Onderzoeksvoorstel

Optimalisatie van de back-upstrategie voor Azure PostgreSQL en MySQL databases bij Forvis Mazars: Een proof of concept met immutabele opslag en automatische back-ups.

Bachelorproef, 2024-2025

Naoufal Bouazzaoui

E-mail: naoufal.bouazzaoui@student.hogent.be
Co-promotor: R. Tetaert (Forvis Mazars, remy.tetaert@mazars.be)

Samenvatting

In deze bachelorproef wordt een optimalisatie van de back-upstrategie voor de Azure PostgreSQL en MySQL databases bij Forvis Mazars onderzocht, de focus ligt voornamelijk op immutabele opslag en automatische back-ups. Het doel is om de back-upstrategie van Forvis Mazars te optimaliseren en het resistent te maken tegen ransomware-aanvallen. Daarnaast wordt er ook een Proof-of-Concept (PoC) uitgevoerd, waarin immutabele opslag wordt geïmplementeerd om ervoor te zorgen dat back-ups onveranderlijk zijn na opslag. Verder worden geautomatiseerde back-ups geïmplementeerd om de back-ups efficiënter te maken en de consistentie van de back-ups te verbeteren. In de state-of-the-art ligt de focus op bestaande back-upstrategieën, zoals cloud back-ups, on-premise back-ups, en offline back-ups. De methodologie omvat een literatuurstudie, een analyse van de huidige back-upstrategie bij Forvis Mazars, en de ontwikkeling van een PoC. De verwachte resultaten zullen de verbeterde beveiliging en efficiëntie van de back-upstrategie aantonen, met als doel het minimaliseren van de risico's op dataverlies en het waarborgen van bedrijfscontinuïteit.

Keuzerichting: System & Network Administrator
Sleutelwoorden: Back-ups, Security, Ransomware

Inhoudsopgave

1. Inleiding

Ransomware-aanvallen zijn één van de meest voorkomende aanvallen dat een organisatie kan treffen de dag van vandaag. Om gegevensverlies tegen te gaan in geval van een aanval moeten bedrijven altijd een back-upplan klaar hebben in geval van een incident. Het niet hebben van een back-upplan of het hebben van een suboptimaal plan kan leiden tot een groot verlies op financieel vlak. Daarnaast kan dit ook zorgen voor het verliezen van cruciale informatie en als laatste kan dit de reputatie van een organisatie sterk doen dalen, aangezien niemand in zee wilt gaan met een bedrijf dat niet goed beveiligd is of niet goed voorbereid is op uitzonderlijke incidenten. Om deze redenen is het van groot belang voor een bedrijf om een doordacht, robuust en veilig back-upplan te hebben.

Het doel van deze bachelorproef is het optimaliseren van het back-upplan van Forvis Mazars. Dit bedrijf maakt gebruik van 2 soorten databases in Azure, eenderzijds een PostgreSQL databank en anderzijds een MySQL databank. Van deze databanken worden er automatische alsook manuele back-ups gemaakt. De automatische back-ups gebeuren door Azure zelf en de manuele back-ups worden per database uitgevoerd. Stel dat applicatie A een nieuwe versie heeft, dan zal er eerst een back-up genomen worden van de database vooraleer de nieuwe versie uitgerold wordt. Echter zijn er nog bepaalde verbeteringen mogelijk, zoals het veiliger opslaan van deze back-ups met behulp van technieken als immutable storage en het instellen van geautomatiseerde dagelijkse, wekelijkse en maandelijkse back-ups. Daarbij is een belangrijk aandachtspunt dat de databanken beter beveiligd moeten worden tegen cyberaanvallen aangezien gegevens in zo'n situatie versleuteld of vernietigd kunnen worden.

De doelgroep van dit onderzoek bestaat uit de

IT-professionals en vooral de systeembeheerders van Forvis Mazars, die verantwoordelijk zijn voor het beheer van de back-ups en de beveiliging van gegevens binnen de organisatie alsook het herstellen van alle gegevens na een incident.

