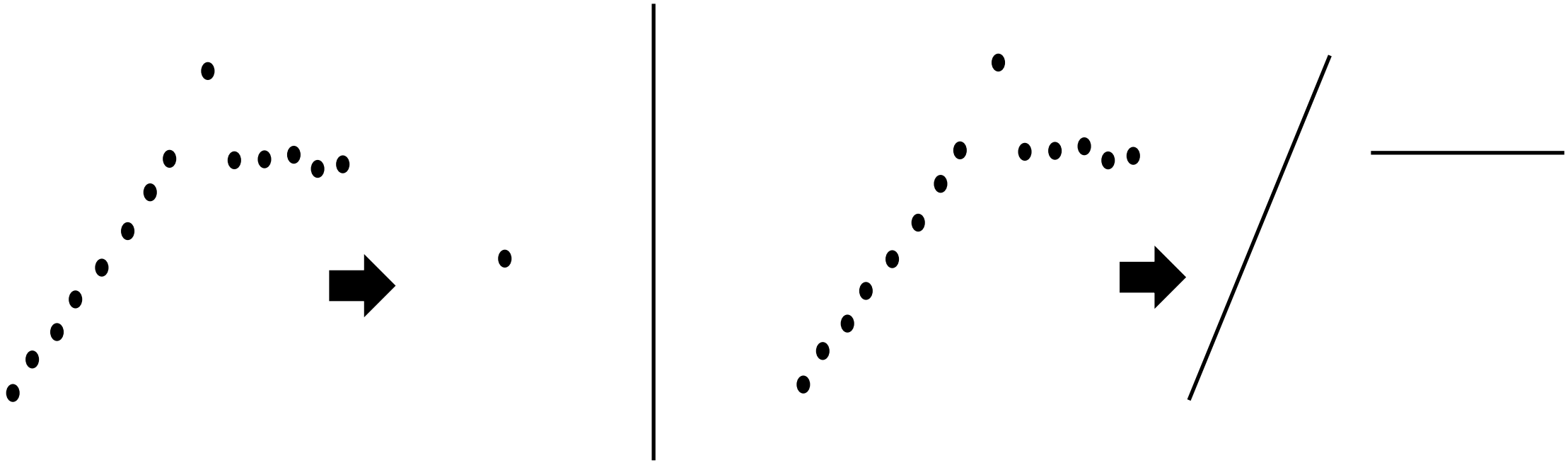
Model-Based Management of Correlated Dimensional Time Series

Søren Kejser Jensen (Aalborg University)
Torben Bach Pedersen (Aalborg University)
Christian Thomsen (Aalborg University)

Motivation

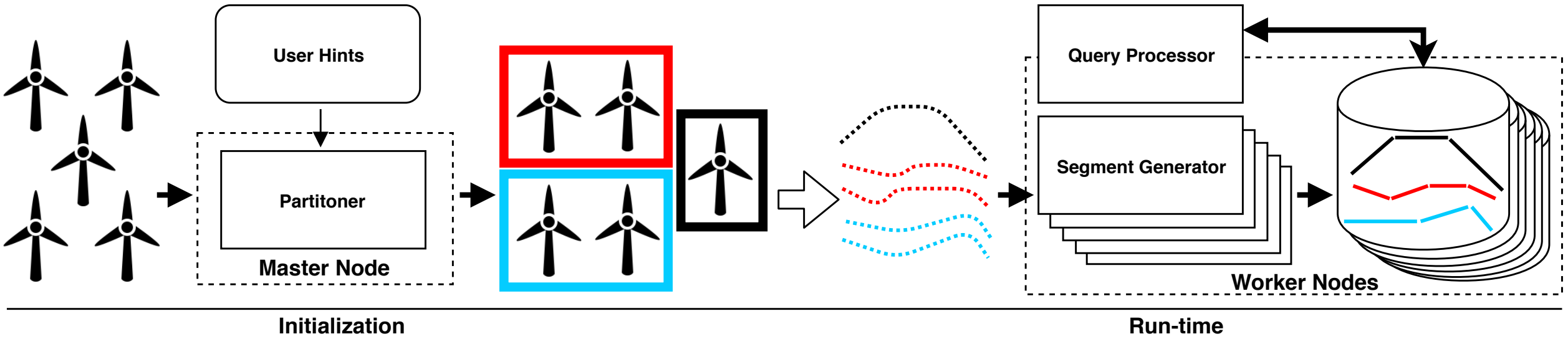
- Large industrial systems produce big amounts of **high-quality sensor data**
- Data is collected as **regular time series** with **only a few gaps** without values

- High frequency could benefit analysis but requires high amounts of storage
- Currently **simple aggregates are stored** with **outliers and fluctuations lost!**



- As a remedy we propose **ModelarDB:** `github.com/skejserjensen/ModelarDB`

System Overview



- **ModelarDB** is a distributed system build using a Master/Worker architecture
- Correlated time series are grouped during initialization and ingested at run-time
- **ModelarDB** stores models for excellent compression and query performance

