PKU Automation

Last updated by | Paul Kelleher | 17 Jun 2020 at 08:00 GMT

- □ Requirements Capture
- □ High Level Solution
- □ High Level Design
- □ Security Engagement
- □ Technical Architecture
- □ Architecture Signoff
- □ Build & Delivery
- □ Post Delivery

Contents

Overview

Executive Summary

Purpose

Scope

It is intended to include all elements required to deliver, secure and support the V3 Platform

Technical Design Elements
☒ Build Platform & IaC
☒ Azure Infrastructure
☒ Azure Subscription Policy
☒ Azure Resources
☒ Access Control Framework
☒ Storage and interdependant mounts
☒ Interfaces to other platforms
☒ User access and data systems
☒ Performance & Capacity Reporting
☒ Cost Management

Security & Resilience
☒ Secure by Design Architecture
☒ Adoption of NCSC Cloud Principles
f ☒ Subscription Level Policy
☒ Resource Level Policy
☒ Adherance to Azure Best Practice
☒ Multi-Zone Infrastructure
☒ Resource Monitoring & Alerting
☒ Alert Groups for Platform Issues
☒ Splunk & Sophos

Process & Support
☒ All Resources as IaC
☒ External Processes Documented
☒ Sub-processes automated
☒ Dependency Maps
☒ Support Scenarios and Resolutions

for the second revision of the monitoring and alerting, it is intended that automated resolution to issues will be sought

g

Requirement

High Level Solution

In principal technology and process propostition and component design

Component	Detail
Connect to Experian sFTP	visdualCron used on a locked down windows machine under AAD/GPO; The machine is located in locked down DMZ Subnet with connections allowed on a one to one allow basis. The experian endpoint is known and previously validated The credentials are controlled and managed by Experian The connection is via SFTP
List and Retrieve Files	The host has been previously verified by pknw1 The data retrieved is zipped and encrypted with a PGP key that is unavailable to this DMZ or its resources
Store Retrieved Files	Files are stored in an Azure storage account The DMZ Server has only a write SASS key The stoprage account is locked down to the DMZ and Processor Subnets
Decrypt Retrieved Files	The Decrypt machine has the only read access to the blob all decryption (which results in PII data) is done in memory and not stored on any non-transient storage. All storage is encrypted Azure Data Bricks is used for validation
Process Files	
Ingest Files	Write files into VCAP via secure connection hard remove files from transient storage with 0000 writes

Data Ingestion & Storage

sFTP Controller Process

- □ VisualCron

DMZ Network Controls

- □ Azure Virtual Networks
- □ Azure Firewall

sFTP Controller Controls

- □ GPO
- □ RBAC AAD

Azure Storage Controls

- □ Uni-direction SAS keys
- □ Network locked with Firewalls
- □ IAM Controls removed all non engineering roles

Decryption and Validation

Data File Load

Data Components

In principal technology and process propostition and component design

Data Retrieval an d initial storage

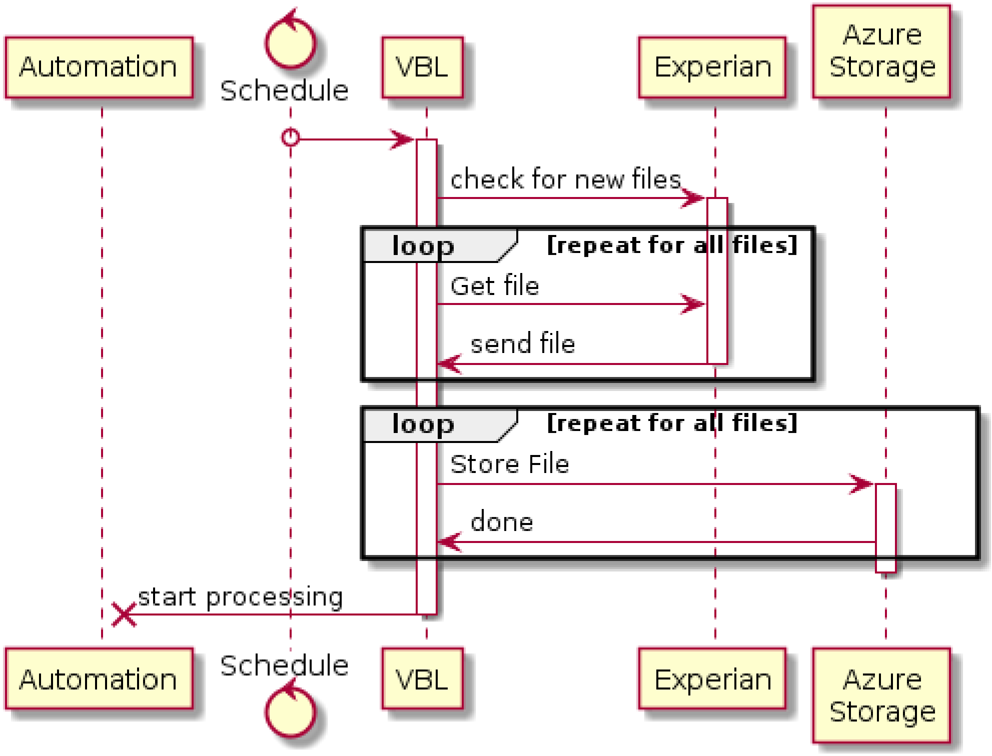
Item	Details
vNet	
sNet	
VM	
VM OS	
Visual Cron	
sFTP credentials	
sFTP Endpoint	
Storage Account	
Storage Account Credentials	
Upload Scripts	
AzCopy	

