Building Scalable Services:
Surviving Pokemon and other Natural Disasters
Jim Mason, Paul Dodd & Vivek Gupta
Pokemon GO
PokeVision

Visits

Tile Requests
Responding to Natural Disasters
Geographically Distributed
Servers
Key Considerations
Key Considerations

- Data Optimization
- Infrastructure Setup
- ArcGIS Server/Service Configuration
- Geographic Distribution
  - Multiple Data Centers
  - Geo Load Balancing
  - Content Delivery Network
Data Optimization - Tips

• Reduce Chokepoints
  - Replicate Databases (LBs, AZURE SQL, AWS RDS)
  - Distribute I/O (DAS, RAID, SAN)
  - Distribute Data/Cache Files (multi-server, CDNs, AWS S3)

• Optimize and Delivery
  - Remove fields, check sizes and types, indexes
  - Reduce Geometry storage
    - Precision (coordinate resolution)
    - Dissolve (multipart and/or simplify)
  - Optimize Raster (Tiled TIF / MRF, remote storage by Proxy)
  - Automate creation and delivery
    - Aggregated Live Feed (ALF) and Mosaic Dataset Configuration Script (MDCS)
The Aggregated Live Feed methodology

• What is it?
  - Processing workflow to handle near real-time data
  - Plus - Python production toolset
  - One more tool in your GIS toolbox

• Why would I use it?
  - Dealing with an unstable data source?
  - Need to reduce Internet load?
  - Internalize content!
  - Would you like to automate content refreshment?
ALF - Toolkit

• Benefits
  - Multi-Processing support
  - Process, Error, and Archive Logging (passive!)
  - Integrated E-mail Altering / Status
  - Production oriented
ALF Methodology – Locally Consumed Feed

- Log Files
- Feed Routine (Aggregator)
- Work FileGDB
- Live FileGDB
- Deployment Logic
- Published Map Service
- ArcGIS Server
- GIS User
- Internet
ALF Methodology – Cloud / Distributed Feed

Internet

Data

Log Files

Feed Routine (Aggregator)

Work FileGDB

Deployment Logic

Zip File

FileGDB Archive

Common Storage
(External or Internal)

Amazon S3

Download and Deployment Logic

Work Folder

Live FileGDB

Published Map Service

ArcGIS Server

Data Deployment
ALFdeployer.py

Feed Aggregator
ALFprocessor.py
ALF Demo
SandyPrecip Routine Started: 22:07:08 Tue 03/19/2013

Importing ArcPy...

    arcpy: v10.1.1
  Configuration: SandyPrecip.cfg, v1.0.0
  Processor: ALProcessor.py, v0.8.0
  Library: ALFlib.py, v1.6.0

Processing: 'Sandy_Oct30_00.shp'

Preparing workspace...
  <directory>: C:\EPC_Sandy\work'
  <gdbName>: HurricaneSandy.gdb'
  <recreate>: True

Removing File Geodatabase...  
Creating File Geodatabase...  
Returned: C:\EPC_Sandy\work\HurricaneSandy.gdb

Importing data...

Executing: CopyFeatures Sandy_Oct30_00.shp C:\EPC_Sandy\Work\HurricaneSandy.gdb\CurrentPrecip 0 0 0
Start Time: Tue Mar 19 12:07:26 2013
Succeeded at Tue Mar 19 12:07:27 2013 (Elapsed Time: 1.00 seconds)

Creating Spatial Index...

Executing: AddSpatialIndex C:\EPC_Sandy\Work\HurricaneSandy.gdb\CurrentPrecip 5 0 0
Start Time: Tue Mar 19 12:07:27 2013
Succeeded at Tue Mar 19 12:07:27 2013 (Elapsed Time: 0.00 seconds)

Deploying...
  Gathering File List...
    Copying From:
      'C:\EPC_Sandy\work\HurricaneSandy.gdb'
    To:
      'C:\EPC_Sandy\live'...
  Updating: 'HurricaneSandy.gdb\a00000001.gdbindexes'
  Updating: 'HurricaneSandy.gdb\a00000001.gdtabell'
ALF Feed Construction – What do I need?

