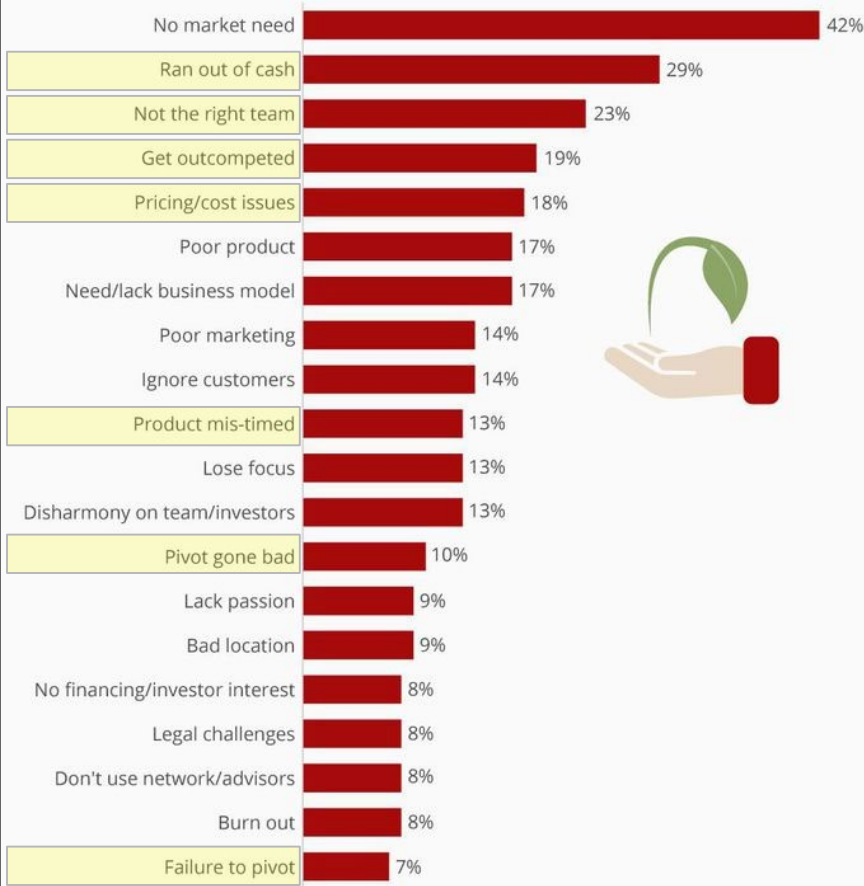


The Top Reasons Startups Fail

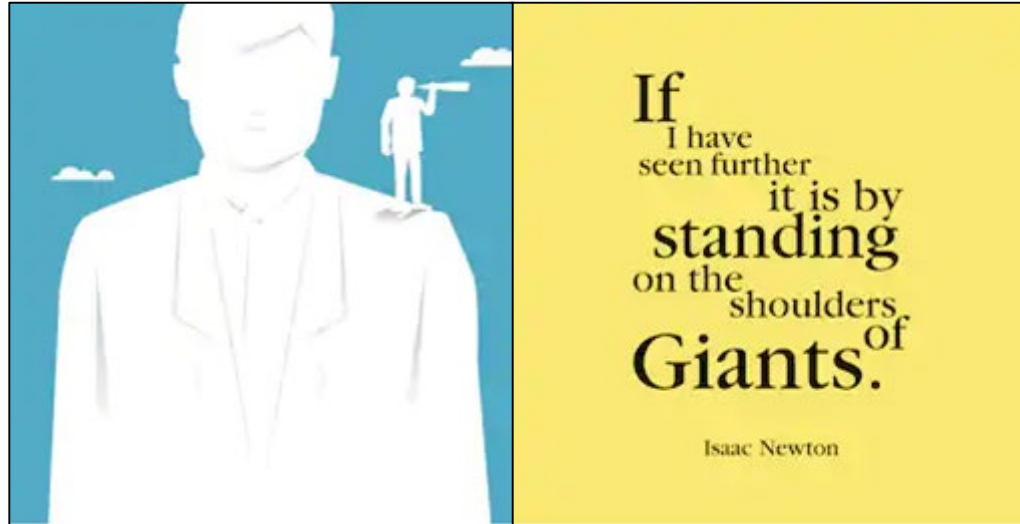
Most frequently cited reasons for startup failure*



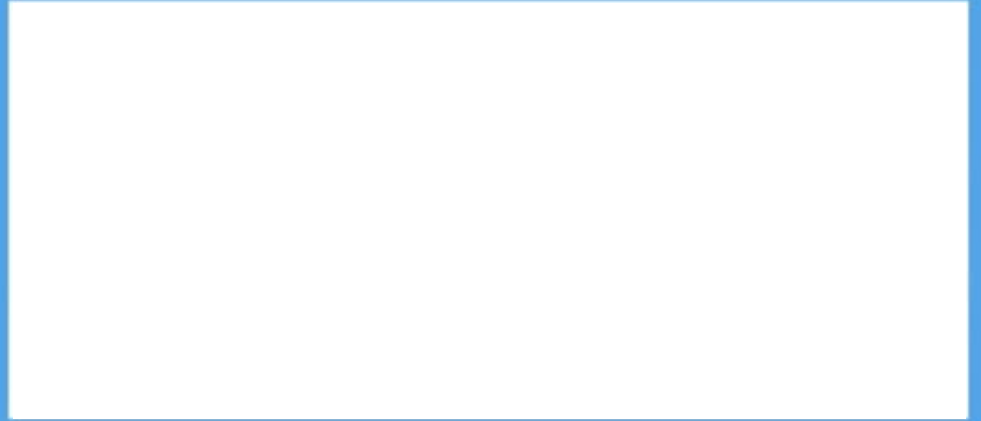
Cloud Native Full Stack

Shoulders of Giants as a Service

Building on previous discoveries



What If You Could Start Over...



Sebastian Nilsson

- In the IT-business since 2003
 - Software since 2007
 - CTO for 3 year at startup
- Microsoft-ecosystem focus
 - Azure Cloud
 - .NET
 - Azure DevOps
- Leadership
 - Agile/Scrum
 - Recruitment (Interviews/screenings)





**WiFi-Problem
Resistant!**

Warning: Powerpointbased presentation!

