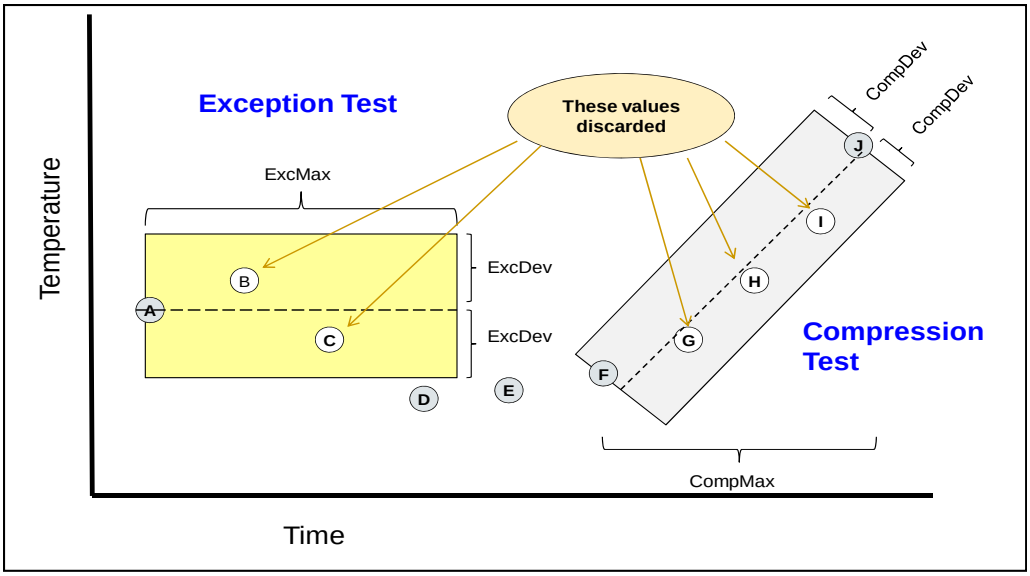
PATTERN DISCOVERY TECHNOLOGIES INC

OSIsoft.

Compression*Insight*™ for the OSIsoft PI System

Data Compression for Industrial Processes

Data historians, such as the OSIsoft PI System, utilize algorithms to compress raw data in order to effectively manage network bandwidth, and efficiently store large quantities of data for downstream analysis tasks. The Boxcar and Swinging Door compression algorithms used by the historian, require the end user to stipulate proper exception and compression deviation to ensure that the data archived closely reflects the characteristics of the original signal. (ExcDev and CompDev are shown in the accompanying diagram). Wide deviation ranges potentially mask subtle changes, leaving too many "straight lines", while narrow ranges (or even turning off exception and compression) will capture more data than is necessary. Establishing these proper compression settings can be a painstaking task, requiring labour intensive input from process engineers or subject matter experts who have an intimate understanding of the instrumentation in the process. Unfortunately, this task seldom registers on their list of priorities. The exercise can be overwhelming when considering thousands of tags. Generally speaking, "rules of thumb" or "default settings" are applied and never revisited, which leaves very little understanding of whether or not the archived data contains sufficient fidelity to accurately capture events that have occurred. The archived data often represents a "different version of the truth" based on these default compression settings.



The end user must attempt to strike a balance between collecting enough data to be meaningful, while also optimizing storage space and network bandwidth. This balance ensures that redundancy in the data is minimized, bandwidth is minimized, storage space is maximized and the performance of data retrieval and analysis tasks are not impacted by an unnecessary amount of data.