De onderzoeksvraag die onderzocht zal worden is: "Hoe kan de back-upstrategie van de Azure PostgreSQL en MySQL databases bij Forvis Mazars worden geoptimaliseerd door het implementeren van automatische back-ups en het veilig opslaan van deze back-ups om gegevensverlies te minimaliseren?" In dit onderzoek wordt onderzocht hoe de bestaande back-upoplossingen kunnen worden verbeterd, zodat Forvis Mazars in geval van een incident goed voorbereid is en geen informatie verliest. De onderzoeksvraag kan onderverdeeld worden in volgende kleinere deelvragen:

- Hoe veilig en betrouwbaar zijn de huidige back-upoplossingen van Forvis Mazars voor Azure PostgreSQL en MySQL databases?
- Welke rol speelt immutabele opslag in het beschermen van back-ups tegen ransomware en andere vormen van dataverlies?
- Wat zijn de belangrijkste uitdagingen bij het integreren van immutabele opslag met Azure cloud back-upsystemen?
- Hoe kan er voor de Azure PostgreSQL en MySQL databases een automatische back-upstrategie worden geïmplementeerd?

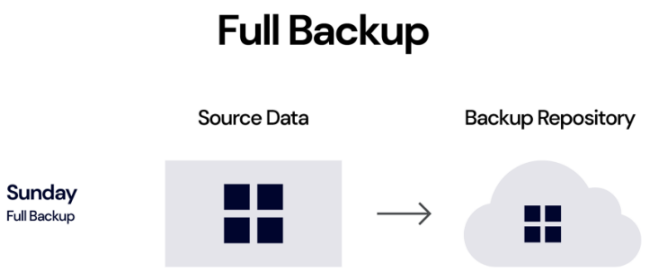
Het uiteindelijke doel van dit onderzoek is om ervoor te zorgen dat de Azure databanken van Forvis Mazars een geoptimaliseerd back-upplan hebben dat veilig en efficient is. Het plan moet immuun zijn tegen ransomware-aanvallen en daarnaast moet het ook geautomatiseerd zijn. Daarbij zal de tijd bij een herstel vanuit een back-up ook onderzocht worden. Om het back-upplan te testen zal er een Proof-of-Concept (PoC) opgesteld worden om alles grondig te testen in een testomgeving.

2. Literatuurstudie

Bedrijven moeten hun data goed beschermen om succesvol te zijn. Echter vormen back-ups van de databases vaak een zwakke schakel in de beveiligingsketen van gegevensbescherming. Hoewel veel organisaties zich richten op het beveiligen van hun actieve databases, worden de back-ups vaak over het hoofd gezien, wat een groot risico met zich meebrengt. Back-ups worden meestal offsite opgeslagen, bijvoorbeeld op tape, en zijn daardoor vatbaar voor verlies of diefstal (Cherry, 2015). Dit maakt het van essentieel belang om back-ups goed te beveiligen, bijvoorbeeld door encryptie. Echter, bij het kiezen van een encryptieoplossing is het belangrijk om een evenwicht te vinden tussen de sterkte van de encryptie en de impact op de prestaties van de server omdat sterke encryptie meer resources nodig heeft en het kan de beschikbaarheid van de back-ups veranderen. In de afgelopen jaren is er een scherpe toename van ransomware-aanvallen gericht op bedrijven, waarbij het aantal getroffen organisaties is gestegen van ruim 2.700 naar bijna 4.900 in slechts twaalf maanden. Deze toename laat zien hoe vastberaden en steeds slimmer ransomwaregroepen worden in hun aanvallen. Wat bijzonder zorgwekkend is voor de bedrijfswereld, is de trend van herhaalde aanvallen op bedrijven, waarbij sommige organisaties binnen korte tijd door meerdere ransomwaregroepen worden getroffen (Dikbiyik e.a., 2024). Dit wijst erop dat cybercriminelen actief profiteren van momenten van kwetsbaarheid om bedrijven in hun zwakste momenten opnieuw aan te vallen, wat de noodzaak voor robuuste preventieve maatregelen benadrukt.