- **Pre-requisites**
  - Python
  - ArcPy
  - Others (.py, 7zip, degrib, …)

- **Optional**
  - IDE – PyScripter, VS, …

- **Components**
  - ALFlib.py
    - Function / Class Library
  - ALFprocessor.py
    - Feed Processor +
    - ‘<feed file>.cfg’ (data processing logic)
  - ALFdeployer.py
    - Archive Distribution
Resources

• AGOL Community Page (ALF samples, scripts) : http://esriurl.com/LiveFeed


• Optimizing Rasters : https://github.com/Esri/OptimizeRasters

• Image Management Workflows : http://esriurl.com/6550

• Author : pdodd@esri.com
System Design
System architectures
High Availability, High Capacity

Traffic Manager

http / https

Loadbalancer

Loadbalancer2
ArcGIS Service Criteria

- Service Design
- Min/Max number of instances
- Max Wait time
- Max Startup time
- Max Usage time
- Recycle
- Low/High Isolation
- Stop/Delete Unwanted Services
ArcGIS Online Hosted Model

• When you have unforeseen demand that cannot be met by running your own infrastructure

• When you need to broaden reach of your content outside of the enterprise firewall

• When you want to use the latest capabilities of the platform

• No need to manage infrastructure, maintenance, monitoring, upgrades etc.
ArcGIS Online Hosted Service

- Frequent update of data
- Can be updated using Desktop or Website
- Control Access for editing
- Export data
- Reference
Geo Load Balancing

- Routes traffic based on location
- Used for:
  - Specialize content
  - Load balancing between data centers
  - Optimize network latency
- Amazon Route 53
- Akamai GTM
CDN [ Content Delivery Network ]

- Network of Geographically Dispersed Servers
- Edge Location
- Cache copies of content
- Close to end users
- Lowers latency
CDN / AWS CloudFront Applications

• Static Content
  - Media Content like audio, videos etc.
  - Software download
  - Web Site files like Images, CSS, JS

• Dynamic Content
  - News, Weather, Advertising etc.

• Live Streaming
  - Sporting events
  - Meetings, Gathering etc.
AWS CloudFront Components

- Distribution
  - Set of Rules
- Origin
  - Source of the content
- Behaviors
- Restrictions, Error pages
- WAF
- Edge Locations
- Price Classes
Behaviors

- Path Pattern Matching
- Origin Selection
- Headers
- Query Strings / Cookies
- Time to Live
- GZIP Compression
- Protocol Enforcement
- SSL certificates
Path Pattern Matching

- Route requests to specific origins
- Set Protocol
- Set Header, Caching options
- Set cookie and Query String Forwarding
- Restrict Access
- Set Compression
AWS CloudFront Benefits

• Scalable
• Secured
  - SSL/TLS delivery
  - DDoS Protection, WAF (Web Application Firewall)
• Robust Real Time Reporting
• Cost effective
  - Pay as you go model
Akamai – Designating Origin Server

If

- Hostname
- is one of
  - server.arcgisonline.com

Behaviors

Origin Server

- Origin Type: Your Origin
- Origin Server Hostname: origin-services-geo.tailover.arcgisonline.com.akcdnz.net
- Forward Host Header: Incoming Host Header
- Cache Key Hostname: Origin Hostname
- Supports Gzip: Yes

⚠️ Due to the security risk, it is strongly recommended that you consult with the security expert at your organization before configuring the Origin Server SSL Certificate Verification settings. If you are the security expert at your organization and you have any questions, contact your account team.
Akamai – Pattern Matching for Edge Caching
CloudFront Reporting: Access Logs

- Log delivered to S3
- Permissions Controlled
- Scheduled delivery
- Metrics
  - Cache Statistics
  - Usage Charts
  - Popular Objects
  - Browser, OS, Devices, Locations, Referrers etc..
Key Considerations (in conclusion)

- Data Optimization
- Infrastructure Setup
- ArcGIS Server/Service Configuration
- Geographic Distribution
  - Multiple Data Centers
  - Geo Load Balancing
  - Content Delivery Network
Please Take Our Survey on the **Esri Events App**!

**Download the Esri Events app and find your event**

**Select the session you attended**

**Scroll down to find the survey**

**Complete Answers and Select “Submit”**