

Profiling Your Data

Angela Henry



Angela Henry

Angela is a Consultant with SQL Data Partners LLC

Data Platform MVP

Local group leader for the PASS User Group in Greensboro, NC

Tweets at @SQLSwimmer

Blogs at SQLSwimmer.wordpress.com



Overview

What

Why

How

What

A data profile is a collection of aggregate statistics about your data

- Distinct Count & Percentage
- Zero, Blank & NULL Percentage
- Min, Max, AVG string lengths
- Numerical Ranges
- Date Ranges

Why

Makes Your Life Easier!

Analyze The Source Data More Effectively

Understand The Source Data Better

Prevent Data Quality Problems Before

They Are Introduced



How

R

Azure Data Catalog

Informatica

Talend Open Studio for Data Quality

SAS

Power BI

SSIS – Data Profiling Task

Demo

Power BI

Currently in Preview

Top 1000 Rows By Default

No String Length Statistics

Incomplete Summary Statistics

SSIS Data Profiling Task

Requirements

Set Up

Analyze

Demo

Review

SQL Server Only Data Source

View Results Within Visual Studio

Use Stand Alone Viewer to View Results

Script Task With XPath Query to Make
Decisions

```
"<?xml version="1.0" encoding="utf-16"?>
<DataProfile xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema"
  xmlns="http://schemas.microsoft.com/sqlserver/2008/DataDebugger/">
```

```
<DataSources />
```

```
<DataProfileInput>
```

```
<ProfileMode>Exact</ProfileMode>
```

```
<Timeout>0</Timeout>
```

```
<Requests>
```

```
<ColumnNullRatioProfileRequest ID="NullRatioReq">
```

```
<DataSourceID>DatabaseConn</DataSourceID>
```

```
<Table Schema="" + @[User::SchemaName] + "" Table="" + @[User::TableName] + "" />
```

```
<Column IsWildcard="true" />
```

```
</ColumnNullRatioProfileRequest>
```

```
<ColumnStatisticsProfileRequest ID="StatisticsReq">
```

```
<DataSourceID>DatabaseConn</DataSourceID>
```

```
<Table Schema="" + @[User::SchemaName] + "" Table="" + @[User::TableName] + "" />
```

```
<Column IsWildcard="true" />
```

```
</ColumnStatisticsProfileRequest>
```

```
<ColumnLengthDistributionProfileRequest ID="LengthDistReq">
```



Feedback Forms

Please fill out and pass to
your room helper before
you leave

References

[Profile Task & Viewer Documentation](#)

[Azure Data Catalog Documentation](#)

[Talend Information](#)

[Power BI Data Profiling](#)

[Automate Multiple Dynamic Tables Blog Post](#)

Contact Information

Twitter: @SQLSwimmer

Blog: sqlswimmer.wordpress.com

Linked In: [\in\angelahenrydba](https://www.linkedin.com/in/angelahenrydba)

Email: henry.angela@gmail.com

