestructureerd overzicht van de gegevens, klaar voor gebruik in rapportages en dashboards.

Medallion architectuur

Binnen Maestro hebben we verschillende standaarden die algemeen bekend zijn in de wereld van Data & Analytics. Zo ook voor een logische opbouw van de verschillende datalagen. Binnen Maestro wordt de Medallion architectuur toegepast waardoor data in drie lagen wordt verdeeld: Bronze, Silver en Gold.

- ▶ **Bronze Layer:** Onbewerkte, originele data wordt hier opgeslagen. De Maestro-accelerator automatiseert het proces om data te ontsluiten uit bronnen en op te slaan in de bronze layer, evenals het doorzetten van de noodzakelijke data naar de volgende laag.
- ▶ **Silver Layer:** De data wordt opgeschoond, gevalideerd, gestructureerd en in hetzelfde format opgeslagen om de betrouwbaarheid en consistentie te waarborgen.
- ▶ **Gold Layer:** Data wordt verder verrijkt, geaggregeerd en geoptimaliseerd voor analytische doeleinden en -rapportages.

De Medallion architectuur zorgt voor efficiënte data-integratie en -transformatie, betrouwbare datastromen en het ondersteunt geavanceerde data-analyses en business intelligence door gestructureerde en verrijkte data te bieden. De opbouw van deze verschillende lagen verdelen wij met Maestro over verschillende Domains en Workspaces. De Maestro Data mesh architectuur (zoals hierboven benoemd) zorgt binnen deze verschillende layers voor de inrichting van de juiste verantwoordelikheden, toegang en samenwerking.

Workspaces en Domains

Met Domains en Workspaces binnen Microsoft Fabric kun je de verschillende resources en standaarden gestructureerd inrichten en beheren op het Maestro platform. Domains maken het mogelijk om data te groeperen en af te schermen. Met deze gestructureerde scheidingen binnen het platform, borg je dat organisatieonderdelen gescheiden toegang en eigenaarschap hebben over hun data en projecten. Workspaces zijn specifieke werkomgevingen binnen een domain waar teams kunnen samenwerken aan data-analyse, rapportages en modellering. Workspaces bieden tools en resources voor data-integratie, transformatie en visualisatie. Deze structuur bevordert de interne samenwerking, verhoogt de efficiëntie en maakt het mogelijk om data governance en beveiliging op maat te beheren, terwijl het overzichtelijk en georganiseerd blijft. Met Maestro hebben wij een standaard structuur ingericht om te werken met Domains en Workspaces. Met deze structuur bieden we de handvatten om de juiste inrichting op te zetten en daarmee een overzichtelijke oplossing te creëren.

Het gebruik van Maestro

“Je wordt als organisatie ontzorgd naar wens, maar je kunt uiteraard ook zelf aan de knoppen zitten van Maestro.”

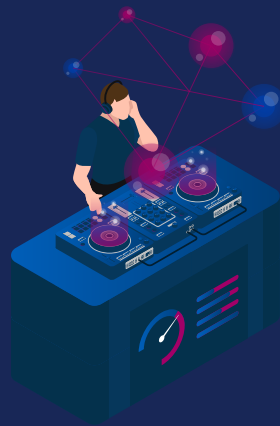
In dit e-book zijn de componenten en ontwerpkeuzes van Maestro met Microsoft Fabric gedetailleerd besproken. Echter, Maestro is geen standaard software-oplossing maar een accelerator die je helpt om snel een volledig en veilig dataplatform op te zetten met Microsoft Fabric en verder te ontwikkelen. Het vormt de kern voor alle analytics binnen een organisatie. Met Maestro kunnen continu nieuwe gegevensbronnen en rapporten worden toegevoegd. Bovendien kan je Maestro BI eenvoudig uitbreiden met de Maestro IoT en/of Maestro AI module. Het Maestro dataplatform wordt aangeboden in een combinatie van technologie en

dienstverlening. De technologische oplossing in de vorm van een dataplatform, is door Axians ontwikkeld en beschikbaar voor klanten. De dienstverlening zorgt voor de garantie dat het dataplatform functioneel en innovatief blijft. Zodat jij je volledig kunt richten op het benutten van data zonder zorgen te hebben over het onderhoud of innovatie conform de Microsoft ontwikkelingen met Fabric. Je wordt als organisatie ontzorgd naar wens, maar je kunt uiteraard ook zelf aan de knoppen zitten van Maestro. Dit stemmen we samen af, kijkend naar de aanwezige expertise en beschikbaarheid van resources binnen jouw organisatie.



Maestro BI

Maestro BI is de generieke analytics-module. Deze bevat functionaliteiten voor het ontsluiten van bronnen richting een ruwe data laag en voor de opbouw van een historische laag op basis van ruwe data.



Maestro IoT

Maestro IoT is een generiek component waarmee je eenvoudig een IoT use case bouwt.



Maestro AI

Met de AI-module maken we het mogelijk om ongestructureerde documenten naar het centrale data lake te brengen, te structureren en te analyseren. Denk bijvoorbeeld aan analyse op duizenden e-mails, documenten of pdf's.

Onze aanpak van Maestro

“Onze aanpak start met het definiëren van data scenario's en het opstellen van een roadmap, waarbij het doel van het Data & Analytics platform centraal staat.”

Onze aanpak start met het definiëren van data scenario's en het opstellen van een roadmap, waarbij het doel van het Data & Analytics platform centraal staat. Zelfs als er al een dataplatform in gebruik is. We benaderen een Maestro traject namelijk idealiter vanuit de eindgebruiker. Onze jarenlange ervaring leert ons namelijk dat we hiermee de grootste successen en hoogste adoptiegraad behalen.

Tijdens een ééndaagse workshop bepalen we samen de lange-termijn visie en prioriteiten. We beginnen vaak met een Minimum Viable Product (MVP), waarbij we het dataplatform uitrollen met één of twee datastromen om de werking van Maestro te demonstreren. Na enkele maanden evalueren we de resultaten en de gebruikservaringen. Vervolgens implementeren en adopteren we Maestro Fabric volledig binnen de organisatie volgens het opgestelde projectplan.

Meer weten over Maestro?

Neem vrijblijvend contact op met Jordy Ravesteijn



Jordy Ravesteijn

BD Manager Data & Analytics

+31 6 2557 8867

jordy.ravesteijn@axians.com

Microsoft & Axians

Met de technologie van Microsoft durven wij iedere uitdaging op het gebied van Data, Analytics en AI aan. Axians en Microsoft werken al jaren samen, op zowel strategisch- als operationeel niveau, om de beste Microsoft oplossingen neer te zetten. Het resultaat:

- ▶ Een lange (meer dan 15 jaar) ervaring met Microsoft solutions;
- ▶ Een bewezen trackrecord met meerdere Gold Partner en Partner of the Year statussen, Advanced Specializations en mooie klanten;
- ▶ Meer dan 700 Microsoft experts bij Axians welke dag in dag uit bezig zijn om de beste technologieën en innovatieve oplossing te leveren;
- ▶ Onze technologieën en oplossingen leveren wij met de meest passende SLA. Daarmee ontzorgen we onze klanten volledig en kunnen zij zich richten op het creëren van business waarde. Onze SLA's zijn flexibel en bewegen met jouw situatie mee. Groeit de kennis in jouw organisatie? Dan wil je misschien wel meer zelf doen en passen we de SLA aan.



The best
of ICT with
a human
touch



axians

Hogeweg 41
5301 LJ Zaltbommel
Tel: +31 88 988 90 10
axians.nl/maestro