# PMU DATA QUALITY MONITORING AT SPP

PRESENTER: MIKE NUGENT

# AGENDA

- Data Quality/Availability Challenges
- SPP's PMU System
- PMU System Monitoring
- PMU Data Quality Monitoring
- What's Next

# DATA QUALITY CHALLENGES

- Data quality and availability is highly variable.  Affected by equipment failures, communications issues, scheduled maintenance, etc..
- Easy to miss when data has gone bad.  Most downstream analytics are set up to ignore bad data.
- Often don't notice data is bad or missing until that data is needed for after-the-fact analysis
- Current system requires someone to be engaged in looking at reports and dashboards to identify potential data quality issues

# PMU SYSTEM OVERVIEW

- SPP recently moved from a single-site architecture to a highly-available dual-site architecture
- SPP's PMU system is built in a corporate environment outside of the ESP
- SPP uses a mix of open-source and vendor-provided tools:
  - PDC:  GPA SIEGate
  - Historian:  GPA openHistorian
    - 1 year full-resolution archive
  - Analytics:  EPG RTDMS
    - 90 day archive
  - Operator UI:  EPG RTDMS

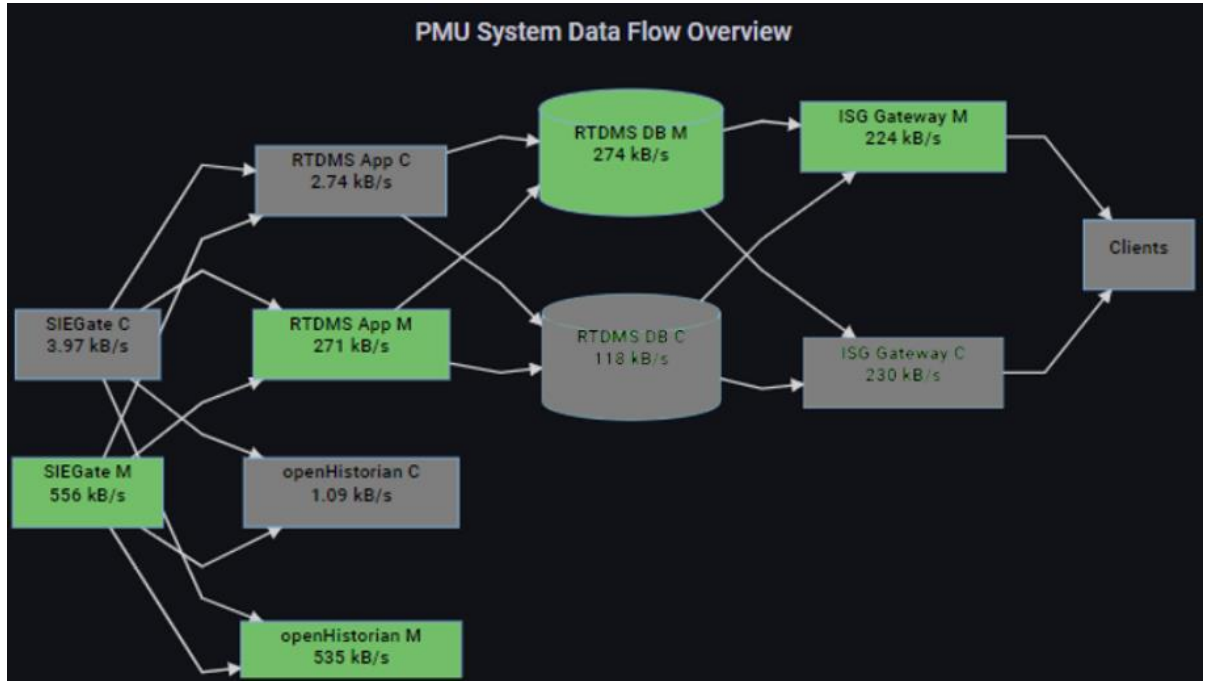

SIEGate

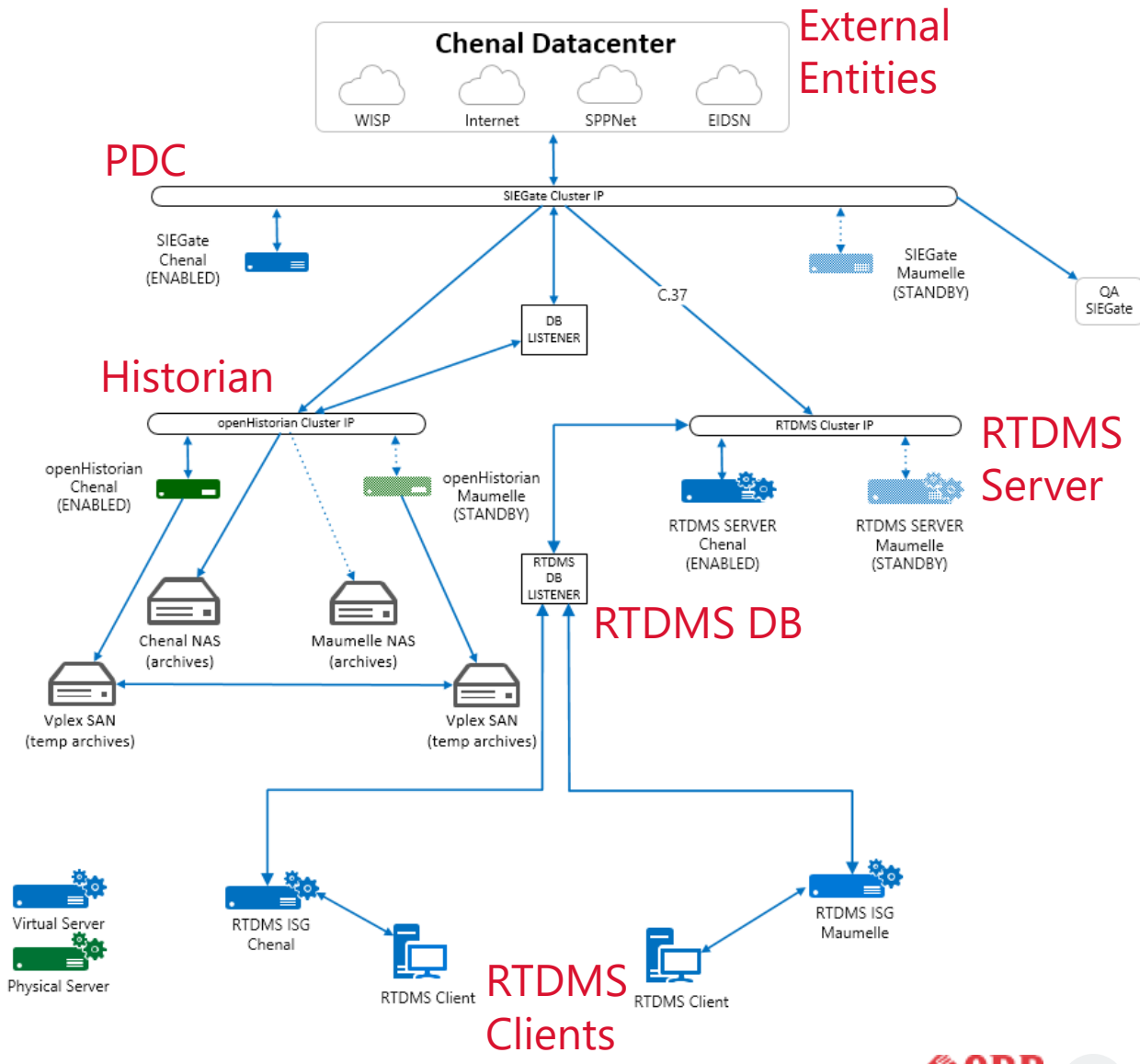
openHistorian 2
High-Performance Measurement Archive