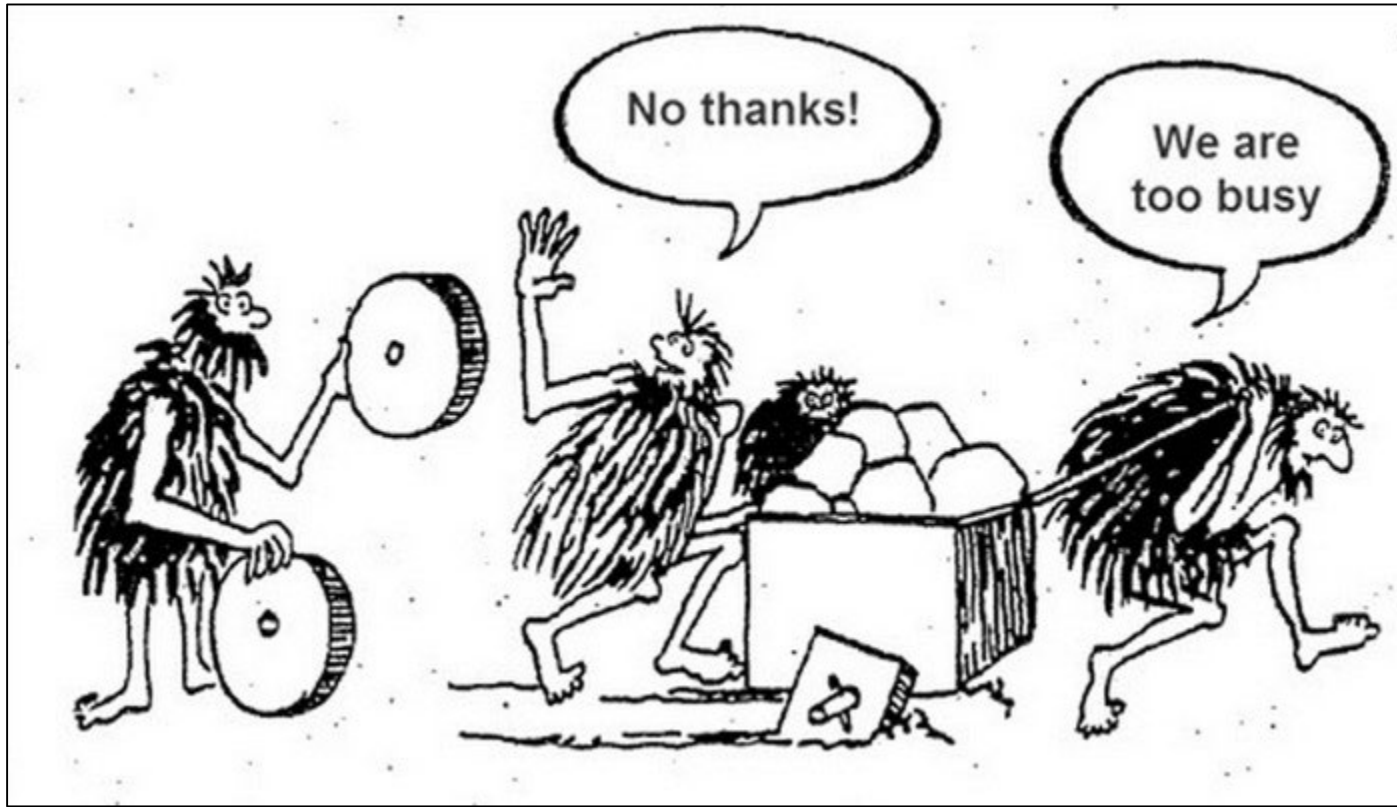


Warning 2: Technology Overload

Agenda

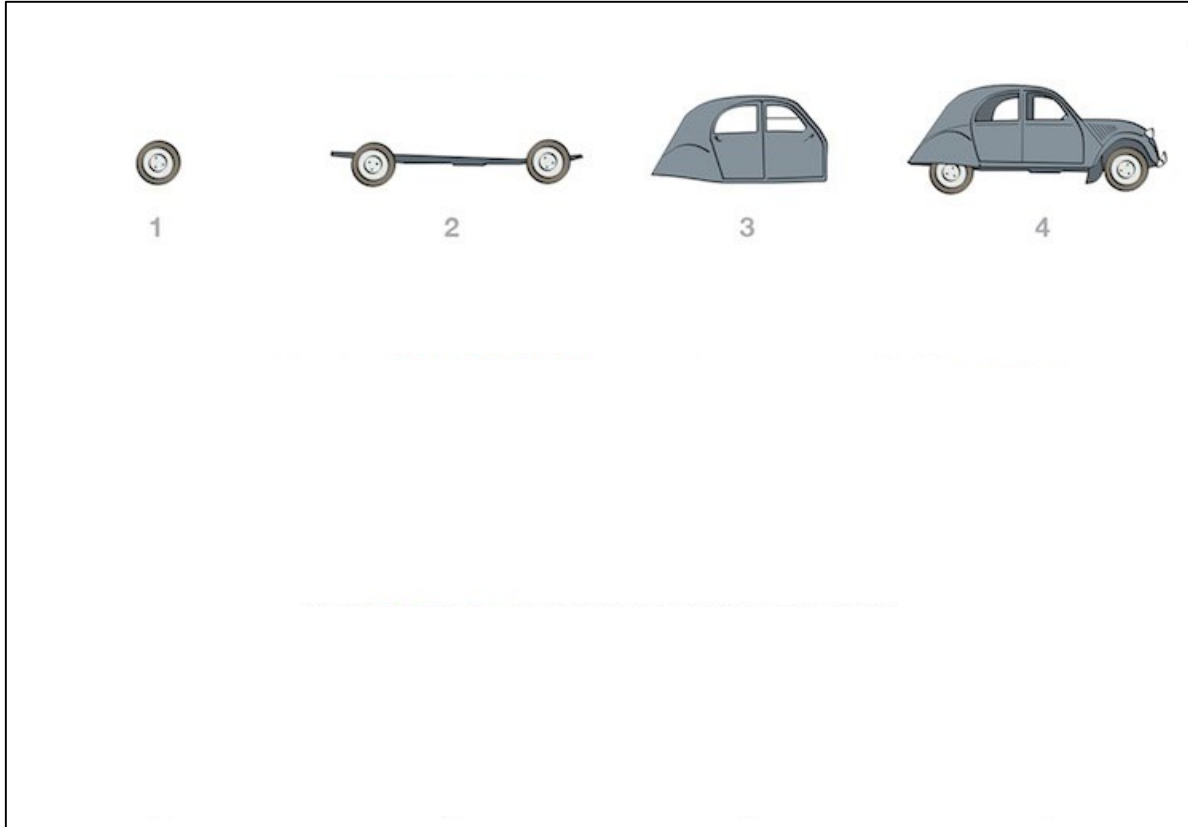
Build a fully functional,
mature startup company

- Domain
 - Architecture
 - DevSecOps (CI/CD)
 - Azure (Cloud)
 - SaaS-services
 - Frameworks/Libraries
-

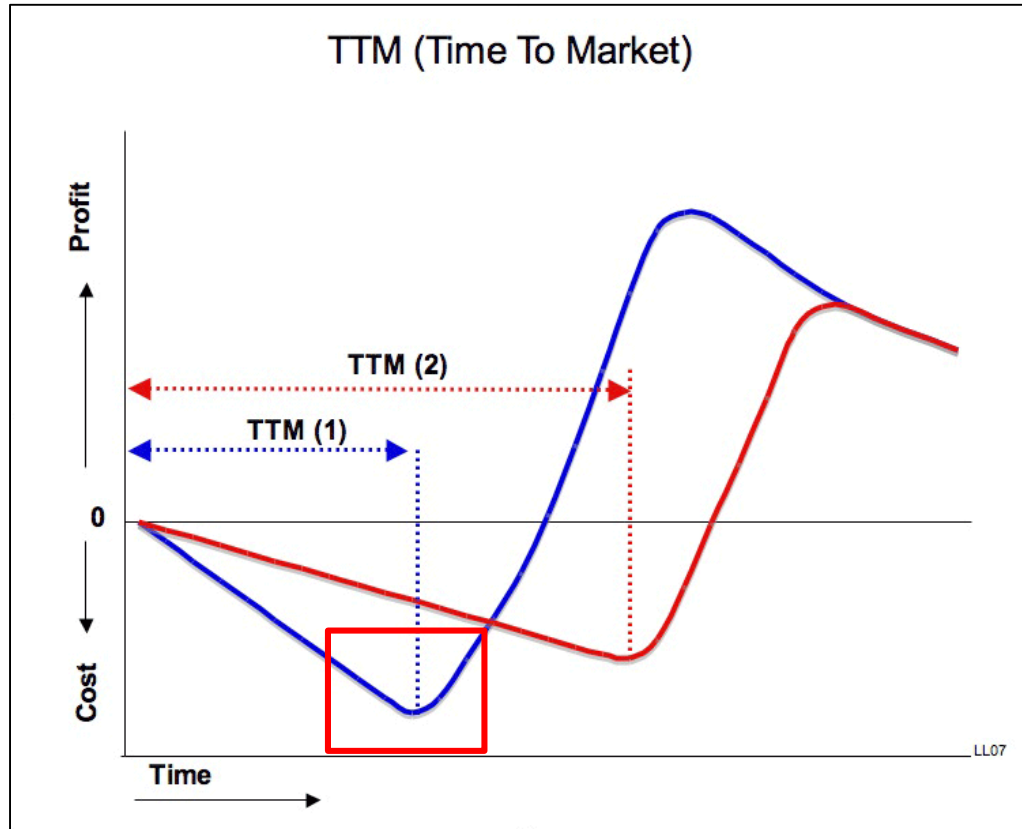


- Focus on **your core business**
- Let the rest be **others' core business**

Minimum Viable Product (MVP)



Time To Market (TTM)



They make all the
LEGO pieces
just the correct size

and snapping them together
a lot of fun .

-- Scott Hanselman



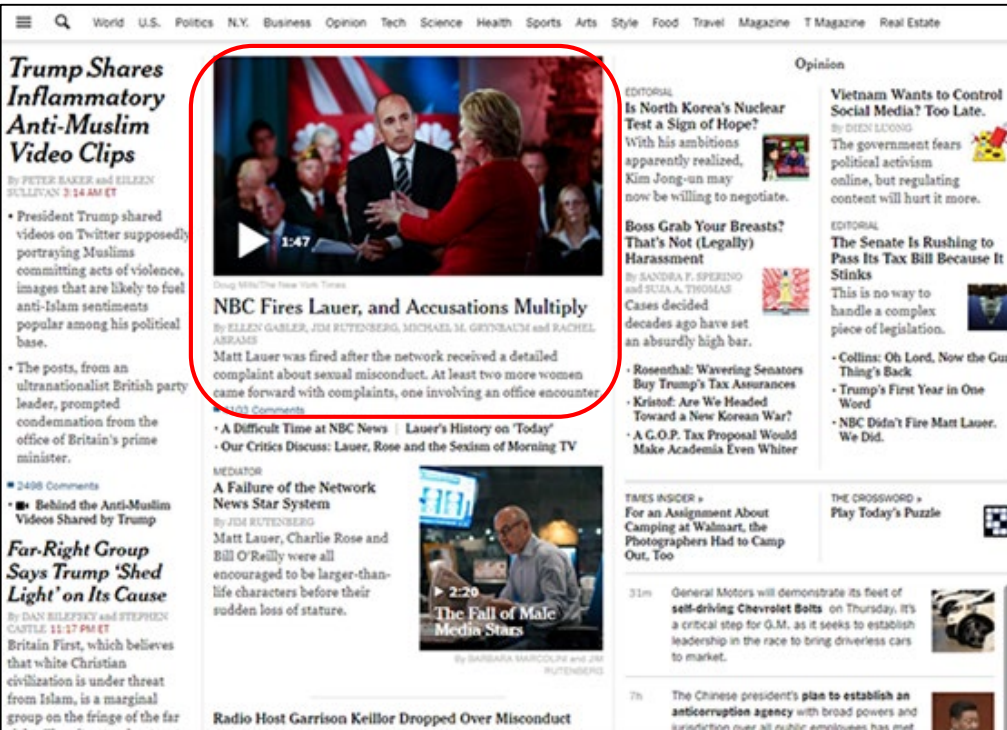
Best Time to Develop Software

- Cloud
 - App Services - Web-sites/APIs
 - Serverless - Azure Functions
 - Azure Service Bus
 - Security - Azure Key Vault
- Azure DevOps
 - CI - Build/Create Artifacts
 - CD - Deployment
- Software as a Service (SaaS)
 - Authentication/Authorization - Auth0
 - Storage/Data - Azure Storage
 - CRM - Airtable
- Frameworks/Libraries/Tools/CLIs

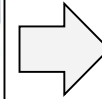
Domain

Web-scraping as a Service

Domain / Problem to solve



The screenshot shows a news website with a navigation bar at the top containing categories like World, U.S., Politics, N.Y., Business, Opinion, Tech, Science, Health, Sports, Arts, Style, Food, Travel, Magazine, T Magazine, and Real Estate. The main content area features several articles. A red circle highlights a video player for the article "NBC Fires Lauer, and Accusations Multiply". The video player shows a man in a suit speaking to a woman in a red dress. Below the video player, there are several other articles, including "Trump Shares Inflammatory Anti-Muslim Video Clips", "Boss Grab Your Breasts? That's Not (Legally) Harassment", "Vietnam Wants to Control Social Media? Too Late.", "The Senate Is Rushing to Pass Its Tax Bill Because It Stinks", "A Failure of the Network News Star System", "The Fall of Male Media Stars", and "Radio Host Garrison Keillor Dropped Over Misconduct".



```
GET
/api/content?url=...
---
{
  "url": "http...",
  "content":
    "NBC Fires Lauer, and
    Accusations Multiply...",
  "image": "http...",
  "format": "text"
}
```




Introducing : Project WebFetch

Scraping URLs

- Core: Getting content from URL
- Core: Transform content to usable format
 - Raw HTML-content
 - Main content as plain text/HTML/Markdown
- SaaS: Login - Auth0
- SaaS: Data-storage - Azure Table Storage
- SaaS: CRM - Airtable
- *Future: API-management - Auth, Limits, Throttling*
- *Future: Cache*

Provide REST API

POST

`/api/v1/scraping/raw-html` Gets the raw HTML from URL.