Impact of Proper Compression

Lower Data Redundancy

Maximize Disk Space

Minimize Network Bandwidth

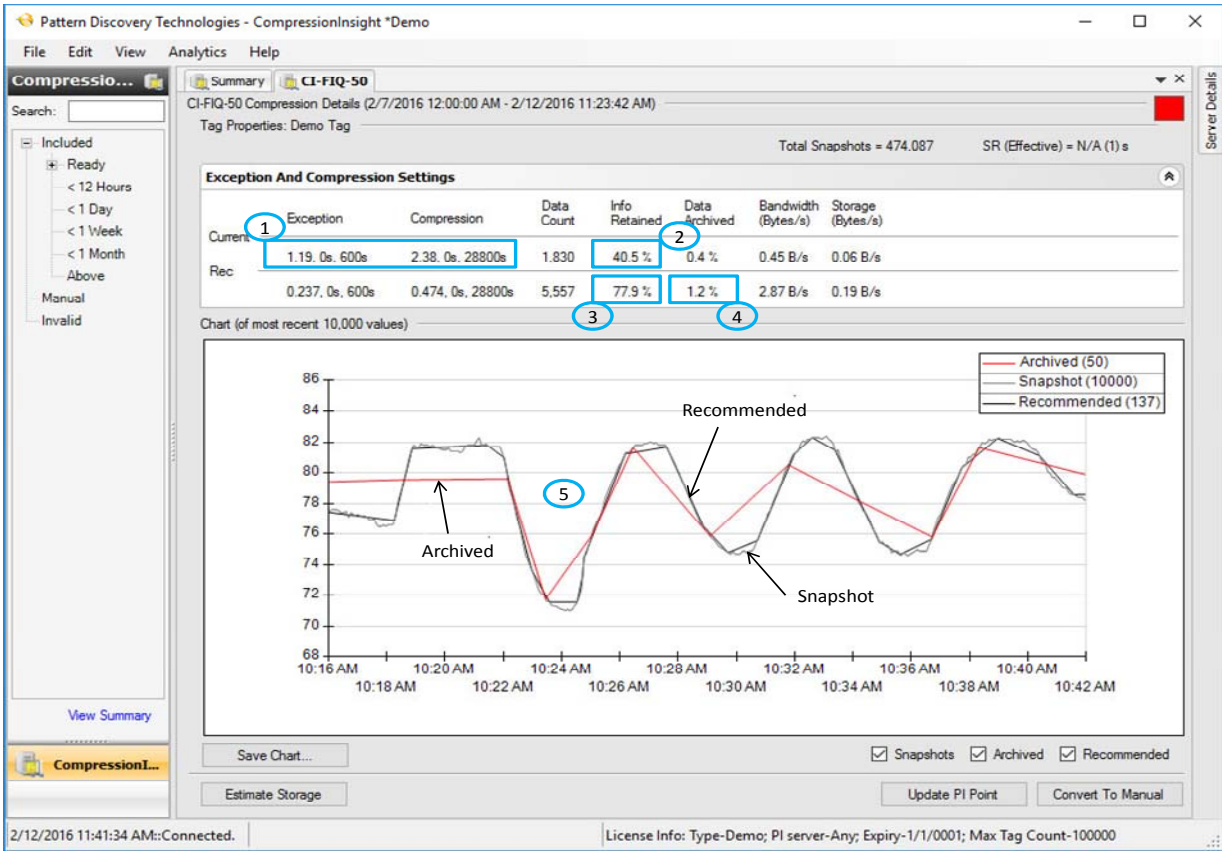
Improve Data Retrieval Performance

Compression*Insight*™ automatically and effortlessly evaluates compression and exception settings and provides feedback and recommendations.

Pattern Discovery Technologies Inc, 554 Parkside Drive, Waterloo, ON, N2L 5Z4
519-888-1001
www.patterndiscovery.com

A Picture is Worth a 1000 Words

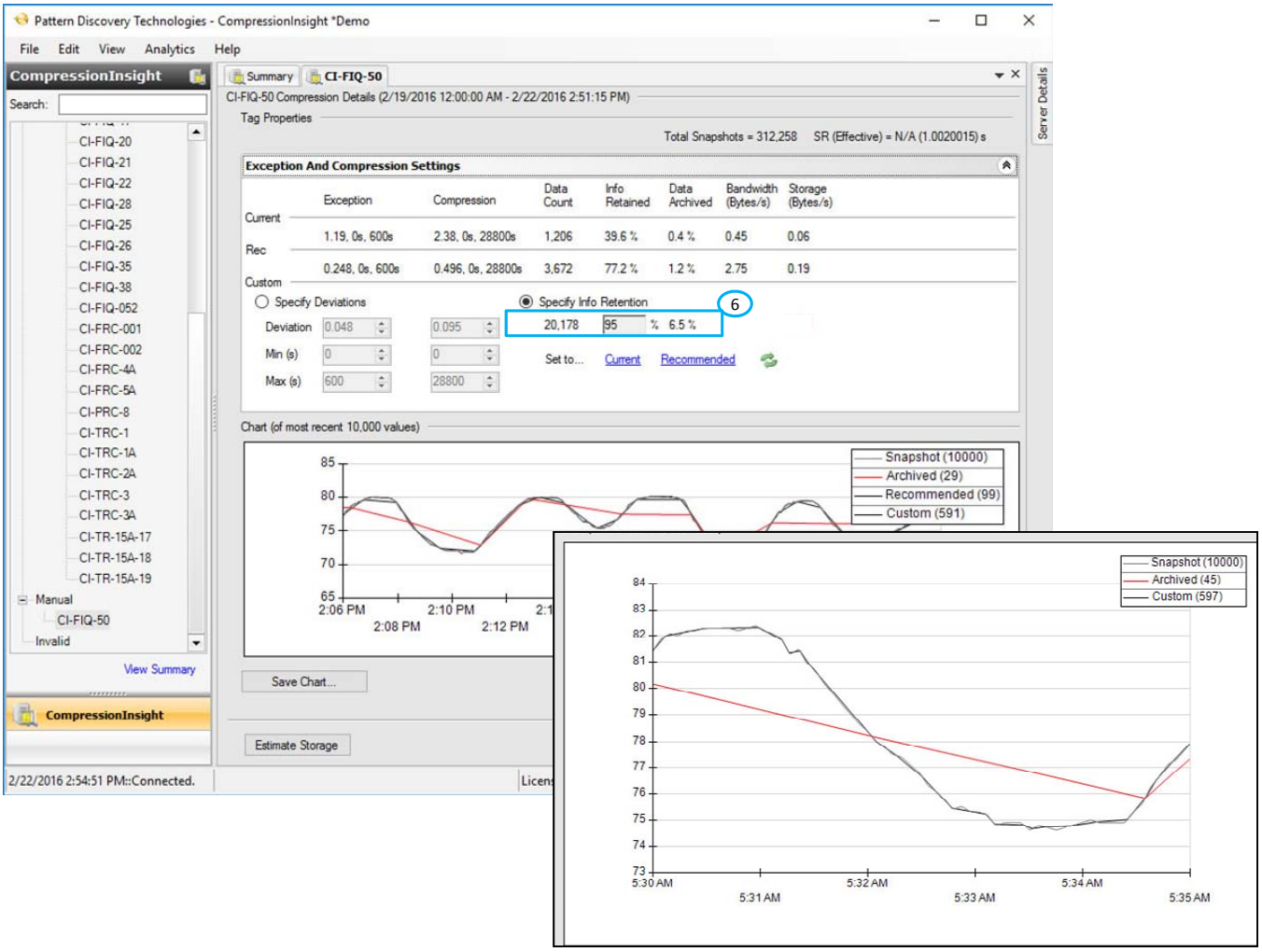
Compression*Insight™* uses advanced statistical analysis to evaluate the ***mutual information*** between raw data snapshots, current archived values, automatic recommendations and any custom settings the user may want to consider. It provides graphical and numerical feedback on the impact a change in the compression settings may have on the final archived value. Up to **5000 tags** can be analyzed simultaneously, with detailed feedback available in summaries, graphs, tables and reports.



