NETWORK COMMUNICATION

Marshall Markham

ABOUT YOU

A quick survey about the class's composition

- Degree Program
- Year of school
- Operating System
- General software engineering experience
  - Jobs and internships
  - Years programming experience
- Stack
  - Relational Databases
  - Spark
  - Hadoop/Amazon S3
  - Python, Pandas, Numpy, Scikit Learn

ABOUT ME

- Data Engineer at Precision Lender
- Data Scientist and Business Analyst at MaxPoint/Valassis Digital
- Master's Degree in Mathematical Statistics from NCSU
- 10 years service in the Army
- Father and soccer fan

PRECISION LENDER

- Located in Cary, NC
- Software for the Finance Industry
- Banking Intelligence

PRECISION LENDER – FLY WHEEL

Making analytics practical for commercial banks

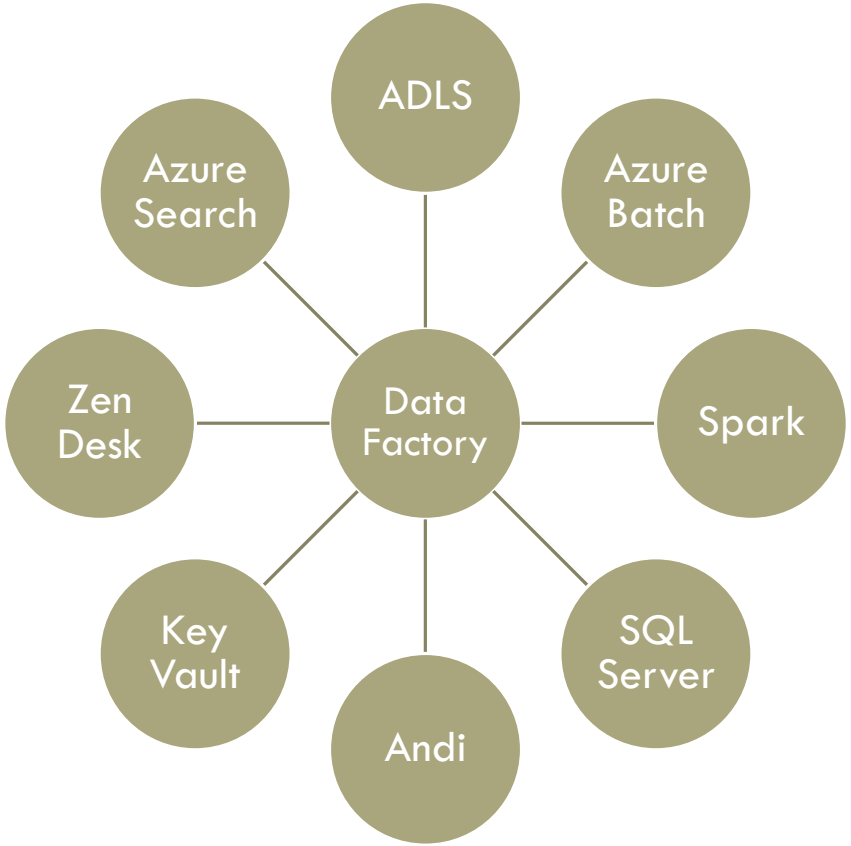
- Users
- Financial data
- Andi
- PrecisionLender pricing application

PRECISION LENDER – L3 (THE DATA)

Presenting financial data to the Bank's analytics team

- Data is both internal and external to the bank
- Presented through a Hive style directory structure
- CSV, Multijson, Parquet, and Avro
- Data schema must resolve across time series

PRECISION LENDER – L3 (THE STACK)



CONNECTING THE STACK

- Microsoft Azure hosted
- Entirely connected through HTTP
- Various connection interactions differ
- Valuable to know:
  - URL construction
  - HTTP
  - Authentication patterns
  - JSON

URL CONSTRUCTION

{SCHEME}://{HOST}[:{PORT}]{PATH}[?{QUERY}]

- Scheme
  - HTTP, HTTPS, HDFS, S3, ADLS
- Host[:Port]
  - Which computer and communication channel
- Path
  - Information hierarchy within the host service
- Query
  - Parameter passing allows us to treat our requests as function calls

EXAMPLE

https://tx11012.prod.cmpny.net/webhdsf/v2/streaming/locations/shapes/0000-abc-123.csv?op=OPEN&buffersize=67108864

- Scheme: https
- Host: tx11012.prod.cmpny.net
- Path: /webhdsf/v2/streaming/locations/shapes/0000-abc-123.csv
- Query: op=OPEN&buffersize=67108864

HTTP VERBS

Best understood as remote CRUD operations on a database

Usually we use GET and POST

- GET : read
- POST : create
- PUT : create/update
- DELETE : delete
- PATCH : partial create/update (rare)
- OPTIONS: metadata about the API

GET VS POST

- Usage is subtler the “read” vs “create”
- Both requests:
  - Send data
  - Result in a response which contains data
- GET sends data in the URL query string
  - Insecure for sending sensitive information (especially over HTTP)
  - Limited in length
  - Simple
  - Response body is not limited in size
- POST sends data in request body
  - Secure
  - Unlimited request body length
  - More complex
  - Response body is not limited size
- POST is used for requests for authentication or depending on complex/lengthy data passing

AUTHENTICATION PATTERNS

Two common authentication patterns: Basic and OAuth 2.0

Basic

- Provide password every time

OAuth 2.0

- Authenticate with password and receive a **bearer token** from an authentication host
- Use bearer token to authenticate to the service host to access API

JSON (JAVA SCRIPT OBJECT NOTATION)

JSON is a common data format for network communication

- Supports basic data types, arrays, and key value lookup
- Hierarchical/Recursive; JSON can contain JSON
- Examples:
  - { “City”: “St. Louis”, “State”: “MO”}
  - {“Location”: {“City”: “St. Louis”, “State”: “MO”}, “Date”: “2019-01-07”}
  - {“RequestDates”: [“2019-01-01”, “2019-02-01”, “2019-03-01”], “RequestType”: “Access”}
  - {“Type”: “Chair”, “QtyType”: “Each”, “Qty”: 4}

FREE DATA

- NASA
- Federal Reserve
- U. S. Census Bureau
  - Of special interest here are GIS services for Geocoding and Reverse Geocoding
- Google
- Twitter
- Data.gov