Electric Power Group

PHASOR RTDMS®

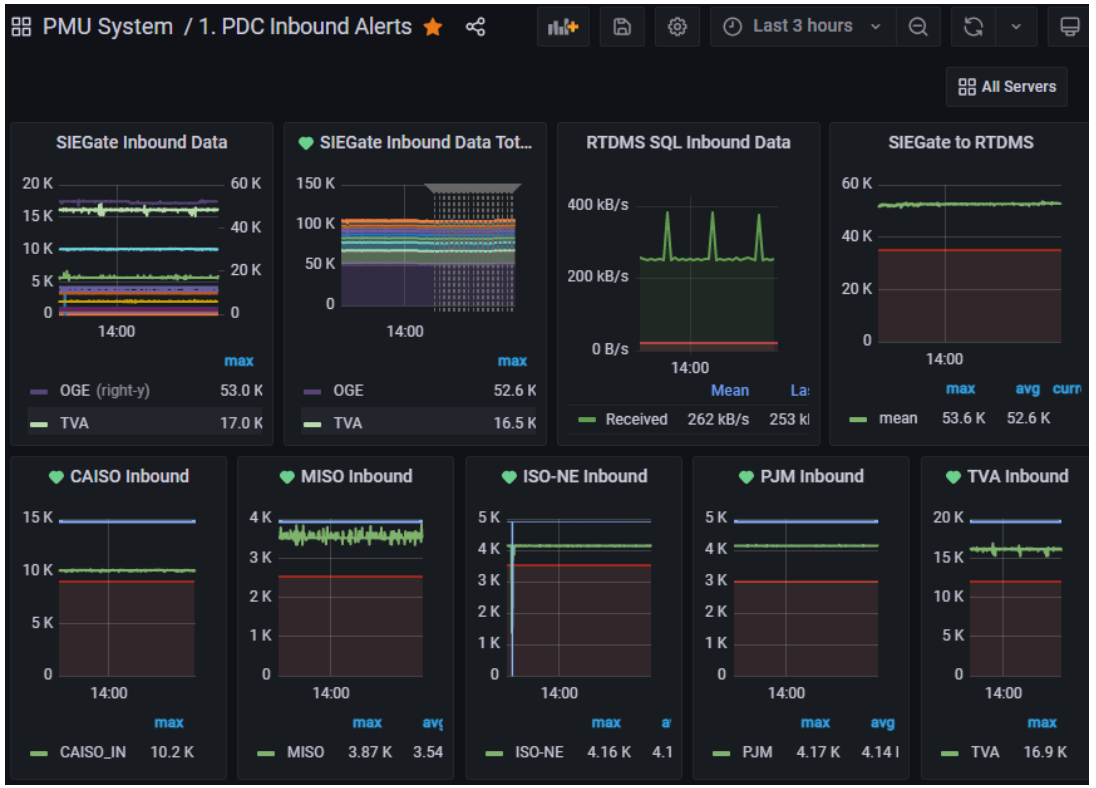
# HIGHLY AVAILABLE PMU ARCHITECTURE



# DATA FLOW MONITORING

- SPP uses measurement volume metric data from SIEGate to monitor and alert for inbound data outages using Grafana.
- Support staff receive automated emails and can begin troubleshooting quickly, depending on criticality
- SPP also monitors PMU devices that are used to calculate mode shapes in the Western Interconnect



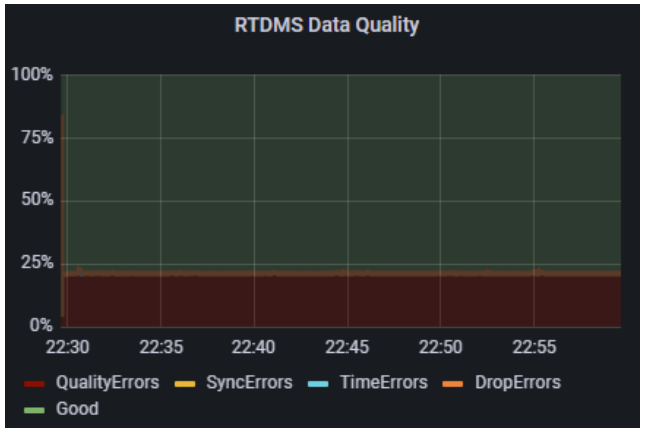

# PMU SYSTEM DASHBOARDS

- SPP uses Windows Performance Counter data to monitor data flow across the PMU system
- Monitoring metrics such as:
  - Network traffic on SIEGate servers
  - Disk I/O on Historian servers
  - Network I/O on RTDMS Database servers