2.1. back-upstrategieën

2.1.1. Full back-up



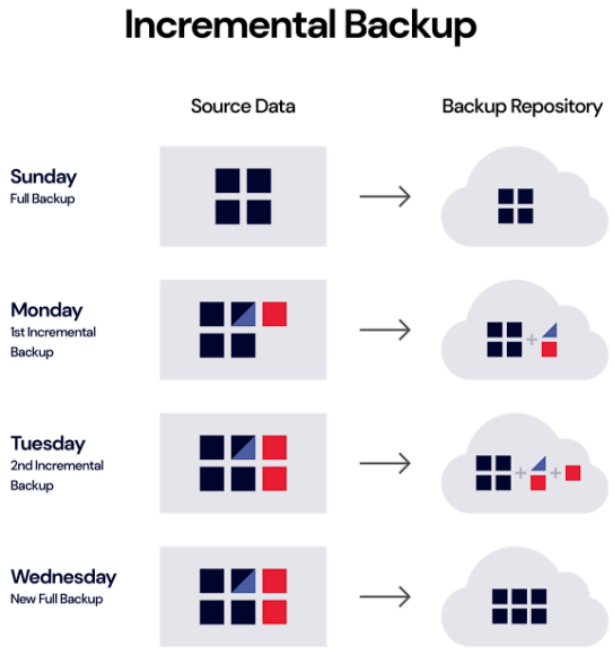
Figuur 1: Representatie van een full back-up (Rivas, 2022)

Een full back-up is een back-upmethode waarbij alle gegevens van een systeem op een specifiek moment volledig worden gekopieerd en opgeslagen. Dit betekent dat elk bestand zonder uitzonderingen wordt gekopieerd, zodat er een exacte kopie van de volledige dataset ontstaat (Beard, 2018). Wanneer er zich een probleem voordoet, zoals het falen van een harde schijf, kan het hele bestandssysteem vanaf deze back-up volledig worden hersteld op een nieuwe schijf. Daarnaast kunnen ook individuele bestanden die verloren zijn gegaan, gemakkelijk worden teruggehaald uit de back-up. Dit soort back-up zorgt ervoor dat alle gegevens veilig zijn opgeslagen, maar heeft wel twee belangrijke nadelen. Ten eerste is het proces van het lezen en schrijven van het volledige bestandssysteem tijdsintensief, vooral bij grote hoeveelheden data. Ten tweede gebruikt het opslaan van een volledige kopie van het bestandssysteem veel opslagruimte, wat inef-

ficiënt kan zijn wanneer de back-ups regelmatig worden gemaakt (Chervenak e.a., 1998).

2.1.2. Incremental back-up

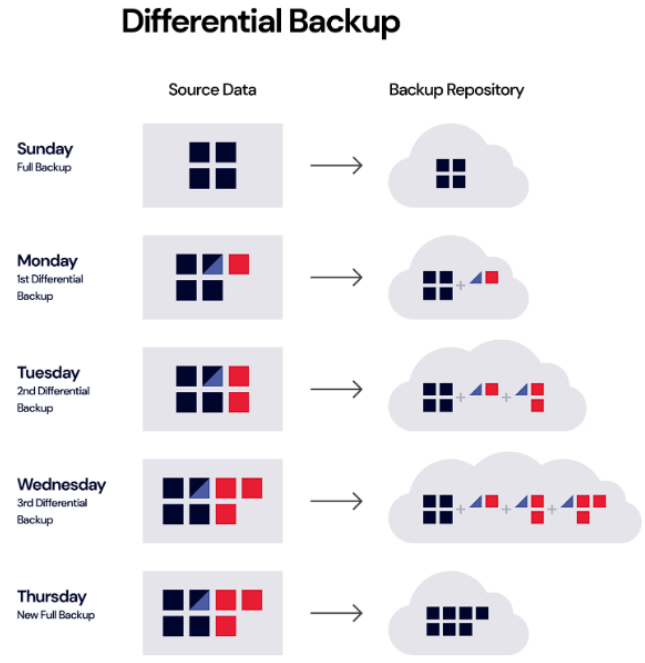
Incremental back-ups zijn een efficiënte methode om alleen gewijzigde data sinds de laatste back-up op te slaan, wat tijd en opslag bespaart. In tegenstelling tot een volledige back-up, die alle data kopieert, richten incrementele back-ups zich enkel op nieuwe of aangepaste bestanden. Dit maakt ze sneller, maar hersteltijden kunnen langer zijn omdat meerdere incrementele back-ups nodig zijn naast de laatste volledige back-up. Recente onderzoeken hebben zich gericht op het optimaliseren van back-upstrategieën, met name voor databasesystemen. Zo zijn er modellen ontwikkeld die bepalen hoe vaak volledige en incrementele back-ups moeten worden uitgevoerd op basis van factoren zoals systeem-betrouwbaarheid, de hoeveelheid dataveranderingen en back-upkosten (Zhao e.a., 2024). Er zijn ook varianten zoals differentiële back-ups, die alle veranderingen sinds de laatste volledige back-up bevatten, waardoor de hersteltijd korter kan zijn dan bij traditionele incrementele back-ups. Daarnaast zorgen moderne geautomatiseerde oplossingen voor continue incrementele back-ups, wat real-time herstelmogelijkheden biedt zonder noemenswaardige belasting van de productieomgeving (Qian e.a., 2010).