POST

`/api/v1/scraping/content-text` Gets the text of the content from URL.

POST

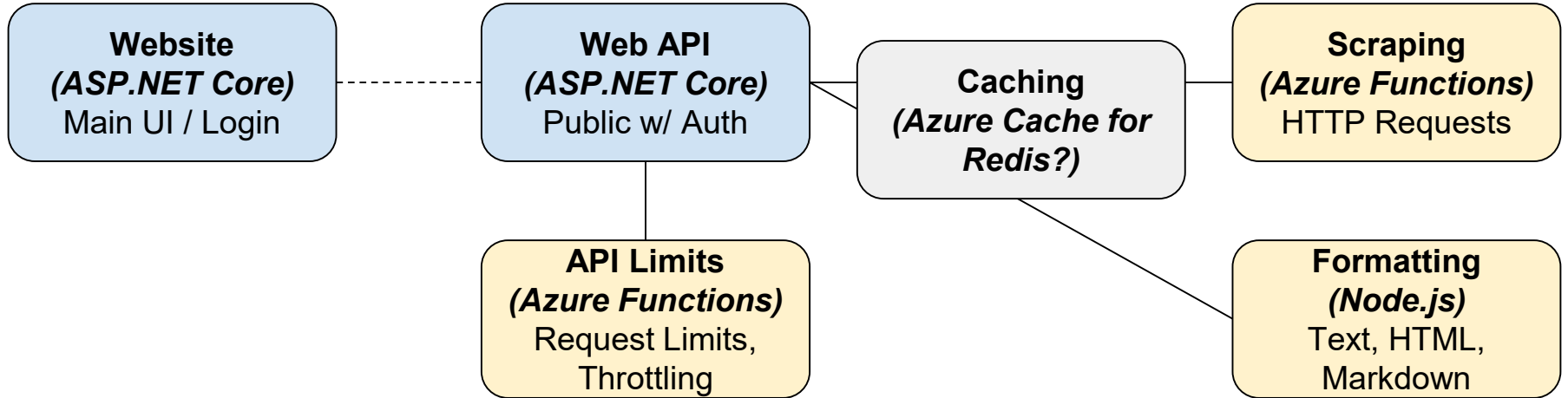
`/api/v1/scraping/content-html` Gets the HTML of the content from URL.

POST

`/api/v1/scraping/content-markdown` Gets the Markdown of the content from URL.

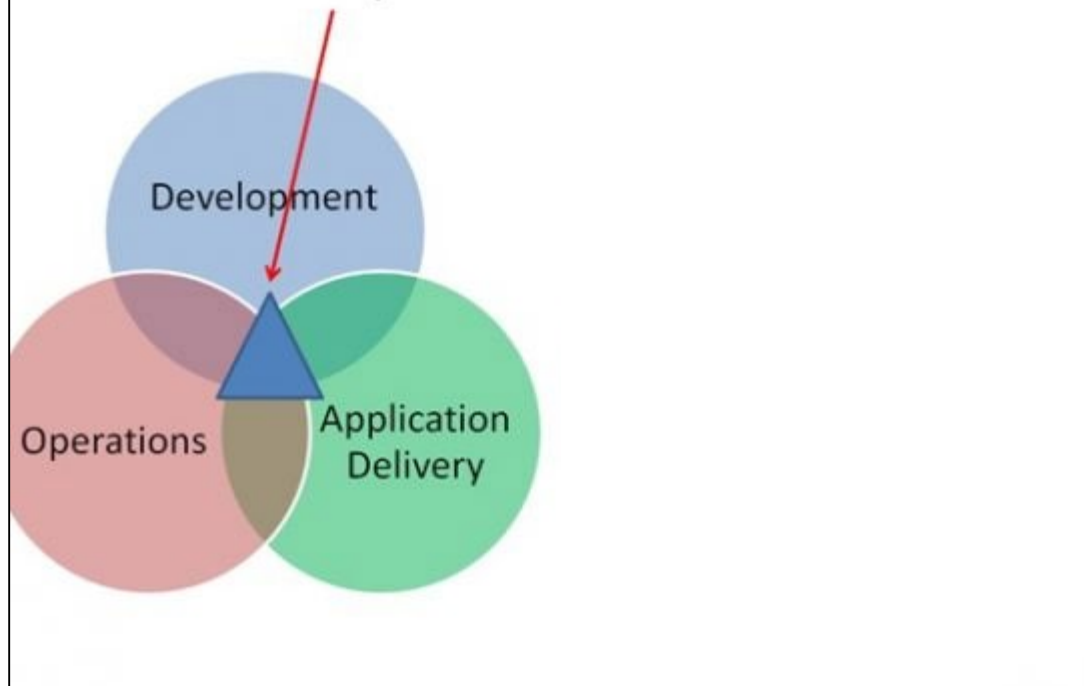
Architecture

Architecture Overview

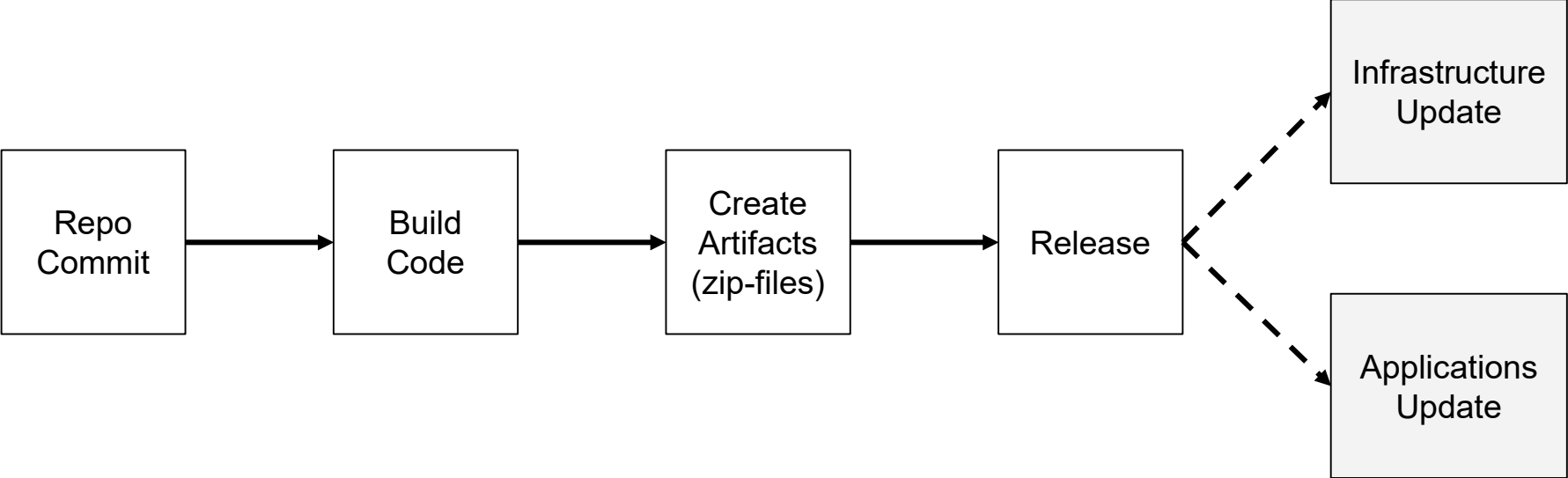


DevSecOps (CI/CD)

DevOps vs. DevSecOps



Dev(Sec)Ops



Azure DevOps


Azure DevOps


Create new project ✕

Project name *
WebFetch ✓

Description

Visibility

 Public
Anyone on the internet can view the project. Certain features like TFVC are not supported.

 Private
Only people you give access to will be able to view this project.

^ Advanced

Version control ?
Git

Work item process ?
Agile

Azure DevOps: Boards- Backlog

WebFetch Team

Backlog Analytics | [+ New Work Item](#) [View as Board](#) [Column Options](#) ...