Ingestion, Storage and Query Processing

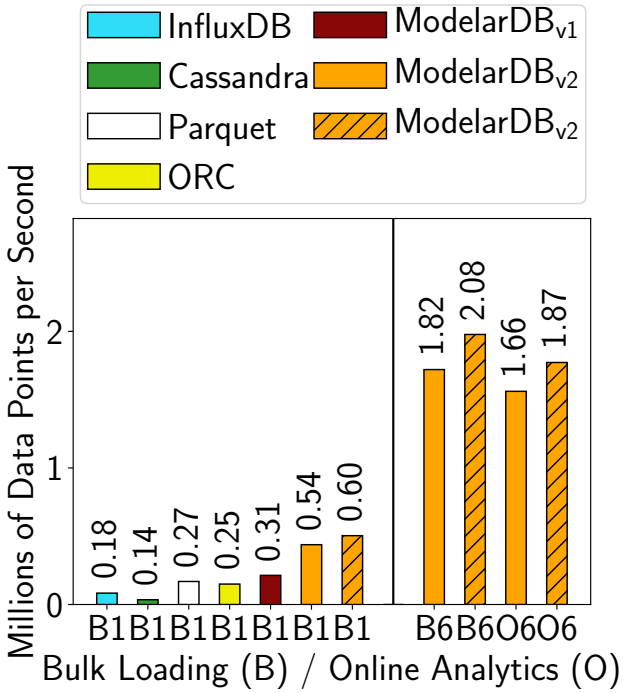


Figure: Ingestion

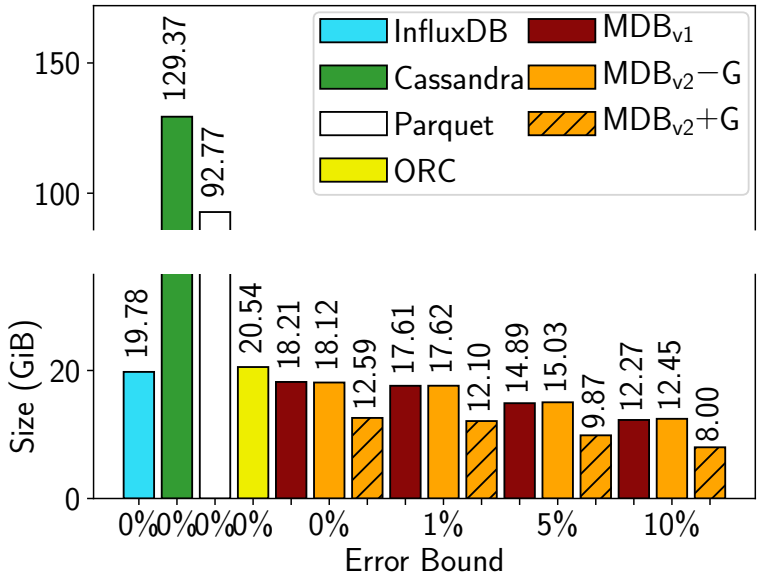


Figure: Storage

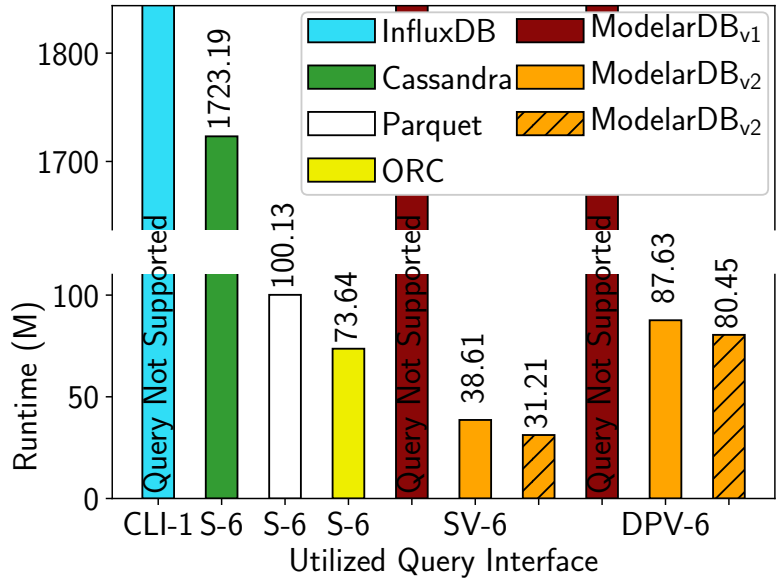


Figure: Large Aggregate Queries

- **ModelarDB has fast ingestion, excellent compression and fast aggregates**
  - Results for a real-life data set from wind turbines (339 GiB in CSV, 60 s sampling interval)
  - Our experiments show that ModelarDB is competitive for small queries and scales linearly
  - The system is extensible and users can implement additional model types through an API

Open Questions

- **Open Question 1:** What set of model types should be used for each data set?
- **Open Question 2:** Can we perform similarity search with multiple model types?
- **Open Question 3:** Can models be fitted to correlated time series at the edge?
- **Open Question 4:** Can the error-bound be inferred from augmented queries?

References

- Søren Kejser Jensen, Torben Bach Pedersen, Christian Thomsen, Time Series Management Systems: A Survey. In *IEEE Transactions on Knowledge and Data Engineering*, Volume 29, Number 11, Pages 2581–2600, November, 2017.

- Søren Kejser Jensen, Torben Bach Pedersen, Christian Thomsen, ModelarDB: Modular Model-Based Time Series Management with Spark and Cassandra. In *Proceedings of the VLDB Endowment*, Volume 11, Number 11, Pages 1688–1701, July, 2018.

- Søren Kejser Jensen, Torben Bach Pedersen, Christian Thomsen, Scalable Model-Based Management of Correlated Dimensional Time Series in ModelarDB". In the C*omputing Research Repository*, arXiv: 1903.10269, 2019.

- Søren Kejser Jensen, Torben Bach Pedersen, Christian Thomsen, Demonstration of ModelarDB: Model-Based Management of Dimensional Time Series. In *Proceedings of the 2019 International Conference on Management of Data*, Pages 1933–1936, 2019.