Figuur 2: Representatie van een incremental back-up (Rivas, 2022)

2.1.3. Differentiële back-up

Een differentiële back-up is een soort back-up waarbij alleen de data die sinds de laatste volledige back-up is veranderd of toegevoegd, wordt gekopieerd. In tegenstelling tot een incrementele back-up, die enkel de veranderingen sinds de laatste back-up opslaat, wordt er bij een differentiële back-up enkel de wijzigingen opgeslagen sinds de laatste full back-up . Dit komt doordat elke differentiële back-up alle wijzigingen sinds de meest recente volledige back-up bevat, waardoor de grootte van de back-up groter wordt naarmate er meer wijzigingen plaatsvinden. Een belangrijk voordeel van differentiële back-ups is de relatief snelle hersteltijd (Beard, 2018). Om data te herstellen, is alleen de laatste volledige back-up en de meest recente differentiële back-up nodig, wat het herstelproces eenvoudiger en sneller maakt dan bij incrementele back-ups. Differentiële back-ups zijn bijzonder nuttig in omgevingen waar een snel herstelproces cruciaal is, zoals bij bedrijven die minimale downtime vereisen. Het nadeel van differentiële back-ups is dat de back-ups groter worden naargelang de tijd tussen de volledige back-ups. Elke nieuwe differentiële back-up bevat namelijk alle wijzigingen sinds de laatste volledige back-up, wat betekent dat deze geleidelijk groter wordt totdat er een nieuwe volledige back-up wordt gemaakt. Daarom is het belangrijk om een goede balans te vinden tussen de frequentie van volledige back-ups en differentiële back-ups.



Figuur 3: Representatie van een differentiële back-up (Rivas, 2022)

2.1.4. On-premise back-ups

Bedrijven staan vaak voor de uitdaging om te beslissen of ze hun data on-premise opslaan of de voorkeur geven aan een cloud-service (Ali e.a., 2024). On-premise back-ups slaan gegevens lo-

kaal op, meestal op fysieke schijven binnen het bedrijf zelf. Deze methode biedt bedrijven volledige controle over hun back-up- en gegevensbeheer. Een belangrijk voordeel van on-premise back-ups is dat data altijd beschikbaar is, zelfs zonder toegang tot het internet, wat nuttig is bij netwerkproblemen (Trovato e.a., 2019). Daarnaast hebben bedrijven volledige eigendom over de beveiliging van hun gegevens, aangezien de opslag lokaal blijft binnen het bedrijf. Hoewel deze methode geen terugkerende kosten aan externe providers met zich meebrengt, brengt het wel risico's met zich mee, zoals schade door brand of overstromingen, en vraagt het om regelmatige onderhoud van de hardware. Het herstelproces is doorgaans sneller dan bij een cloudservice, wat van cruciaal belang kan zijn na een ransomware-aanval of een ander incident.

2.1.5. Cloud-gebaseerde back-ups

Cloud-gebaseerde back-ups zijn een populaire oplossing waarbij data extern wordt opgeslagen bij een bedrijf dat cloudservices aanbiedt. Dit biedt individuen en bedrijven de mogelijkheid om hun gegevens veilig op afstand te bewaren, zonder dat ze hoeven te investeren in fysieke opslagapparaten. Hoewel dit handig is om gegevensverlies te voorkomen bij hardware- of softwarestoringen, of onverwachte rampen, brengt het gebruik van cloud-opslag vaak aanzienlijke kosten met zich mee, vooral op de lange termijn (Obrutsky, 2016). Naarmate je meer opslag nodig hebt is er altijd een mogelijkheid om te opschalen, dit is een groot voordeel van het gebruiken van een cloudservice. Echter, het waarborgen van de veiligheid van deze data is een cruciaal aandachtspunt, vooral omdat cloudproviders vaak niet open zijn over hoeveel kopieën van de data er zijn en waar deze precies worden opgeslagen. Om problemen zoals datalekken en foutieve verwijdering te voorkomen, zijn er nieuwe methoden zoals "assured deletion" ontwikkeld, waarmee klanten zeker weten dat hun gegevens permanent worden verwijderd op verzoek. Hierdoor kunnen bedrijven hun data met zekerheid beheren in de cloud terwijl gevoelige informatie veilig blijft (Rahumed e.a., 2011).