Order	Work Item Type	Title	State	Effort	Business Val...	Iteration Path
	Feature	Azure DevOps Pipelines ...	New			WebFetch\Iteration 1
	User Story	Variable-groups	Active			WebFetch\Iteration 1
	Task	Terraform	Active			WebFetch\Iteration 1
	User Story	Build-setup	Active			WebFetch\Iteration 1
	User Story	Release-setup	Active			WebFetch\Iteration 1
	User Story	Test-plans	New			WebFetch
	User Story	Dev-environment	New			WebFetch
	Feature	Project-structure	New			WebFetch\Iteration 1
	User Story	Applications	New			WebFetch\Iteration 1
	User Story	Infrastructure	New			WebFetch\Iteration 1
	User Story	Domains	New			WebFetch\Iteration 1

Azure DevOps: Boards- Sprint

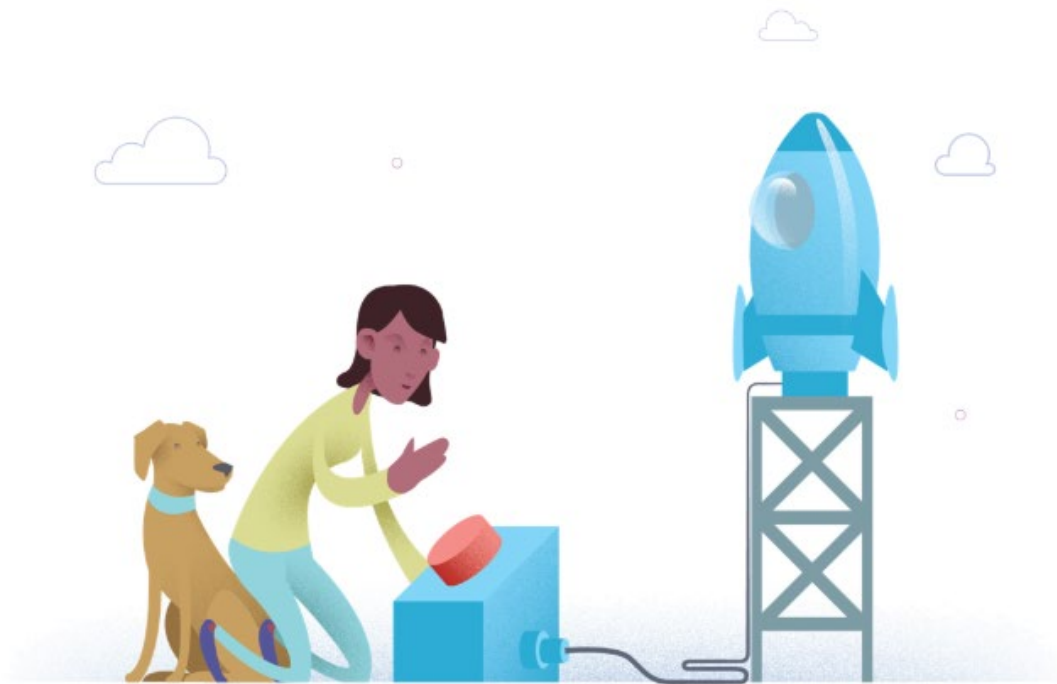
WebFetch Team ☆ 👤

Taskboard Backlog Capacity Analytics | [+ New Work Item](#) [🔗 Column Options](#) 🔍 Iteration 1 👤 Person: All

[^ Collapse all](#)

	New	Active	Resolved	Closed
<div><p>9 Variable-groups</p><p>SN Seb Nilsson</p><p>State ● Active</p></div>	<div><p>+</p></div>	<div><p>10 Terraform</p><p>SN Seb Nilsson</p><p>State ● Active</p></div>		<div><p>13 Global</p><p>SN Seb Nilsson</p><p>State ● Closed</p></div>
<div><p>6 Build-setup</p><p>SN Seb Nilsson</p><p>State ● Active</p></div>	<div><p>12 YAML</p><p>👤 Unassigned</p><p>State ● New</p><p>+</p></div>			

Azure DevOps: Pipelines




Azure DevOps: Pipelines- YAML

```
variables:  
  buildConfiguration: 'Release'  
  
trigger:  
  - master  
  
jobs:  
- job: Job_Infrastructure  
#...  
- job: Job_Applications
```


Azure DevOps: Pipelines- YAML



```
- job: Job_Applications
  steps:
    - task: DotNetCoreCLI@2
      displayName: 'Dotnet Restore'
      inputs:
        command: restore
        projects: '$(slnPath)'
    - task: DotNetCoreCLI@2
      displayName: 'Dotnet Build'
      inputs:
        command: build
        projects: '$(slnPath)'
        arguments: '--configuration $(buildConfiguration) --no-restore'
# ... Test, Publish, Zip, Artifacts
```


Azure DevOps: Build result

 #WebFetch CI_2019.8.23.26 Pipelines: Fixed TF_VAR-prefix
on WebFetch CI



Summary

Triggered by  Seb Nilsson

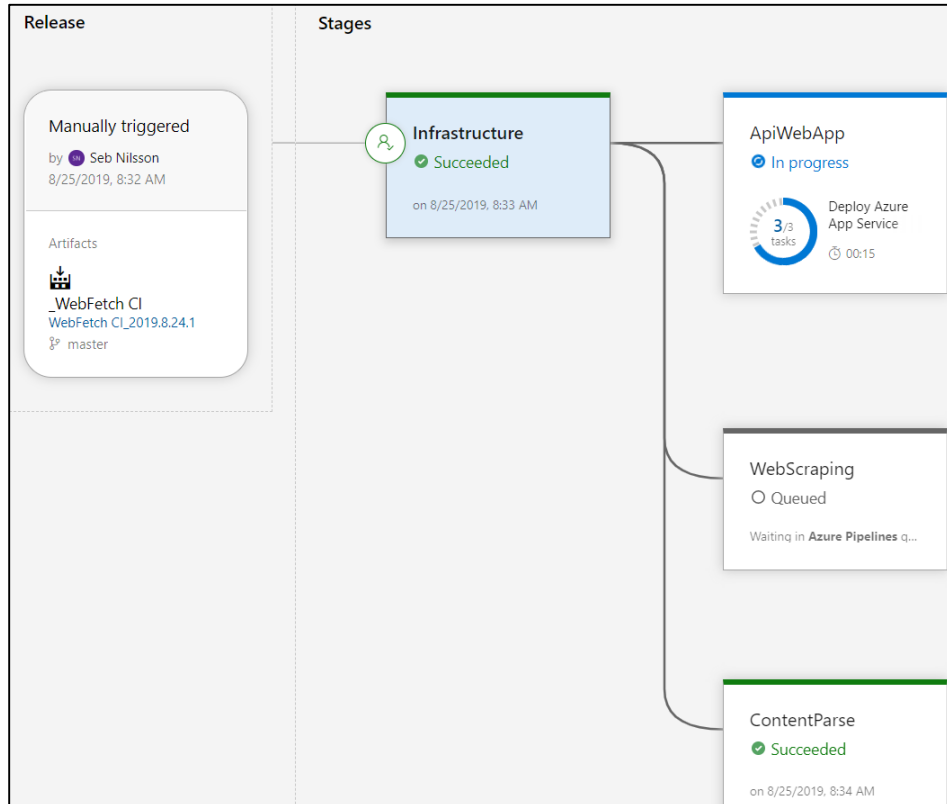
 WebFetch  master d82e733 Duration:

 Today at 5:40 PM 🕒 2m 54s


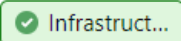
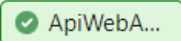
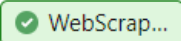
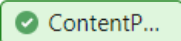

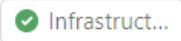






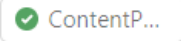
Jobs

Name	Status	Duration
 Infrastructure	Success	🕒 43s
 Applications	Running	🕒 1m 57s

Azure DevOps: Release



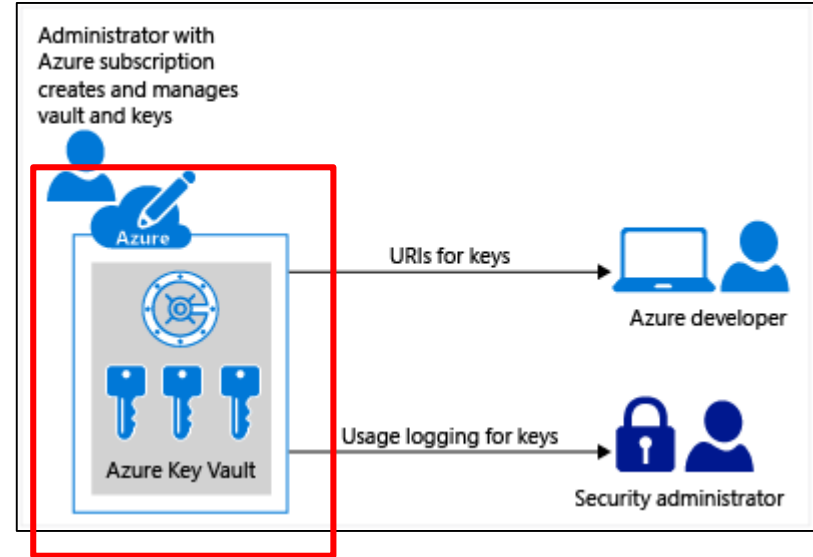
Azure DevOps: Release- Overview

Releases	Deployments	Analytics				All rel
Releases			Created		Stages	
 Release-18 WebFetch CI_20... master			8/25/2019, 8:44:37 AM		   	
 Release-17 WebFetch CI_20... master			8/25/2019, 8:32:13 AM		   	
 Release-15 WebFetch CI_20... master			8/24/2019, 10:41:52 PM		  	

Azure Key Vault- Concept

Securely store and control access to:

- **Secret management**: Tokens, passwords, certificates, API keys, and other secrets.
- **Key management**: Create and control encryption keys that encrypt your data.
- **Certificate management**: Provision, manage, and deploy public and private certificates



