Introduction to Azure Data Factory
Alan Faulkner

- SR BI Consultant, Trainer Pragmatic Works
- Speaker, SQL Saturdays, Code Camps, Webinars, PreCons, 24 Hours of PASS
- Active Member AZSSUG
- SSAS Maestro

@FalconTekNic
www.linkedin.com/in/alandfaulkner
http://falconteksolutionscentral.com
alan.faulkner@falconteksolutionscentral.com
Advanced Data Integration in a Modern Day Data Warehouse

- Why a more modern DW Platform is needed?
- How does data move and integrate in the Modern Day Data Warehouse?
- How do we take data and make sure it gets where it needs to go?
- How does the cloud work with the Modern Day Data Warehouse?
- How do we integrate new types of data and what does that mean?
Traditional Data Integration
Challenges to the New Data Integrator
Today’s Data Landscape
Azure Data Factory Overview
Azure Data Factory Composition

- Data Storage
- Data Processing
- Data Movement
Azure Data Factory – What Can it do?

- Easily work with diverse data storage and processing systems
- Transform data into trusted information
- Monitor data pipelines in one place
- Get rich insights from transformed data.
Application Model

Data Sources (Import from)

Data Connector: Import from source to Hub

Data Hub (Storage & Compute)

Pipeline of Activities

Data Connector: Import/Export among Hubs

Data Hub (Storage & Compute)

Pipeline of Activities

Data Connector: Export from Hub to data store

BI Tools

Data Marts

Data Lake(s)

Dashboards

Apps

Information Production Stages:

Connect & Collect

Transform & Enrich

Publish
Azure Data Factory Core Concepts

- Data Factory – Fully Managed Service
- Linked Service – Link data stores & Compute Services
- Data Set – Named view of data
- Pipeline – Processes Data
Data Factory Concepts

Data Sources
- Game Logs
- Customer Table

Ingest
- Game Log Files
- Customer Table

Transform & Analyze
- Transform, Combine, etc
- Analyze
- Move
- Geocode

Publish
- Visualize

Data Set
- Collection of files, DB table, etc

Activity: a processing step
- Hadoop job, custom code, ML model, etc

Pipeline: a sequence of activities (logical group)
Azure Data Factory Terminology

- **Activity** – Data Processing Step in a Pipeline
- **Data Hub** – A Container For Data Storage & Compute Services
- **Slice** – Logical Time Based Partition of Data Produced
- **Data Management Gateway** – Software that Connects On-Premises Data Sources to Cloud Services for Consumption.
Data Factory Demo

Review Prerequisites

Step 1: Create Azure Data Factory
Step 2: Create Linked Services
Step 3: Create Input and Output Tables
Step 4: Create and run a pipeline
Step 5: Monitor data sets and pipeline

Review More Advanced Data Factory Solution
The Modern Data Warehouse - Today & Future

- Deploy a modern data warehouse as a foundation for leveraging big data.
- Make sense of multi-structured data for new and unique business insights.
- Implement advanced forms of analytics to enable discovery analytics for big data.
- Empower the business to operate in near real time by delivering data faster.
- Integrate multiple platforms into a unified data warehouse architecture.
- Demand high performance and scalability of all components of a data warehouse.
Resources

Introduction to Azure Data Factory Service
• Introduction to Azure Data Factory Service
• Getting Started with Azure Data Factory
• About Execution Policies

Video
• Azure Data Factory Overview

SQL Server Blog
• The Ins and Outs of Azure Data Factory – Orchestration and Management of Diverse Data

JSON Scripting Reference
• Data Factory JSON Scripting Reference

Azure Storage Explorer Download (CodePlex)
• Azure Storage Explorer 6 Preview 3

Azure PowerShell
• How to Install and Configure Azure PowerShell
• Introducing Power Shell ISE
Platinum Level Sponsors

Microsoft  PYRAMID ANALYTICS  CISCO  NEUDESIC

Gold Level Sponsors

Hortonworks  headfarmer  Plus Consulting

cloudera  DBMS_Insights

Venue Sponsor
GRAND CANYON UNIVERSITY

Key Note Sponsor
Pragmatic Works

Pre Conference Sponsor
tyler technologies
Thank You for Attending