High Level Design

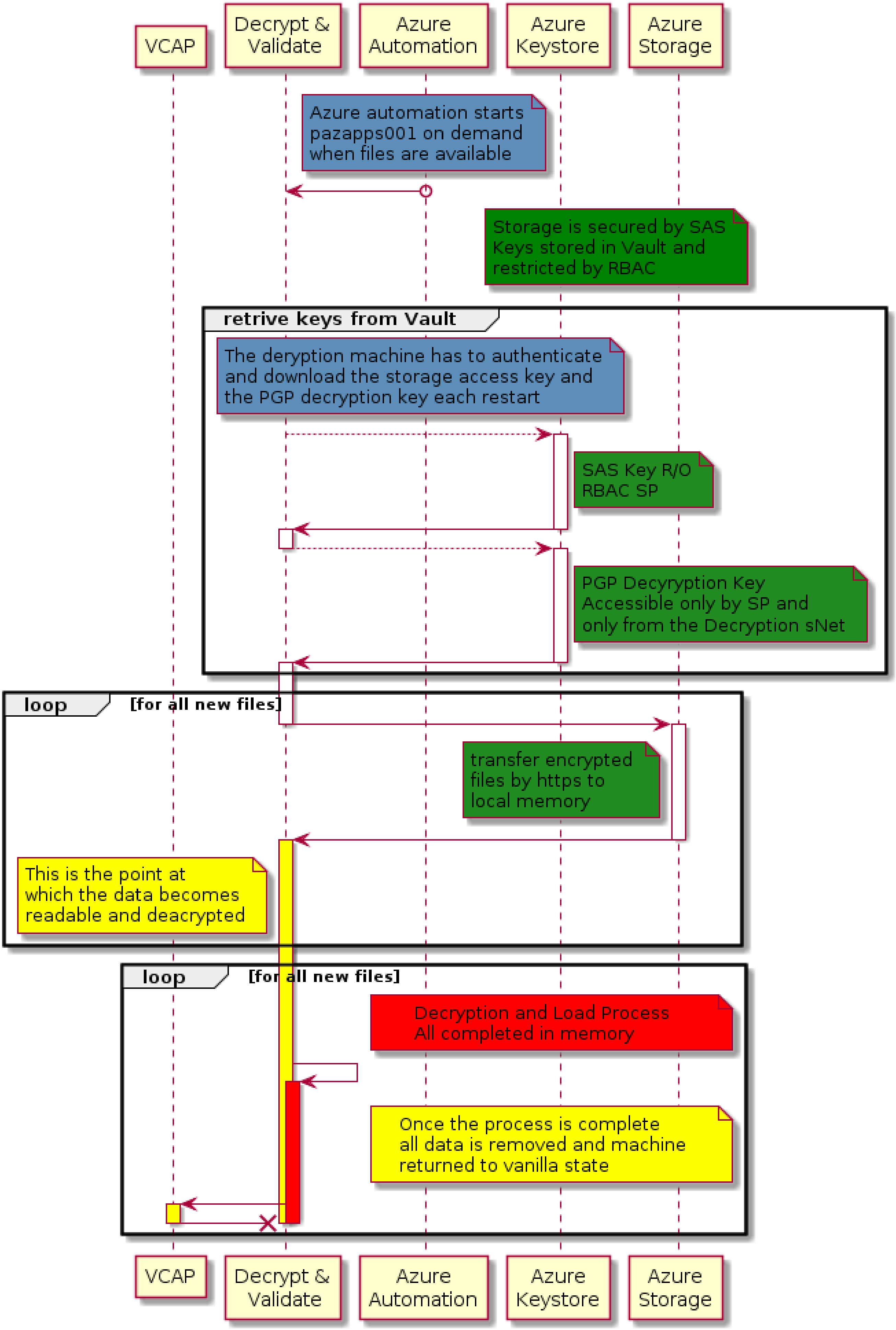
Post proof of concept design based on pknw1 standards and best practice for products