Azure / Cloud

Azure Portal

Azure Portal

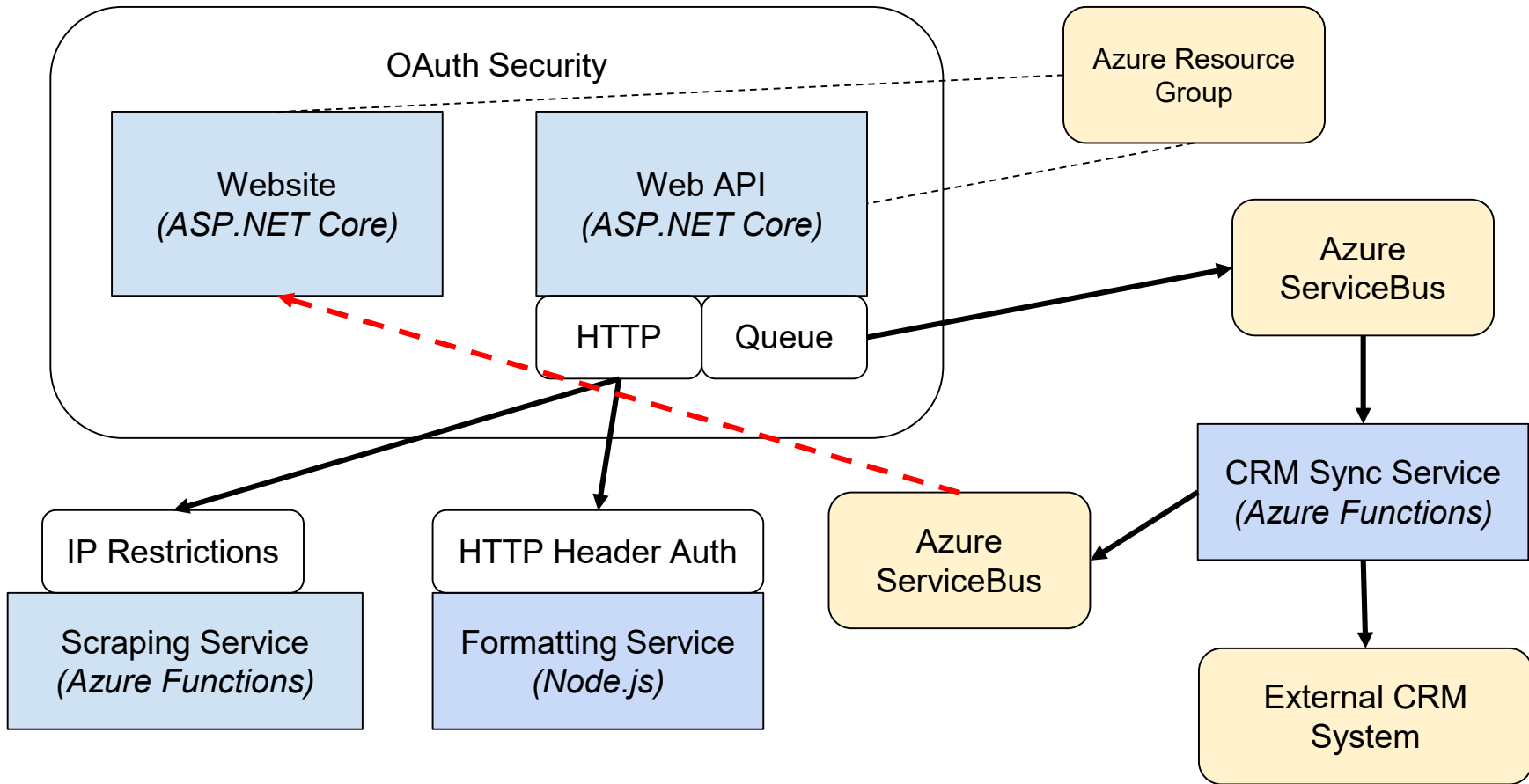
The screenshot displays the Microsoft Azure Portal dashboard for a resource group named 'MYNODEAPP'. The interface is organized into several key sections:

- Resource Group:** A list of resources including Redis Cache, Virtual Machine, Traffic Manager, My SQL Database, Storage Account, Web Apps, and App Service Plan.
- Web Front End:** Contains widgets for 'MyNodeApp WEB APP' (Running), 'MyNodeApp APP SERVICE PLAN' (Available), and a line chart showing 'CPU Percentage and Memory Percentage Past week'.
- Database:** Shows 'MyNodeApp SQL SERVER' (Available) and 'MyNodeApp SQL DATABASE' (Online), along with a chart for 'DTU percentage and Database size percentage past week'.
- Processes:** A table listing 'All runs' with columns for status, start time, and duration.
- Application map:** A flow diagram showing client-side and server-side metrics, such as '96 script errors' and '136.35 ms'.
- Activity Log:** Summary cards for 'MyNodeApp ACTIVITY LOG' (27), 'Sessions MYNODEAPP' (11), and 'Page views MYNODEAPP' (623).
- Browser Performance:** A chart for 'Browsers MYNODEAPP' showing metrics like 'PAGE VIEW LOAD TIME' (1.17s) and 'CLIENT PROCESSING' (490.14ms).
- DTU Quota:** A gauge showing 'CURRENT' DTU at 32 and 'QUOTA' at 55000.
- Cognitive Services:** A chart for 'Total Calls past week' showing a total of 23.2K calls.

Azure Portal

The screenshot displays the Microsoft Azure Portal interface. The top navigation bar includes the search bar and user profile. The left sidebar, titled 'All services', is highlighted with a red box and contains a search field and a list of service categories. The main content area shows a grid of services organized into three sections: GENERAL (15), COMPUTE (23), and NETWORKING (27). Each service entry includes an icon, name, and a star icon for favorites.

Category	Service Name	Star	
GENERAL (15)	All resources	★	
	Subscriptions	★	
	Reservations	★	
	Service Health	★	
	What's new	★	
	Recent	★	
	Resource groups	★	
	Marketplace	★	
	Templates	PREVIEW ★	
	Quickstart Center	PREVIEW ★	
	Management groups	★	
	Cost Management + Billing	PREVIEW ★	
	Help + support	★	
	Tags	★	
	Shared dashboards	★	
COMPUTE (23)	Virtual machines	★	
	Virtual machines (classic)	★	
	Virtual machine scale sets	★	
	Function Apps	★	
	App Services	★	
	Batch accounts	★	
	Service Fabric clusters	★	
	Mesh applications	PREVIEW ★	
	Cloud services (classic)	★	
	Kubernetes services	★	
	Disks	★	
	Disks (classic)	★	
	Images	★	
	OS images (classic)	★	
	Availability sets	★	
	Snapshots	★	
	VM images (classic)	★	
	Citrix Virtual Desktops Essentials	★	
	Citrix Virtual Apps Essentials	★	
	CloudSimple Virtual Machines	★	
	SAP HANA on Azure	PREVIEW ★	
	NETWORKING (27)	Virtual networks	★
		Virtual networks (classic)	★
Load balancers		★	
Application gateways		★	
Local network gateways		★	
Virtual network gateways		★	
Traffic Manager profiles		★	
DNS zones		★	
CDN profiles		★	
ExpressRoute circuits		★	
Network security groups		★	
Network security groups (classic)		★	
Network Watcher		★	
Network interfaces		★	
Public IP addresses		★	
Reserved IP addresses (classic)		★	
Connections		★	
Public IP Prefixes		PREVIEW ★	
On-premises Data Gateways		★	
Route tables		★	
Route filters		★	
Application security groups		★	
DDoS protection plans		★	
Firewalls		★	



Terraform

Terraform

- Infrastructure as code
- Declarative configuration of resources/infrastructure
- Reproducible infrastructure (across environments)
- Configuration stored in version control
- Shows preview of changes
- Automation-friendly
- Supported on Azure DevOps + Azure
- Supports all major cloud-vendors + more
- Open Source

Terraform - Apps

```
resource "azurerm_resource_group" "main" {  
  name      = "${local.name}-rg"      # 'webfetch-rg'  
  location = "${var.azure_location}" # 'West Europe'  
}
```

```
resource "azurerm_app_service" "web_app" {  
  name          = "${local.prefix}web-app"  
  location      = "${azurerm_resource_group.main.location}"  
  resource_group_name = "${azurerm_resource_group.main.name}"  
  app_service_plan_id =  
    "${azurerm_app_service_plan.web_app.id}"
```

Terraform - Azure Key Vault

```
resource "random_string" "keyvault_password" { length = 48 }
```

```
resource "azuread_service_principal_password" "keyvault" {  
  value = "${random_string.keyvault_password.result}"  
}
```

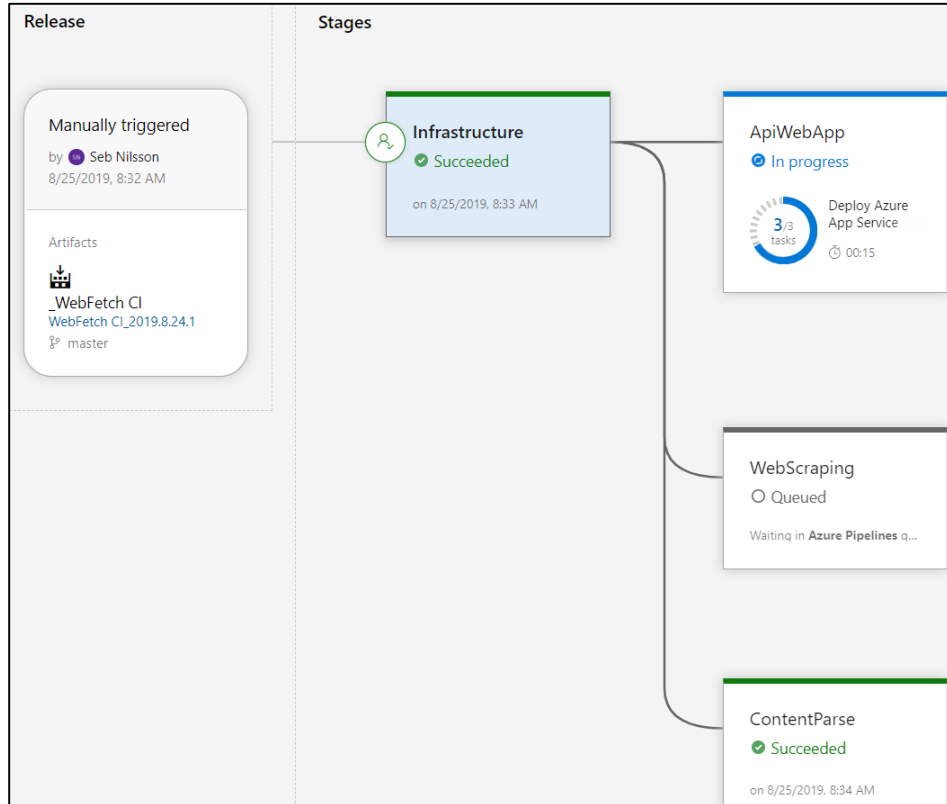
```
resource "azurerms_app_service" "web_app" {  
  app_settings = {  
    KeyVaultClientSecret =  
      "${random_string.keyvault_password.result}"  
    AzureKeyVaultUri =  
      "${azurerms_key_vault.main.vault_uri}"  
  }  
}
```

