Assumptions of traditional BI platforms

1. **Predictable Frequency** - Data is extracted from source systems at regular intervals – typically measured in days, months and quarters.
2. **Static Sources** - Data is sourced from controlled, internal systems supporting established and well-defined back-office processes.
3. **Fixed Models** - Data structures are known and modeled in advance of analysis.
4. **Defined Queries** - Questions to be asked of the data (i.e., the reporting queries) are pre-defined.
5. **Slow-changing requirements** - Rigorous change control is enforced before the introduction of new data sources or reporting requirements.
6. **Limited users** - The consumers of BI reports are typically business managers and senior executives.

“Big Data” factors impacting traditional BI Databases

- **Semi-structured and unstructured data** typical in mobile, social, and sensor-driven applications cannot be efficiently represented as rows and columns in a relational database table.
- **Rapid evolution of database schema** to support new data sources and rapidly changing data structures is not possible in relational databases.
- Performance overhead of JOINs and transaction semantics prevents relational databases from keeping pace with the ingestion of high-velocity data sources.
- **Quickly growing data volumes** require scaling databases out across commodity hardware, rather than the scale-up approach typical of most relational databases.

=> All of this leads to growing use of NoSQL databases and data formats like JSON.
Differences between SQL and NoSQL

- Tables
- Schemas
- Normalization
- JOINs
- Data integrity
- Transactions
- Performance
- Scaling

So.....?

- Which is better?
  - the old technology (data warehouses)
  - the new technology (NoSQL databases)

- When and why?