# DATA QUALITY DASHBOARDS

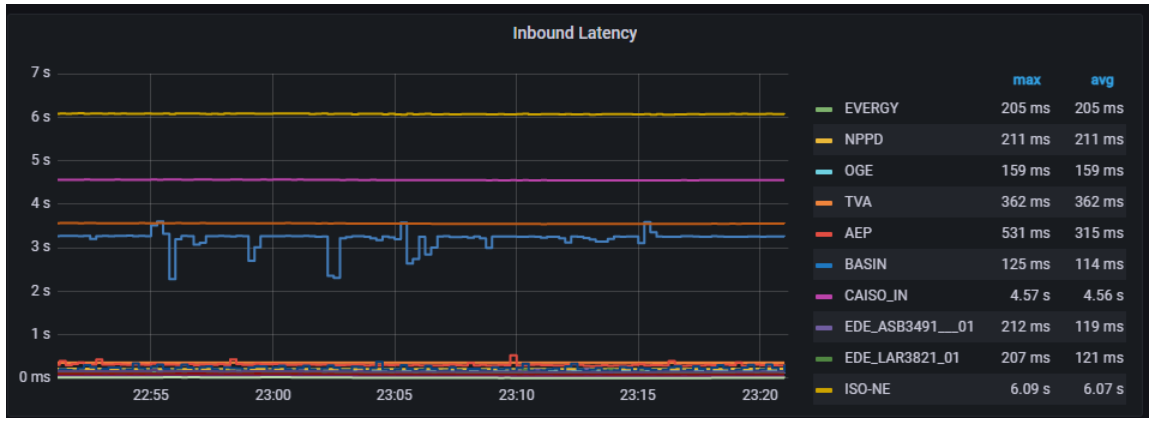
- SPP uses metric data from SIEGate and RTDMS for data quality dashboards

- Monitoring metrics such as:
  - Device-level % of “good” measurements based on PMU quality flags
  - Data completeness
  - Latency of inbound data




SIEGate Measurement Quality — Last 1 hour

| Time | device | Streams | Good | Quality Errors | Time Errors | Device Errors |
|---|---|---|---|---|---|---|
| 2022-09-20 21:41:22 | | 7 | 0% | 0 | 0 | 0 |
| 2022-09-20 21:41:22 | | 7 | 0% | 0 | 0 | 0 |
| 2022-09-20 21:41:22 | | 7 | 46% | 0 | 0 | 0 |
| 2022-09-20 21:41:22 | | 19 | 63% | 0 | 0 | 0 |
| 2022-09-20 21:41:22 | | 15 | 80% | 0 | 0 | 0 |
| 2022-09-20 21:41:22 | | 6 | 83% | 0 | 0 | 0 |
| 2022-09-20 21:41:22 | | 6 | 100% | 0 | 0 | 0 |