Terraform - Output

Terraform will perform the following actions:

```
# azure_rm_app_service.api_web_app will be updated in-place
~ resource "azure_rm_app_service" "api_web_app" {
  app_service_plan_id = "/subscriptions/***/resourceGroups/webfetch-rg/providers/Microsoft.Web/serverfarms/api-web-app-sp"
  ~ app_settings = {
    "APPINSIGHTS_INSTRUMENTATIONKEY" = "fa8f0374-512a-4bd2-b1df-28f6d271ac72"
    "ApplicationInsightsAgent_EXTENSION_VERSION" = "~2"
    ~ "AzureKeyVaultClientId" = "f0e1f0b2-d9fa-4610-84bd-4da2a20e8331" -> "cdff33b1-da7a-46c6-b4f7-61324eb0717f"
    "AzureKeyVaultClientSecret" = "@55r036C
  - "AzureKeyVaultName" = "https://webfetch-kv.vault.azure.net/" -> null
  + "AzureKeyVaultUri" = "https://webfetch-kv.vault.azure.net/"
    "ContentParseWebAppSecret" = "#FSYYH>y,
```

Terraform - Azure DevOps



SaaS-services

Azure Storage

Azure Storage

- Scalable cloud-storage
- File-storage
 - Cross-platform file system
 - Lift & shift - No changes
- Blob-storage
 - Massively-scalable object storage for unstructured data
 - Supports cold/hot/archive
 - Single infrastructure with global reach
- Table-storage
 - NoSQL key-value store
 - For massive semi-structured datasets

Azure Table Storage

The screenshot shows the Azure Table Storage interface. At the top, there is a 'Credits' tab. Below it is a toolbar with icons for Query, Import, Export, Add, Edit, Select All, Column Options, Delete, Table Statistics, and Refresh. The main area displays a table with the following columns: PartitionKey, RowKey, Timestamp, CreatedAt, ModifiedAt, ValidFrom, ValidUntil, and Value. The first row of data is highlighted in blue.

PartitionKey	RowKey	Timestamp	CreatedAt	ModifiedAt	ValidFrom	ValidUntil	Value
auth0j5d6646ffdfc6d10c4f1b3a71	25185599999999999999	2019-09-14T20:50:30.280Z	2019-09-13T22:00:00.000Z	2019-09-14T20:50:30.232Z	2018-12-31T23:00:00.000Z	2019-12-30T23:00:00.000Z	40
















- PartitionKey/RowKey: Azure Table Storage specific
- ModifiedAt: For data-sync
- ValidFrom/ValidUntil/Value : Domain-data

Auth0

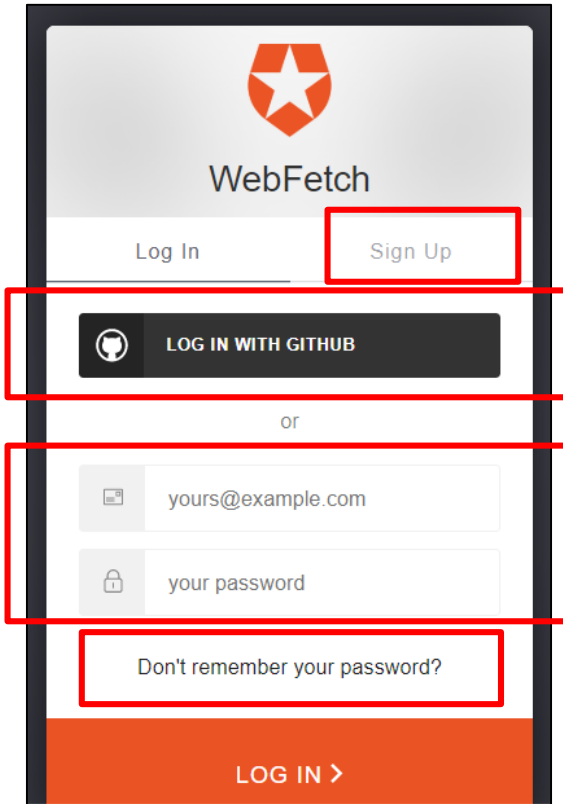
Auth0 - Auth as a Service

- Supports multiple standard protocols
 - OAuth 2.0
 - OpenID Connect
 - SAML, WF-Federation, LDAP
- No additional libs/packages needed
- Supports external identity providers
 - Facebook, Twitter, GitHub, Microsoft, Google, Apple, Dropbox...&many, many more
 - Passwordless
- Single sign on
- Free starting tier
- SaaS - Someone else handles uptime, updates & security

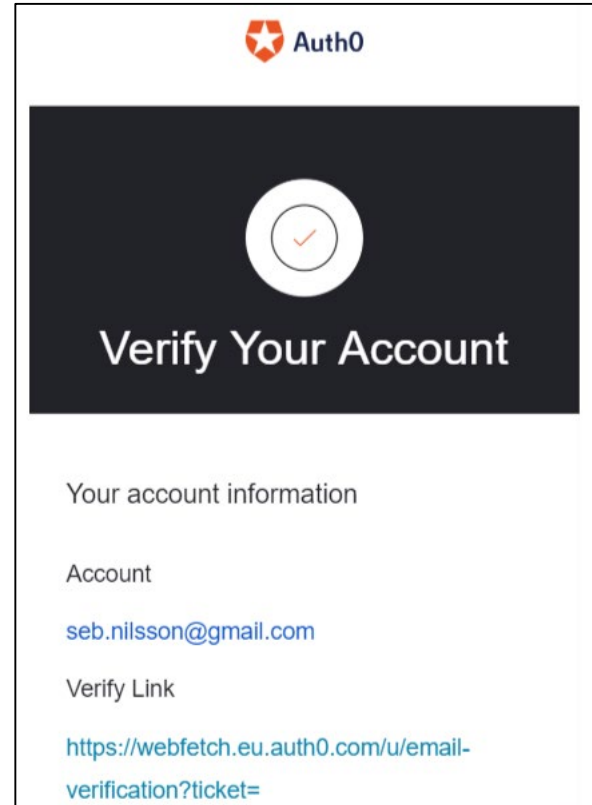
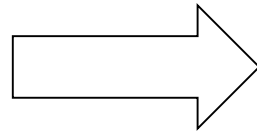
Auth0 - Social connections

 <input type="checkbox"/>	 <input type="checkbox"/>	 Apple <input type="checkbox"/> BETA
 Microsoft <input type="checkbox"/>	 <input type="checkbox"/>	 GitHub <input checked="" type="checkbox"/> TRY >
 Dropbox <input type="checkbox"/>	 Bitbucket <input type="checkbox"/>	 PayPal <input type="checkbox"/>
 <input type="checkbox"/>	 <input type="checkbox"/>	 <input type="checkbox"/>
 <input type="checkbox"/>	 <input type="checkbox"/>	 <input type="checkbox"/>

Auth0 - Login-page



The image shows a mobile-style login page for 'WebFetch'. At the top is the Auth0 logo and the text 'WebFetch'. Below this are two buttons: 'Log In' and 'Sign Up'. The 'Sign Up' button is highlighted with a red box. Below the buttons is a dark grey button with the GitHub logo and the text 'LOG IN WITH GITHUB', also highlighted with a red box. Underneath is the word 'or'. Below that are two input fields: one for an email address containing 'yours@example.com' and one for a password containing 'your password'. Both input fields are highlighted with a red box. Below the password field is a link that says 'Don't remember your password?', also highlighted with a red box. At the bottom of the page is a large orange button with the text 'LOG IN >'.



The image shows the 'Verify Your Account' page. At the top is the Auth0 logo. Below it is a dark grey banner with a white circular icon containing a red checkmark. The text 'Verify Your Account' is centered on the banner. Below the banner, the text 'Your account information' is displayed. Underneath, the word 'Account' is shown, followed by the email address 'seb.nilsson@gmail.com'. Below that is the text 'Verify Link' followed by the URL 'https://webfetch.eu.auth0.com/u/email-verification?ticket='.

Auth0 - GitHub OAuth Application

Register a new OAuth application

Application name *

Something users will recognize and trust.

Homepage URL *

The full URL to your application homepage.

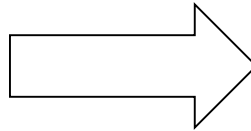
Application description

This is displayed to all users of your application.


Authorization callback URL *

Your application's callback URL. Read our [OAuth documentation](#) for more information.

[Register application](#) [Cancel](#)



auth0

 **auth0user** owns this application.

You can list your application in the [GitHub Marketplace](#) so that other users can discover it.