2.1.6. Offline back-ups

Offline back-ups zijn een traditionele methode waarbij data wordt opgeslagen op fysieke media, meestal externe harde schijven zonder tussenkomst van het internet. Het voornaamste voordeel is dat de data dan beveiligd is tegen online bedreigingen en er geen internettoegang nodig is om aan de data te geraken (Edwards, 2022). Een belangrijk voordeel is dat offline back-ups niet beïnvloed worden door stroomstoringen of internetuitval, waardoor ze een robuuste back-upoptie vormen voor gevoelige data. Echter, in tegenstelling tot on-premise back-ups, die vaak op dezelfde fysieke locatie als de IT-infrastructuur van een bedrijf worden opgeslagen, kunnen offline back-ups eenvoudig meegenomen en elders bewaard worden, waardoor ze extra bescherming bieden tegen fysieke rampen. Toch delen beide methoden het nadeel dat ze kwetsbaar zijn voor schade door ongelukken, diefstal of verlies, en moeten de fysieke apparaten op een veilige locatie opgeslagen worden (James, 2019).

2.1.7. Immutable storage

Immutable storage is een type opslag waarbij data niet meer kan worden gewijzigd of aangepast vanaf het geback-upt is. Dit concept is cruciaal voor het waarborgen van de integriteit van belangrijke gegevens. Het idee achter immutability is dat bepaalde bestanden, nadat ze zijn gemaakt, niet meer mogen worden gewijzigd zonder de juiste autorisatie. Dit biedt een sterke bescherming tegen ongewenste wijzigingen en hierdoor kunnen hackers de gegevens niet aanpassen. Immutable storage speelt dus een belangrijke rol in het beschermen van systemen tegen cyberaanvallen. Bij aanvallen, waarbij hackers volledige toegang verkrijgen, kunnen onbeveiligde systemen worden gemanipuleerd of misbruikt. Immutable storage voorkomt dit, omdat de opgeslagen gegevens niet kunnen worden gewijzigd, zelfs niet door iemand met volledige toegang. Hierdoor wordt de integriteit van de data behouden en is het risico op schade door hackers aanzienlijk kleiner (Hasan e.a., 2005).

3. Methodologie

In de eerste fase van het onderzoek zal er een grondige literatuurstudie worden uitgevoerd rond back-upstrategieën, ransomware-resistentie, en immutable storage met een overzicht van de state of the art van back-upstrategieën en immutabele opslag als deliverable. Wetenschappelijke papers, bedrijfscasussen en technische artikels zullen gebruikt worden om een theoretische basis aan te leggen en om de best-practices te achterhalen. Dit zal ook helpen om de onderzoeksvragen te beantwoorden. Daarnaast zal er een Proof-Of-Concept ontworpen worden waarbij onderzocht zal worden hoe er immutable back-ups gemaakt kunnen worden voor de back-ups van Forvis Mazars. Daarnaast zal er een testomgeving opgezet worden om een ransomware-aanval na te bootsen en het systeem opnieuw op gang te krijgen. De deliverable voor de PoC is een werkende immutable back-upoplossing in Azure die tegen een ransomware-aanval bestendig is. Verder zal er een optimale back-upstrategie opgesteld worden met de state-of-the-art technieken. De literatuurstudie zal ongeveer 4 weken duren, de Proof-Of-Concept zal 4 weken duren en als laatst zal het rapport met de optimale verbeteringen 2 weken

duren.

4. Verwacht resultaat, conclusie

Het verwachte resultaat is dat door de implementatie van immutable storage en automatische back-ups, de back-upstrategie van Forvis Mazars zal worden verbeterd. Vooral de bescherming tegen ransomware en andere bedreigingen zal beter zijn door het gebruik van immutable storage, waarbij back-ups onveranderlijk worden opgeslagen en niet kunnen worden gemanipuleerd. Daarnaast zorgt de automatisering van de back-ups voor een efficiënter beheer, waarbij de manuele taken van het IT-team verminderd worden. Dit kan in de praktijk leiden tot meer consistente back-ups en een verbeterde betrouwbaarheid van het systeem. De resultaten zullen waarschijnlijk aantonen dat de combinatie van deze twee oplossingen zorgen voor een sterkere, efficiëntere en beter back-upstrategie.