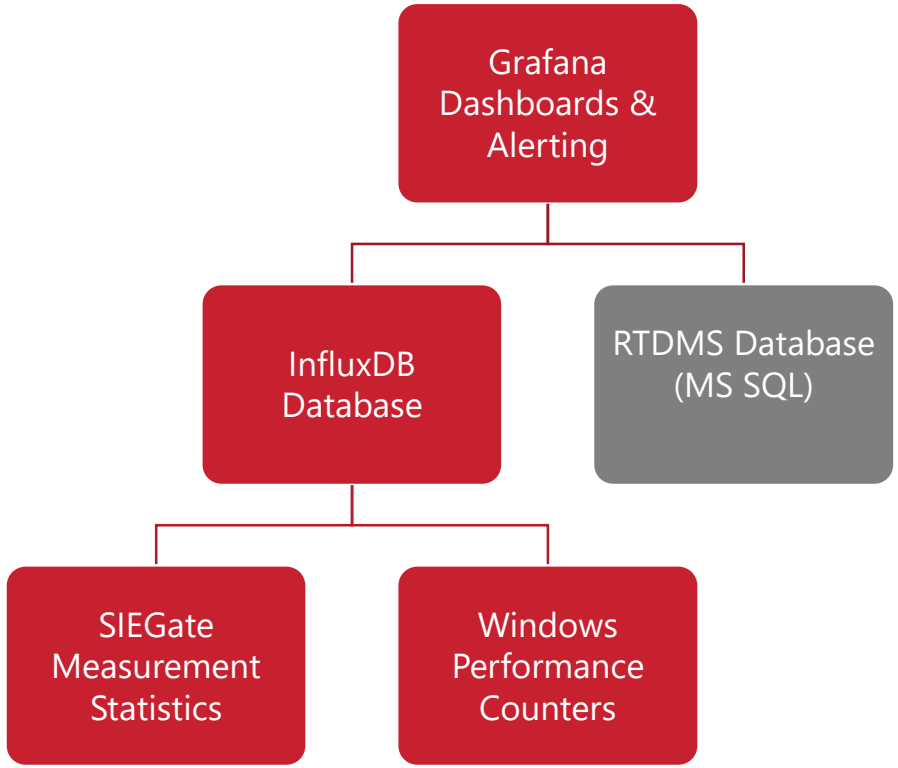



RTDMS Measurement Quality — Last 1 hour

| PMU | QualityErrors | SyncErrors | TimeErrors | DropErrors | Good |
|---|---|---|---|---|---|
| | 98.0% | 0.0% | 0.0% | 2.0% | 0.0% |
| | 98.0% | 0.0% | 0.0% | 2.0% | 0.0% |
| | 98.0% | 0.0% | 0.0% | 2.0% | 0.0% |
| | 98.0% | 0.0% | 0.0% | 2.0% | 0.0% |
| | 2.6% | 0.0% | 0.0% | 15.1% | 82.3% |
| | 9.8% | 0.0% | 0.0% | 2.0% | 88.2% |
| | 0.6% | 0.0% | 0.0% | 2.0% | 97.4% |

# DATA MONITORING BACKEND

- Statistics from SIEGate are written into InfluxDB using a GPA-provided output adaptor
- Windows performance counters are gathered from various servers and written to InfluxDB using custom code
- Grafana dashboards are built using native InfluxDB and SQL Server datasources
- Technologies used:
  - Grafana – open-source dashboarding software supporting multiple data sources
  - InfluxDB – open-source time series database
  - Ostperfmon – custom-developed agent that collects Performance Monitor metrics from remote servers and writes them to InfluxDB
  - SQL Server queries to extract metric data from RTDMS database



# EMAILED REPORTS

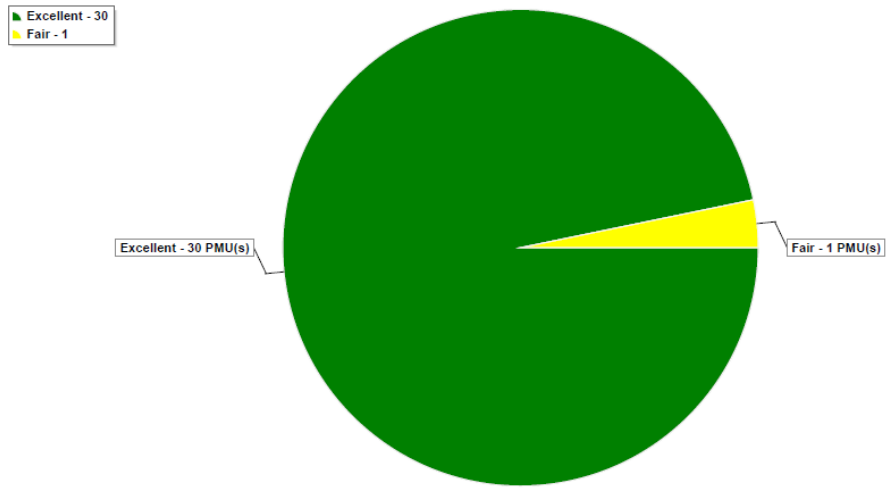
- SIEGate creates a daily PDF-formatted report that is emailed to support staff
  - These reports are nice in that they show recent trends of data completeness

- RTDMS/Gridsmarts sends a daily report for specific Western Interconnect PMU devices.

- Both of these reports are reviewed on a daily/regular basis by support staff.

**5-day Device Data Completeness**

| | 09/15 | 09/16 | 09/17 | 09/18 | 09/19 |
|---|---|---|---|---|---|
| **L4: Good** | 248 | 243 | 250 | 251 | 252 |
| **L3: Fair** | 3 | 8 | 8 | 4 | 4 |
| **L2: Poor** | 3 | 13 | 6 | 8 | 7 |
| **L1: Offline** | 14 | 4 | 4 | 5 | 5 |
| **L0: Failed** | 6 | 6 | 6 | 6 | 6 |
| **Total** | 274 | 274 | 274 | 274 | 274 |



# WHAT'S NEXT

- Refine processes to triage data quality issues:
  - When is it an issue?
  - What is the criticality?
  - Who should be notified?
- Enhanced and intelligent data quality alerting.  Alert support staff when:
  - Large blocks of data are bad
  - Data has been bad for an extended period of time (avoid duplicate alerts)
  - Inbound latency is causing data to drop
- Move away from manual data quality reviews

# QUESTIONS?

Mike Nugent
Lead Engineer – SPP Market Operations Support
mnugent@spp.org