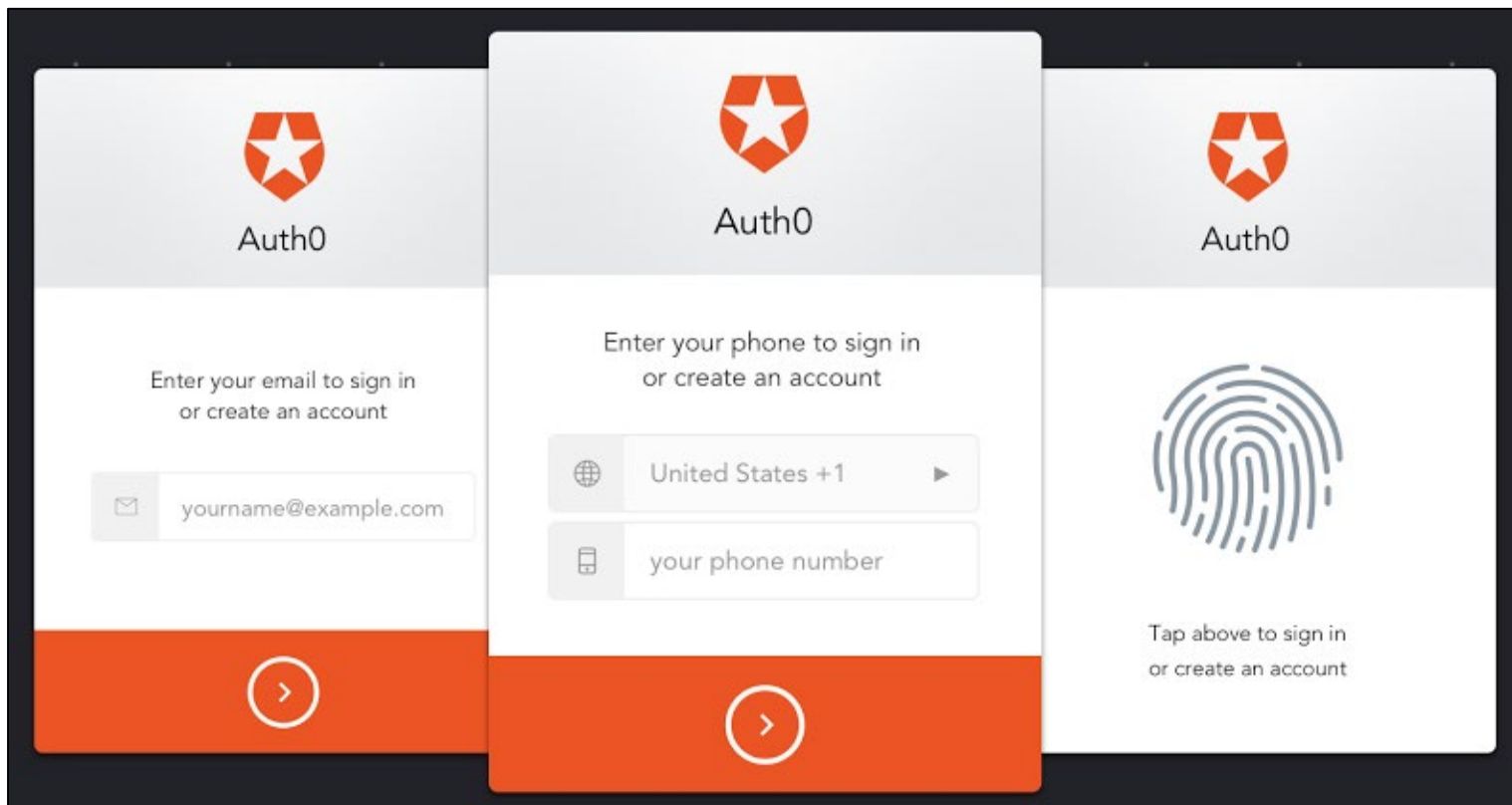
0 users

Client ID
8e17...a064

Client Secret
ea91f4...9517

[Revoke all user tokens](#) [Reset client secret](#)

Auth0 - Don't build this at home



Airtable

Airtable

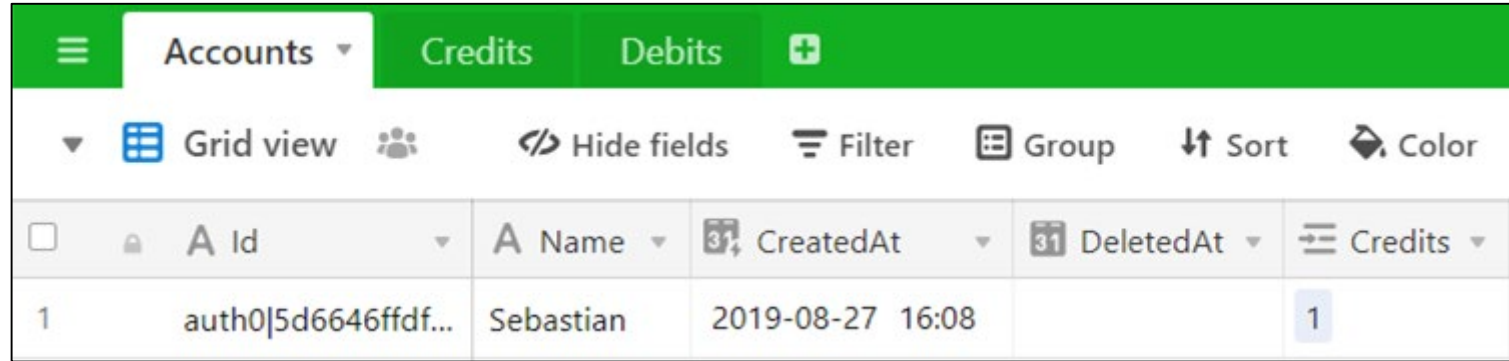
- Database/spreadsheets
- Cloud-based CRM
- Multiple views:
 - Grid
 - Calendar
 - Kanban (Trello)
 - Image-gallery

The screenshot displays the 'Design Project Pipeline' view in Airtable. The interface includes a top navigation bar with 'Design Job Log' and 'Team Members' tabs. Below the navigation, there are filters and sorting options: 'All Projects', '2 hidden fields', 'Filter', 'Grouped by 1 field', 'Sort', 'Color', and a search icon. The main table is organized into sections based on status: 'In-progress' (3 records), 'Blocked' (1 record), 'Need to start' (3 records), and 'Completed' (2 records). Each record includes a task description, priority level (Important, Desirable, Critical), due date, approval status, submitter, and assigned team members.

	Job Request	Priority	Due Date	Approv...	Submitter	Assigned to:	
In-progress STATUS Count 3							
1	Update press kit for company redesign	Important	11/9/2018	✓	Sarah Grimaldi	Al Patino	11/7/
2	Email design for existing customer outreach	Important	11/16/2018		Issra Kadif		11/1/
3	Create mobile support landing page	Desirable	11/25/2018	✓	Shonda Stevens	Minona Writer	11/2/
Blocked STATUS Count 1							
4	New ad creative for ads	Important	11/30/2018		Billy Carpenter	Jack Michaelson Jodi I	11/3/
Need to start STATUS Count 3							
5	Design holiday-themed boxes & packaging	Important	12/9/2018		Patrick Lopez	Jack Michaelson Al Pa	12/8/
6	Presentation for CES 2017 keynote	Critical	12/16/2018	✓	Karina Chen	Al Patino	12/1/
7	New version of setup & installation guide	Desirable	12/16/2018	✓	Timothy Winters	Minona Writer	12/1/
Completed STATUS Count 2							
8	Assets for new software update	Important	11/2/2018	✓	Katrina Dickson	Minona Writer	10/2/

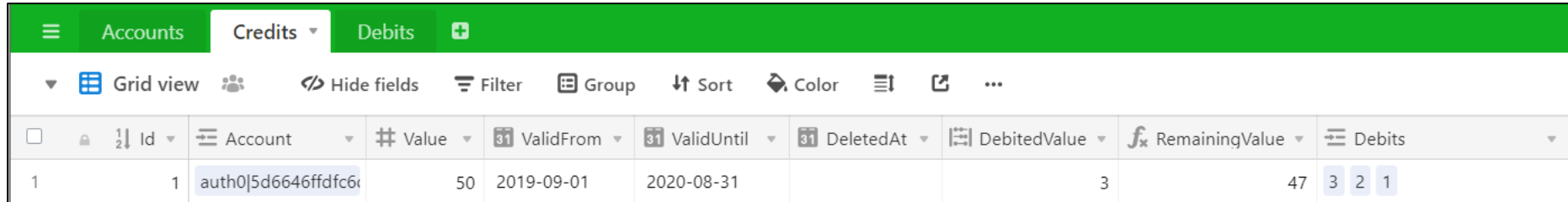
9 records

Airtable - WebFetch-setup



The screenshot shows the top portion of an Airtable interface. The header bar is green and contains a menu icon, the text 'Accounts', and two tabs labeled 'Credits' and 'Debits'. Below the header, there is a toolbar with icons for 'Grid view', 'Hide fields', 'Filter', 'Group', 'Sort', and 'Color'. The table below has the following columns: 'Id', 'Name', 'CreatedAt', 'DeletedAt', and 'Credits'. The first row of data shows '1' in the Id column, 'auth0|5d6646ffdf...' in the Name column, '2019-08-27 16:08' in the CreatedAt column, and '1' in the Credits column.

	A Id	A Name	31 CreatedAt	31 DeletedAt	≡ Credits
<input type="checkbox"/>	1	auth0 5d6646ffdf...	2019-08-27 16:08		1