Referenties

Ali, A., Laghari, A. A., Kandhro, I. A., Kumar, K., & Younus, S. (2024). Systematic analysis of on-premise and cloud services. *International Journal of Cloud Computing*, *13*(3), 214–242. https://doi.org/https://doi.org/10.1504/IJCC.2024.139604

Beard, B. (2018). *Full Backups*. Apress. https://doi.org/https://doi.org/10.1007/978-1-4842-3456-3_1

Cherry, D. (2015). Database Backup Security, 293–311. https://doi.org/https://doi.org/10.1016/B978-0-12-801275-8.00010-5

Chervenak, A., Vellanki, V., & Kurmas, Z. (1998). Protecting file systems: A survey of backup techniques. *Joint NASA and IEEE Mass Storage Conference*, *99*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=4b6cfd832c2eb61c60ae0ab956eed8401f22d3f0

Dikbiyik, F., Gul, F., Tapkan, G., Han, Y., Akdora, O., Budakoglu, G., Ciftci, B., Celik, E. S., Cengiz, S. E., Toprak, G., & Dogan, Y. (2024). State of Ransomware 2024:A Year of Surgesand Shuffling. *Black kite*. https://blackkite.com/wp-content/uploads/2024/05/BlackKite_Report_Ransomware-2024.05.14.pdf

Edwards, B. (2022). Why You Need an Offline Backup. *How-To Geek*. https://www.howtogeek.com/818193/why-you-need-an-offline-backup/

Hasan, R., Stanton, P., Yurcik, W., Brumbaugh, L., Rosendale, J., & Boonstra, R. (2005). The Techniques and Challenges of Immutable Storage with applications in Multimedia. *National Center for Supercomputing Applications*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=578ff4957d4fa2e550ec2a819b6500820d7286cd

James. (2019). Offline backups in an online world. *National Cyber Security Centre*. https://www.ncsc.gov.uk/blog-post/offline-backups-in-an-online-world

Obrutsky, S. (2016). Cloud storage: Advantages, disadvantages and enterprise solutions for business. *Conference: EIT New Zealand*. https://www.researchgate.net/profile/Santiago-Obrutsky/publication/305508410_Cloud_Storage_Advantages_Disadvantages_and_Enterprise_Solutions_for_Business/links/5792976508ae33e89f7cc136/Cloud-Storage-Advantages-Disadvantages-and-Enterprise-Solutions-for-Business.pdf

Qian, C., Huang, Y., Zhao, X., & Nakagawa, T. (2010). Optimal Backup Interval for a Database System with Full and Periodic Incremental Backup. *Journal of Computers*, *5*(4), 557–564. https://doi.org/10.4304/jcp.5.4.557-564

Rahumed, A., Chen, H. C. H., Tang, Y., Lee, P. P. C., & Lui, J. C. S. (2011). A secure cloud backup system with assured deletion and version control. *40th International Conference on Parallel Processing Workshops*, 160–167. https://www.researchgate.net/publication/221617563_A_Secure_Cloud_Backup_System_with_Assured_Deletion_and_Version_Control

Rivas, K. (2022). What's the Diff: Full, Incremental, Differential, and Synthetic Full Backups. *Backblaze*. https://www.backblaze.com/blog/whats-the-diff-full-incremental-differential-and-synthetic-full-backups/

Trovato, F., Sharp, A., & Siman, T. (2019). Cloud, co-location, on-premises and hybrid disaster recovery solutions: Pros, cons, and a cost comparison. *Journal of Business Continuity & Emergency Planning*, *13*(2), 120–135. https://www.ingentaconnect.com/content/hsp/jbcep/2019/00000013/00000002/art00004

Zhao, X., Bu, Y., Pang, W., & Cai, J. (2024). Periodic and random incremental backup policies in reliability theory. *Software Quality Journal*, *32*(3), 1325–1340. https://doi.org/https://doi.org/10.1007/s11219-024-09685-1