Heading

```
File retrieval

@startuml

participant Automation
Control Schedule
participant pknw1
participant Experian
participant "Azure\nStorage" as Storage


Schedule o->pknw1:
    activate pknw1
    pknw1->Experian: check for new files
    activate Experian

    loop repeat for all files

        pknw1->Experian: Get file
        Experian->pknw1: send file
        deactivate Experian
    end

    loop repeat for all files
        pknw1->Storage: Store File
        activate Storage
        Storage->pknw1: done
    end
    deactivate Storage

    pknw1->x Automation: start processing
    deactivate pknw1

@enduml
```



```
@startuml

participant VCAP
participant  "Decrypt &\n Validate" as pazapps001
participant "Azure\nAutomation" as Automation
participant "Azure\nKeystore" as Vault
participant "Azure\nStorage" as Storage






    note over Automation  #5D8EBA
        Azure automation starts
        pazapps001 on demand
        when files are available
    end note
    Automation o-> pazapps001

        note over Vault, Storage #green
            Storage is secured by SAS
            Keys stored in Vault and
            restricted by RBAC
        end note

    group retrive keys from Vault

        note over pazapps001, Vault #5D8EBA
            The deryption machine has to authenticate
            and download the storage access key and
            the PGP decryption key each restart
        end note
        pazapps001-->Vault:
        activate Vault
             note right of Vault #228B22
                 SAS Key R/O
                 RBAC SP
             end note
        Vault->pazapps001:
        deactivate Vault
        activate pazapps001
        pazapps001-->Vault:
        deactivate pazapps001
        activate Vault
             note right of Vault #228B22
                 PGP Decyryption Key
                 Accessible only by SP and
                 only from the Decryption sNet
             end note
        Vault->pazapps001:
        deactivate Vault
        activate pazapps001
    end

    loop for all new files

        pazapps001->Storage:
        deactivate pazapps001
        activate Storage
             note left of Storage #228B22
                 transfer encrypted
                 files by https to
                 local memory
             end note
        Storage->pazapps001
        deactivate Storage
             note left of pazapps001 #yellow
                 This is the point at
                 which the data becomes
                 readable and deacrypted
             end note
             activate pazapps001 #yellow
    end

    loop for all new files

    note over Automation, Storage #red
        Decryption and Load Process
        All completed in memory
    end note

        pazapps001->pazapps001
    note over Automation, Storage #yellow
       Once the process is complete
       all data is removed and machine
       returned to vanilla state
    end note

        activate pazapps001 #red
        pazapps001->VCAP:
        activate VCAP #yellow
        VCAP->x pazapps001:



        deactivate pazapps001
        deactivate VCAP


        deactivate pazapps001
    end group



@enduml
```


VCAP
Decrypt & Validate
Azure Automation
Azure Keystore
Azure Storage
Azure automation starts pazapps001 on demand when files are available
Storage is secured by SAS Keys stored in Vault and restricted by RBAC
retrive keys from Vault
The deryption machine has to authenticate and download the storage access key and the PGP decryption key each restart
SAS Key R/O RBAC SP
PGP Decyryption Key Accessible only by SP and only from the Decryption sNet
loop
[for all new files]
transfer encrypted files by https to local memory
This is the point at which the data becomes readable and deacrypted
loop
[for all new files]
Decryption and Load Process All completed in memory
Once the process is complete all data is removed and machine returned to vanilla state
VCAP
Decrypt & Validate
Azure Automation
Azure Keystore
Azure Storage