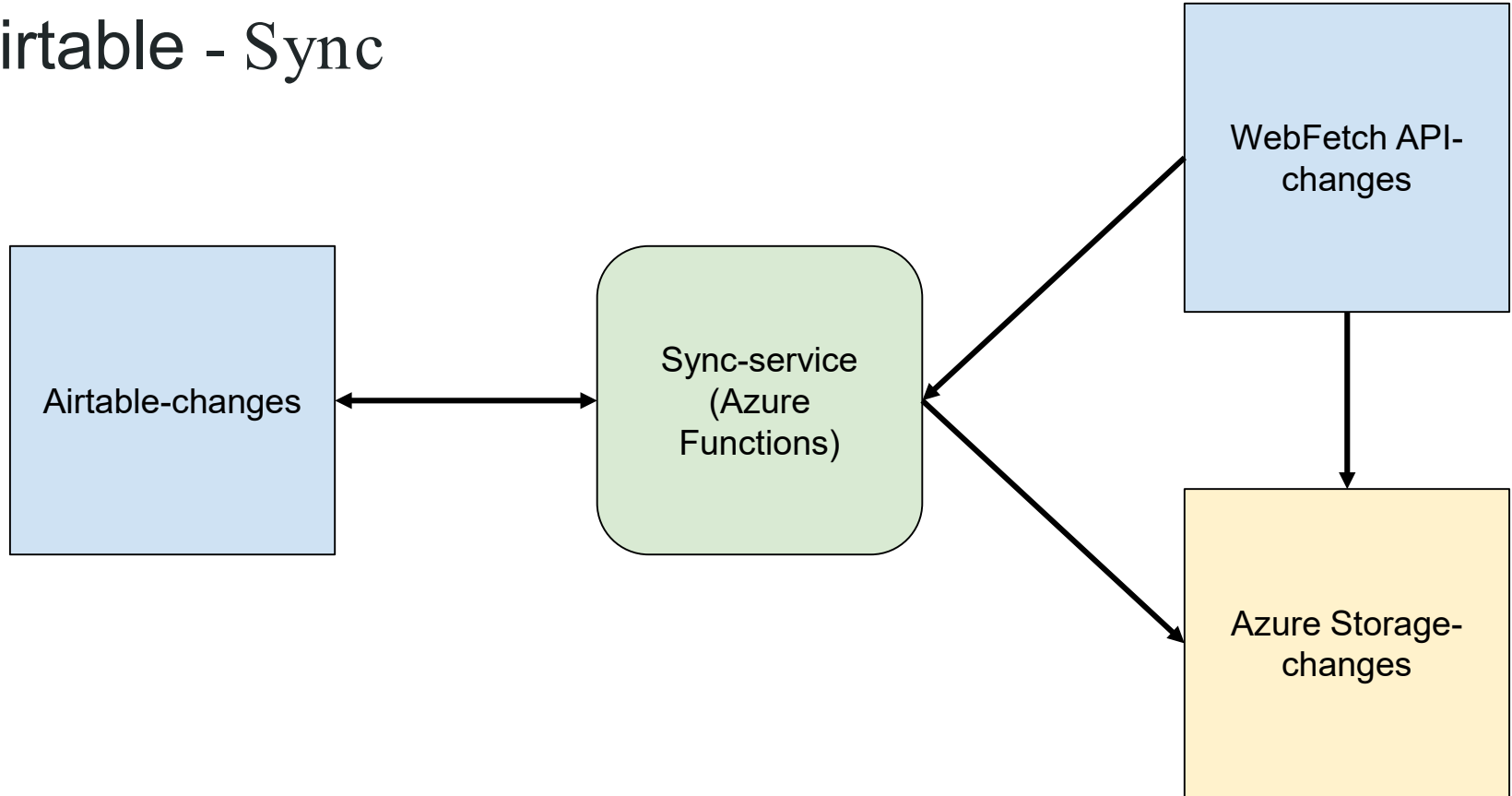
The screenshot shows a more detailed view of an Airtable interface. The header bar is green and contains a menu icon, the text 'Accounts', and two tabs labeled 'Credits' and 'Debits'. Below the header, there is a toolbar with icons for 'Grid view', 'Hide fields', 'Filter', 'Group', 'Sort', 'Color', and a three-dot menu. The table below has the following columns: 'Id', 'Account', 'Value', 'ValidFrom', 'ValidUntil', 'DeletedAt', 'DebitedValue', 'RemainingValue', and 'Debits'. The first row of data shows '1' in the Id column, '1' in the Account column, 'auth0|5d6646ffdfc6...' in the Account column, '50' in the Value column, '2019-09-01' in the ValidFrom column, '2020-08-31' in the ValidUntil column, '3' in the DebitedValue column, '47' in the RemainingValue column, and '3 2 1' in the Debits column.

	Id	Account	# Value	31 ValidFrom	31 ValidUntil	31 DeletedAt	≡ DebitedValue	f RemainingValue	≡ Debits
<input type="checkbox"/>	1	1	auth0 5d6646ffdfc6...	50	2019-09-01	2020-08-31	3	47	3 2 1

Airtable API - Credits

```
{
  "id": "recd821xGdDnWDJg8",
  "fields": {
    "Debits": [
      "recJ6x0VpqSeBECn0", "rec11Bwe9L31t6XHs"
    ],
    "Value": 10,
    "Id": 1,
    "Account": [
      "rec11DbLt5PIKOFTr"
    ],
    "DebitedValue": 2,
    "RemainingValue": 7
  },
  "createdTime": "2019-08-27T14:11:35.000Z"
}
```

Airtable - Sync



cdnjs.com

cdnjs.com

- CDN for JS-libraries
 - No need for local files & package-management
- Host many popular libraries
 - JavaScript, CSS, SWF, images & more
- Supports http2.0/http/https/spdy
- Used by 9,2% of all websites
 - Lots of cached libraries → Speed
- Supported by Cloudflare, DigitalOcean & Algolia

cdnjs.com - Usage

```
<link rel="stylesheet"  
href="https://cdnjs.cloudflare.com/ajax/libs/twitter-bootstrap/  
4.3.1/css/bootstrap.min.css" />  
<!-- ... -->  
<script src="https://cdnjs.cloudflare.com/ajax/libs/jquery/  
3.4.1/jquery.min.js"></script>  
<script src="https://cdnjs.cloudflare.com/ajax/libs/twitter-  
bootstrap/  
4.3.1/js/bootstrap.min.js"></script>
```


Frameworks/Libraries

Swagger
(Swashbuckle)

Swagger - OpenAPI Specification

- Swagger tooling for APIs built with ASP.NET Core
- Automatic API documentation
- UI to explore & test API
- Generates Swagger/OpenAPI JSON
- Enables a self-describing API
- Minimal coding effort

Swagger - OpenAPI Specification

WebFetch API v1 OAS3

[/swagger/v1/swagger.json](#)

API for using WebFetch-features

RateLimit

GET	<code>/api/v1/rate-limit</code>	Gets the rate-limit of the current user
-----	---------------------------------	---

Scraping

GET	<code>/api/v1/scraping/raw-html</code>	Gets the raw HTML from URL.
POST	<code>/api/v1/scraping/raw-html</code>	Gets the raw HTML from URL.
POST	<code>/api/v1/scraping/content-text</code>	Gets the text of the content from URL.
POST	<code>/api/v1/scraping/content-html</code>	Gets the HTML of the content from URL.
POST	<code>/api/v1/scraping/content-markdown</code>	Gets the Markdown of the content from URL.

Swagger - Parameters/Request body

GET /api/v1/scraping/raw-html Gets the raw HTML from URL.

Parameters Try it out

Name	Description
url * required string (query)	The URL to get content from. <i>Default value :</i>
httpMethod string (query)	The HTTP-method to use. <i>Default value : GET</i>

POST /api/v1/scraping/raw-html Gets the raw HTML from URL.

Parameters Try it out

No parameters

Request body application/json-patch+json ▾

Example Value | Schema

```
ContentRequest ▾ {  
  method  
  url*  
}
```

string
string

Swagger - Responses

Responses

Code	Description
200	<p>Success</p> <p>application/json ▾</p> <p>Controls Accept header.</p> <p>Example Value Schema</p> <pre>ContentResponse ▾ { statusCode integer(\$int32) contentType string contentHeaders > {...} headers > {...} content string }</pre>

Swagger - OpenAPI

```
{
  "openapi": "3.0.1",
  "info": {
    "title": "WebFetch API",
    "description": "API for using WebFetch-features",
    "version": "v1"
  },
  "paths": {
    "/api/v1/scraping/raw-html": {
      "get": {
        "tags": [
          "Scraping"
        ],
        "summary": "Gets the raw HTML from URL.",
        "parameters": [
          {
            "name": "url",
            "in": "query",
            "description": "The URL to get content from.",
            "required": true,
            "schema": {
              "type": "string",
              "default": ""
            }
          }
        ]
      }
    }
  }
}
```

Swagger - Controller

```
/// <summary>
/// Gets the raw HTML from URL.
/// </summary>
/// <response code="400">If the URL has no content</response>
[HttpGet("api/v1/scraping/raw-html")]
Public ActionResult<RawHtmlResponse> RawHtml(
    [Required]
    string url,
    string httpMethod = "GET")
{
    // ...
}
```


Polly

Polly

- Resilience & transient-fault-handling library
- Configure policies:
 - Retry
 - Circuit Breaker
 - Timeout
 - Fallback
- Supports HttpClientFactory from ASP.NET Core 2.1

Polly - Implementation

```
services
    .AddHttpClient<IHtmlParseService, HtmlParseService>(
        httpClient =>
        {
            httpClient.BaseAddress = new Uri(webScrapingAppUrl);
        })
    .AddPolicyHandler(
        HttpPolicyExtensions
            .HandleTransientHttpError()
            .WaitAndRetryAsync(
                5,
                retryAttempt =>
                    TimeSpan.FromSeconds(Math.Pow(2, retryAttempt)));
    );
```

Future Steps

Ideas for Future Steps

- Azure Table Storage → Azure SQL Database
- Auth0 → Azure Active Directory B2C / IdentityServer
- ASP.NET Core InMemory Cache → Azure Cache for Redis
- Azure App Service → Docker-containers
- HttpClient (.NET) → PhantomJS/SlimerJS headless browser (Node.js)
- Monitoring with Datadog (configured with Terraform)
- Own Implementation → Azure Application Gateway
- Use Azure EventHub/EventGrid
- Virtual Networks/IP-restrictions

Summary

Time Estimation Comparison

Feature	Cloud Native	Traditional
Auth (Register, Login, Change Password, UI, OAuth, OpenID, ...)	2h	2 weeks
Data Storage (Backup, Redundancy, Scalability, ...)	1h	2 weeks
DevOps-setup (Configuration, Backup, ...)	1h	1-2 day
Machine/Web Server-setup (Configuration, IIS, Backup, ...)	1h	1-2 days
Keeping API-documentation up to date	0,5h	2-3 days
HTTP Retry Policy Logic (Exponential backoff, Retry-count, ...)	0,1h	1-2 days
Synching a CRM	2h	0h
Creating a CRM	0h	50 weeks

2,71€

Monthly Azure-Bill

Takeaways

- Stand on the shoulders of giants -- Don't reinvent the wheel
- Decide on a MVP -- What is “good enough” to start with
- Create a feature-roadmap
- Plan how to switch out fillers with full implementation
- Architecture for switching out parts in the solution
 - Azure Table Storage →SQL Server: **Repositories**
 - Auth0 →Azure ActiveDirectory B2C: **OAuth/OpenID**
 - Architectural changes: **Robust APIs**

Minimize time to market with **smart shortcuts**

Don't re-invent the wheel

Don't do yak-shaving

Thank you!



Questions?

Sebastian Nilsson

Website:

`https://sebnilsson.com`

Twitter:

`@sebnilsson`

GitHub:

`/sebnilsson`