1. Originally archived exception and compression deviation settings (set at approximately 1% and 2% of span – common default value).
2. Calculation of the amount of information retained when compared to the raw snapshot values of the original signal (only 40.5% of the original information is retained).
3. CompressionInsight makes automatic recommendations for over-compressed tags (using an optimization process). The information retained has improved to 77.9% by comparison.
4. Improved data fidelity only increases the data archived to 1.2% of the raw snapshot values (a nominal increase in comparison to the improvement in data fidelity).
5. Results are graphically displayed. The red line is the original archived data, while the recommendations and snapshots are grey and black (evidence that the recommendations more closely follow the snapshots than the originally archived data).

A Platform for Evaluating and Fine Tuning Fidelity

Depending on the nature of the original signal and the importance of the data to downstream analysis tasks, it may be necessary to evaluate higher levels of data fidelity than automatically recommended. It provides feedback on potential changes for under-compressed tags, or where exception and compression may have been turned off. Compression*Insight* provides a means to evaluate various scenarios and reports on information retained, as well as the impact on bandwidth and storage.



6. Information retained is specified as 95% (an increase from the recommended value of 77.2% noted in the earlier example). The data archived increases to 6.5% of the original signal, but a zoom on the graphical display shows that the snapshot and the new custom values are almost completely overlapped (as compared to the original archived value which shows as a straight line in red).

Justifying CompressionInsight

CompressionInsight can be justified by considering savings and improvements in a number of areas:

Process Engineering and Data Admin Time:

As noted earlier, to set compression tags manually requires a significant amount of input from subject matter experts. If one were to consider a 10,000 tag implementation, perhaps 50% would require some form of compression consideration. At a conservative estimate of 10 minutes per tag, this would take 5,000 x 10 minutes = 50,000 minutes or 833 hours. At a nominal cost of $60.00 per hour, cost for tag tuning would be $50,000 in labour alone. This does not include the amount of time required by the data administrator, to go into the system and change tag configurations, using the data administration tools.

Collaboration on Proper Settings:

At times, proper settings may need the input of multiple subject matter experts. In this case, subject matter experts must meet or manually prepare evaluation reports for consideration by team members. Compression*Insight* automatically provides the output required for team collaboration.

Disk Storage and Retrieval Performance:

Large tag counts lead to large storage requirements and storage costs. More importantly, the time to retrieve data for reports and analysis can cause bottlenecks and delays in accessing critical information for process changes or improvements.

Network Bandwidth:

For more remote locations, network bandwidth can be a significant consideration for data communication. Because exception tests are performed at the interface, exception settings can have a significant impact on bandwidth usage.

Data Fidelity:

Confidence in archived data is extremely important to any downstream analysis that will be performed using the data. For this reason, many process engineers request the compression and exception be turned off. The true characteristics of the original signal are lost without proper compression. Compression*Insight* provides concrete evidence and visual feedback, which greatly increases the confidence levels in archived data.

Centralized Compression Evaluation:

For organizations with multiple historians in different locations, all tag tuning can be done from one central location, eliminating the need for travel, extra expenses and associated labour costs.