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
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
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.
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